UNIVERZITET U BEOGRADU
FILOZOFSKI FAKULTET

Goran S. Milovanović

# RACIONALNOST SAZNANJA: METATEORIJSKA I METODOLOŠKA ANALIZA FORMALNIH KOGNITIVNIH TEORIJA

doktorska disertacija

Beograd, 2013.

UNIVERSITY OF BELGRADE
FACULTY OF PHILOSOPHY

Goran S. Milovanović

# RATIONALITY OF COGNITION: A META-THEORETICAL AND METHODOLOGICAL ANALYSIS OF FORMAL COGNITIVE THEORIES

doctoral dissertation

Belgrade, 2013.

Mentor:

Dr Gordana Jovanović, redovni profesor

Filozofski fakultet, Univerzitet u Beogradu

Članovi komisije:

Dr Gordana Jovanović, redovni profesor

Filozofski fakultet, Univerzitet u Beogradu

---

Dr Dejan Todorović, vanredni profesor

Filozofski fakultet, Univerzitet u Beogradu

---

Dr Svetozar Sinđelić, docent (u penziji)

Filozofski fakultet, Univerzitet u Beogradu

---

Disertacija je odbranjena dana _____ 2013. god., u Beogradu.

*Zahvalnost*

Jedna stranica koju imam na raspolaganju ni u kom slučaju nije dovoljna da izrazim zahvalnost svima koji su mi pomogli i podržali me u prethodnom periodu.

Ričard Gonazels sa Univerziteta u Mičigenu i Džon Hej sa Univerziteta u Jorku su ljubazno podelili sa mnom rezultate svojih eksperimentalnih studija. Dejvid Harles sa Virdžinia Komonvelt Univerziteta i Dejvid Šenks sa Univerziteta Koledž London su poslali svoje naučne radove. Pomoć Iris Žeželj sa Filozofskog fakulteta u Beogradu i Ivane Kovačević sa Fakulteta organizacionih nauka u Beogradu sačuvala me je mnogih organizacionih neprilika u eksperimentalnom radu koje bih inače teško izbegao. Bobu Rederu sa Njujorškog univerziteta dugujem zahvalnost za to što me je uveo u (šik!) društvo bejzijanaca. Gregoriju Marfiju sa Njujorškog univerziteta dugujem zahvalnost za mentorski rad tokom dve godine naše saradnje, kao i za gostoprimstvo u Njujorku 2005. Kursevi matematičke statistike Lerija Malonija sa Njujorškog univerziteta bili su od neprocenjivog značaja. Dubravka Pavličić sa Ekonomskog fakulteta Univerziteta u Beogradu pregledala je moje eksperimente merenja monetarnih ekvivalenata i dala korisne sugestije. Milan M. Ćirković je bio inspirativan sagovornik - bez premca. Nadam se da će eksperimenti u ovoj tezi uveriti profesora Dejana Todorovića bar u to da nije potpuno proćerdao vreme uloženo u metodološku obuku studenata psihologije. Filozofu Svetozaru Sinđeliću hvala što je uopšte pristao da bude član komisije - siguran sam da mu je i više nego dosta studenata koji se o problemima saznanja raspisuju po stotinama stranica. Na kraju, svakako najveću zahvalnost dugujem profesorki Filozofskog fakulteta Gordani Jovanović, mojoj mentorki. Svojom hrabrošću i odgovornošću da preuzme mentorat nad doktorskim radom koji je suštinski teorijskog karaktera, na Odeljenju koje nema mnogo simpatija prema takvim poduhvatima, i to u komplikovanoj oblasti debate o racionalnosti, Gordana je postala moj heroj. Nadam se da će jednom moći da kaže da je vredelo stati iza ove teze.

G.S.M.

*Za Teodoru, koja je još mala*

# Racionalnost saznanja: metateorijska i metodološka analiza formalnih kognitivnih teorija

## Apstrakt

Predmet metateorijske i metodološke analize u ovoj tezi je naučni status koncepta racionalnosti saznanja u *kompjutacionoj kognitivnoj psihologiji* (KKP). Na prvom nivou analize, racionalnost se analizira kao predmet proučavanja kognitivne psihologije. U tom proučavanju je moguće doći do zaključka o tome da je neka kognitivna funkcija *racionalna* ili da je *ograničeno racionalna*: ovo empirijsko pitanje je karakteristično za savremenu *debatu o racionalnosti*. Na drugom nivou analize, racionalnost predstavlja teorijski i metodološki koncept koji je *a priori* ugrađen u temelje savremene kompjutacione kognitivne psihologije. *Centralni cilj analize koja se predstavlja u ovoj tezi jeste rasvetljavanje odnosa između ova dva koncepta racionalnosti.*

Analiza je organizovana u šest celina. U I delu teze uvodimo razliku između racionalnosti kao predmeta proučavanja i kao okvira za *a priori* teorijske i metodološke odluke u KKP. Kroz konceptualnu i dijahronu analizu razvoja teorija odlučivanja u društvenim naukama pokazujemo kako dolazi do razvoja savremene suprostavljenosti između normativnih i deskriptivnih teorija u kognitivnim naukama. Uvodimo Andersonovu paradigmu racionalne analize kao centralnu metodološku pretpostavku savremenih racionalnih teorija i dajemo kratak pregled analiza koje slede.

U II delu teze dajemo kompletnu kritičku analizu teorijske strukture kognitivne psihologije kao prirodne nauke o kognitivnim funkcijama. Prvo diskutujemo osnove bihejvioralne metodologije merenja neopservabilnih, internih konstrukata, što je centralni teorijski problem psihologije kao nauke uopšte, i pokazujemo na koji način je ovaj problem povezan sa problemom aksiomatizacije teorija odlučivanja. Zatim detaljno diskutujemo teorijske koncepte KKP kroz tri aktualne paradigme: simbolicističku, konekcionističko-emergentističku i konstruktivističko-enaktivističku. Diskutujemo tipologiju naučnih objašnjenja u ovim teorijskim paradigmama. Pokazujemo da neki delovi naučnog programa standarde simbolicističke KKP uopšte ne podležu mogućnosti falsifikacije. Konačno, definišemo metateorijski okvir za analizu formalnih kognitivnih teorija: skup teorijskih pojmova na osnovu kojih se

formiraju pozicije racionalnih teorija i teorija ograničene racionalnosti u debati o racionalnosti.

U III delu rasprave predstavljamo metateorijsku analizu pet grupa formalnih teorija u savremenoj KKP: teorija odlučivanja u uslovima rizika i neizvesnosti, teorija kauzalnog učenja, teorija pamćenja, teorija suđenja i rezonovanja i teorija koncepata i kategorizacije. Rezultati naše diskusije otkrivaju sasvim nove, do sada neproučene probleme u racionalnoj analizi kognitivnih funkcija. Naše analize ukazuju na to da je teško ili nemoguće tvrditi da kognitivni sistemi postavljaju jedinstvene ciljeve izračunavanja u odnosu na jedinstveno definisane probleme adaptacije, što je osnovna (implicitna) pretpostavka racionalne analize. Ovo pitanje se prirodno odmah povezuje sa pitanjem o postojanju reprezentativnog subjekta ma kog formalnog kognitivnog modela i pitanjem o mogućnosti paralelnog objašnjenja istih bihejvioralnih podataka primenom više različitih modela. Analiza odnosa pojmova racionalnosti, adaptacije i optimizacije u racionalnim analizama otkriva da je teško ili nemoguće razlikovati pojam racionalnosti od pojma optimalnosti, što onda potpuno relativizuje pitanje adekvatnosti normativnih okvira za racionalne analize u psihologiji. Diskutuje se pitanje intuitivne opravdanosti aksiomatskog okvira savremene deskriptivne teorije odlučivanja.

U IV delu teze predstavljamo kratak, jezgrovit pregled istorije ideja koje su vodile razvoju savremene KKP. Prepoznajemo tri istorijske linije koje se sustiču u razvoju KKP u drugoj polovini XX veka. Jedna se odnosi na tendenciju mehanicističkog objašnjenja ljudskih mentalnih procesa koja počinje sa Dekartom. Ta linija ima vrhunac u diskusijama filozofije matematike između intuicionista i formalista koje kao posledicu imaju razvoj ključnih teorijskih koncepata koji grade KKP, poput koncepta izračunavanja ili formalnog sistema. Druga istorijska linija se prepoznaje u kontinuitetu sa razvojem teorije verovatnoće i nastavlja se u savremenoj KKP kroz proučavanje problema odlučivanja i suđenja u uslovima rizika i neizvesnosti. Početak treće istorijske linije koja vodi u savremenu KKP se nalazi u otkriću problema termodinamike i razvoju statističke fizike u drugoj polovini XIX veka a u savremenoj KKP se prepoznaje u sagledavanju kognitivnih problema kao problema optimizacije.

U V delu teze razvijamo matematički model jedne racionalne bejzijanske teorije odlučivanja. Naš cilj je da pokažemo kako je moguće razviti teoriju odlučivanja u kojoj su kognitivni procesi koji vode ka odstupanjima od normativnih odluka izolovani od kognitivnih procesa koji se odnose na samo odlučivanje. Kao modele odlučivanja koristimo Viskuzijevu bejzijansku teoriju perspektivne

reference (Viscusi, 1989) koju proširujemo modelom formiranja verovanja. Teorija poverenja, kako nazivamo rezultirajuću teoriju, može da objasni sve robustne efekte odstupanja od normativnih odluka. Predstavljamo eksperimentalne testove osnovnih pretpostavki teorije poverenja (eksperimenti 1a i 1b) i dva pokušaja selekcije modela odlučivanja kroz eksperimente merenja monetarnih ekvivalenata (eksperimenti 2a i 2b) i eksperimente izbora (eksperimenti 3a i 3b). U ovim eksperimentima, teoriju poverenja poredimo sa kumulativnom teorijom izgleda Tverskog i Kanemana (Tversky & Kahneman, 1992). Pokazuje se da teorija poverenja objašnjava empirijske odluke ispitanika nešto bolje od kumulativne teorije izgleda, uključujući tu i demonstraciju empirijskih efekata za koje kumulativna teorija izgleda uopšte nema eksplanatorne mehanizme. Analiza pretpostavke o homogenosti preferencija pokazuje da ovaj uslov nije zadovoljen u empirijskim odlukama, što znači da kanonička forma teorije izgleda koju su predložili Tverski i Kaneman 1992. nije tačna. Kroz analizu teorijskih struktura dve teorije koje poredimo u V delu uvodimo teorijske pojmove dispozicione i reprezentacione kognitivne teorije.

U VI delu izvodimo tri ključna argumenta kojima pokušavamo da opravdamo sledeći zaključak: racionalnost saznanja je pojam koji treba isključiti iz naučne diskusije kognitivnih funkcija. Naš prvi argument se bazira na analizi mogućnosti da se za svaku dispozicionu kognitivnu teoriju razvije paralelna, odgovarajuća reprezentaciona kognitivna teorija koja će objašnjavati iste bihejvioralne podatke kao i prva. Ukoliko je tako nešto moguće, a mi tvrdimo da vrlo verovatno jeste, onda uopšte nema smisla karakterisati bilo koje opservabilno ponašanje kao racionalno ili ograničeno racionalno. Odluka o racionalnosti saznanja tako postaje tek izbor između dva ekvivalentna deskriptivna jezika. Naš drugi argument se bazira na problemu izbora jedinstvenih ciljeva kognitivnog izračunavanja koji smo prepoznali u III delu. Ukoliko kognitivni sistem ne reprezentuje probleme adaptacije kao jedinstvene kompjutacione probleme, već izvore informacija iz okoline tumači kao da oni zahtevaju različita kognitivna izračunavanja paralelno, optimalno je za takav sistem da problemu adaptacije pristupi primenom mešovitih strategija kako ih definiše teorija igara. Prema tome, svaki kognitivni čin je ujedno čin strategijske interakcije sa okolinom. Da bismo precizno diskutovali ovakvu mogućnost, razvijamo *strategijsko shvatanje kognitivnog sistema* u kome svi kognitivni činovi uzimaju oblik mešovitih strategija odn. distribucija verovatnoće nad različitim formalnim kognitivnim modelima. Pokazujemo zatim da kognitivni sistem koji tako pristupa

problemu adaptacije nužno ostavlja bihejvioralni trag na osnovu koga nije uopšte moguće govoriti o tome da li je on racionalan ili ne. Nekoliko konkretnih primera opravdava ovu tvrdnju. U analizi strategijskih interakcija kognitivnog sistema sa drugim kognitivnim sistemima pokazujemo kako je nemoguće razdvojiti pitanje racionalnosti saznanja od izbora ciljeva koje odgovarajuće kognitivne funkcije treba da zadovolje. Konačno, naš treći argument se nastavlja na diskusiju kreativnih funkcija konceptualnog sistema iz III dela teze i primenom strategijske interpretacije kognitivnih funkcija pokazuje kako je nemoguće postaviti normativne standarde za procese interpretacije simbola uopšte. Pokazuje se da je za konvencionalnu, arbitrarnu prirodu simbola neophodno upravo obrnuto: nivo nestabilnosti u njihovoj interpretaciji koji obezbeđuje mogućnost praktično beskonačnih reinterpretacija. U suprotnom, suštinska odlika semiotičkih sistema - odlika konvencionalnosti sistema u celini i arbitrarnosti na nivou pojedinačnih simbola - ne bi bila zadovoljena. U domenu simboličkih kognitivnih funkcija, tako, nije moguća ni racionalna analiza.

VII i poslednji deo teze daje koncizniji pregled naših rezultata i ponavlja opšti zaključak iz VI dela rasprave: racionalnost saznanja nije pojam naučne teorije, i, u najstrožem smislu, ne zaslužuje da ima naučni tretman unutar diskursa kompjutacione kognitivne psihologije.

*Ključne reči*: aksiomatizacija, intuicija, kognitivna psihologija, matematički model, racionalnost, teorija, viši mentalni procesi

Društveno-humanističke nauke

Psihologija

UDK: 159.9.01:159.953/.56(043.3)

# Rationality of cognition: a meta-theoretical and methodological analysis of formal cognitive theories

## Abstract

The subject of the present meta-theoretical and methodological analysis is the scientific status of the concept of rationality of cognition in *Computational Cognitive Psychology* (CCP). On the first level of analysis rationality is constrained as a subject matter of cognitive psychology, and only on this level of analysis it is possible to reach a conclusion on whether a cognitive function is rational or not. This empirical question is characteristic of the contemporary *rationality debate*. On the second level of analysis, rationality stands as a theoretical and methodological concept which is built *a priori* in the foundations of the contemporary CCP. *The central goal in this thesis is to clarify the relationship between these two concepts of rationality.*

In Part I we introduce the distinction between (*i*) rationality constrained as a subject matter of CCP and (*ii*) rationality constrained as a framework for theoretical and methodological decisions *a priori* in CCP. Through conceptual and diachronic analysis of the development of decision theories in social sciences we demonstrate the development of the contemporary opposition between *normative* and *descriptive theories* in cognitive science. We introduce Anderson's *rational analysis* as a central methodological assumption of rational theories and provide a short overview of the discussions that follow.

In Part II we provide a complete critical analysis of the theoretical structure of CCP as a natural science of cognitive functions. We first discuss the foundations of the behavioural approach to the measurement of internal constructs, which presents the central theoretical problem of scientific psychology in general, and demonstrate the way that this problem is related to the problem of the axiomatization of decision theories. Then we provide a detailed discussion of CCP in three actual theoretical paradigms: symbolicistic, connectionist-emergentistic and constructivist-enactivistic. We discuss the typology of scientific explanation involved in these theoretical paradigms. We demonstrate that some parts of the program of the standard symbolicistic CCP cannot be falsified by means of standard behavioural methodology. Finally, we define a meta-theoretical framework for the analysis of formal cognitive theories: a set of theoretical concepts upon which the positions of

rational theories and theories of bounded rationality are developed in the rationality debate.

In Part III of the discussion we present a meta-theoretical analysis of five groups of formal theories in contemporary CCP: theories of choice under risk and uncertainty, theories of causal learning, theories of memory, theories of judgement and reasoning, and theories of concepts and categorization. The results of our discussion uncover new, previously unstudied problems in the rational analysis of cognitive functions. Our analyses suggest that it is difficult, if not impossible, to assert that cognitive systems formulate unique goals of cognitive computations in front of uniquely defined problems of adaptation, which is a basic (implicit) assumption of rational analysis. This question naturally connects to the questions of $(i)$ the existence of the representative subject of any formal cognitive model and of $(ii)$ the possibility of parallel explanation of same behavioural data by multiple theoretical models. The analysis of the relationship between the concepts of rationality, adaptation and optimization, shows that it is difficult or impossible to differentiate between rationality and optimality in contemporary rational analyses, which complicates additionally the question of the adequacy of normative frameworks. We also discuss the status of intuitive justification of the axiomatic framework of contemporary descriptive theory of choice.

In Part IV of the thesis we present a short overview of the history of ideas that led to the development of the contemporary CCP. We recognize three historical lines that converge in the development of CCP in the second half of the XXth century. The first relates to the tendency to provide a mechanicistic account of human mental processes and dates back to the philosophy of Descartes. This line reaches its peak in the discussion of the philosophy of mathematics between the formalists and intuitionists - the discussion that brings the key theoretical concepts of CCP as its consequences, such as the concept of computation, or the concept of a formal system. The second historical line can be recognized in the continuity of the development of probability theory with the study of the problems of rational choice and judgement under risk and uncertainty in contemporary CCP. The beginning of the third historical line that converges in the contemporary CCP is in the discovery of thermodynamics and statistical physics in the second half of the XIXth century, and it can be recognized in contemporary CCP as an impulse to constrain cognitive problems as optimization problems.

In Part V of the thesis we develop a formal, mathematical model of one

rational, Bayesian decision theory. It is our goal to demonstrate the possibility of a decision theory in which the cognitive processes responsible for deviations from normative decisions stand in isolation from the cognitive processes involved in pure choice. As a decision model we use Viscusi's Prospective Reference Theory (Viscusi, 1989), in which we then incorporate an alternative model of belief formation. The resulting Confidence Theory is able to explain all standard, robust deviations from normative decisions. We present experimental tests of the basic assumptions of the Confidence Theory (experiments 1a and 1b), as well as two attempts at model selection, through experimental measurement of certainty equivalents (experiments 2a and 2b) and choice experiments (experiments 3a and 3b). Confidence Theory explains empirical choices slightly better than the Tversky and Kahneman's Cumulative Prospect Theory (Tversky & Kahneman, 1992) while encompassing some empirical effects that Cumulative Prospect Theory cannot incorporate even in principle. The analysis of the homogeneity of preferences shows that this important condition is not satisfied in empirical choices, which implies that the canonical parametric form of the Cumulative Prospect Theory is not correct. We compare the theoretical, explanatory structures of these two decision models (which we use as representatives of rational and boundedly rational theories in general) and formulate the theoretically important distinction between *dispositional* and *representational cognitive theories*. This distinction is then used in Part VI to formulate one of our main arguments in this thesis.

Part VI presents the key arguments that we use to support the following general conclusion: the rationality of cognition is a concept that should be excluded from the scientific discussion of cognitive functions. The first argument is based on the analysis of the possibility that there is always a representational cognitive theory that matches the behavioural predictions of the respective dispositional cognitive theory. If the former statement is true, and we argue that it is very probably true, than it does not make any sense to characterize any observable behaviour as rational or boundedly rational. The decision of whether cognition is rational or not reflects merely a choice between two equivalent languages of scientific description. Our second argument is based on the problem of choice of uniquely defined goals of cognitive computations that was recognized in Part III. If a cognitive system does not represent the problem of adaptation as a unique computational problem, rather interpreting the source of environmental information as demanding different, parallel cognitive computations, it is optimal for such a system to solve

the adaptation problem by adopting a mixed-strategy as defined in game theory. Accordingly, every cognitive act is always an act of strategic interaction with the environment. To be able to discuss this situation in exact terms, we develop a *strategic view* of the cognitive system, in which cognitive acts take the form of mixed strategies: *probability distributions defined on the space of formal cognitive models*. We show then that a cognitive system which approaches the problem of adaptation strategically leaves a behavioural trace upon which it is impossible to decide whether it was produced by a rational or a boundedly rational system. We discuss several examples to support this observation. In the analysis of strategic interactions between different cognitive systems we also show that it is impossible to separate the question of the rationality of cognition from the question of the selection of goals that the respective cognitive functions are expected to satisfy. Finally, our third argument continues the discussion of creative conceptual functions that began in Part III. This argument shows that from the strategic interpretation of cognitive functions it follows that is not possible to formulate normative standards for the processes of symbolic interpretation. It is shown that just the opposite of stable cognitive strategies is what is necessary to support the conventional, arbitrary nature of symbols - that in turn enables for a potentially infinite number of symbolic interpretations. If cognitive strategies in symbolic interpretations remain fixed, the fundamental conditions of arbitrariness and conventionality are not met. In the domain of symbolic cognitive functions, thus, rational analysis is not possible.

In the last part of the thesis we provide a more concise overview of our results and reiterate the general conclusion reached in Part VI: rationality of cognition is not a concept of a scientific theory, and does not deserve a scientific treatment in the scope of the discourse of computational cognitive psychology.

*Keywords*: axiomatization, intuition, cognitive psychology, mathematical model, rationality, theory, higher mental processes

Social sciences and humanities

Psychology

UDC: 159.9.01:159.953/.56(043.3)

# Sadržaj

# Deo I

# Problem racionalnosti saznanja

Već prema našim intuitivnim shvatanjima, osoba koja tvrdi i veruje da više voli konjak od crnog vina, više voli crno vino od belog vina, i više voli belo vino od konjaka, nije racionalna. Od nekoga ko više voli konjak od crnog vina koje opet više voli od belog vina, očekivali bismo da preferira konjak i u odnosu na belo vino; tvrdnja da belo vino ipak voli više od konjaka čini se iracionalnom. Naravno, ova intuitivna analiza racionalnosti ima važenje samo *ceteris paribus*. Ukoliko neku osobu pitamo o njenim preferencijama prema ovim pićima pre i posle jela, ili u različitim društvenim okolnostima, možemo očekivati da njeni izbori budu različiti. Na prvi pogled neposredniji, jednostavniji primer onoga što kolokvijalno nazivamo iracionalnim bio bi susret sa osobom koja tvrdi i veruje da bezbojne zelene ideje spavaju besno. Izlaganje iracionalnosti ovog tipa moglo bi da se osloni na tvrdnju da navedeni iskaz krši *selekcione restrikcije*, skup ograničenja u upotrebi koncepata u odnosu na njihovo značenje: nešto što je bezbojno ne može ujedno biti i zeleno, ideje nemaju boju, ideje ne mogu da spavaju, spavati se ne može besno. Produbljujući ovo objašnjenje ubrzo ulazimo u čisto psihološku problematiku, nalazeći da se svi termini u navedenoj rečenici odnose na određene koncepte koje reprezentuje kognitivni sistem, da ovi opet nose reference u odnosu na objekte u realnosti, da ih odlikuju izvesne karakteristike čija

pravila kompozicije jesu deo prekršenih selekcionih restrikcija, itd. Ova dva primera sažeto predstavljaju problematiku kojoj će biti posvećene naredne stranice. Oba se odnose na ustanovljavanje racionalnosti, shvaćene kao konzistentnosti nečijih odluka u odnosu prema nekoj minimalnoj, prihvatljivoj koherenciji koju nalaže intuicija, ili prihvatljive kombinacije značenja u postupku provere istinitosti neke tvrdnje, u slučaju semantike; te uopšte, *racionalnosti saznajnog subjekta*, koja je predmet naše analize. Empirijska psihologija je ovu problematiku, poput i mnogih drugih, „nasledila“ iz domena filozofske rasprave uvodeći je u naučnu raspravu njenim određenjem kao predmeta istraživanja i definisanjem empirijske metodologije za njeno proučavanje. Može se pokazati, kao što verujemo, da je preuzimajući problematiku racionalnosti i prenoseći je u domen naučne analize iz filozofije, savremena psihologija u paketu dobila više problema nego što je uobičajeno. Naime, osim racionalnosti ljudskog saznanja, odn. ljudskih *kognitivnih funkcija*, koja je osnovni predmet rasprava o racionalnosti u okvirima kognitivne psihologije, ne treba zaboraviti da je kognitivna psihologija i sama nauka, i da kao takva ona poštuje određeni *credo* naučne racionalnosti. Naš osnovni cilj u raspravi koja sledi je da rasvetlimo problematiku u kojoj se nalazi jedna naučna disciplina koja racionalnost uzima za *predmet* svog proučavanja dok ujedno sama mora da ostane verna *naučnoj racionalnosti* na koju je obavezuje status nauke. Osnovni, duboki paradoks koji motiviše ovaj rad je paradoks u kome jedna naučna disciplina, ako je po sebi i unapred racionalna, uopšte može da dovede u pitanje, ili uopšte postavi tvrdnju o tome da je neki proces saznanja racionalan, ili nije, ili diskutuje da li je racionalan u određenom stepenu. U narednim redovima ćemo postepeno postavljati i analizirati ovaj problem u sve preciznijim terminima.

Dakle, osnovni problem rasprave o racionalnosti koju predstavljamo sastoji se u analizi odnosa dve teorijske pozicije koje ovaj pojam okupira u naučnom diskursu savremene kognitivne psihologije: jedne, u kojoj je on predmet analize, i druge, u kojoj je on deo teorijske, eksplanatorne strukture kognitivne psihologije kao nauke uopšte. Našoj analizi je, naravno, neophodan širi kontekst. Taj širi kontekst u našem slučaju predstavljaće teorijska koncepcija *kompjutacionizma*. Savremena kognitivna psihologija, posmatrana na najopštijem teorijskom nivou analize, jeste *kompjutaciona kognitivna psihologija*. Sa ovom tvrdnjom se neće složiti svi, a deo naše analize racionalnosti biće posvećen obrazloženju ovog stava: kompjutacionizam je danas praktično jedina naučna formulacija problema saznanja. Šta znači da je savremena kognitivna psihologija kompjutaciona? To znači da njom široko dominira tvrdnja da, posmatrani na najopštijem nivou analize, *ljudski saznajni procesi jesu ekvivalentni procesima izračunavanja u nekom formalnom sistemu*. To

dalje znači, kao što ponavljaju neki autori u raspravama o kompjutacionizmu (Fodor, 2000), i to da je shvatanje ljudskih saznajnih procesa kao izračunavanja *naučna formulacija* problema saznanja koja danas nema alternativu; ne nužno i to da je ona jedina i tačna teorija ljudskog saznanja. Onda, kakav je naučni status pojma racionalnosti u kompjutacionoj kognitivnoj psihologiji? Racionalnost je, kao što ćemo pokazati, ugrađena u same temelje naučnog projekta kompjutacione kognitivne psihologije. Preciznije, pokazaćemo da bez obzira na neslaganja o empirijskom statusu racionalnosti ljudskih kognitivnih funkcija, jedna elementarna tvrdnja o racionalnosti saznajnog subjekta jeste deo temelja ove nauke. Ta fundamentalna tvrdnja o racionalnosti saznanja prožimaće celu našu raspravu i toliko je jednostavna da možemo da je postavimo odmah, bez prethodnog uvođenja komplikovanih teorijskih i metodoloških pojmova: *kognitivno je racionalan* subjekt $S$ čije ponašanje $B$ konzistentno svedoči o tome da on dela u skladu sa svojim verovanjima $\psi$, kako bi ostvario svoje ciljeve $G$ u nekoj sredini $E$. Uskoro ćemo detaljno razmotriti strukturu ove tvrdnje. Obeležimo ovaj koncept racionalnosti kao RACIONALNOST$_1$ za potrebe dalje rasprave. Zauzimajući fundamentalnu teorijsku poziciju, RACIONALNOST$_1$ predstavlja *eksplanatorni koncept* čiju ulogu u diskusiji užih, specifičnih teorija o kognitivnim funkcijama čoveka mi nameravamo da proučimo. Na drugom mestu, racionalnost je *predmet istraživanja*, i ovaj drugi položaj koncepta racionalnosti prepoznajemo u naučnom diskursu na onim mestima gde se raspravlja o *racionalnosti kognitivnih funkcija*, kada se koriste termini i koncepti kao što su „*racionalna analiza*“, „*racionalna teorija pamćenja*“, „*racionalno donošenje odluka*“ i drugi. Kada zauzima ovu drugu teorijsku poziciju, pretpostavlja se da je koncept racionalnosti *podložan empirijskoj analizi* u okviru bihejvioralne metodologije koja se koristi u psihologiji, eksperimentalnoj ekonomiji i drugim bihejvioralnim naukama. Za potrebe naše rasprave označićemo racionalnost u ovoj drugoj teorijskoj poziciji kao RACIONALNOST$_2$. Naglašavamo da racionalnost neke kognitivne funkcije, RACIONALNOST$_2$, kao predmet empirijskog istraživanja, nije pretpostavljena: ona može biti ustanovljena ili ne, dokazana ili pobijena, čak je moguće (kao što ćemo videti na primeru odlučivanja) da postoji čitav kontinuum racionalnosti neke kognitivne funkcije. Odnos između racionalnosti kao eskplanatornog, teorijskog koncepta koji određuje strukturu naučnih teorija kognitivne psihologije (RACIONALNOST$_1$) i racionalnosti koja može biti empirijski ustanovljena ili ne u odnosu na neku posebnu kognitivnu funkciju (RACIONALNOST$_2$), jeste centralni odnos koji želimo da rasvetlimo u ovoj tezi.

Razlikovanjem dve teorijske pozicije racionalnosti koje smo označili kao RACIONALNOST$_1$ i RACIONALNOST$_2$ - ili možda dva različita koncepta koji se kriju

pod istim imenom - pitanje o tome na koji način ona racionalnost koja se nalazi u teorijskim temeljima savremene kognitivne psihologije utiče na empirijsku analizu racionalnosti saznajnih fenomena kao predmeta proučavanja te nauke trebalo bi da bude jasno motivisano. Podrazumevajući da je ključni potez u povezivanju empirijskih fenomena sa teorijskim konstrukcijama neke nauke onaj koji se odnosi na konstituisanje tih fenomena odgovarajućom metodologijom, jasno je da će naša rasprava suštinski morati da bude *metodološkog karaktera.* Dva procesa, jedan u kome se značenje i sadržaj prenose sa teorijskih konstrukcija ka empirijskim opservacijama, i drugi, kojim se kroz određeni posmatrački jezik opisom empirijskih fenomena konstituišu teorijski koncepti, nalaze se u fokusu naše rasprave.

Vratimo se sada tvrdnji čija će analiza biti predmet svih budućih diskusija u ovoj raspravi:

*Kognitivno je racionalan subjekt $S$ čije ponašanje $B$ konzistentno svedoči o tome da on dela u skladu sa svojim verovanjima $\psi$, kako bi ostvario svoje ciljeve $G$ u nekoj sredini $E$.*

Naš neposredni cilj je da jasno konceptualno razdvojimo RACIONALNOST$_1$ od RACIONALNOSTI$_2$. U većem delu rasprave koja sledi pokušaćemo da otkrijemo na koji način je motivisano ovakvo shvatanje i demonstriramo da se ono, makar implicitno, nalazi u osnovama svake teorije o kognitivnim funkcijama. Prvo, o terminima koje koristimo. Subjekat $S$ (engl. *Subject*) je *kognitivni akter*, biće opremljeno skupom kognitivnih funkcija koje koristi da bi tokom svoje egzistencije sebi obezbedilo adekvatno saznavanje sveta. Saznanje mora biti adekvatno pošto tokom svoje egzistencije $S$ mora da ispuni određene ciljeve $G$ *(*engl. *Goals) u čiju opravdanost i racionalnost ne sumnjamo,* u nekoj sredini $E$ (engl. *Environment*), koju u početku rasprave tretiramo kao fiksnu, datu, određenu i objektivnu. Sredina $E$ je izvor informacija koje su relevantne za $S$ u odnosu na njegove ciljeve $G$. Tokom interakcija sa sredinom $E$, $S$ formira verovanja $\psi$ koja se odnose na stanja te sredine $E$. Verovanja $\psi$ koja formira $S$ mogu biti, kako nas uči epistemologija, tačna i netačna, opravdana i neopravdana. Konačno, pošto mi nemamo direktan uvid u verovanja $\psi$ nekog subjekta $S$, ne preostaje nam ništa drugo do da se ponašamo kao eksperimentalni psiholozi i o njegovim neopservabilnim verovanjima zaključujemo na osnovu opservabilnog ponašanja $B$. Ponašanje $B$ je rezultat primene *kognitivnih funkcija* kojima je opremljen $S$, i koje proizvode $B$ na osnovu $\psi$, u odnosu prema ciljevima $G$ i sredini $E$: $S(\psi|G,E)\rightarrow B$. Prethodni formalni izraz čitamo kao: „$S$ primenjuje kognitivne funkcije u skladu sa svojim verovanjima $\psi$, *ako su dati ciljevi $G$ i sredina $E$*, da bi proizveo ponašanje $B$." Ako prethodni formalizam opisuje proces koji proizvodi opservabilno ponašanje subjekta, onda naša tvrdnja o racionalnosti

saznanja predstavlja njegovo „čitanje unazad“: kognitivno racionalnim smatramo subjekta čije ponašanje *otkriva* da on dela u skladu sa svojim verovanjima da bi ispunio određene ciljeve u određenoj sredini.

Pretpostavimo, dakle, da se naš kognitivni akter $S$ nalazi u sredini $E$ o kojoj mi imamo objektivna, tačna znanja - na primer, da se on nalazi u nekoj eksperimentalnoj situaciji tipičnoj za laboratorijska istraživanja u kognitivnoj psihologiji - kao i da $S$ ima fiksne, poznate ciljeve $G$ - o čijoj se opravdanosti i racionalnosti mi ne pitamo. Opet u kontekstu poznatih laboratorijskih istraživanja, možemo da pretpostavimo da smo kroz dogovor sa $S$ odredili ciljeve $G$ eksperimentalnom instrukcijom. O ovakvoj situaciji, koja perfektno odslikava proces naučne opservacije u eksperimentalnoj kognitivnoj psihologiji, postoji nekoliko mogućnosti koje želimo da ispitamo u odnosu na proces proizvođenja ponašanja: $S(\psi|G,E)\rightarrow B$. Diskutovaćemo situaciju u kojoj (I.1) su verovanja $\psi$ tačna i opravdana, (I.2) tačna i neopravdana, (II.1) netačna i opravdana i (II.2) netačna i neopravdana. Unutar svake od ovih situacija, moraćemo da se zapitamo ponešto o prirodi kognitivnih funkcija kojima je $S$ opremljen, naime: da li primena tih kognitivnih funkcija na verovanja $\psi$ konzistentno vodi ponašanju $B$ koje zadovoljava ciljeve $G$ u sredini $E$?

I.1. $S$ ima verovanja $\psi$ koja su *tačna* i *opravdana*, te konstituišu ono što po tradicionalnoj definiciji u epistemologiji nazivamo *znanjem*, svesni toga da ta definicija nije bez problema. Ako su $\psi$, do kojih drži $S$, *tačna*, to znači da ona korespondiraju stanjima sredine $E$ koja su nama poznata. Odmah formulišemo I.1a: ukoliko je $S$ opremljen kognitivnim funkcijama koje, za data $\psi$, *konzistentno proizvode ponašanje* $B$ koje zadovoljava njegove ciljeve $G$ u sredini $E$, kažemo da je $S$ kognitivno racionalan. Termin „*konzistentno*“ koristimo umesto reči „*uvek*“ kako bismo kognitivnom sistemu aktera $S$ ostavili mogućnost da nekad pogreši. Pretpostavimo da je kognitivni sistem, tj. skup kognitivnih funkcija aktera $S$, veoma kompleksan i slabo proučen prirodni sistem, koji paralelno mora da zadovolji veći broj ciljeva, u potencijalno složenoj, dinamičkoj sredini, i to oslanjajući se na ograničene resurse (recimo energetske, ili resurse vremena). Smatramo da ne bi bilo pravedno tražiti od $S$ da ostvaruje perfekciju pod takvim uslovima; termin „konzistentno“ onda tumačimo kao da obeležava onu uvek neizvesnu granicu statističke pouzdanosti pod kojom prihvatamo da je neki prirodni proces ustanovljen kao stabilan zahvaljujući njegovom upoznavanju na velikom uzorku posmatranja. Zadovoljićemo se ovako labavim određenjem, svesni toga da matematička statistika nudi daleko preciznije formulacije od naših; toliko vremena ćemo u ovoj raspravi provesti u njenom društvu da je ne želimo uz nas i sada dok postavljamo

fundamentalne koncepte - njih ćemo ionako kasnije prepustiti njenim metodama na elaboraciju. Formulišemo sada I.1b: pod opisanim uslovima da su $\psi$ tačna i opravdana, neka je $S$ opremljen nekim kognitivnim funkcijama koje *ne proizvode konzistentno ponašanje* $B$ tako da ono zadovoljava $G$ u $E$. Termin „konzistentno" uvodimo zbog mogućnosti da kognitivne funkcije aktera ponekad, *makar i sasvim slučajno*, proizvode adekvatno ponašanje. Jasno je da pod ovim uslovima $S$ nećemo smatrati kognitivno racionalnim; njegovo ponašanje $B$ najvećim delom vremena neće otkrivati da on dela u sopstvenom interesu $G$ u odnosu na okolinu $E$. Analiza situacije I.1, pod kojom su $\psi$ tačna i opravdana, obuhvata najjčistije slučajeve racionalnosti i izostanka racionalnosti saznanja koje diskutujemo u ovoj raspravi. Racionalnost saznanja je u njima potpuno određena time da li kognitivne funkcije proizvode ili ne proizvode ponašanja koja zadovoljavaju ciljeve pouzdano informisanog aktera u datoj sredini.

I.2 $S$ ima verovanja $\psi$ koja su *tačna* i *neopravdana*, tako da $\psi$ korespondiraju stanjima sredine $E$, ali je $S$ formirao svoja verovanja $\psi$ kroz neki epistemološki nelegitiman proces. Formulišemo situaciju I.2a, u kojoj kognitivne funkcije na raspolaganju $S$ konzistentno proizvode ponašanje $B$ na osnovu $\psi$ koje zadovoljava njegove ciljeve $G$ u $E$. Pod ovakvim uslovima, *a imajući na raspolaganju samo opservabilno ponašanje B od svih podataka o S*, mi ćemo zaključiti da je $S$ kognitivno racionalan. U situaciji I.2b, pretpostavljamo da kognitivne funkcije ne proizvode ponašanje $B$ koje zadovoljava $G$ u $E$ konzistentno, i ponovo samo na osnovu posmatranja $B$ donosimo odluku da $S$ nije kognitivno racionalan. Interesantno: za sada se čini da *opravdanost verovanja ni na koji način ne određuje* našu odluku o tome da li je saznajni proces u celini racionalan ili nije. Situacija u kojoj $S$ konzistentno formira tačna verovanja $\psi$ koja su neopravdana je svakako egzotična. Spremni smo da se u ovakvim slučajevima oslonimo na nadu da božanstva retko kada pristupaju kognitivnim akterima tako da im pomažu da kroz epistemološki nelegitimne procese konzistentno razvijaju tačna uverenja. Ukoliko, ipak, postoje takvi slučajevi naklonosti, one aktere koji uživaju njihove posledice ćemo nazivati *prorocima I tipa*. Sledeće razmatramo situacije u kojima $S$ raspolaže *netačnim verovanjima* $\psi$ o sredini $E$.

II.1 $S$ ima verovanja $\psi$ koja su *netačna* i *opravdana*, tako da $\psi$ ne korespondiraju stanjima sredine $E$, ali ih je $S$ formirao kroz neki epistemološki legitiman proces. Na primer, $S$ može biti pogrešno informisan o stanjima okruženja $E$ od strane nekog aktera *S'* kome inače veruje. Formulišemo situaciju II.1a u kojoj kognitivne funkcije aktera $S$ u skladu sa verovanjima $\psi$ konzistentno proizvode ponašanje $B$ koje zadovoljava njegove ciljeve $G$ u sredini $E$. Pošto je korespondencija $\psi$ sa $E$

narušena, osmotreno ponašanje $B$ ne zadovoljava ciljeve $G$, i mi donosimo zaključak da $S$ nije kognitivno racionalan, *uprkos tome što su po pretpostavci ovde njegove kognitivne funkcije racionalne*, tj. proizvode ponašanje u skladu sa verovanjima tako da ono zadovoljava postavljene ciljeve u odnosu na ograničenja koje nameće sredina. „*Garbage in - garbage out*“ - poznata sintagma u analizi procesa formiranja verovanja, ovde je na snazi: ukoliko znamo da su verovanja $\psi$ netačna, onda ne možemo da odbacimo kognitivnu racionalnost $S$ na osnovu toga što osmotreno $B$ ne zadovoljava $G$ u $E$. Ali, prvo moramo da znamo da $\psi$ *nisu tačna*, a jedino što imamo na raspolaganju je opservabilno ponašanje $B$. Videćemo da je, pod određenim prepostavkama u savremenim teorijama odlučivanja, na primer, moguće upoznati i subjektivna verovanja $\psi$ samo na osnovu opservabilnog ponašanja $B$. U svakom slučaju, ova situacija opominje na to da nećemo oduzeti nekom $S$ svojstvo kognitivne racionalnosti ako on dela na osnovama netačnih verovanja do kojih je došao na legitiman način. Situacija koju formulišemo sada, II.1b, posebno je složena. U njoj $S$ ima verovanja $\psi$ koja su netačna i opravdana, a opremljen je nekim skupom kognitivnih funkcija koje ne proizvode konzistentno ponašanje $B$ tako da ono zadovoljava $G$ u $E$ na osnovu tih verovanja. Pošto u ovoj situaciji kognitivne funkcije ne proizvode adekvatno ponašanje $B$, odn. za data verovanja $\psi$ proizvode ponašanje koje ne zadovoljava nužno $G$ u $E$, moguće je da osmotrimo ponašanje $B$ koje zadovoljava $G$ u $E$ jer je ono formirano *nepouzdanim* kognitivnim funkcijama na osnovu *netačnih* verovanja $\psi$! Ono što je moguće je *slučajno* „poklapanje“ primene kognitivne funkcije (koja inače ne proizvodi adekvatna ponašanja) sa netačnim verovanjima tako da rezultirajuće ponašanje otkriva kao da $S$ dela da bi zadovoljio $G$ u $E$. Recimo da neko rešava određeni matematički zadatak i konzistentno pogrešno percipira neki broj $a$ kao $-a$. U nekom koraku inferencije koja vodi ka rešenju, taj neko konzistentno pravi grešku koja potire prvobitnu pogrešnu percepciju elementa zadatka i dolazi do tačnog rešenja[1]. Opasnost koja vreba od ovakvog slučaja u analizi racionalnosti saznanja verovatno nije velika: govorimo o nekom kognitivnom akteru koji ima netačna uverenja i raspolaže kognitivnim funkcijama koje ionako ne čine mnogo da razviju ponašanje koje radi u njegovom interesu. Veći deo vremena, tako, ovaj kognitivni akter je krajnje bespomoćan (ili je jednostavno prestao da postoji), ali sigurno su moguće situacije u kojima njegovo ponašanje $B$ svedoči da on dela tako da ostvari svoje ciljeve $G$ u $E$, iako po pretpostavci njegove kognitivne funkcije nisu racionalne, a uverenja jesu pogrešna. Ipak, ako su božanstva sklona da direktnim uticajem *svaki put* koriguju *nepouzdane* kognitivne funkcije aktera $S$ tako da one sada proizvode adekvatna ponašanja iako su verovanja do kojih drži $S$ bila netačna, ostavljeni smo na cedilu, i upoznali slučajeve koje ćemo nazivati

*prorocima II tipa.* Dok proroci I tipa mogu i da ne budu kognitivno racionalni (jer božanstva utiču samo na istinitosnu vrednost njihovih uverenja, ali se ne petljaju u njihove kognitivne funkcije), ponašanje proroka II tipa otkriva da su oni konzistentno kognitivno racionalni (uprkos tome što se, po pretpostavci, njihov kognitivni sistem nalazi u gorem stanju od onog kod proroka I tipa). Još jednom, ni u ovim situacijama se ne čini da opravdanost verovanja $\psi$ ma kako utiče na analizu racionalnosti saznajnog procesa u celini.

II.2 $S$ ima verovanja $\psi$ koja su *netačna* i *neopravdana*, tako da $\psi$ ne korespondiraju stanjima sredine $E$, a $S$ ih je još i formirao kroz neki epistemološki nelegitiman proces. Formulišemo situaciju II.2a u kojoj $S$ raspolaže kognitivnim funkcijama koje konzistentno proizvode $B$ na osnovu $\psi$ tako da ono zadovoljava $G$ u $E$. Pošto $S$ karakterišu netačna verovanja o $E$, njegovo ponašanje $B$ neće zadovoljavati $G$ u $E$, ali mi uprkos tome ne možemo da kažemo da on nije kognitivno racionalan. Još jednom, ne možemo da ne pridamo atribut kognitivne racionalnosti akteru čije kognitivne funkcije rade u njegovom najboljem interesu, ali njegovo ponašanje to ne otkriva jer on dela na bazi pogrešnih uverenja. Ponovimo, savremene teorije o nekim kognitivnim funkcijama poznaju metodologiju kojom je - uz neke prepostavke koje ćemo detaljno diskutovati - moguće steći znanje o neopservabilnim verovanjima $\psi$ aktera $S$. U situaciji II2.b govorimo o akteru $S$ čije kognitivne funkcije ne proizvode konzistentno ponašanje $B$ koje zadovoljava $G$ u $E$; ponovo se nalazimo u situaciji poput II.1b u kojoj nam preti opasnost od proroka II tipa. Naučnicima nenaklona božanstva mogu da utiču na kognitivne funkcije proroka II tipa - koje inače ne proizvode adekvatna ponašanja - tako da konzistentno „poklapaju“ rezultate njihovog rada sa netačnim verovanjima koja on ima *na način da* rezultirajuće ponašanje ipak zadovoljava njegove ciljeve.

Ukoliko poznajemo verovanja nekog kognitivnog aktera i ustanovimo da su ona tačna, te ako taj akter dela u fiksnoj, nepromenljivoj sredini, sa unapred određenim ciljevima koje poznajemo, mi na osnovu njegovog ponašanja možemo da donesemo odluku o tome da li su njegove kognitivne funkcije racionalne ili nisu. Racionalnost koju ustanovljavamo na upravo opisani način nazivamo RACIONALNOST$_1$. Videćemo, kroz rasprave u II i III delu ove teze, da upravo ovako postavljen problem najbolje odgovara načinu na koji se u savremenoj psihologiji tvrdi da su kognitivni sistemi racionalni ili da su oni ograničeno racionalni - što je spor koji predstavlja okosnicu *debate o racionalnosti.* Sud o suštinskoj racionalnosti samih funkcija, međutim, donosimo ako funkcije koje analiziramo konzistentno proizvode ponašanje koje zadovoljava ciljeve aktera u odnosu na data sredinska ograničenja nezavisno od toga ($i$) da li su verovanja tog aktera tačna ili ne, ($ii$) u

kakvom okruženju on dela i (*iii*) kako su određeni ciljevi tog aktera. Dok prva od prethodne dve tvrdnje određuje ono što nazivamo RACIONALNOST$_1$, promenljiva koja uzima vrednost u odnosu na to da li su kognitivne funkcije posmatranog aktera *zaista* racionalne ili nisu određuje koncept RACIONALNOSTI$_2$. Ukoliko se ispostavi da su verovanja kognitivnog aktera netačna, javljaju se problemi u našoj analizi racionalnosti. Ukoliko *znamo* da su njegova verovanja netačna a pri tom imamo konzistentnu empirijsku evidenciju o tome da njegova ponašanja ne otkrivaju da on dela adekvatno, mi možemo da tvrdimo da je on kognitivno racionalan (jer će ispravna primena pogrešnih uverenja u formulisanju ponašanja voditi ka neadekvatnom ponašanju), *ukoliko se prethodno zaštitimo od mogućnosti da njegovo ponašanje ne bude adekvatno iz razloga inherentnih primeni kognitivnih funkcija* odn. kao posledica koincidencije (tj. božanske intervencije kod proroka II tipa). Jedan način da se zaštitimo od takve mogućnosti direktno sledi kauzalnu logiku eksperimenta: ukoliko pri tačnim verovanjima kognitivni akter konzistentno proizvodi adekvatna, a pri netačnim verovanjima neadekvatna ponašanja, on je kognitivno racionalan (osim ako nas božanstva ne varaju sasvim grubo poigravajući se njegovim kognitivnim funkcijama i rezultatima našeg eksperimenta, u kom slučaju ne možemo da znamo ništa o tome jesu li one racionalne ili ne). Ukoliko ovakav eksperimentalni test ne dolazi u obzir - na primer, u slučaju aktera koji uvek ima netačna uverenja - možemo da pokušamo da postuliramo određen nivo koherencije između verovanja ($\psi$) i ponašanja ($B$) kako bismo se zaštitili od takve mogućnosti. Na primer, ukoliko variranjem različitih netačnih verovanja $\psi$ kod tog aktera dobijamo kao ishod različita ponašanja $B$ koja su *regularno povezana* sa tim promenama u verovanjima, saznali bismo bar to da rad njegovih kognitivnih funkcija nije *nasumično* povezan sa vrednostima koje uzimaju njegova uverenja. Uz pretpostavku da bi isti tip regularnosti dobili ako bismo mogli da variramo tačna verovanja $\psi$ istog tog aktera, tj. *da je primena njegovih kognitivnih funkcija konzistentna kada su njegova uverenja tačna i kada nisu*, mogli bismo da ekstrapoliramo naše nalaze u domenu netačnih verovanja u domen tačnih i utvrdimo da li je $S$ kognitivno racionalan.

Međutim, situacije I1.a i I1.b, ukazuju na *čiste* slučajeve analize kognitivne racionalnosti. Ako pouzdano znamo da su verovanja neke osobe tačna, a njena ponašanja konzistentno otkrivaju da ona ne dela u skladu sa svojim ciljevima i osobinama okoline u kojoj deluje (i koju njena tačna uverenja odslikavaju), mi tvrdimo da problem mora da se nalazi u primeni kognitivnih funkcija osobe o kojoj diskutujemo. Na stranu sada stavovi eliminativnih materijalista prema kojima stanja verovanja možda i ne postoje: nama nije poznata dublja intuicija

u izgradnji psihološkog objašnjenja od upravo iznete. Od kliničke psihologije, preko psihologije ličnosti do kognitivne psihologije, upravo opisan rezon omogućava izvođenje zaključaka o racionalnosti ma čijih fundamentalnih mentalnih procesa. U tom smislu, tvrdnja koja predstavlja koncept RACIONALNOSTI$_1$, predstavlja *fundamentalnu pretpostavku* u izgradnji svake kognitivne teorije. Dokle god se krećemo u ontologiji subjekta, njegovih verovanja, ciljeva i okoline sa određenim osobinama u kojoj on mora da dela, priznajući opservabilno ponašanje kao jedini izvor podataka, što je stav koji deli ogromna većina pokušaja da se psihologija tretira kao prirodna nauka[2], mi ne vidimo kako je uopšte moguće izbeći da psihološka analiza neopservabilnih mentalnih funkcija uzme drugačiji oblik do onog koji diktira tvrdnja RACIONALNOSTI$_1$. U kojoj meri i upravo opisana ontologija i tvrdnja RACIONALNOSTI$_1$ opstaju posle oštre kritičke analize kojoj nameravamo da ih podvrgnemo u ovoj raspravi, diskutovaćemo pošto ih podrobno ispitamo u nekim od centralnih oblasti savremene kognitivne psihologije.

Prethodne analize su pokazale da je za analizu racionalnosti saznanja od suštinskog značaja to da li poznajemo verovanja aktera $S$ kao *tačna* ili *netačna*; njihova *opravdanost*, za razliku od njihove istinitosti, kao da ne doprinosi značajno našim odlukama o tome da li je $S$ kognitivno racionalan. Iako problem opravdanosti verovanja možemo više ili manje mirne duše da prepustimo epistemologiji, gde on već decenijama predstavlja centralni predmet rasprave, videćemo kako postepeno usložnjavanje analize racionalnosti saznanja privlači diskusiju procesa *formiranja verovanja* u fokus. U nadi da nas u nastavku rasprave neće proganjati nenaklonjena božanstva koja konzistentno omogućavaju prorocima I tipa da imaju tačna a neopravdana verovanja, ovde se opraštamo od diskusije problema vezanih za opravdanost verovanja. Nasuprot tome, procese formiranja verovanja proučavaćemo u svim delovima rasprave koji slede. Čini nam se da se od epistemologije rastajemo na pošten način; njoj ostavljamo diskusiju *normativnog* pitanja opravdanosti verovanja, dok za nas, u analizi naučnih teorija kognitivne psihologije, zadržavamo pitanje o tome kako neki akter *uopšte* formira svoja verovanja, bez obzira na status njihove opravdanosti.

Ograničenja naše rasprave slede direktno iz formulacije RACIONALNOSTI$_1$. Ciljeve $G$ nekog kognitivnog aktera $S$ ni u jednom trenutku ne dovodimo u pitanje i ne bavimo se time da li su racionalni ili nisu. Rasprava pred nama nije rasprava o racionalnosti, već o racionalnosti saznanja; konceptualna priprema koju smo predstavili planirana je pažljivo da odgovori na potrebe diskusije racionalnosti kognitivnih funkcija kako se ona vodi u debati o racionalnosti već decenijama u kognitivnoj psihologiji i bihejvioralnoj ekonomiji, a u novije vreme i u neurobiologiji.

Postaće jasno tokom rasprave da su ipak neophodne neke minimalne pretpostavke o ciljevima $G$ kognitivnog aktera $S$ da bi rasprava uopšte imala smisla; npr. kao pretpostavka o tome da će kognitivni akter $S$ sa *odgovarajućom motivacijom* bar pokušati *da učini najbolje za sebe* pod ograničenjima koje nameće neka sredina $E$. Tako, kognitivni akteri koji svesno i/ili namerno rade protiv sopstvenih interesa, nisu interesantni za diskusije koje slede.

Smatramo da je za potrebe uvoda u problematiku ove teze najbolje nastaviti sa jednim ilustrativnim primerom. U narednim redovima predstavljamo elementarnu analizu i pregled problema racionalnog odlučivanja, odn. *odlučivanja u uslovima rizika*[3]. Naš cilj je da predstavimo paradigmatičnu problematiku sa kojom se suočavamo u ovoj raspravi. Ne razvijajući još uvek našu metodologiju u potpunosti tokom ove uvodne rasprave o odlučivanju bavićemo se (a) *metateorijskom analizom*, odn. poređenjem različitih teorija analizirajući bitne koncepte koje one koriste na različite načine, (b) *istorijskim pregledom*, kako bismo razumeli razvoj ključnih koncepata, i (c) *metodološkom analizom* koja će nam otkrivati način na koji se konstituišu ključni empirijski fenomeni u diskursima različitih teorija. Analiza racionalnog odlučivanja u ovoj uvodnoj raspravi predstavljaće „model“ naše analize u narednim poglavljima, gde je razvijamo u potpunosti u više oblasti savremene kognitivne psihologije. Verujemo da je ovakav uvod u problematiku teze više primeren od uobičajenog pregleda rezultata iz naučne periodike koji ostavljamo za trenutak kada ćemo prema njima već moći kritički da se odnosimo.

# 1 Naučna zagonetka racionalnog izbora

Diskurs naučne rasprave o ljudskom *odlučivanju u uslovima rizika i neizvesnosti* vekovima prožima fundamentalna rasprava o racionalnosti ljudskog saznanja i ponašanja. Problem odlučivanja, čije prvo egzaktno rešenje - hipotezu očekivane korisnosti - predstavlja Danijel Bernuli još 1738. godine[4] (Bernoulli, 1738), je paradigmatičan primer rasprave o racionalnosti u kojoj su tokom istorije nauke uzeli učešća matematičari, filozofi, ekonomisti, psiholozi, u novije vreme i neurobiolozi. Upravo je tokom proučavanja donošenja odluka u uslovima rizika razvijena cela savremena problematika racionalnosti ljudskog saznanja. U okviru ove diskusije nastao je savremeni spor u kome sa jedne strane nalazimo pristalice normativne racionalnosti i normativnih teorija, a s druge strane pristalice ograničene racionalnosti i deskriptivnih, bihejvioralnih teorija. Šta tačno znači da je neka kognitivna teorija normativa ili deskriptivna, otkrivaćemo postepeno kroz ovu

uvodnu diskusiju; pokazaće se, zapravo, da će veći deo naše rasprave o racionalnosti saznanja direktno ili indirektno biti povezan sa preciznim ustanovljavanjem značenja ovih pojmova.

## 1.1 Hipoteza očekivane korisnosti

Problem odlučivanja u uslovima rizika počinje uočavanjem jednog kockarskog paradoksa. Razni izvori iz teorije verovatnoće i odlučivanja različito objašnjavaju kako je *paradoks Sv. Petrovgrada* (enlg. *The St. Petersburg Paradox*) dobio ime: ili po tome što su ga prvi primetili krupijei u Petrovgradskim kockarnicama u XVI veku (Glimcher, 2004), ili po tome što je klasični rad Danijela Bernulija (Daniel Bernoulli, 1700-1782) u kome je predloženo prvo rešenje paradoksa objavljen u časopisu Carske Akademije nauka u Petrovgradu[5] (Bernoulli, 1738/1954). Sam paradoks je matematički opisao Danijelov rođak Nikolas Bernuli u korespondenciji sa Pjer Rejmon de Montmorom koja je započela 9. septembra 1713. godine. Formulacija problema koja se susreće u najvećem broju današnjih ekspozicija je sledeća: kazino nudi igru u kojoj se baca novčić. Igra traje sve dok novčić ne padne na „pismo“. Sve dok ispada „glava“, igra se nastavlja. Na talon se u početku stavlja 1 dukat, i novčić se baca; ako u prvom bacanju ispadne „pismo“, igra se završava a igrač odnosi talon od jednog dukata. Ukoliko u prvom bacanju ispadne „glava“, iznos na talonu se duplira a novčić se ponovo baca. Svakim bacanjem novčića koje izlazi kao „glava“ iznos na talonu se duplira, tako da igrač čiji novčić izlazi dva puta za redom kao „glava“ i treći put kao „pismo“ osvaja 4 dukata, dok igrač čiji novčić izlazi pet puta kao „glava“ i šesti put kao „pismo“ osvaja 32 dukata. Pitanje koje otkriva paradoks Sv. Petrovgrada glasi: koliku fiksnu cenu bi trebalo da naplaćuje kazino nekome da bi ušao u ovu igru? Svaki put kada igrač sedne za sto da igra ovu igru, koliko bi bila fer cena za učešće u njoj? Formulisano u matematičkim terminima poznatim još od Paskalovog ranog razvoja teorije verovatnoće, pitanje glasi: kolika je *očekivana vrednost igre*, odnosno, kolika je prosečna zarada od ove igre? Pod očekivanom vrednošću igre, kao što samo ime sugeriše, misli se na *prosečnu zaradu* koju ostvaruje igrač u velikom (potencijalno beskonačnom) broju ovakvih ponovljenih igara. Ona se naravno izračunava množenjem određene vrednosti igre ($V$) verovatnoćom ($p$) da se ta vrednost i osvoji, odn. $E = p \cdot V$. Većina ljudi, kada im se postavi ovakvo pitanje, *nudi malu cenu* za ovu igru[6]. Paradoks Sv. Petrovgrada se otkriva kada jednostavnim matematičkim rezonovanjem dođemo do zaključka da opisana igra ima *beskonačnu očekivanu vrednost*, tj. da bi svaki igrač trebalo da bude spreman da uloži ma koliko visok iznos može da priušti da bi u njoj

učestvovao. Naime, na početku igre se na talonu nalazi jedan dukat koji je osvojen sa sigurnošću, te ukoliko u prvom bacanju izađe „pismo“ kojim se igra završava, očekivana vrednost je 1 (verovatnoća, $p$) puta 1 dukat (vrednost, $V$) jednako 1 dukat. Ukoliko u prvom bacanju izađe „glava“, što je događaj sa verovatnoćom $\frac{1}{2}$ u slučaju fer novčića, a u drugom bacanju „pismo“, osvajaju se 2 dukata (jer je iznos na talonu dupliran jednom), sa očekivanom vrednošću $\frac{1}{2} \times 2 = 1$ dukata. Vidimo da

$$E = \sum_{i=1}^{\infty} p_i V_i = 1 \times 1 + \frac{1}{2} \times 2 + \frac{1}{4} \times 4 + \frac{1}{8} \times 8 + ... =$$

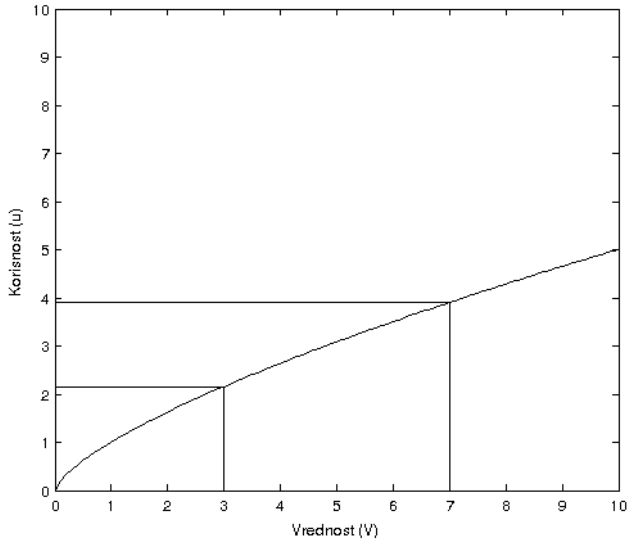
$$= 1 + 1 + 1 + 1 + 1 + ... = \infty. \quad (1)$$

Paradoks Sv. Petrovgrada: u ponudi igre sa beskonačnom očekivanom vrednošću, ljudi su spremni da cenu igre procene kao malu, konačnu vrednost. Rasprava o rešenju ovog paradoksa dovela je nekih od najznačajnijih rasprava u istoriji teorije odlučivanja, ekonomije, a posredno, u novije vreme, i u istoriji psihologije.

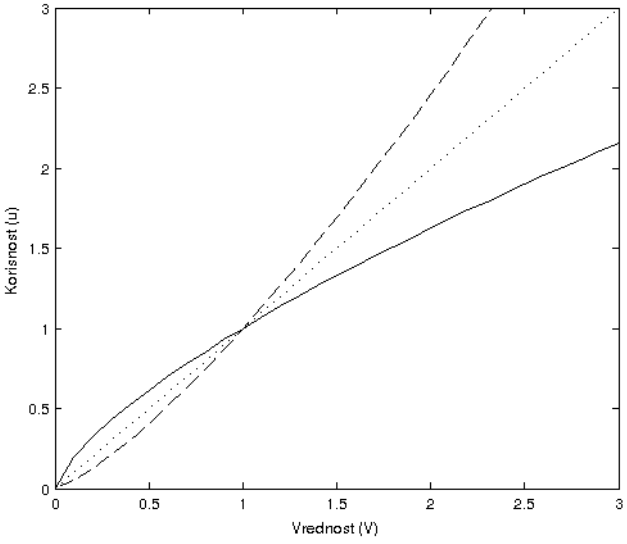
Rešenje koje je ponudio Danijel Bernuli zasniva se na danas dobro poznatoj i prihvaćenoj ideji da ljudi ne ocenjuju dobitke (ili gubitke) na osnovu očekivane vrednosti, već na osnovu *očekivane korisnosti*, gde korisnost u ovim raspravama postaje teorijski termin koji označava psihološku, *opaženu vrednost* neke dobiti (ili gubitka). Možemo slobodno reći da je funkcija korisnosti *psihofizička funkcija u domenu ocene vrednosti*: ona preslikava objektivnu vrednost u opaženu korisnost, $u: V \rightarrow U$, gde u$(\cdot)$ označava funkciju korisnost, $V$ skup vrednosti, a $U$ korisnosti. Hipoteza o očekivanoj korisnosti igra centralnu ulogu u svim potonjim diskusijama racionalnog donošenja odluka. Slika 1a prikazuje jednu tipičnu funkciju korisnosti. Bernulijevo rasuđivanje bilo je sledeće: iako igra kao opisana ima beskonačnu očekivanu vrednost, u ljudskom opažanju, dakle subjektivno, vrednost igre je konačna, jer *preslikavanje vrednosti u korisnost* (opaženu vrednost) *nije linearno.* Funkcija prikazana na Slici 1a. je *nelinearna* i *konkavna.* Bernuli nudi i danas dobro poznato objašnjenje zašto je to preslikavanje nelinearno. Naime, elementarna ekonomska i psihološka intuicija nam svedoči o sledećem: za čoveka čije je ukupno bogatstvo trenutno 10 evra, zarada od 5 evra mora biti značajnija nego za čoveka čije je trenutno ukupno bogatstvo 1000 evra. Prema tome, ukoliko domen funkcije korisnosti obuhvata vrednosti (u valuti u kojoj se dogovorimo da diskutujemo), a kodomen psihološku vrednost odn. korisnost, priraštaj od 5 evra ne može biti isti u slučaju prelaska sa stanja ukupnog bogatstva od 10 na 15 evra kao u slučaju prelaska sa 1000 na 1005 evra; ovaj drugi priraštaj mora biti manji, te funkcija ne može imati

konstantan prvi izvod, i tako ne može biti linearna. Međutim, značaj Bernulijevog objašnjenja seže dublje od ovog intuitivnog nivoa analize, jer ono obuhvata i dobro poznat fenomen *averzije prema riziku*. Averzija prema riziku je fenomen koji se ogleda u činjenici da će ogromna većina ljudi između sigurne dobiti od npr. 50 evra i rizične kocke koja sa verovatnoćom od 50% donosi 100 evra i sa verovatnoćom od 50% ne donosi ništa odabrati sigurnu dobit, iako i sigurna dobit od 50 evra i rizična kocka koju smo opisali imaju istu očekivanu vrednost (očekivana vrednost kocke je u ovom slučaju $E = (\frac{1}{2} \times 0\,EUR) + (\frac{1}{2} \times 100\,EUR) = 50\,EUR$). Fenomen averzije prema riziku se u okviru Bernulijeve hipoteze o očekivanoj korisnosti objašnjava time da, sa konkavnom funkcijom korisnosti kao onom prikazanom na Slici 1a, *korisnost* sigurne opcije od 50 evra jeste veća od očekivane korisnosti *loza* (100 *EUR*, $\frac{1}{2}$; 0 *EUR*, $\frac{1}{2}$). Uopšte, funkcija korisnosti koja pokazuje osobinu averzije prema riziku mora da zadovolji uslov da je *korisnost* sigurnog ishoda one vrednosti koja je jednaka *očekivanoj vrednosti* rizičnog loza *veća od očekivane korisnosti tog rizičnog loza*: $\mathrm{U}(50) > \mathrm{U}(100\ EUR, \frac{1}{2}; 0\ EUR, \frac{1}{2})$ [7], gde sa $\mathrm{U}(\cdot)$ označavamo funkciju korisnosti[8]. Konkavne funkcije korisnosti zadovoljavaju ovaj uslov; konveksne ne. Linearna funkcija korisnosti odgovara fenomenu *neutralnosti prema riziku*. Osoba koja je neutralna prema riziku ne pravi razliku između korisnosti sigurne opcije i rizičnog tiketa koji nosi istu očekivanu vrednost. Familija stepenih funkcija, oblika $u(x) = x^{\rho}$, predstavlja u teoriji odlučivanja veoma popularnu familiju funkcija korisnosti, sa konkavnim funkcijama (averzija prema riziku) za vrednosti $\rho < 1$, konveksnim (sklonost ka riziku) za vrednosti $\rho > 1$, i linearnim (neutralnost prema riziku) za $\rho = 1$ (v. Sliku 1b). I pored elegantne formulacije stava prema riziku (averzije, sklonosti ili neutralnosti) koju stepena funkcija omogućava, ona ni u kom slučaju ne predstavlja nužnu matematičku reprezentaciju tog stava (Wakker, 2008), niti njena primena ostaje bez problematičnih posledica (Wakker, 2010) od kojih ćemo neke diskutovati u ovoj tezi.

Konačno, kako Bernuli stiže do rešenja paradoksa sa kojim smo se upoznali? Zahvaljujući tome što se ljudi u uslovima rizika oslanjaju na očekivanu korisnost, a ne očekivanu vrednost, u donošenju odluka, suma u jednačini (1) koja definiše očekivanu vrednost igre umesto da ima beskonačnu vrednost počinje da konvergira ka nekoj konačnoj vrednosti. Ovo je posledica toga što sa porastom u vrednosti, konkavna funkcija korisnosti daje sve manje i manje (konačne) priraštaje u korisnosti, definišući time još jedan koncept od velikog značaja u istoriji ekonomije, koncept *granične marginalne korisnosti* (engl. *diminishing marginal utility*).




Slika 1a. *Funkcija korisnosti.* Prikazana funkcija korisnosti je stepena funkcija, oblika $u(x) = x^{\rho}$, sa vrednošću $\rho = .70$ koja je odabrana proizvoljno (ali vrednost nije neočekivana u empirijskim istraživanjima). Vidimo da je, sa ovakvom funkcijom korisnosti, dobit od 3 jedinice vrednosti (u valuti u kojoj se dogovorimo da vodimo diskusiju) tek nešto viša od 2 jedinice korisnosti (korisnost bi bila izražena u „unutrašnjim", psihološkim jedinicama), dok je dobit od 7 jedinica vrednosti tek nešto niža od 4 jedinice korisnosti.




Slika 1b. *Funkcije korisnosti koje opisuju averziju, neutralnost i sklonost prema riziku.* Slika prikazuje tri stepene funkcije korisnosti $u(x) = x^{\rho}$ sa vrednošću eksponenta $\rho$. Konkavna funkcija korisnosti sa $\rho < 1$ je data punom linijom, konveksna sa $\rho > 1$ isprekidanom linijom, a linearna funkcija korisnosti sa $\rho = 1$ - tačkastom linijom.

Marginalna granična korisnost novca opisuje fenomen koji bi trebalo da je intuitivno blizak svima: zarada od jednog dolara preko već prethodno zarađenih milion psihološki nije ekvivalentna zaradi od jednog dolara za nekoga ko je prethodno zaradio samo jedan. U drugom slučaju, nečije ukupno bogatsvo je duplirano; u prvom, jedan dolar predstavlja tek kap u moru novca koje neko poseduje. Da sumiramo, uvođenjem konkavne funkcije korisnosti, dakle psihofizičkog preslikavanja objektivne vrednosti u subjektivni, psihološki doživljaj odgovarajuće korisnosti, Danijel Bernuli je (a) bio u stanju da reši paradoks Sv. Petrovgrada, (b) objasni fenomen averzije prema riziku i (c) uvede koncept granične marginalne korisnosti, što su sve fenomeni koji imaju odgovarajuće slike u našoj empirijskoj intuiciji.

Bernulijeva analiza odlučivanja u uslovima rizika predstavlja jedan od prvih slučajeva u istoriji nauke gde je pod jasnu matematičku konstrukciju supsumirano nekoliko empirijskih tvrđenja o ljudskom saznanju i ponašanju. Bernulijeva empirijska tvrđenja još nisu rezultati primene sistematske bihejvioralne metodologije i statističke analize koje su karakteristične tek za XX vek. Međutim, elementarna empirijska intuicija o tome da novac, sa priraštajem, subjektivno počinje da gubi vrednost po jedinici priraštaja, i fenomen averzije prema riziku čine se dovoljno jednostavnim da ljudi Bernulijeve epohe možda nisu ni osećali potrebu da ih sistematski empirijski proverevaju. Možda su ove intuicije toliko jednostavne da je Bernuli mogao da ih uzme i kao *aksiome* za razvoj svoje teorije: osnovne, bezupitne tvrdnje teorije o ljudskim stavovima prema vrednosti[9]? Za razliku od Euklida, koji je aksiomatski pristupio geometriji još oko 300. p.n.e, Bernuli nije našao za shodno da uđe toliko duboko u formalnu izgradnju teorije ljudskog odlučivanja; videćemo uskoro da je taj fundamentalni korak u izgradnji naučne teorije o ljudskom saznanju i ponašanju načinjen oko dve stotine godina kasnije, u XX veku.

Dakle, za razliku od očekivane vrednosti, $E(X) = \sum_{i=1}^{N} p(x_i)V(x_i)$, Bernuli nas savetuje da u evaluaciji svake rizične opcije koristimo očekivanu korisnost:

$$U(X) = \sum_{i=1}^{N} p(x_i)u(x_i) \tag{2}$$

Jednačina očekivane korisnosti (2) igraće prominentu ulogu u svim budućim diskusijama racionalnog izbora.

Na istorijskoj skali problema racionalnog izbora, iz perspektive Bernulijeve analize nalazimo se oko 200 godina pre rada Džona fon Nojmana i Oskara Morgnešterna na razvoju teoriji igara, trenutka u kome ovaj problem doživljava

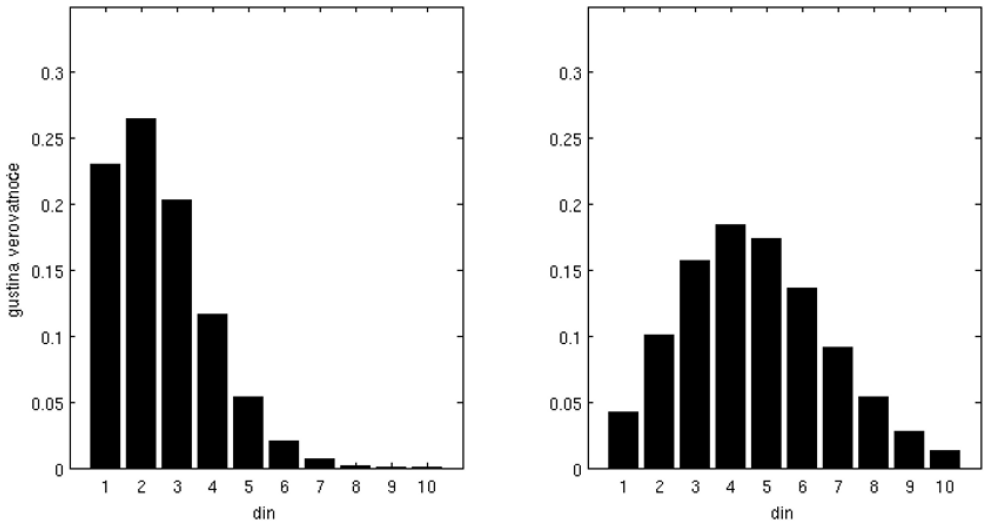
reformulaciju i dobija matematičko ruho u kome ga najčešće proučavamo danas. Začudo, sa izuzetkom retkih autora u istoriji ekonomije, Bernulijev rad nije privlačio značajnu pažnju čitava dva veka (Basset, 1984). Tek polovinom XX veka fon Nojman i Morgenštern reaktualizuju Bernulijev rad i daju savremenoj nauci jasnu perspektivu iz koje ga ova danas sagledava. Iz perspektive rada fon Nojmana i Oskara Morgenšterna nalazimo se na tik od empirijskih uvida francuskog ekonomiste Morisa Alea, čiji će rad uvesti prve *paradokse racionalnog odlučivanja*, fenomene za čijim će objašnjenjem tragati buduće psihološke teorije odlučivanja. Iz iste perspektive, nalazimo se tridesetak godina od teorijske analize sistematskih empirijskih argumenata protiv teorije očekivane korisnosti psihologa Kanemana i Tverskog (Kahneman & Tversky, 1979), analize koja će krajem sedamdesetih godina prošlog veka - bar naizgled - srušiti koncept racionalnog odlučivanja i pridobiti veliku podršku za Sajmonov koncept *ograničene racionalnosti* (Simon, 1955a, 1972).

## 1.2 Aksiomatizacija racionalnog izbora

Džon fon Nojman, američki matematičar, poreklom iz porodice mađarskih Jevreja, smatra se jednim od najkreativnijih i najznačajnijih matematičara XX veka. Njegov doprinos teorijskoj i primenjenoj matematici je ogroman. Jedna od primena koja je fon Nojmana posebno interesovala bila je matematički pristup ekonomiji; poklapanje njegovih interesovanja za strateške probleme, rad na razvoju teorije igara i diskusija sa ekonomistom Oskarom Morgenšternom početkom četrdesetih godina prošlog veka rezultiraće 1944. godine obimnim radom „*Theory of Games and Economic Behavior*“ (von Neumann & Morgnestern, 1944), knjigom koja se danas smatra klasičnim delom ekonomije XX veka (Mirowski, 2002). Od mnogih plodnih doprinosa ove saradnje mi se fokusiramo na onaj koji predstavlja direktan i logičan produžetak Bernulijeve diskusije problema racionalnog izbora. U pitanju je aksiomatizacija racionalnog izbora i razvoj nove teorije očekivane korisnosti, koja danas nosi ime po svojim tvorcima: *fon Nojman-Morgenšternova očekivana korisnost*. Eksplicitan korak u razvoju nove teorije očekivane korisnosti ne nalazi se u originalnom izdanju „*Theory of Games and Economic Behavior*“ već je predstavljen u dodatku drugom izdanju iz 1947.

Izvođenje fon Nojman-Morgenšternove korisnosti u početku nije intuitivno. Mi ćemo u ekspoziciji ove teorije propustiti da predstavimo mnoge tehničke detalje; trenutno je bitno razumeti njenu suštinu i posledice, dok su sami (tehnički složeni) postupci matematičke inferencije u ovom trenutku naše rasprave manje bitni.

Posmatrajmo dve distribucije verovatnoće nad istim skupom vrednosti u dinarima $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ na Slici 2.




Slika 2. *Dve distribucije verovatnoće nad skupom vrednosti* X = *{1, .., 10}*. Funkcija korisnosti fon Nojmana i Morgnešterna definisana je nad ovakvim distribucijama verovatnoće.

Distribucija verovatnoće na levom panelu Slike 2. ima očekivanu vrednost od 2.3 dinara, dok distribucija nad istim skupom vrednosti prikazana na desnom panelu ima nešto veću očekivanu vrednost od 4.7 dinara. Igramo sledeću igru: bira se leva, ili desna distribucija, i onda se u skladu sa specifikovanim verovatnoćama izvlači neki dobitak iz skupa $X$. Distribucija na desnom panelu očigledno nosi veću očekivanu vrednost: ona bi nam u proseku, na duge staze, donela više novca od distribucije na levom panelu. Jedan skup vrednosti, npr. {20 *EUR*, 40 *EUR*} može biti kombinovan sa različitim distribucijama verovatnoće, npr. {.20, .80} ili {.85, .15} da bi se proizveli različiti lozovi, npr. loz koji sa verovatnoćom od 20% donosi 20 evra i sa verovatnoćom od 80% donosi 40 evra, ili loz koji sa 85% donosi 20 evra i sa 15% donosi 40 evra. Fon Nojman i Morgenštern u početku razvoja svoje teorije očekivane korisnosti imaju na umu upravo *problem izbora između lozova* koji sadrže iste vrednosti ali različite raspodele verovatnoća. Naravno, odmah je moguće postaviti pitanje zašto bi neko uopšte postavljao problem odlučivanja *između distribucija verovatnoće*, a ne između samih vrednosti? Upravo to je pomenuti manje intuitivno razumljiv korak u početku razvoja ove teorije.

Uopšte, problem racionalnog izbora možemo da postavimo na sledeći način: neka su $p_1, p_2, ..., p_n$ sve *distribucije verovatnoća* nad istim skupom vrednosti $X$; neka je

$\Delta(X)$ skup svih takvih distribucija, tako da $p_1, p_2, ..., p_n \epsilon \Delta(X)$. Elementi $x \epsilon X$ su same vrednosti (na Slici 2. to je skup $X$ od 1 do 10 dinara). Kako racionalni donosilac odluka bira najbolju distribuciju verovatnoće iz skupa $\Delta(X)$? Na primer, neka su dati loz koji sa verovatnoćom od 20% donosi 20 evra i sa verovatnoćom od 80% donosi 40 evra, (20 *EUR*, .2; 40 *EUR*, .8), i loz koji sa 85% donosi 20 evra i sa 15% donosi 40 evra, (20 *EUR*, .85; 40 *EUR*, .15). Pretpostavljamo da su ovakvi lozovi osnovne jedinice analize i postavljamo pitanje: koje stabilne, konzistentne osobine treba da pokazuje racionalni donosilac odluka pri izboru između ovako definisanih lozova? Kada bira koji od dva ponuđena loza bi pre odigrao, na koje principe se oslanja?

Fon Nojman i Morgneštern matematički pokazuju sledeće: moguće je formulisati jednostavan i intuitivno jasan skup *aksioma racionalnog izbora*, čije značenje se odnosi na *relaciju preferencije*, u oznaci: $\succ$, gde „$p \succ q$“ čitamo kao „$p$ se preferira nad $q$“, i $\succsim$, gde „$p \succsim q$“ čitamo kao „$p$ se preferira ili je indiferentno u odnosu na $q$“, tako da za svakog donosioca odluka koji poštuje te aksiome izvesno postoji funkcija korisnosti $U$, *definisana nad distribucijama verovatnoće u* $\Delta(X)$, preslikavajući svaku distribuciju verovatnoće u neki realni broj, $U : \Delta(X) \to R$, tako da je $p \succsim q$ ako i samo ako je $U(p) \geq U(q)$. Drugim rečima, skup aksioma koji su formulisali fon Nojman i Morgenštern omogućava dokaz da je moguće svaku distribuciju verovatnoće nad jednim istim skupom vrednosti preslikati u neki realan broj, i to preslikavanje je upravo funkcija korisnosti $U : \Delta(X) \to R$, tako da racionalni donosilac odluke bira između dve distribucije verovatnoće na osnovu toga koja od njih ima veću korisnost izraženu tim realnim brojem. Ovu funkciju korisnosti koja pridružuje korisnost ne vrednostima, već distribucijama verovatnoće nad skupovima vrednosti, u daljem tekstu ćemo označavati kao *vNM korisnost*. Ono što se pokazuje u prvom delu izvođenja ove teorije je da postoje aksiomi racionalnog izbora koji garantuju postojanje *funkcije korisnosti nad distribucijama verovatnoće*; ako se podsetimo da je Bernulijeva korisnost definisana nad samim vrednostima, vidimo da postoji fundamentalna razlika između njegovog i pristupa koji uzimaju fon Nojman i Morgenštern. Njihovu funkciju korisnosti zato označavamo kao vNM korisnost da bismo je razlikovali od Bernulijeve korisnosti; prva je definisana nad distribucijama verovatnoće, druga nad samim vrednostima. Međutim, u narednim koracima, fon Nojman i Morgneštern će vratiti dug Bernuliju, pokazujući da njihov konceptualni okvir obezbeđuje i egzistenciju korisnosti nad vrednostima. Pogledajmo sada sistem aksioma[10] koji garantuje postojanje vNM funkcije korisnosti:

A1. *Kompletnost.* Za sve $p$, $q$: ili je $p \succsim q$, ili je $q \succsim p$. Ili više volimo crno vino od belog vina (ili smo indiferentni između ta dva), ili više volimo belo vino od crnog vina (ili smo indiferentni između ta dva).

A2. *Tranzitivnost.* Za sve $p$, $q$, $r$: ako je $p \succsim q$, i $q \succsim r$, onda je $p \succsim r$. Ako više volimo crno vino od belog vina (ili smo indiferentni), i više volimo belo vino od konjaka (ili smo indiferentni), onda više volimo crno vino od konjaka (ili smo indiferentni).

A3. *Kontinuitet.* Za sve $p$, $q$, $r$: ako je $p \succ q \succ r$, onda postoji $\alpha, \beta \in (0,1)$ takvi da $\alpha p + (1-\alpha)r \succ q$ i $\beta p + (1-\beta)r \prec q$. Aksiom kontinuiteta ima tehnički značaj za matematičke inferencije u teoriji odlučivanja i mi nećemo diskutovati njegovu intuitivnu osnovu (iako je ovaj aksiom, kako se najčešće karakteriše u teoriji odlučivanja, lako prihvatljiv posle kraće argumentacije).

A4. *Nezavisnost.* Za sve $p$, $q$, $r$ i bilo koje $\alpha \in [0,1]$: $p \succsim q$ ako i samo ako je $\alpha p + (1-\alpha)r \succsim \alpha q + (1-\alpha)r$. Ukoliko saznamo da je početna preferencija neke osobe takva da više voli plavuše od brineta, a zatim joj ponudimo (a) loz koji joj sa verovatnoćom $\alpha$ obezbeđuje izlazak u grad sa plavušom, a sa 1-$\alpha$ sa crvenkosom devojkom, i (b) loz koji joj sa verovatnoćom $\alpha$ obezbeđuje izlazak u grad sa brinetom, a sa 1-$\alpha$ sa crvenkosom devojkom, ta osoba će odigrati loz (a) pre nego loz (b), što je konzistentno sa njenom početnom pretpostavljenom preferencijom *plavuše* $\succsim$*brinete*. Uskoro ćemo otkriti da kritičko preispitivanje ove intuiciji bliske tvrdnje vodi pravo u pakao poznat pod oznakom istorije teorija odlučivanja druge polovine XX veka.

(u A1-A4. se podrazumeva da su $p, q, r \in \Delta(X)$, gde je $\Delta(X)$skup svih distribucija verovatnoće nad skupom vrednosti $X$).

U razvoju vNM korisnosti, dokazuje se na samo da iz A1-A4. sledi postojanje funkcije korisnosti $U : \Delta(X) \to R$, već i obrnuto, da ukoliko racionalni donosilac odluka ima takvu funkciju korisnosti, onda svi njegovi izbori zadovoljavaju aksiome A1-A4. To znači da vNM funkcija korisnosti *reprezentuje* aksiome racionalnog izbora, i da su te dve konceptualne konstrukcije (zadovoljavanje aksioma i egzistencija vNM funkcije korisnosti) *ekvivalentne*. Dakle, donosilac odluka čije preferencije - koje možemo da posmatramo u nekom kontrolisanom eksperimentu sa izborima u uslovima rizika - zadovoljavaju ovakav sistem aksioma se ponaša *kao da* reprezentuje te preferencije nekom funkcijom korisnosti.

Svi fon Nojman-Morgnešternovi aksiomi su jednostavni i intuitivno jasni, iako neki najčešće zahtevaju diskusiju pre usvajanja. Prva dva aksioma tvrde da su

preferencije *asimetrične* (A1, ne možete u isto vreme voleti više crno od belog i belo od crnog vina) i *tranzitivne* (A2, ako više volite Betovena od Hendla, Hendla od Vagnera, onda više volite Betovena od Vagnera). Aksiom kontinuiteta je, kao što smo rekli, više od tehničkog značaja, ali se i njegova intuitivna osnova demonstrira jednostavnom argumentacijom. Diskusiju aksioma nezavisnosti, težinu čijih posledica smo nagovestili prilikom njegovog uvođenja, ostavljamo za kasnije. Finalni korak u izvođenju ove teorije očekivane korisnosti je dokaz da pod ovim uslovima važi da je *korisnost neke distribucije verovatnoća tj. nekog loza jednaka linearnoj kombinaciji korisnosti samih elemenata* (korisnosti svih $x \in X$ na lozu ponderisanih odgovarajućim verovatnoćama). Izraženo manje formalnim jezikom: korisnost celog loza, npr. ($x$,$p$;$y$,1-$p$) jednaka je sumi korisnosti vrednosti koje on nosi ponderisanih odgovarajućim verovatnoćama: $U(x,p;y,1\text{-}p) = p \cdot u(x) + (1\text{-}p) \cdot u(y)$. Primetimo da ova tvrdnja ima potpuno isti oblik kao jednačina (2).

Korisnost nekog elementa $x \in X$, u oznaci $u(x)$, definiše se kao korisnost cele distribucije verovatnoće nad $X$ kada je $p(x)=1$, tj. kada ta distribucija sadrži samo $x$: fon Nojman i Morgenštern su tako preko svoje vNM funkcije korisnosti nad distribucijama verovatnoće pokazali da njihova aksiomatika racionalnog izbora povlači i postojanje funkcije korisnosti nad vrednostima - tačno kao što je to tvrdio i Bernuli 1738. Za fon Nojmana i Morgenšterna, osnove odlučivanja se nalaze u analizi problema izbora između lozova; konceptualno, ne tretiraju se lozovi kao da su sastavljeni od sigurnih ishoda ponuđenih sa određenim verovatnoćama, već se sigurni ishodi tretiraju kao lozovi koji sa verovatnoćom 1 donose određeni ishod. Intuitivno, vNM funkcija korisnosti, koja omogućava izbor između lozova, je nešto što bismo pre tražili kao *rezultat* teorije racionalnog izbora, polazeći od Bernulijeve korisnosti $u(x)$ nad samim vrednostima. Intuiciji deluje bliže poći od jednostavnijeg ka složenijem u rešavanju problema: ako bismo razumeli osnove korisnosti $u(x)$ definisane nad vrednostima (ovo definiše Bernulijeva funkcija korisnosti), i razumeli zašto i kako ih treba kombinovati sa verovatnoćama $p$, razumeli bismo analitički i zašto se očekivana korisnost lozova računa kao u jednačini (2). Odlučivanje između lozova deluje kao složeniji problem od poređenja korisnosti $u(x)$ prostih ishoda $x \in X$. Kontraintuitivni put koji su odabrali fon Nojman i Morgneštern, odn. da prvo aksiomatizuju vNM korisnost $U(x)$ nad distribucijama verovatnoće (tj. nad lozovima), a onda pokažu da i Bernulijeva $u(x)$ nad vrednostima sledi iz istog konceptualnog okvira, potez je one ingenioznosti koja pokazuje koliko je decentracija od onoga što deluje kao centralni aspekt problema često suštinska

za njegovo rešenje. Za razliku od Bernulijeve rane analize problema racionalnog odlučivanja, fon Nojmanova i Morgenšternova analiza problema se sprovodi na formalno dubljem nivou: aksiomatizacijom tako bitne osobine kao što je sposobnost racionalnog izbora u uslovima rizika, oni su postavili temelje na kojima su nauke o čoveku, makar i samo na terenu teorije, prvi put u istoriji mogle da se ravnopravno porede sa prirodnim naukama.

Teorija očekivane korisnosti fon Nojmana i Morgenšterna nas ne obavezuje na određenu, specifičnu formu funkcije korisnosti. Setimo se da je Bernuli predlagao funkciju korisnosti sa precizno određenim osobinama (nelinaerna, konkavna logaritamska funkcija) koje omogućavaju objašnjenje nekih elementarnih intuicija o ljudskom odlučivanju. Iz teorije fon Nojmana i Morgenšterna slede samo neke globalne osobine koje zadovoljavaju različite specifične funkcije korisnosti, uključujući i Bernulijevu logaritamsku funkciju, i stepenu funkciju koju smo iskoristili na Slici 2. Na primer, funkcija korisnosti $u(x)$ je jedinstvena u odnosu na izbor skale i koordinatnog početka, odnosno ako neka $u(x)$ reprezentuje određeni skup preferencija, onda i $v(x)$ koja je njena pozitivna linearna transformacija, $v(x) = au(x) + b$, $a > 0$, reprezentuje isti taj skup preferencija. Pitanje specifične, empirijski tačne forme funkcije korisnosti posle teorije fon Nojmana i Morgenšterna ostavljeno je kao otvoreno i u teorijskom, i u empirijskom smislu. Njihov razvoj aksiomatske, formalne teorije odlučivanja koja obuhvata Bernulijevu hipotezu kao specijalan slučaj tako pokazuje bitno svojstvo generalizacije prethodnih teorija i odslikava kontinuitet naučne rasprave o ovom fenomenu. Priroda ove formalne teorije odlučivanja koja opisuje način na koji odluku *treba* da donese donosilac odluka u uslovima rizika ukazuje na njen *normativni karakter*, a racionalnost koja je aksiomatizovana u njenim osnovama nazivamo normativnom racionalnošću.

Činilo se da su temelji zasnovani čvrsto: polovinom XX veka, dokaz egzistencije funkcije korisnosti, i činjenica da on počiva na elegantnoj aksiomatizaciji racionalnosti, nisu mogli da izgledaju više ubedljivo. U sintezi sa Bernulijevim zapažanjima, rad fon Nojmana i Morgnešterna zaokruživao je problem racionalnog izbora u koherentnu, egzaktnu celinu. Nauke koje se odnose na ljudsko saznanje i ponašanje napredovale su linijom matematizacije i supsumiranja prethodnih teorija pod nove, generalnije teorije, koja je podsećala na progres kakav su imale prirodne nauke. Međutim, još uvek nije došlo do testiranja ovakvih teorija u odnosu na sistematske eksperimentalne nalaze. Skoro uporedo sa razvojem formalne teorije odlučivanja u ekonomiji i matematici, psihologija je diskutovala bihejvioristička

teoriju, koju će uskoro napustiti, ali tek pošto u sebe čvrsto ugradi stroge bihejviorističke metodološke norme. Nekoliko godina posle objavljivanja teorije očekivane korisnosti, psiholozi i ekonomisti su počeli da se suočavaju sa empirijskim nalazima koji su upućivali na neophodnost promena u samim temeljima elegantne teorije racionalnog izbora.

## 1.3 Paradoksi racionalnog izbora

Podsetimo se na trenutak sledećeg aksioma teorije racionalnog izbora:

A4. *Nezavisnost.* Za sve $p$, $q$, $r$ i bilo koje $\alpha \in [0,1]$: $p \succcurlyeq q$ ako i samo ako je $\alpha p + (1-\alpha)r \succcurlyeq \alpha q + (1-\alpha)r$.

Aksiom nezavisnosti, koji se u nekim teorijskim tretmanima teorije odlučivanja naziva aksiomom supstitucije (Kahneman & Tversky, 1979), predstavlja aksiom od najvećeg značaja u diskusijama teorije očekivane korisnosti. Interpretacija njegove apstraktne forme je sledeća: ukolike dve rizične opcije, $p$ i $q$, za koje važi $p \succcurlyeq q$, i koje su obe dostupne donosiocu odluka sa verovatnoćom $\alpha$, dopunimo istom rizičnom opcijom $r$, sa verovatnoćom 1-$\alpha$, to ne sme da utiče na već ustanovljenu preferenciju $p$ nad $q$, odnosno da je $p \succcurlyeq q$. Primer: ukoliko neka osoba više voli jabuke od breskvi, a nađe se u situaciji izbora u uslovima rizika gde joj se nude opcije ($a$) dobitka jabuke sa verovatnoćom $\alpha$ i kruške sa verovatnoćom *1-*$\alpha$, ili ($b$) breskve sa verovatnoćom $\alpha$ i kruške sa verovatnoćom *1-*$\alpha$, ta osoba će uvek preferirati opciju ($a$) nad opcijom ($b$), jer se one razlikuju samo po tome da se sa istom verovatnoćom $\alpha$ osvoji jabuka (opcija $a$) ili breskva (opcija $b$). Pošto je već ustanovljeno da ta osoba ima preferenciju *jabuke* $\succ$ *breskve*, dodavanje jednoj i drugoj opciji podjednako verovatne mogućnosti osvajanja *kruške* ne sme da utiče na ovu početnu preferenciju. Sasvim je svejedno da li se donosiocu odluka za kog važi $p \succcurlyeq q$ uz rizične $p$, $q$ nudi dopunska mogućnost osvajanje letovanja, kolekcije filmova ili sportskog automobila - minimalni uslov racionalnosti je da to ne utiče na njegove početne preferencije. *Aleov paradoks*, koji sada diskutujemo, protivreči ovoj intuiciji.

*Aleov paradoks* (engl. *The Common Consequence Paradox*). Posmatrajmo sada sledeći primer (Allais, 1953):

Primer 1.

*Opcija A*: sa sigurnošću (100%) dobitak od 100 miliona dolara

*Opcija B*: sa 89% dobitak od 100 milion dolara, sa 10% dobitak od 500 miliona dolara, i sa 1% ništa (0 dolara)

*Opcija A1*: sa 89% ništa (0 dolara) i sa 11% dobitak od 100 miliona dolara

*Opcija B1*: sa 90% ništa (0 dolara) i sa 10% dobitak od 500 miliona dolara

Francuski ekonomista Moris Ale postavio je ovakav i slične zadatke 1951. i 1952. u sklopu rada za dve značajne ekonomske konferencije[11] i našao da je $A$ (sigurni dobitak od 100 miliona dolara) dominanti izbor između $A$ i $B$, dok je $B1$ (90% ništa i 10% dobitak od 500 miliona dolara) dominanti izbor između $A1$ i $B1$. Potonji empirijski rad sa sličnim zadacima dodatno je utvrdio nalaze do kojih je došao Ale. Posmatrajmo sada lozove prikazane u gornjem primeru na sledeći način:

Tabela 1. Originalna formulacija Aleovog paradoksa u razvijenoj formi.

| | 1% | 10% | 89% |
|---|---|---|---|
| *A* | 100M $ | 100M $ | 100M $ |
| *B* | 0 $ | 500M $ | 100M $ |
| *A1* | 100M $ | 100M $ | 0 $ |
| *B1* | 0 $ | 500M $ | 0$ |

U ovakvom prikazu Aleovog paradoksa - koji je prvi iskoristio Leonard Džimi Sevidž (Savage, 1954/72) - jasno je moguće izdvojiti zajedničke elemente obe igre posle čega kršenje aksioma nezavisnosti postaje očigledno. Igra (*A1*,*B1*) nastaje oduzimanjem podjednako verovatne opcije osvajanja 100 miliona dolara od opcija u igri (*A*,*B*).Ono što vidimo kada opcije iz primera ispišemo na ovaj način je da se opcije *A* i *A1*, *B* i *B1*, razlikuju samo po tome što je opcijama *A* i *B* u odnosu na *A1* i *B1* dodata mogućnost osvajanja 1 miliona dolara sa verovatnoćom od 89%. Prema aksiomu nezavisnosti, ispitanici koji u prvom slučaju tvrde $A \succ B$, u drugom slučaju moraju da tvrde $A1 \succ B1$, ali kao što smo videli to nije slučaj. Aleov paradoks predstavlja grubo i očigledno *kršenje aksioma nezavisnosti* koji tvrdi da dodavanje jednako verovatnog ishoda dvema rizičnim opcijama ne sme da utiče na početne preferencije. Ovakav empirijski nalaz ukazuje na to da *ljudsko odlučivanje empirijski odstupa od normativne teorije*, i upućuje na izgradnju nove ili korekciju postojeće teorije tako da ova ispuni *uslov deskriptivne validnosti*, tj. da omogući tačan opis ponašanja u odlučivanju. Za razliku od normativnih teorija koje

opisuju *kako treba* donositi valjane odluke, deskriptivne teorije sebi uzimaju za cilj da objasne empirijsko odlučivanje, koje, kao što pokazuje Aleov paradoks, ne mora biti adekvatno opisano normativnom teorijom.

Pre nego što usvojimo Aleov empirijski nalaz i sagledamo neke posledice po teoriju racionalnog izbora do kojih je doveo potrebno je da naglasimo problem koja će nas pratiti tokom cele diskusije. Naime, nalaz poput Aleovog, bez obzira koliko puta bio repliciran, nije nešto što je moguće dobiti uvek i pod svim uslovima. Variranja u eksperimentalnim rezultatima koja su posledica načina na koji se opcije prikazuju ispitanicima ili promene u drugim eksperimentalnim uslovima fenomen su koji će pratiti razvoj cele debate o racionalnosti; diskusija ovih variranja biće od centralnog značaja za našu analizu racionalnosti saznanja. Rid diskutuje rezultate Konliska (Conlisk, 1989, Read, 2009) koji je ispitivao više varijanti Aleovog eksperimenta, i došao do zaključka da kada se struktura Aleovog paradoksa prikaže tako da se ispitanicima ukaže na zajedničke elemente u obe situacije, dakle:

*Loz I*: Siguran dobitak od 100 miliona dolara

*Loz II*: 1/11 šansi da se ne osvoji ništa; 10/11 šansi da se osvoji 500 miliona dolara

a onda ove skraćene formulacije (*Loz I* i *Loz II*) upotrebe u formulaciji Aleovog paradoksa, nestaju sistematska odstupanja od racionalnog izbora koja je zabeležio Ale, i isti procenat ispitanika bira $A$ nad $B$ kao i $A1$ nad $B1$ iz gornjeg primera. Međutim, isti autor (Read, 2009) u diskusiji Aleovog paradoksa navodi da je ovaj empirijski nalaz robustan i značajan u tom smislu što neki autori (Slovic & Tversky, 1974, prema Read, 2009) tvrde da je ispitanike teško ubediti u to da krše očiglednu racionalnu normu kao što je aksiom nezavisnosti čak i pošto im se paradoks prikaže u razvijenoj formi (Tabela 1) i objasni. Ovakva odstupanja od predikcija teorija odlučivanja koja nastaju kao posledica *promene opisa*, formulacije rizičnih tiketa koje ispitanici porede, nazivaju se kršenjem *deskriptivne invarijantnosti*.

Odmah diskutujemo neke metodološke karakteristike empirijskog istraživanja u ovoj oblasti. S jedne strane, suočavamo se sa pitanjem replikacije eksperimentalnih nalaza u funkciji variranja eksperimentalnih uslova kao što je način prezentacije stimulusa ispitanicima. Možemo da postavimo pitanje *da li se isti empirijski fenomen* (Aleov paradoks u ovom slučaju) *konstituiše nekom metodom* $M_1$ *i nekom drugom metodom* $M_2$. Ukoliko uzmemo pomenuti Konliskov nalaz kao merodavan, jasno je da objašnjenje Aleovog paradoksa može da pođe od (okvirne) psihološke pretpostavke o distribuciji pažnje prema kritičnim informacijama

(verovatnoćama i vrednostima ishoda) u funkciji načina na koji su predstavljene te informacije. S druge strane, suočavamo se sa nalazom koji nam govori o suštinskom narušavanju racionalnosti definisane aksiomatskom teorijom racionalnog izbora, odn. otvorenim neslaganjem ispitanika u nekim istraživanjima da je Aleov paradoks uopšte paradoks i posle objašnjenja načela racionalnog izbora. Metodologija kojim se dolazi do ovog drugog nalaza sigurno nije eksperimentalne prirode i zahteva komplikovaniju komunikaciju sa ispitanicima od one koja je moguća u eksperimentalnim paradigmama. Pojave empirijskih rezultata koji su kontradiktorni u odnosu na normativne osnove racionalnog izbora nastale usled *promena eksperimentalnih procedura* nazivaju se *kršenjima proceduralne invarijantnosti*.

Ovakve situacije otvaraju mnoga pitanja: (a) da li je Aleov paradoks uopšte empirijski nalaz racionalnog izbora ili je posledica manipulacije pažnjom ispitanika, (b) da li je moguće izgraditi teoriju odlučivanja koja će biti deskriptivno validna (uspešno objašnjavati empirijske nalaze) bez inkorporacije psiholoških mehanizama koji nisu specifični samo za odlučivanje (npr. pažnja), (c) kakav je onda smisao takve teorije u odnosu na pitanje normativne racionalnosti, i (d) kakav je odnos između aspekata teorije koji obezbeđuju njenu normativnu validnost, sa jedne, i njenu deskriptivnu validnost, sa druge strane? Uprkos tome što je Tverski još 1969. godine eksperimentalno demonstrirao neka kršenja tranzitivnosti (Tversky, 1969, A2. u pregledu aksiomatike fon Nojmana i Morgenšterna, up. sekciju 1.2) zahtev da teorija odlučivanja poštuje minimalne normativne uslove ostaje na snazi za većinu ekonomista kojima je od suštinskog značaja njen normativni aspekt, odn. mogućnost dolaženja do preporuke o tome šta je najbolje što treba učiniti u nekoj situaciji. Da li, u tom slučaju, ekonomisti, matematičari i psiholozi razvijaju istu teoriju odlučivanja, ili samo koriste isto ime za teorije koje u funkciji različitih potreba mogu biti različitog sadržaja i prirode?

Situacija se dodatno komplikuje kada razmislimo o tome koliko su ovde predstavljeni problemi izbora u uslovima rizika *jednostavni* i postavimo biološki motivisano pitanje zašto bi uopšte priroda razvila klasu kognitivnih sistema koji *sistematski odstupaju* od tako jednostavnog optimalnog rešenja za probleme koji su veoma česti u ekološkom okruženju čoveka? Radovi nekih neurobiologa i ekologa u oblasti ponašanja životinja (kognitivnih sistema izvesno jednostavnijih od ljudskih) ukazuju na to da ono „kao da“ zadovoljava postulate racionalnog izbora u uslovima rizika[12]. Zašto bi proces evolucije kulminirao u razvoju jedne vrste koja sistematski

odstupa od jednostavnog optimalnog rešenja za problem od ogromnog značaja u adaptaciji? Ovo su već neka od suštinskih pitanja na koje ćemo pokušati da damo odgovor u našoj diskusiji racionalnosti saznanja.

Ustanovljavanje empirijskih nalaza poput Aleovog paradoksa i raznih varijacija koje su korišćene u potonjim empirijskim istraživanjima samo je početak problema sa kojima se suočava aksiomatska teorija racionalnog izbora. Uvešćemo sada u diskusiju još neke empirijske fenomene koji svedoče o sistematskim odstupanjima od normativnog okvira.

*Intrazitivnost preferencija.* Ukoliko neka osoba više voli džez od klasične muzike, a klasičnu muziku od kantrija, po tranzitivnosti preferencija očekujemo da ona više voli džez od kantrija; u suprotnom, ta osoba u odlučivanju krši principe normativne racionalnosti. Aksiom tranzitivnosti predstavlja toliko jednostavnu psihološku intuiciju da mu nikakva dopunska interpretacija nije potrebna. Međutim, kršenja ovog aksioma u odlučivanju u uslovima rizika postaju transparentna i česta čim se sa problema odlučivanja između ishoda definisanih na samo jednoj dimenziji pređe na ishode definisane na više dimenzija. Tverski pokazuje da u takvim situacijama ispitanici ne poštuju tranzitivnost preferencija čak i kada se aksiom tranzitivnosti oslabi do probabilitisčke forme (Tversky, 1969). Definišimo da je $P(x,y)$ verovatnoća sa kojom neki ispitanik bira $x$ nad $y$ kada su dostupni $x$ i $y$ u izboru, a $P(y,x)$ verovatnoća da u istoj situaciji bira $y$ nad $x$. Probabilistička forma ovog aksioma, koja se još naziva *slabom stohastičkom tranzitivnošću* (engl. *Weak Stochastic Transitivity*, WST), zamenjuje determinističu tvrdnju A3: za sve $p$, $q$, $r$: ako je $p \succcurlyeq q$, i $q \succcurlyeq r$, onda je $p \succcurlyeq r$ tvrdnjom A3': za sve $p$, $q$, $r$: ako je $P(p,q) > \frac{1}{2} \wedge P(q,r) > \frac{1}{2} \Rightarrow P(p,r) > \frac{1}{2}$ (jasno je da se u ovom slučaju preferencija definiše kao $p \succcurlyeq q \Leftrightarrow P(p,q) \geq \frac{1}{2}$). Vidimo da je u ovoj verziji dopušteno da izbori ne budu deterministički tranzitivni, ali se ograničava da sa verovatnoćom od najmanje $\frac{1}{2}$ osoba poštuje tranzitivan izbor (otud ime *stohastička* tranzitivnost). Tverski je jednostavnim eksperimentima izbora između rizičnih lozova oblika $(x,p)$, gde je $x$ vrednost, a $p$ verovatnoća osvajanja, pokazao da većina ispitanika ne poštuje princip slabe stohastičke tranzitivnosti (Tversky, 1969).

*Kaneman i Tverski o verovatnoći i izvesnosti.* U seriji izvanrednih istraživanja odlučivanja u uslovima rizika koja kulminira objavljivanjem rada „*Prospect Theory: An Analysis of Decision under Risk*“ 1979. godine u časopisu „*Econometrica*“, Kaneman i Tverski su ustanovili više sistematskih empirijskih odstupanja od teorije očekivane korisnosti (Kahneman & Tversky, 1979). Pogledajmo sledeći primer

(prema Kahneman & Tversky, 1979):

Primer 2.

*Opcija A*: 6000$ sa verovatnoćom .45
*Opcija B*. 3000$ sa verovatnoćom .90

*Opcija A1*: 6000$ sa verovatnoćom .001
*Opcija B1*. 3000$ sa verovatnoćom .002

Očekivane vrednosti $A$ i $B$ su iste (2700$), kao i očekivane vrednosti *A1* i *B1* (6$). U prvom slučaju, većina ispitanika bira opciju $B$, koja nosi veću verovatnoću dobitka, dok u drugom slučaju, većina ispitanika bira opciju *A1*, koja nosi veću vrednost. Teorija očekivane korisnosti nema mehanizme kojima može da objasni ovakve preferencije. Ipak, moramo da primetimo jednu metodološku prepreku u konstituisanju ovog empirijskog nalaza: naime, ukoliko formalizam teorije očekivane korisnosti predviđa indiferentnost između $A$ i $B$, kao i između *A1* i *B1*, zašto indiferencija nije ponuđena kao moguć odgovor na izborno pitanje u ovom istraživanju, već je korišćena metoda prinudnog izbora (engl. *forced choice*, ili $A$ ili $B$)? Ukoliko se jedno predviđanje teorije eliminiše iz skupa indikatora na osnovu kojih registrujemo ponašanje ispitanika, odnos između tog registrovanog ponašanja i teorijskog (hipotetskog) konstrukta koji konstitušemo tom metodom neće više biti isti; ostaje otvoreno pitanje da li takvom metodom konstituišemo neki drugi teorijski konstrukt, odn. da li ispitanici donose odluku na osnovu parametara teorije odlučivanja (koja predstavlja hipotezu koja se testira) ili na osnovu nekih drugih parametara.

Sledeći primer dodatno ilustruje ovakva sistematska odstupanje od aksioma nezavisnosti koja se zajedničkim imenom nazivaju *paradoksom zajedničke proporcije,* engl. *The Common Ratio Paradox,* prema Kahneman & Tversky, 1979, koji navode da je Moris Ale takođe prvi konstruisao ovaj nacrt):

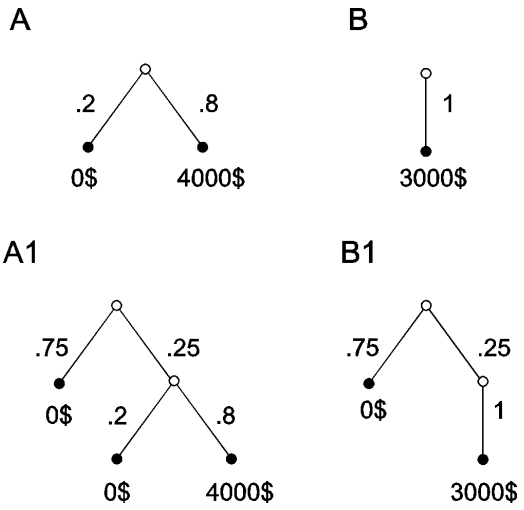
Primer 3.

*Opcija A*: 4000$ sa verovatnoćom .8, 0$ sa verovatnoćom .2
*Opcija B*. Sigurnih 3000$

*Opcija A1*: 4000$ sa verovatnoćom .2, 0$ sa verovatnoćom .8
*Opcija B1*. 3000$ sa verovatnoćom .25

Većina ispitanika u ovakvom zadatku bira opciju $B$ u prvom slučaju i opciju *A1* u drugom slučaju. Međutim, kada razvijemo prikazane opcije na način prikazan na

Slici 3, vidimo da ovakve preferencije predstavljaju kršenje aksioma nezavisnosti.




Slika 3. *Paradoks zajedničke proprocije u razvijenoj formi* (prema Kahneman & Tversky, 1979). Objašnjenje u tekstu.

Empirijske nalaze na osnovu izbora ispitanika u primerima 1. (Aleov paradoks) i 3. (paradoks zajedničke proporcije) Kaneman i Tverski diskutuju pod zajedničkim imenom *efekta izvesnosti* (Kahneman & Tversky, 1979). U oba primera, ispitanici donose odluke pokazujući averziju prema neizvesnosti (izbegavanje rizičnih dobitaka - onih koji su tek verovatni, ne izvesni), odn. sklonost ka izvesnosti. Vrednost od 3000$ u džepu deluje kao bolji izbor od neizvesnih 4000$, čak i sa visokom verovatnoćom osvajanja od 80% (primer 3), iz čega sledi da je skok u šansama za osvajanje ishoda sa 80% na sigurnih 100% na neki način veoma uticajan. Dalje, u odnosu na promene u verovatnoćama ishoda koje Kaneman i Tverski variraju u svojim nacrtima, dolaze do sledećeg zaključka: svi problemi u kojima se konstatuju promene preferencija tipa Aleovog paradoksa i paradoksa zajedničke proporcije, dakle promene koje impliciraju kršenja aksioma nezavisnosti, mogu da se podvedu pod zajedničku formu, odn. zapažanje da ako je izbor između ($y$, $pq$) i ($x$,$p$) indiferentan, onda je ($y$,$pqr$) preferirano u odnosu na ($x$,$pr$), $0 < p,\ q,\ r < 1$. Pojasnimo ovo: ukoliko je prvi rizični ishod, koji nudi vrednost $y$, ponuđen sa nekom verovatnoćom koja je samo frakcija verovatnoće sa kojom je ponuđen drugi rizični ishod $x$ ($y$ se nudi sa $pq$, a $x$ sa $p$, gde je $pq$ očigledno frakcija od $p$), i neka osoba je indiferentna u izboru između ova dva rizična ishoda (što je po očekivanoj korisnosti moguće samo ako je $u(x) < u(y)$), onda će ta osoba preferirati rizični ishod ($y$,$pqr$) u odnosu na rizični ishod ($x$,$pr$), u kome su oba ishoda ponuđena sa verovatnoćama koje su još manje (ali proporcionalne) frakcije prethodnih verovatnoća ($pqr$ je

frakcija $pq$, $pr$ je frakcija $p$; Kahneman & Tversky, 1979). Kako se verovatnoće osvajanja neka dva ishoda u istom lozu proprocionalno smanjuju, tako empirijske odluke ispitanika otkrivaju da u ljudskoj percepciji razlika između tih verovatnoća postaje sve manja. Način na koji ova empirijska generalizacija motiviše određene promene u normativnoj teoriji biće uskoro detaljnije razmotren.

*Različiti stavovi prema riziku u slučajevima dobitaka i gubitaka.* Jedan od najvažnijih empirijskih doprinosa Tverskog i Kanemana jeste nalaz poznat kao *efekat refleksije*. Efekat refleksije predstavlja fundamentalan empirijski fenomen koji je neuklopljiv u okvir teorije očekivane vrednosti pošto podrazumeva *različit tretman dobitaka i gubitaka*. Kao što ćemo videti uskoro, efekat refleksije nije jedina razlika u tretmanu dobitaka i gubitaka koja se javlja u empirijskom odlučivanju. Efekat refleksije se sastoji u sledećem: sva opisana odstupanja preferencija od racionalnog izbora koji definiše teorija očekivane korisnosti *menjaju smer kada se izbor vrši između gubitaka* u odnosu na izbor između dobitaka. Na primer, u izboru između opcija iz prethodnog primera, većina ispitanika koja između $A$: 4000$ sa verovatnoćom .8, 0$ sa verovatnoćom .2, i $B$: sigurnih 3000$, bira $B$, počinje da bira $A$ kada se izbor reformuliše u terminima gubitaka ($A$': gubitak od 4000$ sa verovatnoćom .8, gubitak od 0$ sa verovatnoćom .2; $B$': sigurni gubitak od 3000$). Ključno zapažanje ovde je sledeće: dok donosioci odluka pokazuju averziju prema riziku u domenu dobitaka, oni pokazuju *sklonost ka riziku u domenu gubitaka*. Drugim rečima, sigurni gubitak od 3000$ dolara deluje gore od tek verovatnog gubitka od 4000$, makar i sa visokom verovatnoćom od 80% (koja ipak uključuje mogućnost da se „izvuče“ sa 20% šansi od ma kakvog gubitka); nasuprot ovome, sigurni dobitak od 3000$ je delovao bolje od tek verovatnog dobitka od 4000$ sa istom verovatnoćom. Ovaj nalaz će kasnije biti rafiniran u tzv. četvoročlanu strukturu stavova prema riziku (Tversky & Kahneman, 1992).

*Averzija prema gubicima.* Konačno, Tverski i Kaneman zaključuju da još jedan fenomen različitog tretmana gubitaka i dobitaka mora da se inkoroporira u razumevanje realnog donošenja odluka. To je fenomen *averzije prema gubicima* (engl. *loss aversion*). Averzija prema gubicima u empirijskom odlučivanju znači da je *opaženi efekat gubitaka veći od opaženog efekta dobitaka* i ne treba je mešati sa prethodno diskutovanim različitim stavovima prema riziku u domenu dobitaka (averzija prema riziku) i gubitaka (sklonost prema riziku). Gubitak neke vrednosti ostavlja jači psihološki utisak nego dobitak iste te vrednosti. Preciznije, ukoliko je $u(x)$ funkcija korisnosti, onda $u(x)$ ne može da ima iste karakteristike za dobitke i

gubitke, već ona mora biti *strmija za gubitke*, tako da je $|u(50\,\text{EUR})| < |u(\text{-}50\,\text{EUR})|$, što odgovara empirijskim nalazima; Tverski i Kaneman navode da je u velikom broju empirijskih istraživanja potvrđena ova pravilnost (Kahneman & Tversky, 1979).

Iako se empirijski nalazi koji ne idu u prilog normativnoj teoriji očekivane korisnosti pojavljuju samo nekoliko godina posle fon Nojmanove i Morgnešternove aksiomatizacije racionalnog izbora, bio je potreban niz decenija dok ekonomisti i matematičari nisu počeli da razvijaju alternativne teorije odlučivanja. Ovim alternativnim teorijama postavljen je standard *deskriptivne validnosti*: ukoliko su odstupanja od racionalnog izbora sistematska, zahtev je da teorija odlučivanja mora da obuhvati ove sistematske empirijske nalaze svojim konceptualnim, matematičkim okvirom. Analiza *načina na koji se u deskriptivnim teorijama ta inkorporacija izvodi* predstavljaće jednu od naših kritičkih diskusija debate o racionalnosti. S druge strane, pritisak da pod bilo kojim uslovima teorija odlučivanja mora da se obavezuje na neki minimum logičke odgovornosti odn. normativne standarde, nije prestao da vrši uticaj na razvoje teorija posle kritike očekivane korisnosti. Kompromis koji su naučnici osećali kao neophodan i njihov pristup da novim aksiomatizacijama odgovore na ovaj pritisak biće tema još jedne diskusije koju ćemo voditi u ovom radu. Konačno, invarijantnost empirijskih kognitivnih fenomena u odnosu na promene metoda koje se koriste u njihovom konstituisanju biće još jedna važna tema. Kao što je već diskutovano u slučaju Aleovog paradoksa, nekad je moguće postaviti pitanje da li određeni empirijski fenomen uopšte postoji, zbog mogućnosti grubih promena u registrovanom ponašanju sa promenama u metodi, čime se otvoreno postavlja pitanje o mogućnosti eksperimentalne izolacije, i samim tim, konstitucije, tog empirijskog fenomena. Sva diskutovana metodološka problematika u problemu racionalnog izbora nas samo još jednom podseća na to koliko konstrukcija naučnih teorija nije postupak koji može da se sprovodi nezavisno od odnosa empirijskog fenomena i metode kojom se on konstituiše.

## 1.4 Teorija izgleda

Prema opštem slaganju naučnika u oblasti odlučivanja, *teorija izgleda* (engl. *Prospect Theory*, skr. PT, Kahneman & Tversky, 1979), odn. njena razvijena verzija koja nosi ime *kumulativna teorija izgleda* (engl. *Cumulative Prospect Theory*, skr. CPT, Chateauneuf & Wakker, 1999, Tversky & Kahneman, 1992, Wakker, 2010, Wakker & Tversky, 1993), predstavlja najozbiljniju alternativu teoriji očekivane

korisnosti. Mnogi autori smatraju da je ona u prethodnih trideset godina praktično zamenila teoriju očekivane korisnosti kao standardna teorija odlučivanja u uslovima rizika. Teorija izgleda je veoma komplikovana i, nažalost, nije lako motivisati uvođenje njenih teorijskih konstrukcija bez detaljne analize empirijskih odluka; zato je sada prikazujemo u relativno pojednostavljenom obliku, izbegavajući u potpunosti analizu njenih aksiomatskih osnova. Centralni aksiom kumulativne teorije izgleda diskutovaćemo u III delu naše rasprave u okviru naše kritike debate o racionalnosti u oblasti odlučivanja u uslovima rizika i neizvesnosti.

Kumulativna teorija izgleda spada u grupu tzv. *teorija korisnosti zavisnih od ranga* (engl. *rank dependent utility theories*, skr. RDU, Abdellaoui, 2009). Osnovna osobina ovih teorija, koje ćemo kasnije detaljno diskutovati, jeste ta da se u oceni verovatnoće određenog dobitka ili gubitka na nekom lozu uzima u obzir cela distribucija verovatnoća na tom lozu i relativan položaj tog gubitka ili dobitka u odnosu na druge dobitke ili gubitke koje taj loz sadrži. Tako se svaka opcija na nekom rizičnom lozu evaluira u kontekstu drugih rizičnih opcija, i njena evaluacija zato ne može biti uvek ista. Sledeća suštinska osobina kumulativne teorije izgleda je da se u evaluaciji neke rizične opcije ne uzima u obzir verovatnoća kako je objektivno predstavljena, već *transformacija verovatnoće funkcijom koja preslikava verovatnoću u pondere odlučivanja* (engl. *decision weights*). U kumulativnoj teoriji izgleda (CPT) tvrdi se da postoji *funkcija vrednosti*[13] definisana nad lozovima koja preslikava monetarne vrednosti $X$ na skup opaženih vrednosti (korisnosti) u realnim brojevima, $v : X \to R$, takvi da za loz $(x_i,p_i)$, gde su $x_i \in X$ sve vrednosti na lozu, a $p_i$ odgovarajuće verovatnoće, i $-m \leq i \leq n$, važi:

$$V(f) = V(f^+) + V(f^-) \tag{3}$$

$$V(f^+) = \sum_{i=0}^{n} \pi_i^+ v(x_i) \tag{4}$$

$$V(f^-) = \sum_{i=-m}^{0} \pi_i^- v(x_i) \tag{5}$$

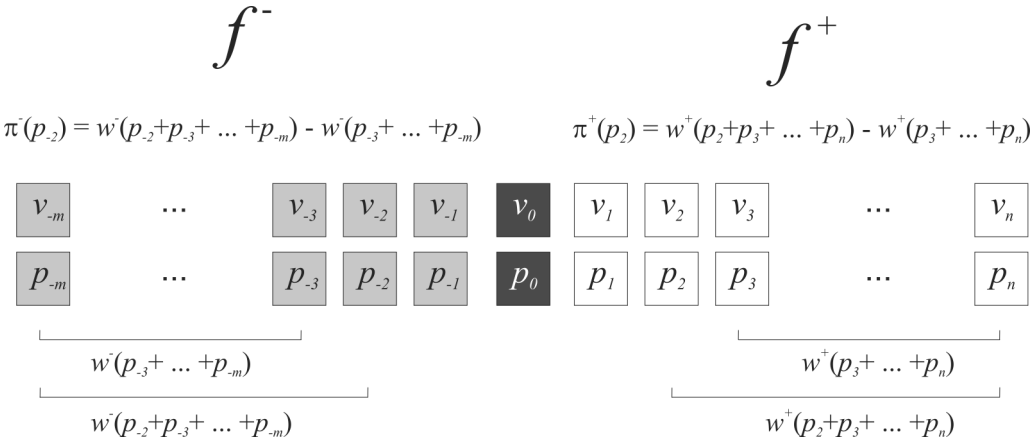
U prethodnim jednačinama, $f^+$označava pozitivni deo loza, odn. sve dobitke koje on sadrži sa odgovarajućim verovatnoćama; $f^-$, jasno, označava negativni deo loza koji okuplja sve gubitke sa odgovarajućim verovatnoćama. Jednačina (3) nam

govori to da će konačna korisnost nekog loza koji uključuje i gubitke i dobitke biti suma korisnosti njegovog pozitivnog i negativnog dela: $V(f^+)$ i $V(f^-)$. Složeni proces ponderisanja verovatnoća u kumulativnoj teoriji izgleda je ilustrovan na Slici 4a. Jednačine (4) i (5) opisuju proces kojim se dolazi do ukupne korisnosti loza u ovoj teoriji: za pozitivni deo loza $V(f^+)$, korisnost je suma proizvoda (a) korisnosti svakog dobitka $x_i$, u oznaci $v(x_i)$, i *pondera odlučivanja* $\pi_i^+$, koji se izračunava za svaki dobitak $x_i$. Korisnost negativnog dela loza, koji sadrži samo moguće gubitke, računa se na isti način, osim što se koriste ponderi odlučivanja $\pi_i^-$za gubitke[14].

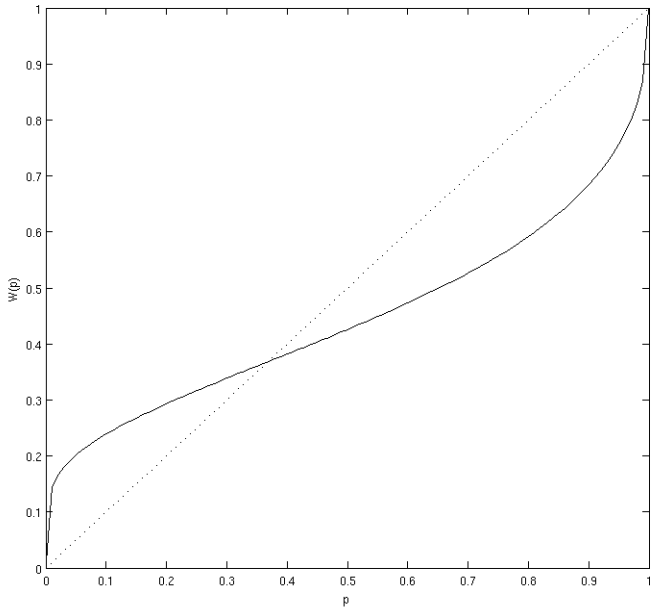
Kao što smo rekli, ponderi odlučivanja se dobijaju tako što se verovatnoće za svaki gubitak ili dobitak transformišu funkcijom ponderisanja verovatnoće. Originalna verzija teorije izgleda iz 1979. godine transformisala je verovatnoće direktno preko ove funkcije. U kumulativnoj teoriji izgleda, ponder odlučivanja za neki *dobitak* $x_i$ se računa na sledeći način: (a) transformiše se dekumulativna verovatnoća prvog sledećeg većeg dobitka od $x_i$ koji se nalazi na lozu, odn. $x_{i+1}$, zatim se (b) transformiše dekumulativna verovatnoća samog $x_i$ odn. dobitka za koji se računa ponder, i konačno se (c) računa razlika između ta dva. Time se dobija razlika između transformisanih dekumulativnih verovatnoća $(i)$ dobitka koji je dobar najmanje kao $x_i$, i $(ii)$ dobitka koji je striktno bolji od $x_i$. Slično se računaju i ponderi odlučivanja za gubitke, osim što se umesto dekumulativnih koriste kumulativne verovatnoće. Sledeća jednačina predstavlja deskripciju ovog procesa za dobitke:

$$\pi_i^+ = W^+(p_i + ... + p_n) - W^+(p_{i+1} + ... + p_n), 0 \leq i \leq n-1 \tag{6}$$

prethodno uzimajući da je $\pi_n^+ = W^+(p_n)$, odn. da je ponder odlučivanja za najveći ponuđeni dobitak na lozu sama transformacija njegove verovatnoće. Funkcija $W^+$ koja transformiše kumulativne verovatnoće $(p_i+...+p_n)$ i $(p_{i+1}+...+p_n)$ u pondere odlučivanja je funkcija ponderisanja verovatnoća; oznaka „+“ iznad funkcije $W$ govori samo to da se odnosi ponderisanje verovatnoće za dobitke. Opis osobina funkcije ponderisanja verovatnoće spada u najveće doprinose teorije izgleda.




Slika 4a. *Proces ponderisanja verovatnoće u kumulativnoj teoriji izgleda.* Na levoj strani je prikazan negativni deo loza, $f^-$, na desnoj pozitivni deo, $f^+$. Prikazane su vrednosti na lozu sa odgovarajućim verovatnoćama. Loz je sortiran od najvećeg gubitka ($-m$) ka najvećem dobitku ($n$) s leva na desno. Slika ilustruje ponderisanje drugog po veličini dobitka na pozitivnom delu loza, $v_2$, i drugog najvećeg gubitka na negativnom delu loza, $v_{-2}$. Ponder verovatnoće za $v_2$, u oznaci $\pi^+(p_2)$, dobija se kada se od transformacije funkcijom ponderisanja $w^+(p_2)$ (prikazane na Slici 4b) dekumulativne verovatnoće dobitka $v_2$ oduzme transformacija dekumulativne verovatnoće prvog sledećeg većeg dobitka od $v_2$, $w^+(p_3)$, kao što slika ilustruje. Simetričan proces određuje ponder verovatnoće za drugi najveći gubitak na negativnom delu loza, $v_{-2}$; u izračunavanju pondera verovatnoće za gubitke prirodno se nameće upotreba kumulativnih, ne dekumulativnih verovatnoća.

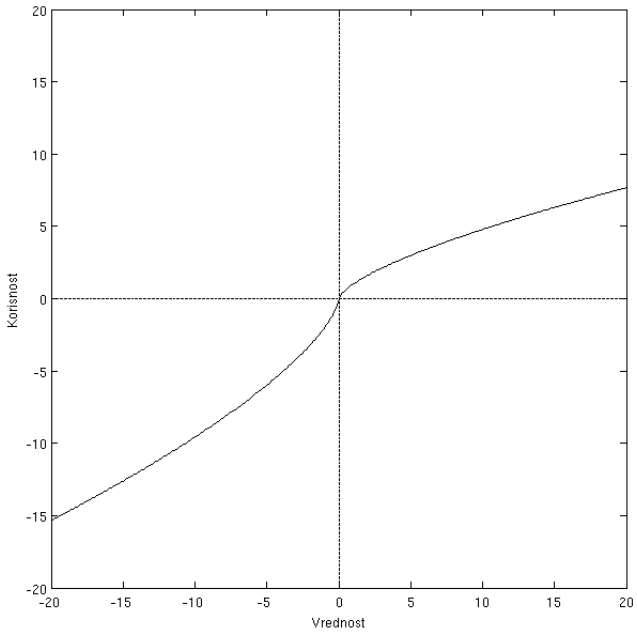



Slika 4b. *Funkcija ponderisanja verovatnoće.* Ispitivano je više matematičkih formi ove funkcije (za pregled up. Fox & Poldrack, 2009). Funkcija na slici je primer jednoparametarske verzije Prelecove funkcije ponderisanja verovatnoće sa vrednošću parametra $\gamma = .43$ (Prelec, 1998).

Slika 4b. prikazuje jednu tipičnu funkciju ponderisanja verovatnoće. Kao što vidimo na Slici 4b, nelinearna funkcija ponderisanja verovatnoće ima tu osobinu da potcenjuje umerene i velike, a precenjuje male verovatnoće. Verovatnoća sigurnog ishoda, kao i verovatnoća nemogućeg ishoda, se ne ponderišu: $W(1) = 1$ i $W(0) = 0$. Takođe, razlike između ekstremnijih, malih i velikih verovatnoća, zahvaljujući specifičnom obliku ove funkcije, imaju veći uticaj na odluku od razlika između verovatnoća u srednjem, umerenom rasponu. U diskusiji kumulativne teorije izgleda ne treba nikad gubiti iz vida proces ponderisanja verovatnoća koji smo prethodno opisali i ilustrovali na Slici 4a: iskustva pokazuju da je najčešći oblik nerazumevanja ove teorije vezan za nepravilnu interpretaciju funkcije ponderisanja verovatnoća. Funkcija ponderisanja verovatnoća ne transformiše verovatnoće, već dekumulativne verovatnoće, relativne doprinose u verovatnoći koje određene vrednosti na lozovima nose u odnosu na njihov položaj na lozu. Ova osobina je inicijalno uvedena u već pomenute RDU modele (modele odlučivanja zavisne od ranga, Quiggin, 1982); Tverski i Kaneman je inkorporiraju u kumulativnu teoriju izgleda tek 1992. Takođe, vrednosti (ordinate, kodomen) funkcije $W(p)$ *nisu verovatnoće*: ova funkcija pokazuje osobinu subaditivnosti, odn. da $W(p) + W(1-p) < 1$: suma pondera komplementarnih verovatnoća generalno je manja od jedan, za razliku od sume samih komplementarnih verovatnoća. Pod pretpostavkom da ljudi u odlučivanju transformišu verovatnoće na ovaj način, moguće je objasniti neka sistematska odstupanja empirijskih odluka od normativne teorije očekivane korisnosti. Ne treba zaboraviti da stavovi prema riziku u teoriji izgleda više ne zavise samo od osobina funkcije korisnosti, već su sada *posledica složenih interakcija između osobina funkcije korisnosti i osobina funkcije ponderisanja verovatnoće.* Takođe, za naše dalje diskusije, bitno je imati na umu da su osobine funkcije ponderisanja verovatnoće ustanovljene dosledno formalnim, matematičkim putem: Kaneman i Tverski su do nje došli postepenom formalizacijom karakterističnih osobina izbora u kojima ljudi odstupaju od normativne teorije, i na osnovu dobijenih formalnih stavova o izborima *konstruisali* funkciju ponderisanja verovatnoće kojom možemo da objasnimo empirijski osmotrena odstupanja (Kahneman & Tversky, 1979). Ona dakle nije bazirana na nekoj dubljoj intuiciji o ljudskom odlučivanju; njen oblik, u kome konveksni region smenjuje konkavni, i to asimetrično, kako primećuje Prelec, „*...ne izgleda kao oblik koji bi neko nacrtao osim ako na to nije primoran jakim empirijskim dokazima*“ (citirano prema Prelec, 1998, naš prevod).

Pored funkcije ponderisanja verovatnoće i transformacije kumulativnih umesto

običnih verovatnoća, teorija izgleda donosi još jednu promenu u odnosu na teoriju očekivane korisnosti. Već diskutovani empirijski fenomen averzije prema gubicima operacionalizuje se u ovoj teoriji modifikacijom same funkcije korisnosti (odn. funkcije vrednosti, u terminologiji Kanemana i Tverskog) koja je, prema teoriji izgleda, *strmija za gubitke nego za dobitke*. Ova ideja nalazi se još u prvoj verziji teorije iz 1979. i zadržana u kumulativnoj teoriji izgleda iz 1992. Slika 5. prikazuje funkciju vrednosti koja se koristi u teoriji izgleda.




Slika 5. *Funkcija vrednosti teorije izgleda.* Kao i funkcija korisnosti teorije očekivane korisnosti, funkcija vrednosti teorije izgleda preslikava (objektivne) vrednosti u (subjektivne) korisnosti. Funkcija je strmija za gubitke nego za dobitke, reprezentujući tako empirijski fenomen averzije prema gubicima. Funkcija na slici je nastala modifikacijom stepene funkcije korisnosti sa eksponentom $\rho = .68$ uvođenjem koeficijenta averzije prema gubicima $\lambda = 2$ (up. jednačinu (8)); tako je subjektivni doživljaj određenog gubitka dva puta jači u odnosu na subjektivni doživljaj dobitka iste vrednosti.

Polazeći od stepene funkcije korisnosti, što je uobičajena pretpostavka u empirijskm analizama i normativnih i deskriptivnih teorija, $u(x) = x^{\rho}$, modifikacija koja obuhvata fenomen averzije prema gubicima sada nalaže da:

$$v(x) = x^{\rho}, x \geq 0 \tag{7}$$

odn.

$$v(x) = -\lambda(-x)^{\rho}, x < 0 \qquad (8)$$

Kao što je u teoriji očekivane korisnosti $u(x)$ definisana nad vrednostima a $U(X)$ nad lozovima, u teoriji izgleda je $v(x)$ definisana na vrednostima, a $V(X)$ nad lozovima. Jednačine (7) i (8) definišu funkciju preslikavanja vrednosti u korisnosti pod pretpostavkom da je stepena funkcija njena dobra aproksimacija. Baš kao i teorija očekivane korisnosti, teorija izgleda ne obavezuje na neku specifičnu parametarsku formu ove funkcije; više funkcionalnih formi može da zadovolji uslove koje teorija izgleda postavlja.

Kombinovanjem osobina funkcija vrednosti i ponderisanja verovatnoća teorija izgleda može da objasni veliki broj empirijskih anomalija racionalnog izbora koje smo diskutovali u prethodnoj sekciji; na prvom mestu, teorija izgleda otklanja Aleov paradoks, odn. pruža objašnjenje kako dolazi do ove empirijske anomalije koja je potresla svet normativnih teorija. Međutim, tvrdnja da teorija izgleda ima *de facto* veću prediktivnu validnost od teorije očekivane korisnosti ne može da se napravi bez izvesnih ograda. Kao što je već diskutovano, empirijske anomalije racionalnog izbora poput Aleovog paradoksa nije moguće dobiti uvek, svakom procedurom, i sa svim ispitanicima. Direktnih poređenja dva modela odlučivanja ima malo (npr. Hey & Orme, 1994, Blavatskyy, 2011); neka od njih pokazuju da u eksperimentima sa većim brojem izbora više nego često nije moguće konstatovati superiornost deskriptivnog nad normativnim modelom.

Teorija izgleda je aksiomatizovana (Chateauneuf & Wakker, 1999, Tversky & Kahneman, 1992, Wakker & Tversky, 1993), ali njen *aksiomatski okvir nije ni iz daleka tako intuitivan* kao aksiomatski okvir teorije očekivane korisnosti. Zaista, u kasnijim analizama ćemo videti kako aksiomatizacija teorije izgleda počiva na prilično komplikovanim teorijskim konstrukcijama. Kako u matematičkom, tako i u psihološkom smislu, ono što teoriju izgleda čini problematičnom jeste upotreba neaditivnih pondera odlučivanja umesto običnih verovatnoća. Kaneman i Tverski su u originalnoj verziji teorije izgleda iz 1979. godine bili eksplicitni po tom pitanju: ponderi odlučivanja su karakteristični za samo odlučivanje i nisu posledica neke psihofizičke funkcije opažanja verovatnoće (Kahneman & Tversky, 1979). U aksiomatizaciji kumulativne teorije izgleda iz 1992. je zato iskorišćen koncept *kapaciteta* koji generalizuje uobičajene verovatnoće, pružajući tako odgovarajuću formalnu deskripciju pondera odlučivanja (Tversky & Kahneman, 1992, Wakker &

Tversky, 1993).

Teorija izgleda motivisala je ogroman broj teorijskih i eksperimentalnih istraživanja. Ona danas predstavlja jednu od najuticajnijih psiholoških teorija uopšte.

## 1.5 Normativna i deskriptivna objašnjenja

U prethodnim redovima pratili smo razvoj jednog naučnog objašnjenja kroz faze početne matematičke deskripcije, preko stroge formalizacije, sistematskog eksperimentalnog testiranja i korekcije do tada važeće matematičke teorije. Paralelno, razvoj teorije odlučivanja u uslovima rizika ocrtava paradigmatičan *razvoj deskriptivne kognitivne teorije kroz uvođenje postepenih korekcija za empirijska odstupanja od normativnih standarda.* Na prvi pogled, teorija racionalnog izbora od XVIII do XX veka prolazi kroz faze tipične za razvoj i testiranje naučnih teorija u ma kojoj nauci. Kraći kritički osvrt na istoriju ovog problema, međutim, ukazuje na to da proučavanje ovog problema tokom istorije opisuje neobičnu putanju.

Svest o povezanosti normativnih i deskriptivnih obaveza koje teorija racionalnog izbora mora da ispuni karakteristična je za Bernulijevu analizu. Posvećenost traganju za formalnim rešenjem koje će zadovoljiti minimalne zahteve intuicije i logičke konzistentnosti je karakteristika razvoja u prvoj polovini i sredinom XX veka, kada nastaju aksiomatske teorije racionalnog izbora poput fon Nojman-Morgenšternove ili Sevidžove (Savage, 1954/1972). Ipak, tek dve ili tri decenije eksperimentalnog rada bile su dovoljne da mnogi odustanu od koncepta normativne u korist koncepta ograničene racionalnosti. Od tada, proučavanje donošenja odluka - a kao što ćemo uskoro videti, i naučna psihologija saznanja uopšte - podeliće se u dva tabora, od kojih će jedan braniti osnove normativne racionalnosti, dok će drugi tvrditi da one ne opisuju adekvatno psihološku, bihejvioralnu realnost i da ih se treba odreći u korist deskriptivnih teorija. To su dve osnovne pozicije u savremenoj *debati o racionalnosti.* Da bismo rasvetlili kako je došlo do podele u ove dve struje, i pod kojim uslovima je empirijska analiza vodila neke autore, poput Tverskog i Kanemana, ka teorijama drugačije prirode od normativnih, moraćemo detaljno da proučimo *metodološke i teorijske pretpostavke konstrukcije kognitivne psihologije kao nauke.* Te temelje savremene kognitivne psihologije predstavićemo i analizirati u narednim poglavljima, a onda se vratiti problemu racionalnog odlučivanja u psihologiji, ekonomiji i matematici u pokušaju da razumemo odnos normativnog i deskriptivnog na jedan nov način.

Najbitnije odlike Bernulijevog rešenja su sledeće: prvo, njegova hipoteza očekivane korisnosti *opisuje ponašanje* oslanjajući se na hipoteze o tome kakav *stav* ljudi zauzimaju prema riziku. Dakle, već Bernulijeva analiza odlučivanja jeste psihološka analiza, bez obzira na to što u svoje vreme on nije imao na raspolaganju sistematsku eksperimentalno-psihološku metodologiju. Dalje, Bernulijev pristup je bio *deskriptivan*: njegova hipoteza ne polazi od idealizovanog donosioca odluka već se razvija da bi objasnila empirijske fenomene averzije prema riziku i opadajuće marginalne korisnosti novca, koji su karakteristika realnih donosioca odluka. Bernulijeva analiza odlučivanja, istorijski možda prvi pokušaj da se matematički opišu zakoni ljudskog ponašanja, nosi karakteristike i normativne i deskriptivne teorije. Ona je normativna u tom smislu što predstavlja jedan matematički formalizam kojim se tvrdi *kako treba* donositi odluke, a deskriptivna pošto pruža objašnjenje realnih, empirijskih fenomena.

Aksiomatska analiza fon Nojmana i Morgenšterna predstavlja formalno dublji korak od Bernulijeve. Potez koji vodi od Bernulija ka fon Nojmanu i Morgenšternu obuhvata formalizaciju nekih osnovnih, veoma jednostavnih intuicija o odlučivanju u uslovima rizika. Povezivanje tih intuicija (odn. aksioma racionalnog izbora) sa hipotezom o očekivanoj korisnosti tehnički je složeno, ali to je očekivana cena za poduhvat koji pretenduje na tako temeljno zasnivanje oblasti odlučivanja. Međutim, aksiomatizacija racionalnog izbora sprovodi se nezavisno od sistematskih bihejvioralnih eksperimenata; videli smo kako se u susretu sa ovim istraživanjima temelji racionalnog izbora lako zaljuljaju. Ne treba zaboraviti da se aksiomatska teorija racionalnog izbora udaljila od empirijskog posmatranja više nego Bernuli koji takođe nije raspolagao sistematskim posmatranjem ponašanja: dok Bernuli specifikuje precizno funkciju korisnosti da bi bio u stanju da objasni intuitivno poznate empirijske fenomene, aksiomatska teorija fon Nojmana i Morgenšterna obezbeđuje tek njenu egzistenciju i ne karakteriše njenu specifičnu formu. Averzija prema riziku, na primer, ne predviđa se kao *nužna* pojava u okviru aksiomatske teorije racionalnog izbora (iako funkcija korisnosti u EU *može* da opisuje i averziju prema riziku). Dakle, u odnosu na normativni okvir racionalnog izbora, empirijsko posmatranje ponašanja uvodi i neke *a posteriori* kriterijume za selekciju matematičkog modela ovog fenomena.

Konačno, ono što odlikuje savremeni razvoj teorije odlučivanja, jeste jedan neobičan kompromis između normativnih zahteva i deskriptivne validnosti. Detaljnu raspravu ćemo ostaviti za kasnije, kada budemo razmatrali elemente aksiomatizacije

teorije izgleda. Međutim, već smo napomenuli da takve aksiomatizacije nisu intuitivne, samorazumljive i prihvatljive kao aksiomatizacija koju su postavili fon Nojman i Morgneštern; potrebno je solidno prethodno iskustvo deskriptivne analize odlučivanja samo da bi se razumelo značenje ključnih aksioma teorije izgleda. Koliko je cena ovakvog pristupa aksiomatizaciji deskriptivnih, psiholoških teorija odlučivanja previsoka, podjednako (*i*) u smislu odricanja od prethodno ostvarene koherencije u proučavanju problema odlučivanja koliko i (*ii*) u odnosu prema teorijskim fundamentima kognitivne psihologije kao nauke, zapitaćemo se u jednoj od centralnih diskusija u ovoj raspravi.

Zadržimo se još na trenutak na postupku aksiomatizacije. Naravno, svaka aksiomatizacija je korisna: ona omogućava najviši stepen egzaktnosti i jasan uvid u konzistentnost i ograničenja neke teorije. Takođe, postupak aksiomatizacije, istorijski potvrđeno, nije bez neobičnosti. U XIX veku, ruski matematičar Nikolaj Ivanovič Lobačevski i mađarski matematičar Janoš Boljaji otvorili su put do tada nepoznatim teorijama geometrije odlukom da izmene problematične aksiome klasične geometrije. Modifikujući tzv. peti Euklidov postulat, odn. aksiom paralelnosti, a onda proučavajući osobine geometrije koja se dobija kao rezultat ovog eksperimenta, matematika je otkrila svet neeuklidskih geometrija. Posledica ovog rada bila je revolucionarna promena u našem razumevanju geometrije, a potez kojim je otkrio nove teorije uticao je na filozofiju matematike verovatno više nego ma koji drugi događaj u njenoj istoriji. Dakle, razni eksperimenti u okviru aksiomatske analize jesu mogući, i nekad veoma plodni. Međutim, mi verujemo, i pokušaćemo da pokažemo, da je savremena aksiomatizacija deskriptivnih teorija odlučivanja proces u kome kao da nedostaje ono strpljenje matematičara koji su *vekovima* pokušavali da intuitivno složen peti postulat izvedu iz drugih, jednostavnijih stavova, pre pokušaja da se radi sa njegovim modifikacijama.

Jednom uspostavljen sistem aksioma, sistem koji obezbeđuje složenije formalne konstrukcije, smatra se skupim, dragocenim postignućem, i od njega se po pravilu ne odustaje bez krupnih razloga. Tverski i Kaneman slede formalni metod u razvoju teorije izgleda: oni formalizuju matematički jednu po jednu bitnu osobinu empirijskih odluka koje krše principe racionalnog izbora, i na osnovu tih formalnih tvrdnji konstruišu osobine teorijskih objekata koji su im potrebni, kao što je funkcija ponderisanja verovatnoće. Na taj način teorija izgleda stiče empirijsku validnost, a zadržava formalnu zasnovanost. Mi se nadamo da ćemo daljim analizama u ovoj tezi uspeti da pokažemo da je, uz pomoć jedne drugačije, klasičnije i konzervativnije

formalne analize odlučivanja, moguće razviti teorijsku konstrukciju koja zadržava empirijsku validnost bez toliko odstupanja od jednostavnih tvrđenja normativne teorije. Konačno, očekujemo da nam metateorijska analiza različitih (normativnih i deskriptivnih) pristupa izgradnji kognitivnih teorija omogući dublji uvid u prirodu šireg pitanja o tome da li je ljudsko saznanje racionalno i načina da se na to pitanje odgovori.

Naredni redovi pružaju sistematizaciju problema racionalnosti saznanja, predstavljaju ciljeve našeg istraživanja i skicu diskusija koje predstoje.

# 2 Problem racionalnosti saznanja

Problem racionalnosti saznanja predstavlja predmet našeg istraživanja. Njegov obim postaje jasan kada se diskusija odnosa normativnog i deskriptivnog u oblasti poput racionalnog izbora proširi na druge kognitivne funkcije. U našoj raspravi ta generalizacija će biti ograničena na određene domene viših i simboličkih kognitivnih procesa. Pored ove generalizacije problema, neophodno je uputiti i na određeno tehničko značenje koje racionalnost dobija u ovim diskusijama. To značenje najbolje se ogleda u razvoju metateorijske i metodološke paradigme racionalne analize koju pruža Džon Anderson (Anderson, 1991a, 1991b) sledeći teorijske osnove Dejvida Mara, postavljenje u projektu kompjutacionog proučavanja viđenja (Marr, 1982). U naredim redovima predstavljamo *domen rasprave*, odn. oblasti kognitivne psihologije u okviru kojih diskutujemo problem racionalnosti saznanja, zatim uvodimo metodologiju *racionalne analize* i konačno precizno definišemo *ciljeve* našeg istraživanja u odnosu na ovako definisan predmet.

## 2.1 Domen rasprave

Rasprava o racionalnom izboru predstavlja paradigmatičan primer debate o racionalnosti kognitivnih procesa. Ono zbog čega smo je izabrali za temu uvodne diskusije ove teze je, pre svega, stepen njene formalizacije: on je najviši među svim postojećim kognitivnim teorijama. Osim odlučivanja, mnoge druge kognitivne funkcije se proučavaju sa stanovišta odnosa ponašanja koja proizvode prema određenim normativnim standardima. U oblastima van odlučivanja, stepen formalizacije teorija varira, a stepen stroge aksiomatizacije se retko sreće; ipak, ne izostaju matematički formulisani normativni standardi. U našoj diskusiji

racionalnosti saznanja, kao što je već rečeno, pažnju ćemo usresrediti na *više i simboličke kognitivne procese*. Podela na „niže“ (senzorni procesi, pažnja, percepcija do određenog nivoa složenosti) i „više“ kognitivne procese ili fenomene je prema našem mišljenju samo nesrećno terminološko rešenje. Ono je pre posledica neophodne (sociolingvističke) podele rada među naučnicima u ogromnoj oblasti kao što je kognitivna psihologija, nego što odslikava realnu strukturu problema. Postoji više razloga za izbor „viših“ procesa - kao što su odlučivanje, učenje i opažanje kauzalnih odnosa, pamćenje, kategorizacija, indukcija, rezonovanje i razumevanje značenja - kao predmeta naše analize. U nekim od navedenih oblasti debata o racionalnosti je već eksplicitno razvijena. S druge strane, neke od ovih oblasti proučavanja, videćemo, otkrivaju veoma inspirativne probleme kada na njih pokušamo da primenimo metodologiju racionalne analize. Rad koji bi obuhvatao raspravu o racionalnosti kognitivnih funkcija *uopšte* daleko bi po obimu morao da nadmaši rad koji mi predstavljamo, a po stepenu ekspertize autor svakako ne bi mogao da odgovori na takav zahtev. Postavlja se pitanje, onda, koje kognitivne funkcije ne uključiti u raspravu. Mi izostavljamo iz diskusije istraživanja u oblasti percepcije. Istraživanja perceptivnih sistema su obimna najmanje koliko i istraživanja viših i simboličkih kognitivnih funkcija, i zaista je teško odlučiti se na analizu debate o racionalnosti koja bi obuhvatala i jedne i druge. S druge strane, izostavljamo psiholingvistiku, jednu od najrazvijenijih disciplina savremene kognitivne psihologije, ne samo zbog izuzetne teorijske i empirijske širine ove oblasti, već i zbog specifičnosti gledišta razvijenih tokom njene istorije. Neretko, analiza teorijskih pozicija u psiholingvistici zahteva poznavanje više nego jedne nauke: ona skoro po pravilu obuhvata problematiku lingvistike, često genetike, ponekad antropologije i sociologije. Jedina naša analiza koja se delom odnosi na psiholingvističku problematiku odnosiće se na diskusiju o normativnom okviru za psihološku teoriju značenja. Ovo ograničenje, naravno, ne znači da nećemo koristiti argumente ili primere iz oblasti percepcije i psiholingvistike kada oni mogu značajno da doprinesu našoj diskusiji viših i simboličkih procesa.

III deo ove teze predstavlja kritičku diskusiju savremene debate o racionalnosti u oblasti viših i simboličkih kognitivnih procesa, tako da ćemo ovde pružiti samo pregled problematike koja će biti detaljno diskutovana tamo.

*Odlučivanje u uslovima rizika i neizvesnosti.* O normativnim i deskriptivnim pristupima ovoj oblasti već je dovoljno rečeno tokom uvodne diskusije. U daljem razvoju diskusije ćemo zato imati prilike da se posvetimo detaljima aksiomatizovanih

teorija odlučivanja i finesama koje se nalaze u samoj srži debate o racionalnosti.

*Opažanje i učenje kauzalnih odnosa.* U kognitivnoj psihologiji, savremena debata o racionalnosti u domenu kauzalnih odnosa grana se u dva problema: (a) da li je ljudsko *rezonovanje* o kauzalnosti normativno racionalno, i (b) da li je način na koji ljudi *uče* kauzalne odnose normativno racionalan. Problemu učenja kauzalnih odnosa ćemo posvetiti više pažnje. Taj problem se tradicionalno vezuje za pokušaje rešenja čuvenog *Hjumovog problema kauzalne indukcije*: da li je na osnovu podataka o kovarijaciji (korelaciji) fenomena moguće zaključiti o postojanju ili nepostojanju kauzalnog odnosa među njima? Do poznih decenija XX veka smatralo se da je načelno nemoguće rešiti Hjumov problem (prisetimo se tradicionalne metodološke „mantre“ da korelacija ne implicira kauzalnost), kada se došlo do saznanja da postoji određen skup pretpostavki o modelima kauzalnih odnosa pod kojima je moguće oceniti verovatnoću da između varijabli postoji kauzalni odnos - i to samo na osnovu poznavanja njihove kontingencije. *Teorija kauzalnih modela* (koju koristimo kao sinonim za *teoriju kauzalnih mreža*) tako se nameće kao normativna teorija kauzalnog učenja, i u tradiciji eksperimentalnog rada u ovoj oblasti suprostavlja tzv. *asocijacionističkim teorijama* poput čuvenog Reskorla-Vagner modela.

*Epizodička memorija.* Pitanje racionalnosti ljudske memorije može da se postavi u odnosu na stepen u kome kognitivni sistem optimalno koristi raspoložive informacije o prošlosti. Tako se racionalna analiza memorije susreće sa problemom *optimalnog pristupa* memorijskim tragovima (engl. *retrieval*), a ne sa problemom veridičnosti koji *prima facie* određuje normativni kriterijum za pamćenje. Anderson je sistematski pristupio nekim od markantnih empirijskih fenomena memorijskog pristupa i pokušao da razvije racionalnu teoriju pamćenja koja će biti glavna meta naše analize.

*Rezonovanje* i suđenje. Normativnu osnovu za deduktivne procese i testiranje hipoteza predstavljaju sistemi formalne logike, od kojih ćemo se mi ograničiti mahom na raspravu o racionalnosti ljudskog rezonovanja u odnosu na pravila zaključivanja iskaznog računa, odn. računa prvog reda. Diskusija deduktivnog rezonovanja i testiranja hipoteza izuzetno je značajna zbog odnosa „kanoničkog“ normativnog okvira u logičkom računu i „alternativnog“ normativnog okvira koji u novijim radovima predstavljaju teorija verovatnoće i teorija informacija. Funkcije rezonovanja i donošenja sudova duboko su povezane u istraživačkim paradigmama u debati o racionalnosti i zato smo odabrali da ih diskutujemo u istoj sekciji (iako kognitivne funkcije donošenja sudova prožimaju sve probleme viših kognitivnih

procesa koje obrađujemo).

*Funkcije kategorizacije konceptualnog sistema.* Zbog kompleksnosti koja je vezana za procese učenja koncepata i njihove kategorizacije, do sada nije eksplicitno tvrđeno da postoji jedna normativna teorija učenja koncepata, iako je moguće više matematičkih, probabilističkih modela posmatrati kao normativne u odnosu na ove procese. S druge strane, više principa ponuđenih u naučnoj periodici mogu da se tretiraju kao normativni u odnosu na problem optimalne organizacije konceptualnog sistema u odnosu na zahteve sredine za adaptacijom. Problem interakcije funkcija pamćenja i učenja sa funkcijama reprezentacije koncepata u ovoj oblasti ograničava razvoj teorije koja bi govorila samo o konceptualnom znanju, što dodanto usložnjava raspravu.

*Kreativne funkcije konceptualnog sistema.* Problemi *interpretacije* i *kreiranja karakteristika* predstavljaju najsloženija pitanja u oblasti konceptualne organizacije. Verujemo da ćemo uspeti da pokažemo da trenutno razumevanje ovih funkcija u okvirima teorijskih paradigmi kognitivne psihologije ne obezbeđuje nikakav normativni okvir za njihovu analizu. Diskutovaćemo i pitanje da li je uopšte moguće postaviti normativni okvir za analizu ovih fenomena.

U svim navedenim domenima viših i simboličkih kognitivnih funkcija, predmet naše analize su uvek *formalne kognitivne teorije*: to ne znači ništa drugo do toga da govorimo o *matematičkim teorijama određenih kognitivih procesa.* Našu analizu fokusiramo na jedan specifičan podskup ovakvih teorija: na *formalne teorije razvijene na kompjutacionom nivou analize* (po Marovom razlikovanju nivoa kognitivne teorije, Marr, 1982). To znači da u analizu ne uključujemo ma kakve matematičke formulacije kognitivnih teorija, već samo one koje su direktno motivisane rešavanjem nekog određenog adaptivnog cilja. Da bismo razumeli osobine formalnih teorija kompjutacionog nivoa, teorije na koje se fokusira savremena debata o racionalnosti, moramo reći nešto o *metodologiji racionalne analize* koja je strateški povezana sa razvojem ovakvih teorija; ova metodološka paradigma pratiće nas tokom svih diskusija u ovoj tezi.

## 2.2 Racionalna analiza

Možda deluje neobično da kognitivna psihologija, koja se kao nauka sama obavezuje racionalnim principima, a za svoj predmet uzima saznajne sposobnosti ljudskog uma, u svom teorijsko-metodološkom arsenalu poseduje metodologiju

po imenu *racionalne analize*. Naime, za ljudski um, čijih je kognitivnih moći jedna posledica i naučna racionalnost, ne bi li ona morala da bude postulat koji se zapravo ne može dovesti u pitanje? Ne diktira li RACIONALNOST$_2$ da RACIONALNOST$_1$ bude nužna? Kao što smo videli na primeru problema racionalnog izbora, empirijski nalazi pokazuju da ljudske odluke često mogu da se kose i sa veoma jednostavnim racionalnim principima. Pošto verujemo da bi bila bila nekorektna pretpostavka o tome da kognitivni sistemi koji konstruišu racionalne naučne teorije (RACIONALNOST$_1$) bili *a priori* racionalni, za razliku od nekih drugih kognitivnih sistema koji predstavljaju empirijske objekte analize (RACIONALNOST$_2$), ostaje nam samo da se zapitamo kako je takav dvostruki položaj koncepta racionalnosti saznanja moguć?

U teorijskom i metodološkom okviru racionalne analize, racionalnost dobija jedno specifično značenje koje odgovara deskriptivnim analizama problema poput problema odlučivanja u uslovima rizika i neizvesnosti. Prema preporukama metodologije racionalne analize, koju uvodi Anderson (Anderson, 1991a, 1991b) oslanjajući se na teorijske postavke Dejvida Mara (Marr, 1982), osnovna pretpostavka od koje treba poći u analizi kognitivnih sistema je *princip racionalnosti*:

„*Kognitivni sistem optimizuje adaptaciju ponašanja organizma*“

(Anderson, 1991a).

Dakle, polazeći od toga da svako ponašanje ima određenu vrednost u adaptaciji, osnovni princip racionalne analize kognitivnih fenomena pretpostavlja da će funkcionisanje kognitivnog sistema omogućiti izbor *optimalnog ponašanj*a, odn. takvog ponašanja koje, u *skladu sa ograničenjima* (dostupnosti informacionih, energetskih i drugih resursa) ima *maksimalnu adaptivnu vrednost*. Na osnovu ovog principa racionalne analize jasno vidimo zašto ljudsko odlučivanje u uslovima rizika nije racionalno: bez obzira na deskriptivnu moć teorije izgleda, donosilac odluka koji se rukovodi njenim, a ne normativnim principima teorije očekivane korisnosti, sigurno donosi *suboptimalne odluke* u odnosu na verovatnoće i visine dobitaka i gubitaka sa kojima je suočen u svom okruženju. S druge strane, normativni karakter teorije očekivane korisnosti implicira njenu optimalnost: ne postoji način da u datim uslovima (pod datim ograničenjima) zaradite više (ili izgubite manje) do da se rukovodite principima racionalnog izbora.

Metodologija racionalne analize podrazumeva sprovođenje logičnog sleda koraka u izgradnji kognitivne teorije nekog fenomena. Ti koraci su sledeći:

1. *Precizno odrediti koji su ciljevi kognitivnog sistema.*
2. *Razviti formalni model okruženja na koji je kognitivni sistem adaptiran.*
3. *Postaviti minimalne pretpostavke o tome koja je cena izračunavanja da bi se ciljevi (1) rešili u tom okruženju.*
4. *Razviti optimalnu bihejvioralnu funkciju na osnovu (1) - (3).*
5. *Empirijski testirati predikcije bihejvioralne funkcije.*
6. *Ukoliko predikcije bihejvioralne funkcije nisu potvrđene, ponavljati proces racionalne analize dok one to ne budu.*

Dakle, racionalnost, shvaćena kao optimizovanje ponašanja da bi ono imalo maksimalnu adaptivnu vrednost pod datim ograničenjima, podrazumeva postojanje *ciljeva*: besciljna racionalnost, racionalnost sama po sebi, u ovom kontekstu ne znači ništa. Ona je moguća samo u odnosu na neki kriterijum koji je *norma* uspešnog rešavanja problema adaptacije. Drugo, optimalnost ponašanja se postiže tako što se pronalazi najbolja moguća funkcija koja odgovara *strukturi okruženja* u kome se adaptacija odvija; *cena* izračunavanja koju kompjutacioni sistem plaća u jedinicama raspoloživih resursa da bi pronašao takvu funkciju mora se uzeti u obzir prilikom ocene optimalnosti njegovog odgovora. Tipična ekonomska logika optimizacije prožima metodologiju racionalne analize u potpunosti.

Rekli smo da se Andersonova racionalna analiza oslanja na ranije teorijske postavke Dejvida Mara. Veza između racionalne analize i metodoloških principa koje je za projekat razvoja kompjutacione teorije viđenja predložio Dejvid Mar (Marr, 1982) je u akcentu koji obojica stavljaju na *analizu okruženja* u kome se odvija adaptacija i *specifikaciju problema* koje kognitivni sistem treba da reši. Mar je, suočen sa kompleksnošću problema u svom radu na kompjutacionoj teoriji viđenja, razvrstao različite *nivoe* na kojima je moguće formulisati neku kognitivnu teoriju. Danas ih po njemu nazivamo *Marovim nivoima.* Dejvid Mar je razlikovao je tri nivoa analize kognitivnih funkcija: *nivo 3*, ili *kompjutacioni nivo*, na kome postavljamo pitanje o tome *koji problem* kognitivni sistem pokušava da reši nekim skupom kognitivnih procesa i *zašto* pokušava da reši upravo taj problem; preciznije, na kompjutacionom nivou analize se postavlja pitanje o tome *koju funkciju kognitivni sistem pokušava da izračuna* da bi njene rezultate upotrebio u adaptivne svrhe. Na nivou 2, koji se još naziva i *algoritamskim nivoom*, proučavamo neposredne *algoritme* i *reprezentacije* za koje pretpostavljamo da ih kognitivni sistem koristi

u izračunavanju funkcije koje smo opisali na nivou 3. Nivo 1 analize problema saznanja je *implementacioni nivo*, gde prema Marovom programu istraživanja u kognitivnoj psihologiji proučavamo neposredno otelotvorenje algoritama opisanih na nivou 2 u neurofiziološkom supstratu kognitivnog sistema (Marr, 1982). Anderson vidi postupak racionalne analize kao postupak koji se odvija na Marovom kompjutacionom nivou, ili nivou 3 (Anderson, 1991a, 1991b). U tom smislu, formalne kognitivne teorije koje će biti predmet naših analiza će biti teorije koje su operacionalizovane, ili ideje koje bi makar u principu mogle da postanu operacionalizovane teorije, na Marovom kompjutacionom nivou analize.

Racionalnost shvaćena kao optimizacija adaptacije ponašanja u okviru racionalne analize, u potpunosti odgovara jednostavnoj formulaciji sa čijom ćemo se analizimo sretati još mnogo puta tokom naše rasprave: *kognitivno je racionalan subjekt* S *čije ponašanje B konzistentno svedoči o tome da on dela u skladu sa svojim verovanjima* $\psi$, *kako bi ostvario svoje ciljeve G u nekoj sredini E.* U ovoj tvrdnji, dovoljno je dodati da je ispravnost verovanja $B$ jedna vrsta resursa (informacionih), te da akcije subjekta $S$ jesu ograničene raspoloživim resursima u smislu efikasnosti, da bi se ona proširila u optimalnost Andersonove racionalne analize. U kojoj meri, i na koji način, ovako shvaćen koncept racionalnosti odgovara analizama kognitivne psihologije kao pokušaja razvoja prirodne nauke o saznanju, jedno je do osnovnih pitanja na koje tražimo odgovor u raspravama koje slede.

## 2.3 Ciljevi

Osnovni cilj naše rasprave je da se odredi *naučni status koncepta racionalnosti saznanja u kompjutacionoj kognitivnoj psihologiji.* U tom smislu, racionalnost, kako je shvaćena u temeljima programa kompjutacionističke kognitivne psihologije, u okvirima metodologije racionalne analize, ali i kako je još mogla biti shvaćena u alternativnim teorijskim koncepcijama, osnovni je predmet našeg rada. Videćemo da bez detaljne analize veza između teorijskih pretpostavki naučne teorije kognitivne psihologije i pretpostavki *merenja*, odn. eksperimentalne, bihejvioralne metodologije kojom se konstituišu ključni empirijski fenomeni, nije moguće adekvatno odgovoriti na ovo pitanje.

Specifični ciljevi naše rasprave su sledeći:

A. *Odrediti naučni status koncepta racionalnosti koji se nalazi u osnovama standardne paradigme kompjutacione kognitivne psihologije.* Specifično, kritičkom

diskusijom standardne teorije kompjutacione kognitivne psihologije i raznih pravaca unutar nje, treba utvrditi koji se status pridaje racionalnosti ljudskih kognitivnih stanja i procesa u ovom teorijskom okviru. Takođe, potrebno je odrediti prirodu eksplanatorne moći koncepta racionalnosti u odnosu na druge koncepte kompjutacionih kognitivnih teorija.

B. *Odrediti naučni status koncepta racionalnosti u odnosu na savremene empirijski motivisane diskusije o racionalnosti kognitivnih procesa.* Specifično, metateorijskom i logičko-metodološkom analizom teorija i empirijskih nalaza *pro et contra* racionalnosti viših i simboličkih kognitivnih procesa, potrebno je ustanoviti zašto je i kako moguće da različite teorije, oličene u različitim matematičkim modelima, sa različitim implikacijama i različitim parametrima, objašnjavaju empirijske nalaze dobijene primenom sličnih ili istih metodologija. Analiza odnosa prirode eksperimentalne metodologije u kognitivnoj psihologiji i načina na koji se empirijski nalazi povezuju sa bitnim teorijskim konceptima igraće ključnu ulogu u ispunjenju ovog cilja.

C. *Odrediti istorijski razvoj koncepta racionalnosti koji rezultira savremenim shvatanjem ovog pojma u dominantnoj paradigmi kompjutacione kognitivne psihologije.* Ispunjenje ovog cilja istraživanja treba da obezbedi relevantan kontekst za razumevanje koncepta racionalnosti koji je karakterističan za kompjutacionu kognitivnu psihologiju.

D. *Odrediti granice analize racionalnosti saznanja u okviru kompjutacione kognitivne psihologije.* Najvažniji cilj je ovaj: potrebno je ustanoviti odnos između racionalnosti kao osnovne pretpostavke izgradnje kompjutacione kognitivne teorije uma ($\textsc{Racionalnost}_1$) i racionalnosti kao predmeta empirijskog istraživanja ($\textsc{Racionalnost}_2$) pod pretpostavkama jedne tako izgrađene teorije uma. Preciznije, potrebno je odrediti granice i mogućnosti empirijske analize racionalnosti u odnosu na situaciju u kojoj se racionalnost javlja na dva navedena nivoa analize. U tom smislu, ispunjenje ovog specifičnog cilja povezano je sa ispunjenjem cilja opisanog pod B.

E. Uzimajući u obzir centralno mesto koje po važnosti koncept racionalnosti zauzima u kompjutacionoj kognitivnoj psihologiji, na osnovu rezultata svih prethodnih analiza doneti *ocenu njenog položaja u proučavanju ljudskog saznanja u prirodnim naukama.*

# 3 Sinopsis

Izlaganja i diskusije koje slede organizovani su na sledeći način.

U II delu predstavljamo i kritički diskutujemo *naučnu paradigmu kompjutacione kognitivne psihologije*. Standardna *simbolicistička paradigma* kompjutacione kognitivne psihologije predstavlja sve ljudske saznajne sposobnosti kao oblike izračunavanja u nekom formalnom sistemu simbola. Teorijsku konstrukciju standardne paradigme obezbeđuju (a) kompjutacionistička pretpostavka, oličena u hipotezi fizičkih sistema simbola i kompjutacionoj teoriji uma, (b) metodologija kojom se eksperimentalni, bihejvioralni rezultati povezuju sa parametrima i arhitekturom postuliranih, neopservabilnih kognitivnih (pod)sistema u okviru opšteg kibernetičkog pristupa *analize crne kutije* odn. inženjerisanja uma unazad (engl. *reverse-engineering*). Suštinske logičko-metodološke analize koje sprovodimo u ovoj tezi proučavaju odnos bihejvioralne metodologije i postuliranih neopservabilnih podsistema ljudskog saznajnog aparata; zato je izlaganje u II delu veoma značajno za razumevanje diskusija koje slede. U ovom delu rada diskutujemo i odnos standardne paradigme sa dve alternativne kompjutacione koncepcije: konekcionizmom (i sa njim povezanim emergentizmom) i konstruktivizmom (oličenim u enaktivističkoj paradigmi). Zaključne diskusije u ovom poglavlju predstavljaju uvide u ograničenja analize racionalnosti u paradigmi kompjutacione kognitivne psihologije i uvode metateorijski konceptualni okvir za analizu normativnih i deskriptivnih teorija.

U III delu predstavljamo kritičku diskusiju *debate o racionalnosti*. Pružamo kritičku, metateorijsku i metodološku diskusiju tipičnih normativnih i deskriptivnih teorija u oblasti viših i simboličkih kognitivnih funkcija: odlučivanja, kauzalnosti, epizodičkog pamćenja, suđenja i rezonovanja, kategorizacije i učenja koncepata i kreativne funkcije simboličkog sistema. Ovo poglavlje, zahvaljujući širini problematike koju obuhvata, osvetljava sve probleme analize racionalnosti saznanja u oblasti viših i simboličkih kognitivnih funkcija.

IV deo izlaganja otvara mogućnost da iz istorijske perspektive sagledamo naučni status savremene debate o racionalnosti. Dok izlaganje i diskusije u prethodna dva dela teze obezbeđuju kritički uvid u savremenu debatu o racionalnosti, u IV delu pokušavamo da obogatimo naše razumevanje određivanjem pogodnog istorijskog konteksta. Upravo u ovom poglavlju počinjemo razvoj suštinskih argumenata protiv mogućnosti dosledne i smislene racionalne analize u domenu viših i simboličkih kognitivnih funkcija. Proučavamo proces *naturalizacije uma*,

odn. *razvoja razumevanja ljudskog uma kao prirodnog sistema*, obraćajući posebnu pažnju na vezu između razvojnog puta koncepta racionalnosti i načina na koji je on praktično operacionalizovan u raspravama matematičara, filozofa i logičara u prvoj polovini XX veka. Pokazuje se da savremeno shvatanje racionalnosti kao optimalne alokacije kompjutacionih resursa u izračunavanju optimalnog adaptivnog odgovora ima duboke korene u razvoju prirodnih nauka od naučne revolucije XVII do kompjutacionističke revolucije XX veka.

V deo rada je eksperimentalne i logičko-metodološke prirode. U ovom delu pokušavamo da razvojem jedne specifične bejzijanske teorije odlučivanja u uslovima rizika „racionalizujemo" neke od fenomena koje savremene deskriptivne teorije tretiraju kao posledice ljudske ograničene racionalnosti u ovom domenu. Predstavljamo originalne rezultate primene standardne eksperimentalne metodologije u odlučivanju (eksperiment *ocene monetarnih ekvivalenata* i eksperimente *izbora*) i diskutujemo metodološke pristupe dizajnu ovakvih eksperimenata. Na osnovu prikupljenih podataka i podataka do kojih su došla druga istraživanja, sprovodimo i analiziramo proces selekcije modela teorije odlučivanja. Poređenje različitih teorijskih strategija u objašnjenju odlučivanja i analiza načina na koji se eksperimentalni podaci inkorporiraju u te teorijske, eksplanatorne strategije obezbeđuju ključne uvide u naučni status koncepta racionalnosti saznanja.

VI deo rada predstavlja završnu metateorijsku i metodološku analizu pojma racionalnosti saznanja u okviru kompjutacionističke paradigme. Analiziramo način izgradnje teorija kognitivnih funkcija o čijoj se racionalnosti raspravlja u nameri da otkrijemo kako je moguće paralelno postojanje normativnih i deskriptivnih teorija koje često podjednako dobro objašnjavaju iste eksperimentalne rezultate različitim eksplanatornim mehanizmima. Diskutujemo strategiju racionalne analize (koja po pravilu rezultira zaključkom da su ljudske kognitivne funkcije racionalne) i pokazujemo da su zaključci *pro et contra* racionalnosti bazirani na prethodnim *metateorijskim* odlukama o lokusu eksplanatorne moći u psihološkim teorijama. Postavljamo fundamentalno pitanje o stabilnosti kognitivnih funkcija koje su meta racionalne analize i primenom teorijskog koncepta *mešovite strategije* (poreklom iz teorije igara) u analizi forme kognitivnih funkcija pokazujemo da racionalna analiza, konsekventno izvedena, nikada ne može da dovede do selekcije određenih normativnih ili deskriptivnih modela kao adekvatnih. Diskutujemo i neke mogućnosti rešenja teorijsko-metodološkog problema koji nastaje kao posledica ovog zaključka, kao i pitanje da li je koncept racionalnosti saznanja, zbog svoje inherentne

povezanosti sa strategijom izgradnje naučne teorije o kognitivnim funkcijama, koncept koji uopšte može da ima interpretaciju *unutar diskursa kognitivne psihologije strogo određene kao prirodne nauke.*

VII deo rada predstavlja sintezu celokupne rasprave i predstavlja pokušaj da se pruže odgovori na pitanja postavljena u ovom uvodnom delu rasprave.

# Deo II

# Racionalnost u kompjutacionoj kognitivnoj psihologiji

> „*... polje u psihologiji - malo poznato širokoj publici, a takođe odsutno iz Roamojona, ali polje koje skoro potpuno dominira univerzitetskim scenama - jeste akademska psihologija.* [...] *Radeći najviše u veštačkim eksperimentalnim okruženjima i oslanjajući se svesrdno na pronalaske dobijene u radu sa laboratorijskim pacovima i univerzitetskim zamorcima, istraživači iz ove oblasti su metodično, ciglu po ciglu, sagradili jednu građevinu koja sadrži pregrađene odaje ispunjene spoznajama o tome kako ljudi vide trodimenzionalne objekte, kako uče da kucaju na pisaćoj mašini, kako pamte spiskove reči, kako elektrošokovi mogu da utiču na ponašanje, kao i čitav niz drugih problema koji su takve prirode da se mogu eksperimentalno ispitivati.* [...] *Štaviše, ovo zanemiravanje se može i opravdati*“.
>
> Hauard Gardner, 1980, u prikazu rasprave između Čomskog i Pijažea „*Teorije jezika, teorije učenja*“ (Piatelli-Palmarini, 1975), održane u u opatiji de Roajomon 1975, prvobitno objavljenom u časopisu „*Psychology Today*“.

Prethodni citat iz teksta Hauarda Gardnera najbolje ilustruje motivaciju za *metodološki karakter rasprave o racionalnosti* u okvirima kompjutacione kognitivne psihologije koja sledi. Filozofski i teorijski rad na postavljanju temelja

kompjutacione kognitivne psihologije skoro u potpunosti ignoriše problem merenja. Autori koji se posvećuju analizama osnova kognitivne psihologije kao nauke po pravilu propuštaju analizu onih koraka koji povezuju teorijske konstrukte sa opservabilnim ponašanjem. Doprinos rasprave koja sledi vidimo upravo u njenom fokusiranju na analizu odnosa prirode bihejvioralne metodologije, karakteristične za eksperimentalnu („akademsku“, kako kaže Gardner) psihologiju, i prirode teorijskih pojmova kompjutacione kognitivne psihologije.

Poreklo kompjutacione kognitivne psihologije je u diskusijama u oblastima veštačke inteligencije, lingvistike i filozofije uma koje karakterišu prve decenije druge polovine XX veka. Ove diskusije se, s druge strane, oslanjaju na rezultate postignute u matematici i filozofiji matematike u prvoj polovini veka. Bihejvioralna metodologija, zajednička eskperimentalnoj psihologiji, bihejvioralnoj ekonomiji i drugim naukama o ponašanju, poreklom je starija od te teorijske konstrukcije i oslanja se na dva izvora. Prvi je razvoj metodologije eksperimentalne psihologije koji seže duboko u XIX vek - do prvih psihofizičkih eksperimenata. Drugi je analiza mogućnosti povezivanja bihejvioralnih opservacija sa zaključcima o neopservabilnim teorijskim konceptima koja je karakteristična za fundamentalne rasprave u ekonomiji, teoriji verovatnoće i filozofiji matematike. Nažalost, ove linije razvoja naučne misli u kognitivnoj psihologiji - jedna tipično teorijska i druga tipično eksperimentalna, metodološka - nikada nisu u potpunosti konvergirale. U IV poglavlju rada ćemo detaljnije diskutovati odnos ove dve linije razvoja naučne misli, ali je već sada potrebno da budemo svesni ovog diverziteta tradicija koji već dugo otežava sintezu teorijskih i metodoloških rasprava u psihologiji saznanja.

Našu sintezu teorijske i metodološke rasprave o fundamentima kompjutacione kognitivne psihologije oslanjamo na kritički pregled dve bitne ideje. Prvo poklanjamo pažnju ideji o mogućnosti merenja neopservabilnih subjektivnih verovanja koju razvija britanski matematičar i filozof Frenk Remzi. Verovanje (engl. *belief*) je centralni, suštinski koncept epistemologije i kognitivne psihologije; suvišno je, dakle, napominjati koliki je značaj preciznog definisanja metode merenja ovog koncepta. Remzijev rad je prvi u seriji pokušaja egzaktnog povezivanja opservabilnog ponašanja sa dedukcijom internih, neopservabilnih subjektivnih funkcija koje se odnose na verovanja. On postavlja fundamente na kojima grade de Fineti (1970/74), fon Nojman i Morgenštern (1947), i konačno Sevidž (1954/72). Posle uvođenja ove bitne metodološke ideje, koja je kao po pravilu izostavljenja iz teorijskih i istorijskih pregleda čiji su autori psiholozi, predstavljamo osnove

kompjutacione kognitivne psihologije kroz tri teorijske koncepcije: simbolicizam, emergentizam i kognitivni konstruktivizam[15]. Tek uzeti zajedno, metod za merenje neopservabilnih konstrukata na osnovu opservabilnog ponašanja, i kompjutaciona teorija kognitivnih procesa, predstavljaju naučnu paradigmu savremene psihologije saznanja.

# 4 Metod: merenje subjektivnih verovanja

Kao što će se ponoviti mnogo puta tokom naše rasprave, i ovaj put smo u situaciji da diskusija problema izvorno postavljenih van psihologije ima neposredan značaj za psihološku problematiku. Rad britanskog matematičara, filozofa i ekonomiste Frenka Remzija (1903 - 1930) tradicionalno je zapostavljen u teorijskim i metodološkim raspravama u okvirima psihologije, iako predstavlja kamen temeljac teorije merenja u ma kojoj naučnoj disciplini koja se oslanja na bihejvioralnu metodologiju, tako tipičnu upravo u psihološkim istraživanjima. Remzi je bio jedan od prvih matematičara koji su tokom prve polovine XX veka bili zainteresovani za razvoj tzv. *subjektivističke teorije verovatnoće*. Danas razlikujemo dva osnovna filozofska stava u pogledu prirode pojma verovatnoće: frekvencionistički i subjektivistički. Frekvencionističko gledište, starije od subjektivističkog, vezuje se za same početke matematičkog mišljenja o verovatnoćama i verovatnoću konceptualizuje kao objektivnu meru oličenu u graničnoj vrednosti proporcije $\frac{m}{n}$, gde je $m$ broj „povoljnih događaja“, a $n$ - broj „ukupnih posmatranja (Ellis, 1843). Ukoliko neko želi da izračuna verovatnoću da će kiša padati onda kada on napusti svoj dom sa kišobranom u rukama, potrebno je da podeli broj puta kada je padala kiša ($m$) sa brojem puta kada je izašao iz kuće sa kišobranom ($n$). Kada broj posmatranja teži beskonačnom, kao granična vrednost ove proporcije dobija se, prema frekvencionističkom gledištu, tačna verovatnoća događaja. U frekvencionističkom okviru, verovatnoća je potpuno objektivna: ona nikako ne zavisi od subjektivnih uverenja o ma kojim stanjima sveta u kome se verovatnoća računa. Nasuprot ovom gledištu, subjektivističko gledište (Fishburn, 1986), koje je zainteresovalo Remzija, jeste da je verovatnoća *mera stepena uverenja* (ili verovanja, engl. *degree of belief*; mogli bismo reći i: *snage*, ili *intenziteta verovanja*) koju neki saznajni subjekt ima u odnosu na određeno verovanje (npr: „*Danas će padati kiša*“). Subjektivističko gledište dopušta da razni akteri mogu da imaju različite stepene verovanja u odnosu na ista verovanja; tek posle akumulacije empirijskih

podataka koji opovrgavaju ili potvrđuju njihova verovanja, njihovi stepeni verovanja mogu da konvergiraju ka istoj meri. Ta mera stepena verovanja, kao što smo rekli, za subjektiviste jeste verovatnoća. U čuvenom eseju „*Istina i verovatnoća*“, iz 1926. godine, Remzi je postavio pitanje da li je, i na koji način, moguće meriti stepene verovanja u terminima verovatnoće (Ramsey, 1926). Ignorisanje ovog eseja u razmatranjima istorije psihologije čini se posebno arogantnim ako uzmemo u obzir da je sam Remzi bio potpuno svestan značaja psihološkog aspekta svog rada; on u uvodu u razvoj metodologije merenja verovanja navodi: „*Mi moramo, dakle, da pokušamo da razvijemo čisto psihološki metod merenja verovanja*“[16] (Ramsey, 1926). Kakav je to „čisto psihološki“ metod merenja subjektivnih verovanja koje predlaže Remzi? Uskoro ćemo se uveriti da je taj metod upravo metod koji bi birao akademski psiholog obrazovan da primenjuje standardnu bihejvioralnu metodologiju u psihološkim istraživanjima.

Sledeći primer kojim ilustrujemo Remzijev rezon je preuzet iz njegovog rada „*Istina i verovatnoća*“. Prepostavimo da je neka osoba krenula na put, i da se sad nalazi na raskrsnici; sigurna je da, od dva puta ispred nje, jedan jeste ispravan i vodi ka njenom cilju, ali nije potpuno sigurna koji. Kojim god putem da krene, udaljava se od drugog puta i od polazne tačke; između dva puta koja može da izabere, nalazi se polje kojim često prolaze meštani koji dobro poznaju kraj. Recimo da ta osoba, na osnovu bilo kojih pretpostavki od kojih je mogla da pođe, ili na osnovu puke intuicije, odabrala da krene jednim od dva puta i vidi šta se dešava. Ako vidi nekog čoveka u polju, može da odluči da skrene sa puta i potroši nešto vremena da bi došla do njega i pitala ga koji put vodi ka mestu na koje želi da stigne. Pretpostavimo da smo u stanju da kvantitativno opišemo sledeće parametre ove situacije: neka je $f(d)$ gubitak (u novcu) koji osoba snosi ukoliko skrene sa puta u polje hodajući $d$ metara do meštana kojeg je videla i koji će joj sigurno reći koji put je pravi; neka je $r$ vrednost koja za nju predstavlja dostizanje cilja (birajući pravi put), a $w$ ona vrednost koju stiče dolazeći na drugo mesto (birajući pogrešan put); konačno, neka je $p$ mera njenog *subjektivnog verovanja* da se trenutno nalazi na pravom putu.

Polazeći od teorije očekivane vrednosti, koju je u XVIII veku kritikovao Bernuli, i pažljivo navodeći ograničenja svog postupka koja se odnose na to što ne počiva na hipotezi o očekivanoj korisnosti, Remzi razmišlja na sledeći način: na duge staze, u potencijalno beskonačnom ponavljanju učešća u ovakvoj situaciji, koja strategija se osobi iz primera više isplati? Ukoliko je $n$ broj puta u kojima ona rešava ovakav problem, dolazimo do sledećih pravilnosti:

($i$) Ukoliko osoba nikad ne skrene sa puta koji je unapred odabrala da bi se raspitala za to koji je put pravi, ukupna vrednost za nju je

$$npr + n(1-p)w \tag{9}$$

Objasnimo: $n$ puta je moguće da se zaradi vrednost $r$ stizanja na pravo mesto sa verovatnoćom (subjektivnim uverenjem) $p$ da je na pravom putu i to naravno mora da se sabere sa $n$ puta vrednošću $w$ da je na pogrešnom putu, gde je uverenje osobe da je na pogrešnom putu komplement njenog uverenja $p$ da je na pravom putu, dakle 1-$p$.

($ii$) Ukoliko svaki put odluči da skrene sa puta ka polju da bi se raspitala za to koji je put pravi, ukupna vrednost za nju je

$$nr - nf(d) \tag{10}$$

U drugom slučaju, ne moramo da množimo članove sa subjektivnim uverenjima tj. verovatnoćama da je osoba na pravom ili pogrešnom putu, pošto ona svaki put pita koji je put pravi i saznaje to; $n$ broj puta da se stekne vrednost stizanja na pravo mesto $r$, dakle, treba samo umanjiti $n$ brojem puta u kojima ona snosi gubitak hodanja $d$ metara, u oznaci: $f(d)$.

Posle preuređivanja ovih matematičkih tvrdnji, lako dolazimo do zaključka da postoji relacija koja povezuje ($i$) stepen subjektivnog uverenja $p$ da se osoba nalazi na pravom putu sa ($ii$) opservabilnim varijablama u situaciji koju analiziramo: $p = 1 - \frac{f(d)}{r-w}$. Ako pretpostavimo da su zarade $r$ i $w$ konstantne, kao i da je $r > w$ (jer $r$ je nagrada za stizanje na pravo mesto), možemo da zaključimo sledeće: (1) verovatno niko neće biti spreman da potroši vrednost $f(d)$ - skrećući sa puta po tačnu informaciju - veću od iznosa $r$-$w$, jer time gubi bilo kakvu dodatnu vrednost osvajanja veće nagrade $r$, što izraz $\frac{f(d)}{r-w}$ zadržava između 0 i 1, i (2) vrednost $p$ je direktno povezana sa izborom vrednosti $d$ metara skretanja sa puta: što veće skretanje sa puta je neko spreman da podnese da bi saznao sa sigurnošću koji je put pravi, to je njegovo subjektivno uverenje da je na pravom putu manje. Ono što je suština Remzijevog pristupa je sledeće: ukoliko je moguće kvantitativno odrediti parametre situacije u kojoj se neka osoba nalazi, i to na prvom mestu kvantitativno izraziti relevantne dobiti i gubitke za tu osobu, kao i parametre njenih akcija („cenu“ informacije u gornjem primeru, tj. razdaljinu $d$ u metrima skretanja sa puta da

bi se pitalo), *onda je moguće posmatranjem ponašanja te osobe otkriti meru njenog određenog subjektivnog verovanja.* U ovom primeru konkretno: sa fiksnim nagradama $r$ i $w$, subjektivno uverenje da se nalazimo na pravom putu je direktno vezano za dužinu skretanja sa puta koju smo spremni da žrtvujemo da bismo saznali koji je put pravi. Pošto je postupak skretanja sa puta i hodanja $d$ metara opservabilan, *sledi da na osnovu opservabilnog ponašanja možemo da saznamo nešto o neopservabilnim konceptima.*

Vrednostima $r$, $w$ - tj. visinama nagrada - možemo eksperimentalno da manipulišemo, i pod pretpostavkom da uvek imamo na raspologanju neko opservabilno $f(d)$ - u gornjem primeru, to je cena dobijanja tačne informacije - možemo da ocenimo intenzitet neopservabilnog verovanja $p$ na osnovu gore navedenog izraza. Uopšte, Remzi konstatuje da posmatranjem realnih izbora koje neka osoba donosi možemo u terminima verovatnoće da ocenimo meru njenog stepena verovanja da su određena stanja sveta takva kakva jesu. Da iskoristimo u svrhe ilustracije primer koji će nekoliko decenija kasnije posle Remzijevog rada navesti Leonard Džimi Sevidž u čuvenim „*Osnovama statistike*“ (Savage, 1954/1972), u kojoj takođe razvija subjektivističku teoriju verovatnoće: uzećemo dva jajeta iz frižidera, jedno više braon i drugo više bele boje. Sve troškove koji mogu da nastanu tokom ispitivanja vaših uverenja ćemo vam nadoknaditi (npr. dobićete dva jajeta istog kvaliteta ukoliko budemo morali da ih polupamo). Pitamo vas: koje od dva jajeta je dobro, a koje pokvareno? Posle vašeg odgovora, razbićemo ih oba i ustanoviti da li ste bili u pravu; ako ste pogodili koje jaje je pokvareno, zaradićete 1 dolar. Kao što kaže Sevidž, pod ovakvim uslovima ispitivanja, ukoliko osoba odluči da se kladi na braon jaje u pokušaju da zaradi dolar, to više nego dobro korespondira sa uobičajenim načinom govora da je *za tu osobu subjektivna verovatnoća* da je braon jaje pokvareno *veća od subjektivne verovatnoće* da je pokvareno belo jaje. Polazeći od ovog intuitivnog pojma verovatnoće, koji naziva kvalitativnom verovatnoćom (i koja se, očigledno, meri na ordinalnoj skali), Sevidž pokazuje kako je pod skupom aksioma koji predlaže moguće dokazati da kvalitativne verovatnoće sa intuitivno jasnim karakteristikama možemo reprezentovati kvantitativnim verovatnoćama koje poznajemo iz matematike (tj. verovatnoćama koje poštuju Kolmogorovljeve aksiome tzv. *apsolutne verovatnoće*, Kolmogorov, 1933/56).

Remzi takođe razvija aksiomatski sistem koji obezbeđuje inferencije ka merama intenziteta subjektivnih uverenja osobe na osnovu njenih opservabilnih izbora, tako da ta mera u potpunosti matematički odgovara konceptu verovatnoće. Remzijeva

formalna teorija subjektivističkog zasnivanja verovatnoće za nas je manje relevantna; njegov pionirski rad u teoriji verovatnoće je kasnije unapređen radom Bruna de Finetija (de Finetti, 1970/74) i konačno prevaziđen radom Sevidža (Savage, 1954/1972). Sevidžova teorija subjektivne očekivane korisnosti je od posebnog značaja za psihologe: ona se nalazi u direktnom kontinuitetu sa Remzijevom originalnom idejom, ali u jedinstvenom teorijskom okviru obuhvata i njegove ideje o merenju subjektivne verovatnoće i fon Nojmanovu i Morgnešternovu koncepciju merenja korisnosti i reprezentacije preferencija. Ono što je bitno za našu diskusiju o osnovama kognitivne psihologije je da analiziramo prirodu elementarne strategije koju predlaže Remzi i koju drugi kasnije razvijaju u rafinirane metodologije za merenje verovanja. Pre svega, od suštinskog značaja je povezivanje koncepta izbora sa konceptom *merenja neopservabilnih, latentnih, hipotetičkih konstrukata.* Subjektivna verovanja o verovatnoćama - ili o ma čemu drugom što makar i u principu može da se opiše kvantitativno - su neopservabilna, privatna, dostupna samo introspektivnim naporima neke osobe. Ključno za razvoj bihejvioralne metodologije jeste postaviti principe koji omogućavaju merenje takvih neopservabilnih koncepata[17]. Remzijev rad predstavlja ranu analizu ovakve mogućnosti, i kroz određenu matematičku formalizaciju, on pokazuje kako je na osnovu posmatranja izbora (preferencija) moguće donositi ocene o neopservabilnom stepenu uverenja osobe čiji se izbori posmatraju. Ako se sada prisetimo diskusija iz I poglavlja, uvidećemo da je strategija aksiomatizacije izbora u uslovima rizika fon Nojmana i Morgenšterna potpuno ista - sa tom razlikom da oni pretpostavljaju da su verovatnoće objektivno date, ali *neopservabilnu funkciju korisnosti* neke osobe *izvode na osnovu opservabilnih preferencija* te osobe. Zato se za fon Nojmanov i Morgenšternov rad kaže da postavlja temelje *reprezentacione teorije merenja*: neopservabilna funkcija korisnosti u njihovoj teoriji reprezentuje one izbore koji zadovoljavaju vNM aksiomatiku. Remzijev esej „*Istina i verovatnoća*“ iz 1926. zapravo ostaje nezapažen sve do drugog izdanja „*Teorije igara i ekonomskog ponašanja*“ iz 1947 u kojem je predstavljena aksiomatika izbora u uslovima rizika. Tek Sevidžov rad ujedinjuje ova dva pristupa, pokazujući uslove pod kojima je moguća identifikacija i subjektivnih korisnosti i subjektivnih verovatnoća u jedinstvenom matematičkom okviru (Savage, 1954/72, Fishburn, 1986).

Druga bitna odlika Remzijeve analize mogućnosti merenja neopservabilnih subjektivnih uverenja je veza između prirode njegove metode i prirode metodologije racionalne analize koju smo predstavili u I delu. Već Remzijeva analiza pretpostavlja

da je moguće posmatranjem ponašanja izmeriti neopservabilni stepen verovanja ako je osoba čije se ponašanje posmatra *motivisana*: podsetimo se, za merenje subjektivnih verovanja je potrebno poznavati vrednosti dobitaka koji se nalaze na krajevima alternativnih puteva u Remzijevom primeru, kao i cenu skretanja sa odabranog puta da bi se stekla ključna informacija za dostizanje cilja veće vrednosti. Sevidž, decenijama posle Remzija, u izgradnji teorije subjektivne očekivane korisnosti poklanja izuzetnu pažnju ovoj činjenici, aksiomatski osiguravajući u svom sistemu to da osoba neće izmeniti svoje preferencije ukoliko pokušava da pogodi koje je jaje pokvareno za nagradu od 1$ ako joj ponudimo nagradu od 10$. Pretpostavka - koja je u Remzijevom eseju eksplicitna - da će svako, ispravno motivisan, pokušati da *osvoji što više određene vrednosti* je moguće u nekoj situaciji, neophodna je, i garantuje plauzibilnost inferencija o neopservabilnim stepenima verovanja na osnovu opservabilnih izbora: ova pretpostavka daje „logiku“ celoj ovoj ideji merenja subjektivnih verovanja. Na Marovom kompjutacionom nivou 3 analize kognitivnog sistema, zadatak naučnika je da ustanovi *šta* kognitivni sistem pokušava da izračuna: on zato mora da zna kako je organizam čiji kognitivni sistem proučava motivisan, odn. šta u određenoj sredini za njega predstavlja adaptivnu vrednost koju treba steći ili uvećati. Bez pretpostavke o okolini sa ograničenim resursima, u kojoj je funkcija kognitivnog sistema evolutivno određena kao optimizacija ponašanja u odnosu na distribuciju tih resursa, ni Marova analiza po nivoima kognitivne teorije, ni Andersonova racionalna analiza, ni matematička operacionalizacija metodologije merenja neopservabilnih verovanja, nemaju smisla. Vidimo koliko je više puta pomenuto shvatanje racionalnosti, sadržano u iskazu da je *kognitivno racionalan subjekt $S$ čije ponašanje $B$ konzistentno svedoči o tome da on dela u skladu sa svojim verovanjima $\psi$, kako bi ostvario svoje ciljeve $G$ u nekoj sredini $E$*, prisutno u samim metodološkim osnovama kognitivne psihologije i drugih nauka o ponašanju. Upravo to prisustvo ovog principa u *metodološkim osnovama* - dakle, na mestu gde biramo *sredstva izgradnje naučne teorije*, ne još u *sadržaju koncepata koje teorija opisuje* - motiviše metodološki karakter naše diskusije racionalnosti saznanja. Tokom daljih rasprava u ovoj tezi - analizom značenja i značaja onog poteza u kome *ponašanje subjekta $S$ otkriva* da on dela u skladu sa svojim verovanjima i ciljevima - pokazaćemo koliko je priroda metode od suštinskog značaja za razumevanje samog problema racionalnosti saznanja.

# 5 Teorija: kompjutaciona kognitivna psihologija

Posle ove „rasprave o metodi“, koja je i iz istorijskih (njeno zapostavljanje u teorijskom diskursu) i konceptualnih (logički sled naše diskusije) razloga morala da prethodi narednim redovima, počinjemo da uvodimo teorijsku koncepciju *kompjutacione kognitivne psihologije* (u daljem tekstu: KKP). Diskurs KKP je ogroman: on se proteže kroz više disciplina, poput psihologije, filozofije uma, filozofije jezika, kompjutacione lingvistike, matematičke logike i kompjuterskih nauka. Naš prvi cilj u narednim redovima jeste da pružimo elaboriran kritički uvid u fundamente KKP, i tako osvetlimo sve one njene osobine koje mogu da budu relevantne za raspravu o racionalnosti saznanja unutar teorijskog diskursa koji ona definiše. Naš sledeći cilj biće da ukažemo na neke prethodno zapostavljene probleme u analizi racionalnosti saznanja pod KKP. Pokazaćemo da je zapostavljanje tih problema posledica izostanka pažljive analize odnosa između bihevioralne metodologije merenja internih hipotetskih konstrukata (verovanja) i teorijskih osnova KKP.

## 5.1 Standardna paradigma

Kompjutaciona kognitivna psihologija je teorijska konstrukcija koja počiva na nekoliko međusobno povezanih, zavisnih konceptualnih konstrukcija. Smatramo da je najbolji način za njeno uvođenje slediti redom ekspoziciju teorije folk-psihologije (engl. *Folk Psychology*, skr. FP), kompjutacione teorije uma (engl. *Computational Theory of Mind*, skr. CTM), hipoteze o fizičkim sistemima simbola (engl. *Physical Symbol Systems*, skr. PSS) i hipoteze o jeziku uma (engl. *Languge of Thought Hypothesis*, skr. LOTH). Redosled našeg izlaganja narušava logički red: folk-psihologija zapravo počiva na teorijama i hipotezama nabrojanim posle nje, ali izlaganje ideja folk-psihologije smatramo najpogodnijim da motivišemo izlaganje fundamentalnijih ideja. Pojam formalnih sistema, jedan od najvećih doprinosa matematike i logike XX veka, igra prominentnu ulogu u razvoju standardne paradigme KKP.

### 5.1.1 Folk-psihologija: um kao formalni sistem verovanja

Pretpostavite da se nalazite u hodniku nekog fakulteta čiji stakleni spoljašnji zid gleda na ulicu. U hodniku se nalazi mašina za kafu i druge napitke. Neka osoba, verovatno student, prilazi mašini i posmatra spisak pića koja mogu da se dobiju.

Hvata se za džep, vadi novčanik, pregleda ga kratko, zatim zatvara novčanik i vraća ga u džep. Potom napušta zgradu fakulteta, i vi je vidite kroz staklo kako odlazi do kioska na ulici, i zatim se vraća do zgrade fakulteta. Uskoro, vidite je ponovo u hodniku, prilazi mašini za kafu i sokove, vadi novčanik, ubacuje nekoliko novčića i podiže jednu flašicu mineralne vode. Na osnovu posmatranja ovakvih događaja, um zrele, psihološki normalne osobe, u stanju je da odmah sastavi eksplanatorni narativ koji daje objašnjenje za sve korake koji su opisani u ovoj situaciji. Prvo, osoba koja prilazi mašini za kafu, sokove i vodu, to verovatno čini zato što joj je potrebna kafa, ili je žedna. Na osnovu toga da li ona kupuje kafu, ili vodu, možemo da zaključimo preciznije o kom motivu je reč. Dalje, na osnovu čina pregleda novčanika, napuštanja zgrade, odlaska do kioska, i povratka posle kog ubacuje sitan novac u slot, naš um proizvodi deo eksplanatornog narativa koji tvrdi da je osoba verovatno bila bez sitnog novca, da je rezonovala da novac može da rasitni u radnji, da je to učinila, da bi se vratila do mašine da kupi šta joj je potrebno.

Folk-psihologija (engl. *Folk Psychology*, skr. FP) počiva na ideji da je naučna teorija o tome kako funkcioniše ljudski um uopšte *samo naučna elaboracija svakodnevnog rezonovanja* koje smo ilustrovali prethodnim primerom. Za naučnu elaboraciju, potrebni su naučni koncepti, i mi ćemo uskoro videti kako se naučni koncepti neophodni za FP nalaze u matematičkoj logici, teoriji verovatnoće i kompjuterskim naukama. Osnovna jedinica analize u FP, i osnovno mentalno stanje čije odnose sa drugim mentalnim stanjima ona analizira, jeste stanje *verovanja*. Verovanja mogu biti različite forme. Neka od njih su verovanja o motivacionim stanjima: da smo gladni, da smo žedni, da nešto ne razumemo, da smo besni, da nešto želimo da postignemo. Ovakva verovanja predstavljaju naše *ciljeve*, *želje* i *motive*. Na osnovu pretpostavke da se svi ljudi rukovode sličnim postupcima u zadovoljavanju svojih ciljeva, želja i motiva, polazeći od toga na koji način mi koristimo raspoloživa znanja o svetu i drugim osobama, mi smo u stanju i da analiziramo ponašanje drugih onako kao što smo činili u prethodnom primeru. Druga vrsta verovanja predstavljaju naš odnos prema *mogućim stanjima sveta*. Mi možemo da verujemo da je mačka na krovu, ili da je Napoleon kralj Francuske. Mi verujemo da je neophodno ubaciti novac u mašinu za kafu da bismo mogli sa nje da podignemo kafu. Osoba iz primera se oslanja na svoja verovanja o tome da će unošenje vode redukovati njenu žeđ, da je za kupovinu na automatu potreban sitan novac, da će novac naći u svom novčaniku, da novac može da usitni u nekoj prodavnici, itd. Mi, pošto delimo zajednička znanja sa tom osobom, i možemo da pravimo hipoteze o

njenim internim stanjima na osnovu opservacije njenog ponašanja, ujedno možemo i da objasnimo njeno opservabilno ponašanje pozivanjem na različita stanja verovanja i relacije među njima. Prema FP, uopšte sve što se tiče objašnjenja ponašanja se svodi na inferenciju i razumevanje odnosa verovanja o motivima (ciljevima, željama) i stanjima sveta. Znanja koja imamo o svetu koji nas okružuje omogućavaju da razvijemo nove motive: da bismo kupili kafu, moramo da imamo sitan novac, ako nemamo sitan novac, motivisani smo da ga usitnimo, itd. Relacija *kauzalnosti* je ključna u pravilima na osnovu kojih se vrše inferencije u analizi *u suštini teleološki određenog ponašanja* u okviru FP: ide se na kisok *da bi* se usitnio novac, novac se sitni *da bi* se iskoristio u automatu, kafa se kupuje *da bi* se podigao nivo budnosti pred ispit, i sve „da bi“ veze u prethodnim tvrdnjama opravdane su verovanjima u određene kauzalne regularnosti sveta. Poznavanje kauzalnih pravilnosti u svetu, odn. poznavanja veza uzroka i posledica u njemu, i kauzalnih pravilnosti koja povezuju verovanja, ljudsko ponašanje i okolinu je ključan proces u postavljanju FP objašnjenja ma kog čina. Zahvaljujući poznavanju kauzalnih odnosa između realizacije određenih propozicija, naših mentalnih dispozicija i propozicionih stavova, mi smo u stanju da ostvarimo tri suštinski bitne adaptivne funkcije: *kontrolu okoline*, *objašnjenje okoline*, i *predviđanje događaja u okolini*. Ako ne želimo da se prehladimo, zimi ne treba da stavljamo led u sokove i vodu: poznavanje ovog kauzalnog odnosa nam omogućava *kontrolu* sopstvenog zdravstvenog stanja. Ukoliko želimo da iznerviramo Partizanovca, verovatno je dovoljno da provedemo veče u baru sa njim prepričavajući mu sve uspehe Crvene zvezde. Ukoliko vidimo nekoga da to čini, možemo sa određenom verovatnoćom da *predvidimo* da će uspeti da iznervira navijača Partizana. Naša drugarica se razbolela zato što je pila hladno pivo na koncertu u februaru: ovo *objašnjenje* nečijeg zdravstvenog stanja ponovo dugujemo poznavanju kauzalnih pravilnosti sveta.

FP analiza postaje još bliža empirijskoj realnosti kada se postavi u probabilističkim terminima. Jedan razlog za to je što tačnost inferencija o socijalnom i fizičkom svetu zavisi od mnogobrojnih okolnosti, a drugi taj što subjekti najčešće ne poseduju kompletnu informaciju o svetu. Osoba iz primera je mogla da bude zaposlena u marketing agenciji koja istražuje rasprostanjenost modela automata za kafu i sokove raznih proizvođača. Njoj je takođe mogao da bude neophodan sitan novac, pa je zato otišla do kioska da rasitni novčanicu. Ali, njena kupovina mineralne vode je način da se odgovori na pitanje u upitniku koji primenjuje, naime, da li je registrovani automat ispravan ili nije; ona uopšte nije bila žedna. Kako

razlikovati ovaj eksplanatorni narativ od prethodnog? Prvo, na osnovu dopunskih informacija, ako možemo da ih prikupimo. Na primer, osobe koje rade ovakva istraživanja za marketing agencije najčešće nose neka obeležja tih agencija, npr. traku oko vrata. Ako imamo takvu informaciju, možemo da isključimo jedan od dva eksplanatorna narativa. Međutim, ako su sva naša verovanja, i sva verovanja drugih koja analiziramo, *probabilistička*, tj. sadrže i meru stepena uverenja da je neko stanje sveta takvo i takvo, onda i eksplanatorne sheme koje razvijamo uzimaju probabilističku formu. Tako je događaj koji objašnjava prvi eksplanatorni narativ toliko verovatan da bismo bili iznenađeni da saznamo da je drugi scenario zapravo tačan: on ima daleko nižu verovatnoću da se ostvari od prvog. Možda je vreme da se podsetimo da zahvaljujući radu Remzija i drugih poznajemo i metodologiju koja nam omogućava da kroz kontrolisane eksperimente izvedemo meru stepena (neopservabilnih) verovanja, što nam dalje omogućava da testiramo hipoteze o tome u koju od alternativnih eksplanatornih shema neka osoba veruje u nekoj situaciji.

U FP razlikujemo *propozicije* i *stavove*. Propozicije su tvrdnje o svetu koje mogu biti tačne ili netačne: „mačka je na krovu“, „čaša je na stolu“, „Napoleon je osvojio Evropu“, „ljudi su smrtni“. Stavovi su mentalne dispozicije, *subjektivni odnosi prema propozicijama*: „*verujem da* je čaša na stolu“, „*smatram da* Napoleon jeste osvojio Evropu“, „*voleo bih* da je mačka na krovu“, „*želim da* je čaša na stolu“. Upravo navedene tvrdnje su primeri *propozicionih stavova*, koji se dobijaju kombinacijom stavova i propozicija: „voleo bih da je mačka na krovu“ je kompozicija mentalne dispozicije „*voleti da* je situacija takva i takva“ i propozicije „*mačka je na krovu*“. Dakle, postoje stanja sveta, i naši mogući odnosi prema stanjima sveta: kada se neki moguć odnos prema određenom stanju sveta realizuje, mi smo u stanju određenog verovanja, i realizuje se objekat koji se naziva propozicionim stavom. Folk-psihologija se često naziva i *psihologijom propozicionih stavova*[18].

FP, kao što vidimo, opisuje domen mentalnog i ponašanje upotrebom predikata koje koristimo svakodnevno, u prirodnom jeziku, i koji korespondiraju našim željama, ciljevima, činovima i stanjima sveta. Na dubljem nivou opisa, FP pretpostavlja da je domen mentalnog *realizacija određenog formalnog sistema*. Ideja o formalizaciji mentalnog otvara put ka realizaciji standardne paradigme kao naučne teorije u *simboličkom, matematičkom smislu*, dok kompjutaciona teorija uma, videćemo uskoro, predstavlja predlog o njenoj realizaciji u *fizičkom smislu*, odn. omogućava naturalističku teoriju uma kao specifične vrste prirodnog sistema. Standardna paradigma KKP pretpostavlja mogućnost deskripcije našeg

mentalnog života kao određenog *formalnog sistema* koji je dovoljno moćan da reprezentuje sve neophodne *varijable* i *relacije*, sredstvima *formalnog jezika* tog sistema. Šta su formalni sistemi? Prema Haugelandu (Haugeland, 1981, 1985), svaki formalni sistem se može razumeti kao igra manipulacije tokenima[19], i takva igra ima tri suštinske odlike: (i) *ona je igra manipulacije tokenima*, (ii) *ona je diskretna*, i (iii) *ona je konačna*. Sve ove tri odlike imaju specifično značenje u teoriji formalnih sistema. Prva odlika definiše suštinu formalnih sistema: oni su skupovi određenih tokena i određenih pravila manipulacije tim tokenima. Postoje dozvoljene i nedozvoljene manipulacije: u engleskom jeziku, pozicioniranje objekta pre subjekta u rečenici je nedozvoljena manipulacija, u šahu, pomeranje pešaka za tri polja je nedozvoljena manipulacija. Formalna pravila zapisana u egzaktnom, simboličkom jeziku, predstavljaju ograničenja skupa mogućih manipulacija tokenima kao osnovnim, atomarnim elementima sistema. Druga odlika formalnih sistema je da su oni *diskretni*. Ne postoji nikakva kontinuiranost u formalnom sistemu: svaka njegova primena može da se svede na niz isključivih, zasebnih odluka, i celokupno njegovo funkcionisanje može da se opiše kroz niz takvih odluka. Svaki token i tip tokena se tretiraju kao individualne, zasebne, nedeljive celine na koje se određeno pravilo formalnog sistema ili primenjuje, ili ne: ova isključivost u izboru i precizno razgraničenje različitih tokena i tipova tokena jeste suština diskretne prirode formalnih sistema. Na kraju, formalni sistemi su *konačni*: oni ne zahtevaju beskonačne resurse niti primenu sredstava koja su van formalnog sistema koji posmatramo. Svako pravilo manipulacije tokenima u okviru formalnog sistema proizvodi samo neku novu konfiguraciju tokena u tom sistemu i to konačnim sledom koraka.

Prema strogoj, u matematici uobičajenoj definiciji formalnih sistema svaki formalni sistem je skup koji sadrži

(*i*) *skup simbola* koji se koriste u izgradnji rečenica u jeziku tog formalnog sistema,

(*ii*) *skup sintaksičkih pravila* (gramatiku) prema kojima se formiraju rečenice u jeziku tog formalnog sistema,

(*iii*) *skup aksioma* - polaznih, elementarnih tvrdnji tog formalnog sistema čija se istinitost smatra očiglednom, i

(*iv*) *pravila inferencije* kojima se iz određenih tvrdnji (u početku, isključivo aksioma) grade druge tvrdnje, odn. *teoreme* tog formalnog sistema (Herre & Schroeder-Heister, 1998).

Formalni sistem tako čini jedan jezik za izražavanje istina o određenom domenu za koji on pruža *egzaktnu deskripciju*. Svi formalni sistemi su apsolutno apstraktni: oni se međusobno razliku samo po *strukturi* koju opisuju sve četiri nabrojane komponente - *bez obzira na to šta ta struktura uopšte znači*. Čuveni formalni sistem Peanove aritmetike, na primer, obuhvata četiri grupe aksioma od kojih se suštinski samo jedna odnosi na osobine prirodnih brojeva (preostale tri obuhvataju logički račun sudova, upotrebu kvantifikatora i pravilo inferencije *Modus Ponens*, i kao takve su karakteristične za mnoštvo matematičkih teorija, Rucker, 1982/2005). Međutim, aksiome Peanove aritmetike, na najvišem niovu apstrakcije, uopšte ne moramo nužno da posmatramo kao da su *o prirodnim brojevima*. Oni su „prazne konture", pružaju *samo* formu, i u slučaju ovog sistema ispostavlja se da je ta forma pogodna za izražavanje tvrdnji o domenu realnosti koji intuitivno, psihološki poznajemo kao prirodne brojeve. U svojoj suštini, Peanova aritmetika je, dakle, samo jedna struktura relacija. Za nju su prirodni brojevi samo *model* - realnost na koju ona može da referira, i to na način koji zadovoljava našu empirijsku intuiciju.

Formalni sistemi su uvek skupovi pravila manipulacije *simbolima*. U deskripciji, svaki token - kakva god da je njegova fizička realizacija - jeste instanca nekog simbola, nekog tipa tokena. Fizička realizacija šaha zahteva tablu, sto, igrače, njihove pokrete, itd. Međutim, šah, analiziran matematički, logički, *formalno*, ne zahteva nikakvu fizičku realizaciju. On postaje apstraktna matematičko-logička konfiguracija odnosa između simbola za figure i striktnih pravila manipulacije tim simbolima. Svaka šahovska igra postaje jedna procesija konfiguracija simbola na apstraktnom objektu šahovske table koja je striktno uređena pravilima igre. Pretpostavljajući da svi mentalni procesi, u idealnom slučaju, jesu podložni formalizaciji, standardna paradigma KKP predstavlja suštinski *simboličku teoriju uma*.

Svaka matematička teorija jeste formalni sistem. Svaka naučna teorija koja je potpuno matematizovana je formalni sistem. Svaka striktno uređena deskripcija *ma kog stanja sveta* koja odgovara gore diskutovanim principima - da je diskretna i konačna manipulacija nad skupom nekih elemenata - jeste neki formalan sistem. Formalni sistemi su najapstraktniji teorijski entiteti koje analiziraju sistemi matematičke logike, koji su uvek i sami formalni sistemi. Kompletan skup propozicija i stavova, pravila njihove kompozicije, kauzalnih zakona koji njima upravljaju i osnovnih elemenata - objekata sveta i našeg mentalnog života - može da se tretira kao formalan sistem. Takav sistem, kada bismo poznavali njegovu konkretnu formu, predstavljao bi *formalizaciju* FP: on je njeno matematičko, logičko

ruho koje podrazumeva primenu elemenata matematičke logike i teorije verovatnoće u naučnom opisu odnosa između motiva, verovanja i ponašanja. Postupak kojim se ma koji sistem pod analizom, pa i naš mentalni život i ponašanje, opisuje u terminima određenog formalnog sistema, jeste formalizacija.

FP, koju smo upravo diskutovali, predstavlja pre jednu teorijsku paradigmu i istraživački program nego konkretnu psihološku teoriju. U okviru FP, postoje različita shvatanja o tome na koji način bi domen mentalnog mogao da bude adekvatno formalizovan. Veliki broj ovih diskusija pripada oblasti filozofije uma i manje su relevantne za našu diskusiju racionalnosti saznanja[20].

O neposrednoj, egzaktnoj realizaciji FP u nekom formalnom sistemu, odn. o konkretnom formalnom sistemu koji bi predstavljao tačan opis simboličke aktivnosti ljudskog uma, savremena nauka nema mnogo toga da kaže. Postoje mnogi formalni, matematički sistemi koji bi „u principu“ mogli da budu model FP koji odgovara ljudskom umu. Ipak, zbog nepregledne kompleksnosti ljudskog uma i ponašanja koje je on u stanju da proizvede i kontroliše, u psihologiji se po pravilu ograničavamo na proučavanje relevantnih podskupova funkcija i ponašanja. Na ovom mestu korisno je povući razliku između psihološkog objašnjenja na *personalnom* i *subpersonalnom* nivou (Bermudez, 2005). Analize i psihološka objašnjenja na personalnom nivou se odnose na um i ponašanje koje on proizvodi i kontroliše *u celini*: šta neka osoba uopšte čini u nekoj situaciji, da li nešto uopšte opaža kao situaciju jedne ili druge klase fenomena, kojim redosledom organizuje akcije u nekom složenom ponašanju i sl. Rečnik koji koristimo na ovom nivou opisa je molaran u odnosu na rečnik psihološkog objašnjenja na subpersonalnom nivou analize i to jeste rečnik FP. Na subpersonalnom nivou objašnjenja mi zapravo koristimo vokabular savremene KKP: kompjutacione module, specifične kognitivne funkcije i njihove matematičke forme, arhitekture neuronskih mreža i druge eksplanatorne mehanizme. Očigledno, funkcionisanje ljudskog mentalnog sistema opisano na personalnom nivou zavisi od mehanizama i zakona opisanih na subpersonalnom nivou. Generalni pristupi u KKP koji bi mogli da predstavljaju rešenja za realizaciju FP razvijeni su na subpersonalnom, ne na personalnom nivou; oni, dakle, tek pružaju objašnjenja za kognitivne funkcije koje bi mogle da *podrže* realizaciju FP na personalnom nivou. Većina takvih teorija uopšte nisu razvijene pod okriljem psihologije: Perlova teorija kauzalnih modela (Pearl, 2000), na primer, koju ćemo kao normativnu osnovu za kauzalno rezonovanje i učenje diskutovati u III poglavlju, predstavlja veoma moćnu formalizaciju kauzalnih odnosa i probabilističkog rezonovanja uopšte, razvijena je u

kompjuterskim naukama. Formalizacije inspirisane uticajnim rezultatima Čomskog u oblasti kompjutacione lingvistike predlažu istraživački program koji bi vodio otkriću formalnog „jezika uma“ (Fodor, 1975, 2008): hipoteza o jeziku uma, koju ćemo detaljno diskutovati, opet predstavlja „podršku“ kognitivnog sistema sa subpersonalnog nivoa analize na osnovu koje bi na personalnom nivou bila moguća FP. Neki poznati generalni pristupi, poput Andersonove ACT-R arhitekture (Anderson, 1996), razvijeni su direktno za potrebe psihologije. Savremeni bejzijanski probabilistički modeli karakteristika su najnovijih razvoja u KKP; slobodno možemo reći da savremenom KKP dominira primena probabilističkih modela, kao i da najnoviji radovi ukazuju na mogućnost da kauzalni probabilistički modeli predstavljaju kandidate za „veliku teoriju“ viših kognitivnih procesa KKP uopšte (Kemp & Tenenbaum, 2008, Tenenbaum, Kemp, Griffiths & Goodman, 2011).

Bez obzira na neposrednu realizaciju celokupne FP, koja je pitanje tolike kompleksnosti da je moguće da nikad ni ne dobijemo odgovor na njega, ne postoji *a priori* prepreka zašto stvari ne bi mogle da funkcionišu na način predložen u FP. *Prima facie*, FP predstavlja plauzibilnu hipotezu. Neki od najznačajnijih filozofa uma smatraju da je FP daleko više od toga. Polazeći od toga koliko je FP analiza ponašanja koju svakodnevno koristimo da bismo razumeli fizički i socijalni svet oko sebe suštinski značajna za naš život uopšte, i koliko je njena upotreba omniprezentna u ljudskom društvu, Fodor tvrdi:

*„... jer ako zdravorazumska intencionalna psihologija zaista propadne, to bi bila, bez poređenja, najveća intelektualna katastrofa u istoriji naše vrste: ako smo toliko grešili o umu, to je onda naše najveće pogrešno uverenje o ma čemu. Propast natprirodnog, na primer, ne može da se poredi; teizam nikada nije postigao da bude tako intimno povezan sa našom mišlju i našom praksom - posebno našom praksom - kao što je to objašnjenje putem verovanja i želje. Ništa [...] nije blisko jezgru naših kognitivnih sposobnosti kao intencionalno objašnjenje. Bićemo u velikoj, velikoj nevolji ako budemo morali da ga se odreknemo*“ (Fodor, 1987).

I zaista, *intencionalno objašnjenje* - da ljudi čine nešto jer imaju namere da nešto učine, i da na osnovu poznavanja kontigencije i kauzalnosti između namera, znanja, motiva i postupaka možemo da razumemo socijalni svet oko sebe (dok nam je poznavanje kauzalnih zakona dovoljno da razumemo fizički svet) - je toliko suštinsko našoj svakodnevnoj egzistenciji da je teško zamisliti naš svet i naš um bez njega. Međutim, to još uvek ne znači i da je intencionalno objašnjenje jedina moguća *naučna* teorija psihologije.

### 5.1.2 Kompjutacionizam: um kao fizički sistem simbola

Šta je to *izračunavanje*? Izračunavanje (engl. *Computation*) je jedan od toliko apstraktnih pojmova naučne misli da je njegovo definisanje uvek neraskidivo povezano sa diskusijom o tome šta ono jeste. Matematički, izračunavanje je egzaktno definisano, na način koji danas prihvata ogromna većina matematičara. Međutim, za analizu primene ovog koncepta u empirijskoj nauci kakva je psihologija, reći da je izračunavanje ekvivalentno rešavanju skupa problema koji mogu da se postave i reše Tjuringovom mašinom, Lambda računom (Kleene, 1967/1987), celularnim automatom klase 4 (Wolfram, 2002), ili nekim drugim, ekvivalentnim formalnim sistemom, nije od mnogo pomoći. Intuitivna analiza značenja ovog koncepta će biti značajnija; veza sa matematičkim formalizmom će postati jasna ako razumemo suštinu ideje izračunavanja onako kako su je razumeli matematičari, logičari i filozofi prve polovine XX veka, omogućavajući tim svojim uvidima jednu od najvećih naučnih revolucija u istoriji nauke uopšte. Ako možemo da razumemo ideju i značaj izračunavanja u kontekstu savremene naučne misli, razumećemo i šta znači da je *univerzalno izračunavanje* moguće u različitim, formalno ekvivalentnim sistemima, od kojih smo neke pobrojali u ovom paragrafu.

Posmatrajmo jednostavan problem kao što je množenje pozitivnih celih brojeva. Formalni zapis 5×4, kao što smo naučili u osnovnoj školi, se zapravo *prepisuje* u sekvencu: 5+5+5+5. Po zakonu komutativnosti množenja, zapis ekvivalentan zapisu 5+5+5+5 je: 4+4+4+4+4. Množenje pozitivnih celih brojeva je procedura koja je *egzaktno* i *konačno* definisana: uzeti jedan od dva broja u zapisu 5×4, i onda sabrati toliko puta drugi broj u tom zapisu sa samim sobom. Ovo *uputstvo* iscrpno - *bez ostatka u objašnjenju značenju onoga što treba učiniti* - definiše operaciju množenja. To da je uputstvo toliko precizno u odnosu na značenje svih postupaka i elemenata nad kojima se ti postupci primenjuju mi označavamo kao *egzaktnost*. Kada smo opisali ovo uputstvo, odn. ovaj *sled instrukcija* za množenje, mi smo definisali jednu *efektivnu proceduru*. Drugi matematički izraz za efektivne procedure jeste *algoritam*. Naša procedura je spisak instrukcija, koraka koje treba izvesti nad određenim *podacima*. Podaci su konkretni pozitivni, celi brojevi u našem primeru. Moguće efektivne procedure nad pozitivnim celim brojevima su posledica njihove strukture, tj. algebarskih osobina skupa celih brojeva, koje su precizno definisane pravilima kao što je komutativnost njihovog sabiranja i množenja.

Množenje pozitivnih celih brojeva je jednostavan primer, ali već on omogućava analizu svih suštinski osobina pojma izračunavanja. Pretpostavimo da proučavamo

mnogo složeniji problem od množenja. Na primer, proučavamo rešenja neke komplikovane diferencijalne jednačine čija bi primena bila značajna u inženjeringu ili nauci. Ukoliko do rešenja jedne takve jednačine možemo da dođemo procedurom koja je konačna i egzaktna, mi onda poznajemo algoritam za njeno rešavanje. Iz srednje škole znamo da za neke diferencijalne jednačine poznajemo efektivne procedure. Bilo koji problem koji može biti postavljen na toliko egzaktan način da je za njega uopšte moguće tražiti egzaktno rešenje u vidu efektivne konačne procedure jeste algoritamski problem. Algoritmi rešavaju algoritamske probleme, a ceo opisani postupak u kome neki algoritam - neki niz preciznih, konačnih instrukcija - operiše nad nekom strukturom, transformišući je u druge strukture dok nisu zadovoljeni uslovi rešenja, jeste *izračunavanje*.

Suštinski uvid za razumevanje pojma izračunavanja i pojma algoritma, za koje vidimo da su praktično ekvivalenti, jeste uvid u njihovu *simboličku prirodu*. Alan Metison Tjuring, britanski matematičar kome dugujemo najveći doprinos razumevanju pojma izračunavanja i njegovoj potencijalnoj primeni u razumevanju ljudske inteligencije, naglašavao je da pod izračunavanjem podrazumeva totalnu specifikaciju svih postupaka koje neka osoba izvodi rešavajući ma koji problem (Turing, 1936). Ovakva aktivnost je uvek simboličke prirode. Namerno smo u primeru množenja prirodnih brojeva iskoristili termin „prepisivanje“: zapis 5×4 se *prepisuje* u sekvencu: 5+5+5+5 (ili 4+4+4+4+4). Sve što se događa u nekom izračunavanju doslovce i bez ostatka jeste *prepisivanje jednog niza simbola u drugi niz simbola prema egzaktnim pravilima*. Medijum izračunavanja dakle uvek jeste *jezik*. Kakav jezik? *Formalan jezik*: jezik nekog formalnog sistema, koji obuhvata skup simbola koji mogu da se koriste, pravila formiranja rečenica, i u kome su izraženi aksiomi koji predstavljaju osnovne istine tog sistema i pravila inferencije koja definišu prenos istinitosti tvrdnji sa aksioma na kompleksnije konstrukcije (teoreme). Posle ovog izlaganja, moć i opštost koncepta izračunavanja bi morali da postanu jasni ako zamislimo formalne sisteme *ma kog reda kompleksnosti*, ali sa definisanim efektivnim procedurama za izvođenje rečenica u njima i efektivnim procedurama za ustanovljavanje njihove istinitosti. Podižući kompleksnost formalnog jezika proizvoljno, dok se krećemo u svetu efektivnih procedura, konačnosti i egzaktnosti, ma koliko složena empirijska realnost koja može biti deskriptivno obuhvaćena takvim jezikom, jeste *izračunljiva*: mi možemo da je razumemo, kontrolišemo, i predviđamo događaje u njoj izračunavanjem.

Svaka naučna teorija mora bar u principu biti podložna matematičkoj deskripciji.

Ova tvrdnja je jednostavno deo definicije diskursa nauke. Ako svaka matematička tvrdnja jeste deo formalnog sistema u kome su istinitosti svih tvrdnji izračunljive, a nauka obuhvata totalitet empirijske realnosti, *svet je izračunljiv.* Upravo iznet stav možemo da razumemo kao najkraću formu filozofije kompjutacionizma, ili kompjutacionističkog stava, koji je implicitno ili eksplicitno karakterističan za epohu u kojoj živimo. Da sve nije izračunljivo, i da kompjutacionizam nije, niti može da bude totalna teorija sveta, već samo jedna metafora, jedna naučna paradigma u Kunovom smislu, jedna metafizika, *zeitgeist* karakterističan za našu epohu, videćemo u IV poglavlju ovog rada.

Koncept *univerzalnog izračunavanja* je od suštinskog značaja za razumevanje kompjutacionizma i njegove empirijske primene. Sledeći prethodno iznetu ideju izračunavanja, jasno je da matematički možemo da konstruišemo različite sisteme koji izračunavaju (rešavaju) različite probleme. Matematički objekti koji su dovoljno apstraktni za generalizaciju u ovom slučaju su *funkcije.* Sve matematičke funkcije su preslikavanja iz domena jedne ili više promenljivih u kodomen druge promenljive. Dobar način da se u kontekstu izračunavanja razmišlja o funkcijama je da se domen shvata kao *input*, a kodomen kao *autput* (engl. *output*) nekog izračunavanja. Istraživanja Tjuringa i drugih pokazala su da je moguće definisati formalne sisteme koji su u stanju da izračunaju *ma koju matematičku funkciju koja uopšte možemo da razumemo kao izračunljivu.* Tjuringova mašina, koja nosi ime po svom tvorcu, jeste takav formalan sistem. Sve prethodno navedene formalizacije (Lambda račun Alonza Čerča, Volframova klasa 4 celularnih automata i druge) su matematički *ekvivalentne* Tjuringovim mašinama. U stvari, *svi sistemi univerzalnog izračunavanja su međusobno ekvivalenti*: svaki je u stanju da simulira rad svakog drugog, i samim time, svaki u stanju da izračuna sve što i ma koji drugi. Saznanje da postoje ovakvih formalni sistemi otvorilo je mogućnost da apstraktna ideja izračunavanja izvrši snažan uticaj na razvoj kibernetike kao opšte teorije sistema, kompjutacione lingvistike i konačno, kompjutacione kognitivne psihologije. Inspiracija je, naravno, sledeća: ako postoje sistemi univerzalnog izračunavanja, onda je *možda inteligencija upravo jedan takav sistem.* Formalni sistemi sposobni za univerzalno izračunavanje predstavljaju ujedno i matematičku, egzaktnu definiciju izračunavanja: izračunavanje je ono što se u takvim sistemima izvodi, izračunljivo je ono što oni mogu da izračunaju: oni jesu izračunavanje. Fizičar i kompjuterski naučnik Stiven Volfram razvija ovu ideju još dalje, sugerišući bitnu hipotezu o empirijskoj realnosti, naime, da univerzalna izračunavanja u svetu nisu retkost,

već da se sistemi sposobni za univerzalna izračunavanja, zapanjujuće jednostavne strukture, nalaze svuda oko nas, mnogo češće nego što bismo očekivali u odnosu na kompleksnost procesa koje su oni u stanju da proizvedu (up. *princip kompjutacione ekvivalencije*, Wolfram, 2002).

Na koji način su ideje formalizacije i izračunavanja omogućile razvoj KKP? Odgovor je: omogućavajući konstrukciju njenog teorijskog jezgra. Kompjutaciona teorija uma i - njoj veoma slična - hipoteza o fizičkim sistemima simbola predstavljaju to konceptualno jezgro. One objašnjavaju kako je moguće da određeni formalni sistem bude implementiran u nekom *prirodnom sistemu* tako da domen ljudskog mentalnog života i ponašanja može da se opiše kako ga opisuje folk-psihologija.

Kompjutaciona teorija uma (engl. *Computational Theory of Mind*, skr. CTM) je filozofska tvrdnja o prirodi ljudskog uma (Horst, 2011). Sva dosadašnje razmatranja o FP i kompjutacionizmu možemo da posmatramo kao jedan predlog teorije *ljudskog rezonovanja*. Počivajući na formalnom sistemu određene snage, ljudski um je u stanju da *reprezentuje* one situacije iz svog socijalnog, biološkog i fizičkog okruženja koje mogu da se formalizuju u jeziku tog sistema. Pošto je u stanju da vrši operacije nad takvim reprezentacijama, odn. - kako tvrdi kompjutacionizam - da vrši izračunavanja po određenim algoritmima u okviru tog sistema, ljudski um postaje sposoban da predviđa moguće događaje, objašnjava ono što se dogodilo, i određuje akcije kojima kontroliše tokove događaja u svojoj okolini. Međutim, sama ideja o ekvivalentnosti ljudskog uma (ili nekih njegovih delova) i nekog formalnog sistema nije dovoljna da objasni *prirodu* ljudskog uma koji tako funkcioniše. Nedostaje teorija o tome kako bi neki materijalni, fizički, biološki supstrat - neuronske mreže našeg mozga, na primer - mogao da funkcioniše na način ekvivalentan izračunavanjima u nekom formalnom sistemu. Nedostaje, za naučnu psihologiju suštinska, *teorija o prirodi mentalnih stanja*, koja bi objasnila kako ona mogu da budu ekvivalentna formalnim izračunavanjima. CTM je takva teorija.

Da bismo razumeli CTM, moramo ponovo da se upustimo u diskusiju o formalnim sistemima. Rekli smo da su sva izračunavanja u formalnim sistemima egzaktna i konačna. Uveli smo još jednu suštinsku, ne toliko očiglednu odliku formalnih sistema, a to je da je odnos između izračunavanja i objekata izračunavanja (simbola na koje se izračunavanja odnose) *odnos čisto sintaksičke prirode*. To znači da su izračunavanja u nekom formalnom sistemu *potpuno nezavisna od interpretacije tog sistema*, odn. od značenja elemenata nad kojima se izračunavanja vrše. Ako

su u igri šaha formulisana pravila upravljanja pešacima, onda je igraču šaha koji ih poznaje sasvim svejedno da li su pioni od drveta, mermera, ili se šah igra hartijicama na kojima su odštampane siluete figura na osnovu kojih ih prepoznajemo. Ma kakva promena značenja elemenata nad kojima procesi formalnih izračunavanja operišu *ne utiče na prirodu tih procesa*; njihov odnos prema elementima izračunavanja, prema tokenima ili tipovima tokena, je sintaksičke prirode, upravo kao što sintaksa (gramatika) prirodnog jezika jeste skup onih pravila koja manipulišu elementima tog jezika bez obzira na njihovo značenje.

Osnovno pitanje koje se postavlja je, naravno, šta se u tom slučaju dešava sa značenjem? Značenje, koje je fundamentalna psihološka kategorija, mora na neki način da bude inkorporirano u ovakav teorijski sistem. Odgovor kompjutacionizma na ovaj problem je možda teorijski najmoćnija osobina ovog sistema. Posmatrajmo *strukturalna značenja*, odn. složena značenja koja uključuju odnose više simbola. Određena konstelacija figura na šahovskoj tabli svakako ima određeno značenje za igrača. Međutim, evolucija te konstelacije je posledica poštovanja striktno formalnih, sintaksičkih pravila igre. Osobina neke konfiguracije simbola koja u ljudskom umu razvija ovakvu ili onakvu interpretaciju i time konstituiše određeno značenje jeste posledica sintaksičkih odnosa između tih simbola. Pretpostavimo da se igrač u šahu nalazi u situaciji koju interpretira kao tešku po sebe. On planira određeni sled poteza i analizira moguće reakcije protivnika da bi poboljšao svoju situaciju. Pretpostavimo da on posle nekoliko poteza uspe da ostvari stratešku prednost nad protivnikom. U formalnoj igri, kakva šah jeste, on to uspeva isključivo formalnom manipulacijom simbolima: naime, ništa drugo ne definiše moguće šahovske poteze i samim time strategije igre do formalnih pravila izračunavanja. Konstelacija simbola - šahovskih figura - koja je interpretirana kao teška, interpretirana je kao teška zbog toga što o njenoj težini svedoče čisto sintaksički odnosi između simbola, koji određuju moguće i nemoguće poteze i izlaze iz situacije; nova konstelacija, koja se interpretira kao prednost, takođe se interpretira kao prednost jer analiza čisto sintaksičkih odnosa ukazuje na neke nove moguće poteze i razvoje; put koji vodi od jedne ka drugoj konstelaciji simbola takođe je put primene formalnih pravila manipulacije simbolima, i ništa više. Kao posledicu ove analize dobijamo suštinski uvid u značaj formalizacije nekog sistema za teoriju ljudske psihologije: *ako se pobrinemo za sintaksu reprezentacionog sistema, njegova semantika će se sama pobrinuti za sebe* (Haugeland, 1981, 1985). Drugim rečima, odnosi između konstituenata u strukturalnim značenjima strikno su vezani za sintaksičke odnose među znacima:

promena sintaksičke strukture u interpretaciji bez ostatka determiniše promenu semantičke strukture. Naravno, ovo podrazumeva da postoji egzaktno određen skup tokena koji predstavljaju konstituente mogućih strukturalnih značenja na koja se argument odnosi, i kao što ćemo videti kasnije, ogroman problem u proučavanju psihološke semantike leži u izostanku odgovora na pitanje o postojanju tako određenog skupa elementarnih tokena (Schyns, Goldstone & Thibaut, 1998).

Precizno određene ove povezanosti sintakse i semantike u nekom formalnom sistemu počiva na matematičkom pojmu *interpretacije* formalnog sistema. Formalni sistemi, sami po sebi, čisto su simboličke konstrukcije. Igra šaha, jednom formalizovana, može da uzme ma koju fizičku realizaciju, uključujući i neku koja ni iz daleka ne liči na šah. Dok se ista formalizacija odnosi na dva različita (realna) sistema, od kojih za oba kažemo da su *modeli* istog formalnog sistema, među njima ne postoji prava razlika u formalnom smislu. Formalni sistemi dobijaju semantiku tek u svojim modelima. Suština veze između sintakse i semantike se očitava kada se o istinitosti nekog izraza u formalnom sistemu razmišlja u terminima nekog njegovog konkretnog modela. Kakva god bila realizacija modela nekog formalnog sistema, istinitost koja se pravilima inferencije nasleđuje kroz sve tvrdnje formalnog sistema (formirane tim pravilima inferencije) će se *nužno* poklapati sa istinitošću odgovarajućih interpretacija u modelu. Semantika formalnog sistema jeste sintaksa nekog njegovog modela, neke njegove interpretacije: tek u modelu ima smisla govoriti o istinitosti, a istinitost odnosa u modelu jeste determinisana sintaksičkim odnosima u formalnom sistemu čiji je to model. U procesu formalizacije ma kog sistema, rešiti sintaksički problem zaista znači rešiti i semantički problem.

Priroda procesa formalizacije, koja otkriva da se doslednim poštovanjem sintaksičkih principa značenja praktično samoregulišu, odn. da regularno prate čisto sintaksički, formalni tok manipulacije simbolima, omogućava CTM da konstituiše značenje kao psihološku kategoriju. Ako ljudski um funkcioniše kao formalni sistem, onda se i interpretacije konstituisane različitim odnosima između simbola tokom formalnih izračunavanja regulišu prateći pravilnosti tih izračunavanja. Ovo čvrsto obavezuje CTM na *sintaksičku teoriju* značenja, odn. teoriju koja sve semantičke fenomene mora da svede na odnose između simbola.

Prvi korak u postavljanju CTM je ovo vezivanje semantike za sintaksu. Proces formalizacije bilo kog sistema koji analiziramo pokazuje kako je to vezivanje moguće. Sada preostaje drugi i poslednji potez: povezivanje čistog formalizma, jedne čisto simboličke strukture, sa kauzalnim zakonima. Naučni zakoni su kauzalni i mi duboko

verujemo u to da pravilnosti odnosa uzroka i posledica uređuju prirodu. Ako je um prirodni sistem, mora postojati naturalistička, kauzalna teorija uma. CTM to jeste.

Suštinski, CTM je *mehanicistčka teorija* uma. Teorijski potez koji povezuje proces formalizacije i formalna izračunavanja sa kauzalnim zakonima sastoji se u uviđanju mogućnosti da se formalna izračunavanja mehanizuju, a prethodno diskutovani koncept univerzalnog izračunavanja pokazuje da je to moguće. Tjuringova mašina, ili drugi ekvivalentni sistemi izračunavanja, koji sami predstavljaju i definiciju izračunavanja, mogu da se projektuju kao automatski formalni sistemi. Štaviše, oni mogu da se projektuju kao sistemi koji su doslovce *mehaničke mašine*. Univerzalna Tjuringova mašina je apstraktni koncept u tom smislu reči što zahteva jedan beskonačan element („traku“ sa koje Tjuringova mašina očitava i na koju upisuje simbole), ali sama formalna izračunavanja svakako mogu da budu realizovana i u konačnom fizičkom sistemu. Mi danas dobro poznajemo mnoge takve sisteme: naime, svaki kompjuter, i mnogi drugi uređaji koji nas okružuju, jesu materijalni, fizički izraz formalnog izračunavanja. Ako smo mi to u stanju da izvedemo u supstratu silikona, koristeći poluprovodnička integralna kola, nema nikakvih prepreka za hipotezu da su priroda i proces biološke evolucije to mogli da izvedu kroz dugotrajnu selekciju i mehanizam slučajnih mutacija u supstratu neuronskih mreža našeg mozga. Zakoni koji omogućavaju da fizički sistemi poput kompjutera obavljaju formalna izračunavanja u onim formalnim sistemima u kojima to nama odgovara su kauzalne prirode: to su jednostavno zakoni fizike. Postoje li kauzalni zakoni čija primena omogućava formalna izračunavanja? Da, i na taj način je moguća veza kauzalnosti i formalizacije. Pošto su značenja već vezana sa sam proces formalizacije, postaje jasno na koji način CTM jeste naturalistička, kauzalna teorija uma: kauzalni procesi mehaničkog izračunavanja ekvivalentni su operacijama formalnih izračunavanja u formalnim sistemima, a intepretacija njihovih stanja zavisna samo od sintaksičkih odnosa među njihovim elementima. Primetimo da iz jedinstvenog skupa hipoteza u CTM okviru sledi mogućnost i *prirodne* i *veštačke inteligencije*.

CTM je u potpunosti naturalistička, naučna hipoteza. Ona se duboko oslanja na kauzalne zakone fizike, tvrdeći da je moguće formalne sisteme realizovati u fizičkom supstratu, iz čega sledi da regularnost procesa formalnog izračunavanja može da uzme kauzalnu formu. Ova tvrdnja je očigledno tačna. CTM nema nikakvih problema sa primenom matematičke deskripcije, što je takođe uslov da neka teorija bude naučna. Naprotiv, moguće formalizacije koje bi predstavljale pogodan opis

predikata FP i funkcionisanje ljudskog uma, pod pretpostavkom da se svi problemi kojima se on bavi mogu formalizovati, su bezbrojne; njihova implementacija u nekom kompjutacionom mediju ne predstavlja nikakav problem, baš kao što ne predstavlja nikakav problem u savremenim kompjuterima. CTM objašnjava na koji način je značenje, kao fundamentalan fenomen psihičkog života čoveka, moguće u umu koji realizuje formalni sistem verovanja. Odnos CTM i FP je u potpunosti kauzalan „odozdo na gore“: formalni sistem uma počiva na kauzalnim zakonima CTM, odn. mehanicističkim procesima izračunavanja u fizičkom supstratu, dok FP, kao teorija uma višeg reda, počiva na tom formalnom sistemu. Dok je FP izražena u predikatima svakodnevne psihologije, bliska našim opisima svakodnevnih situacija u životu sredstvima prirodnog jezika, i dok formalni sistem koji je realizuje precizira te opise u svom formalnom jeziku, CTM bi, kada bi bila kompletno, detaljno razvijena naučna teorija uma, bila izražena u terminima formalnih izračunavanja. Za tako nešto bilo bi neophodno detaljno poznavanje kompjutacione arhitekture ljudskog centralnog nervnog sistema. CTM je način da se pokaže kako je moguć Marov (implementacioni) nivo 1 analize kognitivnih procesa, i kako je moguće da on bude dosledno povezan sa nivoom 2 (algoritamske obrade tj. kognitivnih procesa) i kompjutacionim nivoom 3 (odnosa između zadatka koji postavlja okolina i formalizma koji povezuje input i autput organizma koji taj zadatak rešava). Kao što smo već rekli, sve primene kompjutacionizma pod CTM su, zbog ogromne kompleksnosti sa kojom se suočavamo, ograničene ili na proučavanje određenih podsistema ljudskog kognitivnog sistema, ili na formulaciju apstraktnih kognitivnih arhitektura koje u principu rešavaju mnoge kognitivne probleme (Anderson, 1983) ali su daleko od toga da predstavljaju kompletne, integralne sisteme veštačke inteligencije.

Ovakva razmatranja, koja su posledica razvoja matematičke logike u prvoj polovini XX veka, naučne revolucije koja je razvila koncepte formalnih sistema i izračunavanja, dovela su do danas preovlađujućeg *kompjutacionističkog stava* u psihologiji i kognitivnim naukama uopšte. Njega je, kao i svaki koncept velike opštosti i apstraktnosti, moguće izraziti na više načina. *Hipoteza o fizičkim sistemima simbola* (engl. *Physical Symbol Systems*, skr. PSS) je jedan način da se ovaj stav predstavi. Iako je u osnovi PSS hipoteza formulisana za potrebe kompjuterskih nauka i istraživanja veštačke inteligencije, njen značaj za KKP je očigledan. Formulišu je 1976. Njuvel i Sajmon u klasičnom radu „*Computer Science as Empirical Inquiry: Symbols and Search*“ (Newell & Simon, 1976) na sledeći način:

*„Fizički sistem simbola ima nužna i dovoljna sredstva za opšte inteligentno ponašanje.“*

Šta su PSS? Definicije Njuvela i Sajmona otkrivaju potpuno konceptualno poklapanje sa shvatanjem uma kao prirodnog sistema u CTM koje smo do sada diskutovali. Pre svega, takvi sistemi počivaju na zakonima fizike: u kompjuterskim naukama, oni su proizvod nekog inženjerskog procesa. U psihologiji, rekli bismo da je taj inženjerski proces prirodna, biološka evolucija. Simboli su *fizičke strukture* koje mogu da uđu u složene odnose, tvoreći *izraze*, odn. složene strukture simbola. Struktura simbola sadrži tokene u određenim odnosima. Sistem, pored toga što uvek sadrži određene strukture simbola, sadrži i procese koji operišu nad tim strukturama. Fizički sistem simbola je mašina koja vremenom razvija jedne strukture simbola iz drugih struktura simbola primenjujući operacije nad njima. *Označavanje* je jedna suštinska osobina fizičkih sistema simbola: neki izraz označava neki objekat ukoliko sistem koji sadrži taj izraz može ili da utiče na objekat za koji izraz stoji ili da proizvede ponašanje koje zavisi od tog objekta. Pristup ma kom objektu se u okviru sistema odvija preko izraza koji ga označava. Druga suštinska osobina fizičkih sistema simbola je *interpretacija*. Sistem interpretira određenu simboličku strukturu ukoliko ta struktura označava proces dostupan samom sistemu, i ako sistem koji sadrži takvu simboličku strukturu može da izvede taj proces. Osobina interpretacije je od ključnog značaja za razumevanje fizičkih sistema simbola: ona ukazuje na sposobnost da se reprezentuju sopstveni procesi, i tako upravlja sopstvenim akcijama i ponašanjem. Ova elementarna rekurzija u fizičkim sistemima simbola osnova je sveukupnog inteligentnog ponašanja koje su oni u stanju da proizvedu. PSS, potpuno konzistentno sa CTM, pruža objašnjenje o tome kako *mehanički kompjutacioni sistemi mogu da imaju osobine semiotičkih sistema*, i kako na osnovu tih osobina mogu, preko sopstvenih kompjutacionih procesa, da određuju svoje akcije prema okolini. Iako je PSS suštinski hipoteza o mogućnosti veštačke inteligencije, ona je i hipoteza o mogućnosti KKP: još jednom, ne postoji *očigledan* razlog koji bi isključio mogućnost da prirodni umovi funkcionišu na upravo opisan način.

Fodorova hipoteza o jeziku uma (engl. *Languge of Thought Hypothesis*, skr. LOTH, Fodor, 1975, 2008) predstavlja teorijski najrazvijeniju implementaciju kompjutacione teorije uma. Kao što samo ime kaže, LOTH tvrdi da ljudski um poseduje karakterističan *interni jezik* koji omogućava formalizaciju onih problema koji su tokom evolucije oblikovali naš um. Taj interni jezik predstavlja strukturu

reprezentativne mašinerije našeg uma, koji operacijama nad strukturama tog jezika transformiše simboličke izraze iz jednih formi u druge. Um tako rešava nametnute probleme, što mu omogućava kontrolu ponašanja i okoline, razumevanje i predviđanje. Propozicioni stavovi poput „$S$ veruje da $P$“, „$S$ želi $G$“ i sl. uvek obuhvataju (*i*) subjekat propozicionog stava, odn. identifikuju kognitivnog aktera čiji se propozicioni stav posmatra, (*ii*) rečenicu $P$ koja predstavlja objekat propozicionog stava, i (*iii*) određeni predikat koji dovodi u vezu $S$ sa $P$: „veruje“, „želi“, „namerava“ i sl. Rečenica $P$ je u jeziku uma reprezentovana mentalnom reprezentacijom $p$: ovakve elementarne reprezentacije su osnova semantike jezika uma u tom smislu da ako mentalna reprezentacija $p$ reprezentuje $P$ uzimamo da $p$ *znači* $P$. Svaki *predikat* kojim se subjekat $S$ dovodi u vezu sa objektom propozicionog stava $P$ u jeziku uma ima dodeljenu jedinstvenu psihološku relaciju R. Ova dodela jedinstvene psihološke relacije u LOTH je fundamentalna osnova za razlikovanje različitih verovanja: ono što pravi razliku između propozicionih stavova „$S$ *veruje* da $P$“ i „$S$ *želi* da $P$“ jeste u tome što um povezuje $S$ i $P$ u ova dva različita stava dvema fundamentalno različitim psihološkim relacijama. Tako, $\mathrm{R}_1(S,p)$ može da bude realizacija propozicionog stava „$S$ veruje da $P$“, a $\mathrm{R}_2(S,p)$ realizacija stava „$S$ želi da $P$“. Dakle, raznolikost psiholoških relacija u inventaru LOTH postoji na elementarnom nivou na kome su one jedinstvene i nedeljive: svaka psihološka relacija nosi specifičan identitet i samo primena različitih relacija na iste konstituente ($S$, $p$) određuje različite propozicione stavove koji ih obuhvataju.

Tradicionalno shvatanje verovanja kao odnosa subjekta i propozicije u LOTH je prošireno iz diadne (*subjekt*, *propozicija*) u trijadnu (*subjekt*, *mentalna reprezentacija*, *propozicija*) relaciju. U ovakvom teorijskom okviru, „$S$ veruje da $P$“ je tačno ako i samo ako $\mathrm{R}(S,p)$ gde je $p$ mentalna reprezentacija stanja opisanog rečenicom $P$, odn. $p$ znači $P$, a R psihološka relacija „verovati“. LOTH tvrdi da su svi mentalni procesi kauzalne sukcesije tokena ovakvih mentalnih reprezentacija, realizovanih na potpuno isti način na koji sistemi formalnog izračunavanja realizuju inferencije koje odgovaraju određenim formalnim jezicima. LOTH tvrdi da interni jezik uma ima kombinatornu sintaksu koja određuje pravila prepisivanja jednih simboličkih izraza u druge: kauzalni procesi generisanja jednih simboličkih izraza od drugih odslikavaju pravila te kombinatorne sintakse, i zato su kauzalne operacije nad simbolima osetljive samo na njihovu sintaksičku strukturu ali ne i na njihovo značenje. Izuzetno značajno za razumevanje LOTH je da kombinatorna sintaksa jezika uma omogućava izgradnju složenijih simbola kombinacijom jednostavnih na

taj način da je *značenje složenih simbola iscrpno definisano značenjem njihovih konstituenata*; ovu osobinu smo već upoznali razmatrajući formalne sisteme uopšte. Konačno, LOTH tvrdi da su mentalne reprezentacije definisane *funkcionalno*: neka određena psihološka relacija R je određena svojom ulogom u kauzalnim procesima transformacije simbola, u tom smislu što tačno određen skup procesa može da operiše nad strukturama koje sadrže relaciju R, dok ostali procesi to ne mogu. Na primer, proces koji povezuje inferenciju od (*i*) „*S veruje da će dohvatiti bokal ako pruži ruku*“ i (*ii*) „*S veruje da će oboriti svećnjak ako pruži ruku*“ ka (iii) „*S veruje da će oboriti svećnjak ako pokuša da dohvati bokal*“ je validan. Međutim, proces koji povezuje inferenciju od (*i*) „*S želi da dohvati bokal*“, (*iia*) „*S veruje da će dohvatiti bokal ako pruži ruku*“, i (*iib*) „*S veruje da će oboriti svećnjak ako pruži ruku*“ ka (*iii*) „*S želi da obori svećnjak*“ ne mora da bude validan. Predikati koji definišu različite propozicione stavove, kao „verovati“ i „želeti“, moraju da budu definisani funkcionalno, odn. kroz skupove operacija transformacije simbola koji na njih *mogu da se primene*: procesi koji omogućavaju prvu inferenciju bi omogućili i drugu kada bi bili primenjeni na predikat „želeti“ kao što su bili primenjeni na predikat „verovati“. To znači da između ovakva dva predikata postoji funkcionalna razlika: oni učestvuju u različitim formalnim izračunavanjima, dakle uzimaju različite *uloge* u tim procesima, te su jedinstveno definisani skupom procesa izračunavanja koji mogu da ih prepoznaju.

Standardna paradigma KKP koju smo predstavili se često označava kao *simbolicistički pristup*: ona pretpostavlja ljudski um eksplicitno reprezentuje okolinu i probleme koje rešava simbolički, i da njegova aktivnost jeste vrsta *eksplicitne manipulacije simbolima.* Videli smo kako ideja da se semantika veže za čisto sintaksičke operacije u procesu formalizacije, a sintaksičke operacije zatim vežu za one kauzalne procese koji mogu da iznesu formalna izračunavanja, vodi ka mogućnosti potpuno naturalističke, mehanicističke teorije uma čija je aktivnost suštinski simboličke prirode. Rasprava o ulozi verovanja kao fundamentalnog mentalnog stanja u ovoj analizi direktno je povezana sa analizom racionalnosti saznanja: uskoro ćemo pokazati kako je ova rasprava fundamentalno ograničena kada postavimo pitanje šta možemo da znamo o tome u šta tačno veruje neki subjekat *S koji je racionalan ako njegovo ponašanje otkriva da on konzistentno dela u skladu sa svojim verovanjima B kako bi ostvario svoje ciljeve G.* Pre toga, dužni smo da poklonimo pažnju teorijskim gledištima koja u okviru kompjutacionizma pružaju alternative standardnom simbolicističkom pristupu.

## 5.2 Kritike standardne paradigme i alternativna shvatanja

Sve kritike simbolicističkog pristupa se fokusiraju na jednu ili više od sledećih tvrdnji:

(*i*) *Kognitivni sistem eksplicitno reprezentuje simbole.*

(*ii*) *Kognitivni sistem eksplicitno manipuliše simbolima putem procesa formalnih izračunavanja* (odn. sledeći formalna pravila).

(*iii*) *Lokus eksplanatorne moći psihološkog objašnjenja je u konceptu reprezentacije*: neko ponašanje koje posmatramo jeste takvo kakvo je zato što je sistem čije ponašanje posmatramo reprezentovao problem koji rešava tako kao što ga je reprezentovao.

Emergentizam, oličen u konekcionističkom pokretu i teoriji dinamičkih sistema, i različite forme konstruktivizma, od kojih ćemo posebno razmotriti enaktivističku teoriju, udaljavaju se od simbolicizma onoliko koliko odbacuju tvrdnje (*i*), (*ii*) i (*iii*). Konekcionisti većinom ne prihvataju (*i*) i (*ii*); pristalice teorije dinamičkih sistema takođe odbacuju (*i*) i (*ii*) ali češće od konekcionista odbacuju i (*iii*), a konstruktiviste, posebno pristalice teorije enaktivizma, praktično definiše neprihvatanje nijedne od tri navedene tvrdnje. Pre razmatranja ovih alternativa simbolicističkom pristupu, dajemo kratak pregled Drajfusove kritike standardne paradigme iz 1972. godine; sve teorijske alternative simbolicizmu i danas mogu da se posmatraju kao pokušaji odgovora na pitanja koja je tada postavio Drajfus.

### 5.2.1 Hjubert Drajfus: generalna kritika standardne paradigme KKP

Verujemo da je i danas najbolji način da se sistematizuje kritika standardne KKP predstavljena u kultnoj knjizi „*Šta računari ne mogu*“ filozofa Hjuberta Drajfusa (Dreyfus, 1972/77). Iako se mnogo toga dogodilo u istraživanjima veštačke inteligencije i prirodnog uma od prvog izdanja Drajfusove knjige (1972. godine), suština kritike ovog programa se nije promenila; Drajfusa danas sa pravom smatramo za filozofa koji je uz Džona Serla i druge postavio sistematsku i suštinsku kritiku standardne paradigme KKP.

Drajfus sistemazije kritiku simbolicističkog programa u kognitivnoj psihologiji i veštačkoj inteligenciji eksplicirajući četiri pretpostavke na kojima on počiva, i onda redom kritikujući svaku od njih. Prema Drajfusu, *biološka pretpostavka* standardne paradigme KKP je da na nekom nivou analize ljudski mozak obrađuje informacije diskretnim operacijama. Inspiracija za ovakvu pretpostavku svakako se nalazi u

neurofiziološkoj tvrdnji da neuroni funkcionišu po principu „sve ili ništa“ u prenosu bioelektričnih signala kroz nervni sistem. *Psihološka pretpostavka* standardne paradigme sastoji se u tvrdnji da um vrši obradu informacija u skladu sa formalnim pravilima. *Epistemološka pretpostavka* je, prema Drajfusu, pretpostavka o tome da celokupno znanje može da se formalizuje, odn. izrazi u formi deskripcije logičkim funkcijama. Konačno, *ontološka pretpostavka* na kojoj se temelji standardna KKP je da svaka informacija o svetu može da se shvati kao situaciono slobodna i određena, odn. da svet može da se shvati kao niz činjenica od kojih je svaka logički nezavisna od ostalih.

*Biološka pretpostavka*, kojoj Drajfus posvećuje najmanje pažnje, prema njemu može da se odbaci iz razloga nepotpunosti: iako neuroni funkcionišu po diskretnom principu u tom smislu reči što poštuju određeni prag ispod kog ne proizvode signal nikad i iznad kog proizvode signal uvek, to ni iz daleka nije dovoljno da se opravda tvrdnja da je rad centralnog nervnog sistema suštinski diskretna obrada informacija. Masovna interakcija neurona povezanih u kompleksne neuralne mreže je ono što je nosilac funkcionisanja nervnog sistema: diskretna priroda signala u takvom sistemu, odn. frekvenca emisije signala određenog skupa neurona, može da bude posledica i graduiranih, (u idealizovanom modelu) kontinuiranih procesa. Suštinska diskretnost u obradi informacija bi postojala kad bi svaki emitovani nervni impuls mogao da se dovede u vezu sa nekim tačno određenim algoritamskim korakom, međutim, sve što znamo o nervnom sistem upućuje nas na to da je ponašanje koje on proizvodi posledica složenih interakcija ogromnog broja elemenata i procesa, a ne korak po korak izvođenih operacija koje se u krajnjoj liniji svode na sled impulsa pojedinačnih nervnih ćelija. Analiza i odbacivanje biološke pretpostavke predstavlja suštinsku inspiraciju za razvoj *konekcionističke teorije*, koju danas najčešće posmatramo u kontekstu teorije *emergentizma*: emergentizam je upravo tvrdnja da su kognitivne strukture i fenomeni makroskopske (globalne) posledice masovnih mikroskopskih (lokalnih) interakcija između elemenata kompleksnih sistema kakav je centralni nervni sistem, te da složeno ponašanje čoveka i viših organizama možemo da razumemo tek ako ga posmatramo iz takve perspektive.

*Psihološka pretpostavka*, srž standardne paradigme KKP, tvrdi da postoji nivo opisa ljudskog uma koji je izomorfan operacijama određenog formalnog sistema sa funkcionalno definisanim predikatima. Drajfus kritikuje ovu pretpostavku sa empirijskog i apriornog stanovišta, od kojih je za našu diskusiju ključna prva kritika. Empirijski, naučni program simbolicizma u kognitivnoj psihologiji nije uspeo da

objasnu svu kompleksnost i fleksibilnost viših kognitivnih procesa, a u razvoju veštačke inteligencije nije uspeo da simulira takve procese. Uspesi ovog pristupa su parcijalni, i ogledaju se upravo u analizi onih problema koji se lako analiziraju u terminima diskretnih elemenata i koraka, koji čak i samo analoški odgovaraju opisu putem formalnih sistema. Veoma zanimljiv je, sa ovog stanovišta, tzv. *Moravecov paradoks*, koji se odnosi na nalaz dosadašnjih istraživanja u kognitivnim naukama da je za simulaciju viših kognitivnih procesa, npr. rezonovanja, potrebno mnogo manje kompjutacione moći nego za simulaciju nižih, senzomotornih procesa (Agrawal, 2010). Iz njega sledi da je, nasuprot očekivanjima naučnika simbolicističke orijentacije, relativno lako razviti veštački formalni sistem koji uspešno igra šah, ali je veoma teško razviti npr. veštački sistem koordinacije oko-ruka. U vreme prvog izdanja Drajfusove knjige, šahovski programi su bili tek na začetku razvoja, dok danas oni rutinski pobeđuju čoveka i čak svetske šampione (Hsu, 2002). Međutim, mnogi problemi su u međuvremenu nastali na neočekivanim mestima, i to upravo tamo gde se pretpostavljalo da jednostavniji procesi igraju bitne uloge: na prvu loptu, aktivnost kao što je hodanje ne deluje složeno kao aktivnost dokazivanja teorema ili igranja šaha. Ali upravo senzomotorne aktivnosti jesu one za koje je karakteristična potreba za masovnom paralelnom obradom informacija u kojoj kontekst i međusobni odnosi potencijalno ogromnog broja parametara moraju da se uzmu u obzir. Pitanje kako centralni nervni sistem artikuliše ogroman broj stepeni slobode koji se javlja u opisu kinematike senzomotornog sistema prvi je postavio ruski fiziolog Nikolaj Bernštajn (up. formulaciju problema u npr. Sporns & Edelman, 1993), po kome ono danas nosi ime *Bernštajnov problem*. Bernštajn je pokazao da zbog ogromnog broja stepeni slobode koji se nalaze u opisu sistema motorne artikulacije centralni nervni sistem može da ostvari upravljanje ovakvim sistemom na potencijalno beskonačan broj načina. Ovo zapažanje je nazvano problemom „*inverzne dinamike*“: iz date trajektorije tačaka relevantnih u matematičkom opisu nekog pokreta ne sledi jedinstveno rešenje za upravljanje senzomotornim sistemom čiji se pokret opisuje. Drajfus tvrdi da se pod simbolicističkom paradigmom mahom analiziraju kognitivni problemi koji se relativno lako „rastavljaju“ na skupove elemenata čija pravila kompozicije prepoznajemo, dok se zanemeruje realna kompleksnost okruženja u kome funkcioniše ljudski kognitivni sistem.

*Epistemološka pretpostavka* standardne paradigme, prema Drajfusu, jeste tvrdnja o tome da je celokupne znanje moguće formalizovati, odn. da bez obzira

na to da li ljudski um funkcioniše (svesno ili nesvesno) u diskretnim koracima formalnih izračunavanja, *inteligentno ponašanje može da se opiše jezikom nekog formalnog sistema*, pa prema tome i simulira mehanički, kao što to tvrdi standardna, simbolicistička paradigma KKP. Suštinska razlika između psihološke i epistemološke pretpostavke je u tome što se psihološkom pretpostavkom standardne paradigme tvrdi da um *funkcioniše* kroz eksplicitnu, internu, mentalnu primenu odgovarajućih formalnih pravila, dok epistemološka pretpostavka tvrdi samo da je takva pravila uvek moguće formulisati. Kritika epistemološke pretpostavke nije jednostavna, iako je od svih koje Drajfus razmatra najbolje utemeljena u istoriji novije filozofije. U III poglavlju posvetićemo pažnju problemima psihološke semantike, gde nas diskusija značenja i komunikacije suočava sa najozbiljnijim problemom u analizi racionalnosti saznanja uopšte. Kritika epistemološke pretpostavke, pokazaćemo tada, neposredno je vezana za pitanje racionalnosti saznanja, kada se ono diskutuje kao pitanje mogućnosti formulisanja normativne osnove semantičkih interpretacija.

*Ontološka pretpostavka*, pretpostavka o tome da je univerzum, pa prema tome i ljudsko ponašanje, uopšte moguće posmatrati i analizirati u terminima činjenica koje su *kontekstualno nezavisne*, pretpostavka je na kojoj je počivala rana filozofija Vitgenštajna, izražena u klasičnom delu „*Tractatus Logico-Philosophicus*“ (Wittgenstein, 1921/1987). Drajfus problematizuje analizu i na ovom nivou, postavljajući pitanje da li naši apriorni analitički okviri uopšte omogućavaju da se pitanje o prirodi ljudskog saznanja postavi na odgovarajući način; ipak, analiza ontološke pretpostavke u Drajfusovom radu mnogo je više relevantna za filozofsku raspravu nego za raspravu one prirode koju mi vodimo.

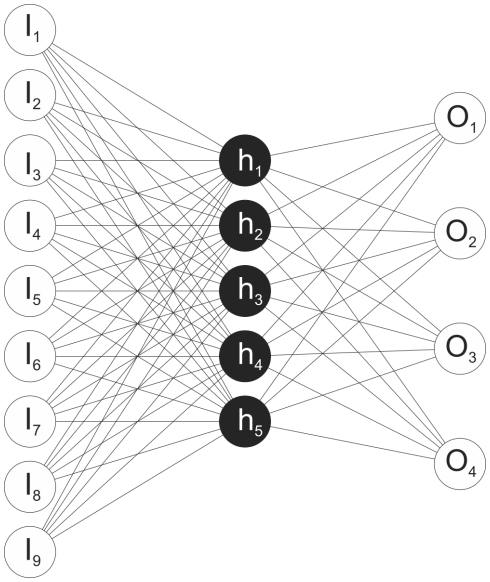
### 5.2.2 Emergentizam i konekcionistički modeli

Drajfusova kritika biološke pretpostavke nas upućuje na to da smo, za razliku od problema koji su neposredno podložni formalizaciji diskretnim elementima, u praksi često suočeni sa problemima čiji su prirodni, praktični opisi kontinuirani. Samo paralelna obrada i integracija *svih* bitnih informacija može da garantuje adaptivnu aktivnost u ovakvim slučajevima, i samo kompjutacioni sistemi sposobni za masovnu, paralelnu obradu i integraciju informacija predstavljaju odgovarajuće modele kognitivnih procesa koji rešavaju ovakve probleme. Analogija sa nervnim sistemom, u kome diskretne jedinice - neuroni - paralelnim radom proizvode kompleksne nervne procese, procese koji upravljaju složenim biofizičkim sistemima kao što je senzomotorna koordinacija, predstavljala je osnovu za razvoj

*konekcionizma* (engl. *Connectionism*), teorijskog pokreta koji je formalne sisteme simbolicističkog pristupa pokušao da zameni modelima *neuronskih mreža*.

Neuronske mreže predstavljaju klasu kompjutacionih sistema koji fleksibilno menjaju svoje ponašanje u funkciji promene konteksta. Ovi modeli dolaze do rešenja problema koji im se nameće kroz evoluciju sopstvenog unutrašnjeg stanja i proces učenja, prilagođavajući svoje reakcije promenama u strukturi okoline na koju uče da reaguju. U kontekstu konekcionizma, najčešće proučavan model neuronske mreže je model *višeslojne mreže sa povratnom propagacijom signala* (engl. *Multilayer Backpropagation Network*, McClelland, Rumelhart & the PDP Research Group, 1986a, 1986b). Da bismo razumeli konceptualnu razliku između konekciozma i simbolicističkog pristupa, sažeto analiziramo ovaj model čija je arhitektura prikazana na Slici 7.

Slojevi u višeslojnoj neuronskoj mreži sadrže *čvorove*, koji predstavljaju elementarne jedinice obrade informacija analogne nervnim ćelijama. Mreža prikazana na Slici 7. je troslojne arhitekture: ona sadrži sloj input čvorova, sloj autput čvorova, i između njih sloj koji se naziva *skrivenim* - u analogiji sa realnom neurofiziologijom, on sadrži interneurone, jedinice koje povezuju input i autput čvorove. Iz arhitekture mreže vidimo da signali sa svih input čvorova konvergiraju ka svim skrivenim čvorovima, a sa njih divergiraju ka svim autput čvorovima. Za potrebe primera, posmatrajmo svih devet input čvorova kao reprezentacije signala o prisustvu ili odsustvu neke od devet mogućih osobina određene grupe objekata. Dakle, svaki input signal u mrežu jeste složaj aktivacije ovih devet ulaznih čvorova, i svaki takav složaj aktivacije reprezentuje određeni objekat. Na primer, ako složaj aktivacije na input čvorovima predstavlja vektor [1,0,1,0,0,0,1,1,1], onda on kodira prisustvo prve, odsustvo druge, prisustvo treće itd. karakteristike objekta koji se predstavlja na ulazu u mrežu. Kodiranje objekata na inputu definiše promene u sredini u kojoj se mreža nalazi. Svaka neuronska mreža uči da se prilagođava promenama svoje okoline. Cilj mreže sa ovakvom arhitekturom je da nauči da razvrstava sve moguće objekte koje joj prikazujemo u jednu od četiri kategorije. Svaki od autput čvorova, obeleženih kao $O_1$ ,$O_2$, $O_3$, $O_4$ na Slici 7, predstavlja jednu od te četiri kategorije.




Slika 7. *Višeslojna neuronska mreža sa povratnom propagacijom signala.* Čvorovi mreže u levom redu, obeleženi slovom „I“, su input čvorovi koji reprezentuju signale koje sistem prima iz okoline. Čvorovi obeleženi slovom „O“ su autput čvorovi, koji predstavljaju odgovor mreže na signale iz sredine. U sredini, slovom „h“ su obeleženi skriveni (engl. *hidden*) čvorovi, čija funkcija obezbeđuje paralelnu integraciju informacija iz više input čvorova i osetljivost na kontekst kodiranjem intenziteta odgovora na složene kombinacije različitih inputa u sistem. Signali iz sredine se propagiraju kroz mrežu od input ka autput čvorovima, dok se signali korekcije greške povratno propagiraju u obrnutom smeru.

Ukoliko objekat koji predstavljamo na input čvorovima spada u treću od četiri kategorije, mreža treba da nauči da na složaj aktivacije koji predstavlja taj objekat odgovori maksimalnom aktivacijom čvora $O_3$. Sve veze među čvorovima karakterišu *težine*, koje određuju intenzitet signala koji će neki čvor primiti od prethodnika ukoliko ga on aktivira. Suština prilagođavanja ovakvih sistema impulsima iz sredine je u postepenoj modifikaciji težina veza među čvorovima. Pretpostavimo da input čvorovi neke neuronske mreže kodiraju osobine različitih vrsta životinja: prvi čvor je aktivan ukoliko životinja ima kljun, drugi ukoliko ima kandže, treći ukoliko ima krzno, četvrti ukoliko može da pliva itd. Možemo da zamislimo mrežu koja preko velikog broja ovakvih input čvorova na ulazu prima reprezentaciju neke od velikog broja životinja. Prepostavimo dalje da mreža treba da razvrsta sve životinje koje joj prikažemo u ribe, gmizavce, ptice i sisare na osnovu karakteristika predstavljenih na ulazu u sistem. Višeslojne neuronske mreže sa povratnom propagacijom signala uče ovakve klasifikacije kroz sled koraka. Na svaki složaj signala sa inputa prvo reaguju neuroni skrivenog sloja, koji dalje prenose aktivaciju na neurone autput sloja. Svi neuroni u mreži imaju definisanu *funkciju aktivacije*, koja može uzeti različite matematičke forme. Ta funkcija determiniše intenzitet izlaznog signala iz određenog

čvora na osnovu intenziteta ukupne aktivacije koju taj čvor prima. Kada se neuroni autput sloja aktiviraju, na osnovu fidbeka koji im se pruža - znajući koji složaj signala na inputu označava koju životinju i kojoj vrsti ona pripada - računa se *greška* u aktivaciji svakog od njih. Ukoliko je na ulazu karakteristikama predstavljena neka ptica, a najveću aktivaciju na izlazu ima čvor koji označava sisare, taj čvor dobija fidbek visoke greške. Praktično, fidbek u treningu neuronske mreže je razlika između očekivane aktivacije autput čvorova i njihove aktivacije posle prikazanog inputa. *Algoritmom povratne propagacije signala* (engl. *Backpropagation*, McClelland, Rumelhart & the PDP Research Group, 1986a), koriguju se težine veza između svih čvorova u neuronskoj mreži relativno u odnosu na njihov doprinos signalu greške na svakom autput čvoru. Mreži se prikazuju svi objekti koje treba da nauči da razvrsta, definisani nizom karakteristika na ulazu, a težine veza između čvorova u mreži se posle svake prezentacije sukcesivno podešavaju na osnovu vrednosti koje računa algoritam povratne propagacije, dok odgovor mreže ne postigne maksimalan nivo tačne klasifikacije.

Postoje suštinske razlike između ovakvog pristupa dizajnu kompjutacionih sistema koji ostvaruju kognitivne funkcije i sistema karakterističnih za simbolicistički pristup. Prvo, neuronske mreže *ne grade eksplicitne simboličke reprezentacije* sredinskih objekata. Reprezentaciju nekog objekta u neuronskoj mreži predstavlja *niz brojeva* koji su deskripcija težina veza između svih čvorova kroz sve slojeve mreže. Nigde ne postoji eksplicitna reprezentacija nekog određenog objekta izvedena putem liste predikata koji bi se na njega odnosile; nigde nema formalnih pravila kompozicije predikata koja, ukoliko su ispunjena, obezbeđuju inferenciju da je neki objekat pripadnik neke kategorije. Celokupno „znanje“ neuronske mreže je u *strukturi težina veza* koje postoje između *neinterpretabilnih elementarnih jedinica* - čvorova mreže, tj. neurona. Kategorizacija nekog input signala u neku od kategorija je spontani proces koji odgovara realnosti zato što se propagacija složaja signala sa inputa prelama kroz težine veza, prethodnim procesom korekcije grešaka podešenih tako na autputu proizvedu složaj signala koji odgovara tačnoj klasifikaciji datog inputa. Svo „znanje“ neuronske mreže je implicitno, sadržano u strukturi težina veza koja je naučena tokom procesa prilagođavanja autput reakcije sredinskim promenama. Strukture težina veza se u konekcionističkom pristupu nazivaju *distribuiranim reprezentacijama*, čime se naglašava njihova suštinska razlika u odnosu na eksplicitne, simboličke reprezentacije.

Ovakav pristup ne samo da se suštinski razlikuje od simbolicističkog, već

omogućava modeliranje nekih realnih osobina ljudskog kognitivnog sistema koje se pod simbolicističkom paradigmom teško modeliraju. Većina odraslih osoba dobro zna da kit i delfin nisu ribe, već vrste sisara. Međutim, svi su svesni da između kitova i delfina, s jedne, i riba, s druge strane, postoje mnoge morfološke i bihejvioralne sličnosti koje su posledica života u praktično istom ekosistemu. Ovakve graduirane, nejasne slučajeve u kategorizacijama, simbolicistički pristup mora eksplicitno da modelira, npr. reprezentacijom nužnih i dovoljnih karakteristika i bez mnogo tolerancije za morfološku sličnost nekih sisara i riba koja mora eksplicitno da bude označena kao nebitna karakteristika. To dovodi do daljih problema u reprezentaciji, jer je činjenica da organizmi slični po biološkim karakteristikama *najčešće* jesu slični i po svojim morfološkim karakteristikama, što se kosi sa prethodnim izolovanjem morfologije kao nesuštinske u klasifikaciji životinja. Simbolicistički sistem može da reši ovaj problem samo implementirajući direktno logiku „pravila i izuzetaka“, i markirajući predikate koji opisuju morfološke osobine kao značajne za klasifikaciju u nekim problemima, ali beznačajne za isti proces u drugim problemima. Konekcionistički pristup prirodno rešava ovakve probleme, bez potrebe za eksplicitnom intervencijom u reprezentacionom sistemu. Ako četiri čvora na izlazu neuronske mreže predstavljaju četiri kategorije životinjskih vrsta, dovoljno je da mreža uspe da nauči da najjača aktivacija za signal koji predstavlja osobine delfina treba da bude na čvoru za sisare, što ne isključuje mogućnost da na reprezentaciju delfina autput signal na čvoru za ribe ima višu vrednost od signala na čvoru za ptice ili gmizavce. Na taj način, odgovor neuronske mreže u klasifikaciji nas informiše da je u pitanju pripadnik sisara, koji nije sasvim tipičan, već deli neke bitne karakteristike drugih kategorija. Tako su neuronske mreže, zahvaljujući distribuiranim reprezentacijama koje razvijaju, u stanju da prirodno predstave ambivalentnost u kategorizaciji. One pokazuju visok stepen tolerancije nejasnih situacija, što je posledica globalne, paralelne integracije informacija koju izvode, bez potrebe da eksplicitno simbolički reprezentuju i formalno izračunavaju sve detalje u potencijalno ogromnoj količini informacija koju kognitivni sistem sreće u okolini.

Primer koji smo diskutovali je pojednostavljen do nivoa koji nam omogućava da predstavimo suštinske karakteristike konekcionističkog pristupa, bez diskusije relativno složene matematičke forme algoritma povratne propagacije signala. Napredne arhitekture ovakvih mreža obuhvataju rekurzivne neuronske mreže, primenjene u psiholingvistici (Elman, 1990, 1991, 1993) i psihološkoj semantici

(Rogers & McClelland, 2004). Pored višeslojne mreže sa povratnom propagacijom, karakteristične za okosnicu konekcionističkog pristupa oličenu u PDP grupi (McClelland, Rumelhart & the PDP Research Group, 1986a, 1986b) i kasnijim razvojima koji su konekcionizam približili teoriji dinamičkih sistema (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, Plunkett, 1996), razne druge mrežne arhitekture su korišćene u modeliranju različitih problema kognitivnih nauka (v. Haykin, 1999, za pregled velikog broja različitih arhitektura). Specifične vrste hibridnih simboličko-konekcionističkih sistema takođe su našle primenu u modeliranju viših kognitivnih procesa (up. Hummel & Holyoak, 2003 za ingenioznu primenu hibridnog modela u analoškom mišljenju, v. Wermter & Sun, 2000, Sun, 2001, za uvod u hibridne modele).

Konstatovali smo da je jedna od suštinskih osobina neuronskih mreža - osobina za koju ćemo uskoro videti je svojstvena i drugim kompjutacionim sistema - ta da je njihovo kompleksno ponašanje proizvod masovne interakcije između elementarnih, neinterpretabilnih jedinica. Te neinterpretabilne jedinice u konekcionističkim modelima su neuroni, čvorovi neuronskih mreža. Oni nose funkciju elementarne obrade informacija u sistemu, ali sami za sebe, posmatrani izolovano, ne nose nikakvu semantički interpretabilnu odliku. Semantički interpretabilne odlike ovakvih sistema posledica su evolucije odnosa između ovih mikroskopskih elemenata sistema. Evolucija njihovih odnosa, promena težine veza među neuronima sa vremenom, razvija se do stabilne strukture tih odnosa u kojoj mreža reaguje regularno i interpretabilno na impulse iz sredine. Semantički intepretabilne osobine ovakvog sistema *pripadaju sistemu u celini*: tek sve težine veza, uzete zajedno sa arhitekturom mreže, predstavljaju objašnjenje regularnih i interpretabilnih reakcija sistema, tj. njegove mogućnosti da tačno klasifikuje strukture inputa na autputu. Dakle, ukupno stanje sistema, posmatranog na makroskopskoj ravni, posledica je izuzetno složene kauzalne interakcije njegovih gradivnih elemenata, posmatranih na mikroskopskoj ravni: kažemo da je semantička interpretabilnost ponašanja konekcionističkih sistema *emergentna karakteristika* interakcije njihovih elemenata. Teorijski pravac u okviru kojeg se najčešće interpretiraju konekcionistički i sve drugi dinamički modeli u kojima se interpretabilni fenomeni javljaju kao posledica masovnih, paralelnih interakcija između njihovih elemenata naziva se *emergentizmom*[21] (engl. *Emergentism,* Bedau & Humphreys, 2007). Emergentizam se suštinski razlikuje od simbolicističkog pristupa upravo po objašnjenju porekla semantičke interpretabilnosti ponašanja

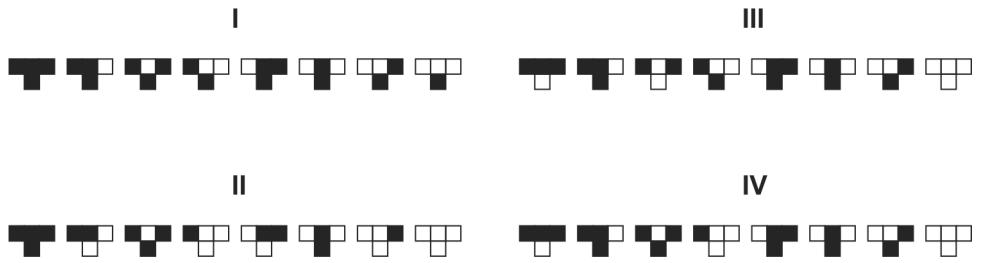
nekog složenog sistema. Dok je u simbolicističkim sistemima semantika inherentna sistemu koji eksplicitno reprezentuje elementarne predikate i razvija složene strukture formalnim procesima kompozicije simbola, u emergentnim sistemima se semantički interpretabilne odlike spontano pojavljuju kao proizvod procesa evolucije interakcija između delova sistema koji su po sebi neinterpretabilni na relevantnom nivou analize. Na nivou mikroskopske analize konekcionističkih modela, sami čvorovi mreže nemaju nikakvu relevantnu interpretaciju. Međutim, na relevantnom nivou analize, na kome posmatramo proces učenja odgovarajuće klasifikacije inputa od strane celog sistema, semantički interpretabilne celine - odgovori, tj. strukture signala na autputu neuronske mreže - pojavljuju se kao posledica interakcije većeg broja elementarnih, neintepretabilnih jedinica koje grade sistem. Proces propagacije greške, kojim se fino podešavaju težine veza među neuronima, predstavlja proces *zadovoljavanja ograničenja* (engl. *constraint satisfaction*) koji se paralelno odvija kroz veliki broj elemenata sistema. Mreža ne može da nauči ispravnu klasifikaciju ukoliko težine veza u njoj evoluiraju u skladu sa odgovorima na samo jednu strukturu signala na input sloju: zato su podešavanja težine veza u algoritmu povratne propagacije rezultati velikog broja „kompromisa“ između različitih reakcija mreže, različitih grešaka koje je proizvela u tim reakcijama, i različitih struktura inputa na koje je reagovala. Zahvaljujući poštovanju doprinosa svih informacija koje utiču na sistem paralelno, distribucija greške kroz sistem vremenom dovodi do stabilne konfiguracije težine veza u kojoj mreža postiže *optimalan odgovor* u odnosu na promene u svojoj sredini. Analiza rada mreže na nivou pojedinačnih neurona ne može da objasni njenu evoluciju ka globalnim adaptivnim stanjima: tek *dinamička analiza evolucije sistema u celini*, analiza evolucije svih odnosa među elementima mreže, omogućava razumevanje pojave emergentnih karakteristika na nivou celog sistema. Teorija dinamičkih sistema koja sredstvima matematičke analize proučava vremensku evoluciju kompleksnih sistema našla je primenu u emergentizmu i konekcionizmu kao prirodan jezik za opis emergentnih sistema (Elman, 1990, 1995, 1998, van Gelder, 1996). Međutim, nije nužno da veza između globalnih osobina ovakvih sistema i njihovih elemenata na mikro-nivou opisa uopšte može da se eksplicira. Zato bi trebalo govoriti o emergentnim odlikama u užem i širem smislu reči[22]. U širem smislu reči, možemo da kažemo da je neka globalna odlika sistema emergentna ako nije očigledna posledica interakcija njegovih elemenata ali nju jeste moguće teorijski ispratiti sa globalnog na lokalni nivo opisa, i ustanoviti zakonitosti interakcija lokalnih elemenata čijih su globalne odlike posledice. U užem smislu

reči, emergentna odlika (kako je shvata savremena filozofija nauke) je nesvodiva na opis na nivou interakcije lokalnih elemenata u sistemu; prirodu globalnih odlika opisuju zakoni koje je Brod nazvao *trans-ordinalnim*, i sugerisao da oni treba da budu korišćeni kao fundamentalni prirodni zakoni za objašnjenje nekih kvaliteta makroskopskih struktura u prirodi pri punoj svesti o tome da ti zakoni nisu svodivi na zakone nižeg nivoa (McLaughlin 1997/2008). Termin *supervinijencija* (engl. *supervenience*) se koristi u diskusijama emergentizma da označi odnos u kome se nalaze globalne, makroskopske karakteristike nekih sistema u odnosu na interakcije njihovih lokalnih elemenata na mikroskopskom nivou opisa: globalne odlike su supervinijentne nad odlikama lokalnim elemenata sistema.

Suštinski, emergentizam, konekcionizam i teorija dinamičkih sistema, uprkos nekim razlikama u shvatanjima različitih autora, danas čine jednu teorijsku, konceptualnu celinu koja stoji nasuprot standardnoj simbolicističkoj paradigmi KKP. Interesovanje za emergentističku paradigmu u KKP naglo je poraslo osamdesetih godina XX veka, posle uvođenja modela neuronskih mreža u mejnstrim kognitivne psihologije, iako je emergentizam kao model naučnog objašnjenja odavno poznat i istoriji (Ernst Najgel pronalazi izvor savremenog emergentizma u „*Sistemu logike*“ Džona Stjuarta Mila, delu objavljenom još 1843, prema McLaughlin 1997/2008) i filozofiji nauke (za izvanrednu analizu i kritiku ideje emergentizma up. Hempel & Oppenheim, 1965/2008).

Neočekivane i duboke uvide u prirodu kompjutacionih sistema koji razvijaju emergentna svojstva omogućilo je proučavanje klase sistema poznatih kao *celularni automati* (engl. *cellular automata*) u drugoj polovini XX veka. Fascinantna osobina celularnih automata je da oni u svojoj vremenskoj evoluciji mogu da razviju ponašanje *ma koje složenosti* kao emergentnu karakteristiku *banalno jednostavnih interakcija* između osnovnih elemenata. Celularne automate su prvi proučavali fon Nojman i Ulam. Fon Nojman je, radeći na problemu konstrukcije samoreprodukujućih automata, na Ulamovu sugestiju razvio matematički formalizam celularnih automata kakav poznajemo danas, razvijajući pri tom prvi veštački samoreprodukujući sistem, celularni automat koji je mogao da iz elemenata sredine kojom je okružen konstruiše novi celularni automat kao svoju kopiju (Waldrop, 1993, Coveney & Highfield, 1995). Najveći doprinos radu na celularnim automatima dugujemo Stivenu Volframu koji je prvi formulisao sistematske zaključke o kompleksnoj evoluciji celularnih automata najjednostavnije moguće strukture (Wolfram, 1983).

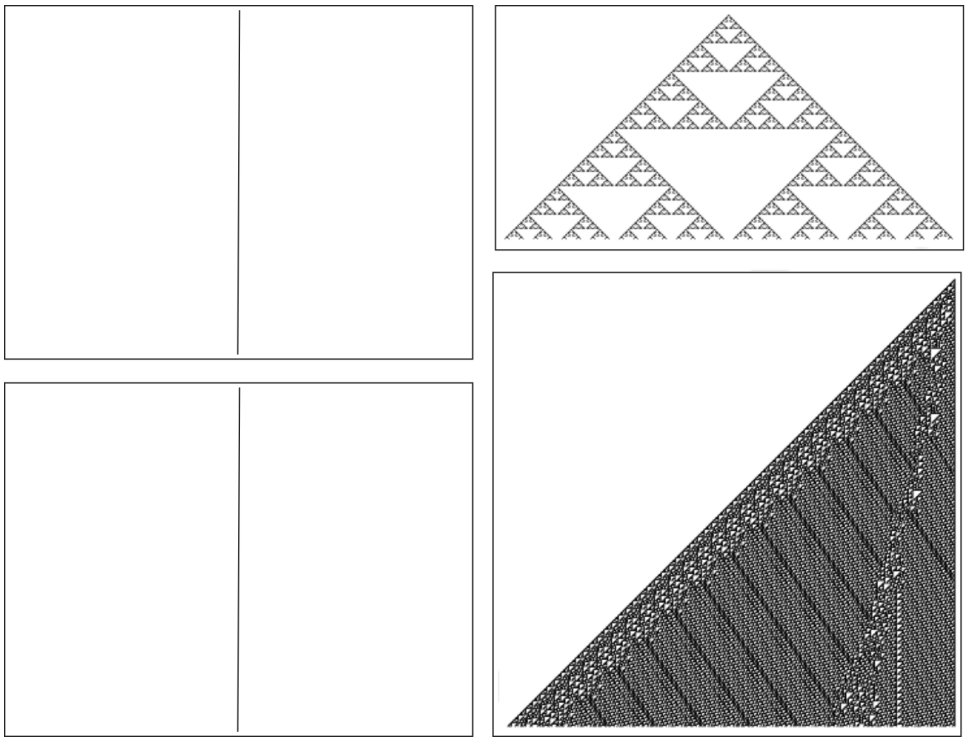
Elementarni jednodimenzionalni celularni automat se sastoji od niza jedinica („ćelija“) koje uzimaju samo dva stanja, „1“ ili „0“, najčešće vizuelno predstavljenih kao niz kvadratića crne ili bele boje. U stanju „1“, jedinica nekog celularnog automata je aktivna, u stanju „0“, nije aktivna. Ponašanje ovog sistema reguliše pravilo koje određuje da li je stanje bilo koje ćelije u automatu u trenutku $t$ aktivno isključivo na osnovu (a) stanja njenog levog suseda u vremenu $t$-1, (b) sopstvenog stanja ćelije u vremenu $t$-1, i (c) stanja njenog desnog suseda u vremenu $t$-1. Dakle, stanje celokupnog celularnog automata u nekom diskretnom trenutku vremena je bez ostatka determinisano njegovim stanjem u prethodnom diskretnom trenutku vremena. Pravila za četiri elementarna jednodimenzionalna celularna automata prikazana su na Slici 8. Svaki red na slici prikazuju korak vremenske evolucije pojedinačne ćelije nekog celularnog automata: pravila jednoznačno determinišu da li je ćelija u narednom koraku evolucije aktivna ili ne u zavisnosti od toga kakva je prethodna aktivnost njenih suseda i nje same.




Slika 8. *Pravila evolucije za četiri elementarna jednodimenzionalna celularna automata*. Svaka konfiguracija na slici pokazuje da li je ćelija nekog celularnog automata aktivna ili ne u sledećem diskretnom koraku vremenske evolucije u odnosu na sopstvenu prethodnu aktivnost i aktivnost svog levog i desnog suseda. Rimski brojevi iznad pravila predstavljaju Volframove klase kojima pripadaju odgovarajući automati (objašnjenje u tekstu). Poštujući Volframovu notaciju pravila, redom Volframovih klasa su prikazana pravila 255, 164, 90 i 110 (Wolfram, 2002).

Svaki jednodimenzionalni celularni automat treba posmatrati kao prsten: levi sused prve ćelije u nizu je poslednja ćelija s desna, čiji je opet prva ćelija desni sused. Međutim, automate prikazujemo „razmotane“ u niz ćelija, tako što svaki sukcesivan korak njihove evolucije prikazujemo red za redom. Prvi red definiše inicijalne uslove, drugi red je rezultat primene pravila određenog automata na prvi red, itd. Posmatrajući vremensku evoluciju celularnih automata „red po red“ prepoznajemo razvoj određenih globalnih ili lokalnih struktura koje predstavljaju njihove emergentne karakteristike. Slike 9a i 9b prikazuju vremensku evoluciju celularnih automata sa pravilima prikazanim na Slici 8 (po Volframovoj notaciji, to su pravila: 255, 164, 90 i 110, Wolfram, 2002). Vremensku evoluciju celularnih

automata posmatramo odozgo na dole, dakle primenjujući pravila na prvi red, na vrhu dijagrama (Slike 9a i 9b), i onda sukcesivno na naredne redove, uvek uzimajući prethodni red kao početni. Na Slici 9a, evolucija svih celularnih automata počinje redom neaktivnih ćelija sa samo jednom aktivnom ćelijom u sredini. Na Slici 9b, nasuprot, evolucija svih celularnih automata počinje od reda ćelija čija je aktivacija posledica slučajnog procesa.



Slika 9a. *Vremenska evolucija četiri elementarna jednodimenzionalna celularna automata*. Inicijalni uslovi za sva četiri automata na ovoj slici je niz neaktivnih ćelija sa jednom aktivnom ćelijom u sredini. Poštujući Volframovu notaciju, gore levo: pravilo 255 (klasa I), dole levo: pravilo 164 (klasa II), gore desno: pravilo 90 (klasa III), dole desno: pravilo 110 (klasa IV).

Na Slici 9a, oba automata na levoj strani, od kojih gornji nastaje primenom pravila $255^{23}$ a donji 164, evoluiraju veoma jednostavno. Struktura njihovih pravila proizvodi monotonu evoluciju, koja se na vremenskom dijagramu prepoznaje kao vertikalna linija: ona predstavlja jednu jedinu ćeliju koja ostaje aktivna posle svake ponovoljene primene pravila. Na istoj slici, sa desne strane vidimo daleko složenije evolucije celularnih automata. Gornji automat, čiju evoluciju kontroliše pravilo 90, vremenom razvija strukturu koja vizuelno podseća na trougao Sjerpinskog. Očigledna je rekurzivna forma, detalj trougla koji pravilno ispunjavaju drugi trouglovi i koji sledeći istu shemu ispunjava ceo univerzum sa evolucijom ovog celularnog automata. Takođe na desnoj strani Slike 9a, dole, prikazana

je evolucija veoma interesantnog celularnog automata sa pravilom 110. Tokom vremenske evolucije ovog automata postaju vidljive lokalne strukture, koje neko vreme održavaju istu formu, ponekad se susrećući jedna sa drugom i ulazeći pri tom u kompleksne interakcije.

Slika 9b prikazuje iste celularne automate kao Slika 9a, sa tom razlikom što na ovoj slici posmatramo vremensku evoluciju automata koja počinje od slučajnih inicijalnih uslova. Svaki od automata starovan je nizom slučajno, nasumično aktivnih i neaktivnih ćelija. Na levoj strani, vidimo sada razliku u evoluciji gornjeg (pravilo 255) i donjeg (pravilo 164) automata: dok gornji automat odmah postiže veoma jednostavno, monotono stabilno stanje posle koga nema više nikakvih promena (sve ćelije postaju aktivne i održavaju aktivnost), donji automat evoluira u preoznatljiv niz aktivnih ćelija koje ostaju aktivne sa vremenom. Na desnoj strani, gore (pravilo 90) vidimo koliko celularni automati određenog tipa mogu biti osetljivi na inicijalne uslove: automat čija je vremenska evolucija od samo jedne aktivne ćelije na početku razvijala pravilnu matematičku strukturu rekurzivnih osobina (Slika 9a, gore desno), sa slučajnim početnim uslovima počinje da proizvodi evoluciju koju odlikuje potpuna slučajnost (Slika 9b, gore desno). Forme trouglova su prepoznatljive u ovoj evoluciji, ali njihov raspored i veličina su potpuno nepredvidljivi; automat kroz evoluciju ne postiže nikakvo stabilno stanje karakteristično za dva automata sa leve strane. Konačno, dole desno na Slici 9b, automat sa pravilom 110, pokazuje praktično istu vremensku evoluciju startovan od slučajne konfiguracije kao na Slici 9a gde je startovan od samo jedne aktivne ćelije.

Bez prethodnog pokušaja kompjuterske simulacije ovakvih modela teško je pomisliti da sistemi čije ponašanje regulišu tako jednostavna pravila kao ona prikazana na Slici 8. mogu da razviju vremensku evoluciju tolike kompleksnosti kakvu pokazuju automati sa pravilima 90 i 110 na slikama 9a i 9b. Automat 90 praktično može da posluži kao generator slučajnih brojeva - njegova evolucija je toliko nepredvidljiva i kompleksna da predviđanje njegovog stanja uopšte nije moguće. Jedino što nam omogućava da uopšte saznamo nešto o budućoj evoluciji automata sa ovakvim pravilom je da poznajemo tačno inicijalno stanje koje mu je zadato i zatim eksplicitno izvedemo sve korake do onog koji nas interesuje.

Stiven Volfram je čisto fenomenološkom analizom vremenskih evolucija najrazličitijih tipova celularnih automata došao do zaključka da oni svi mogu da se podele u četiri fundamentalne klase. Pravilo 255, prikazano gore levo na prethodnim slikama, pripada Volframovoj klasi I celularnih automata: posle

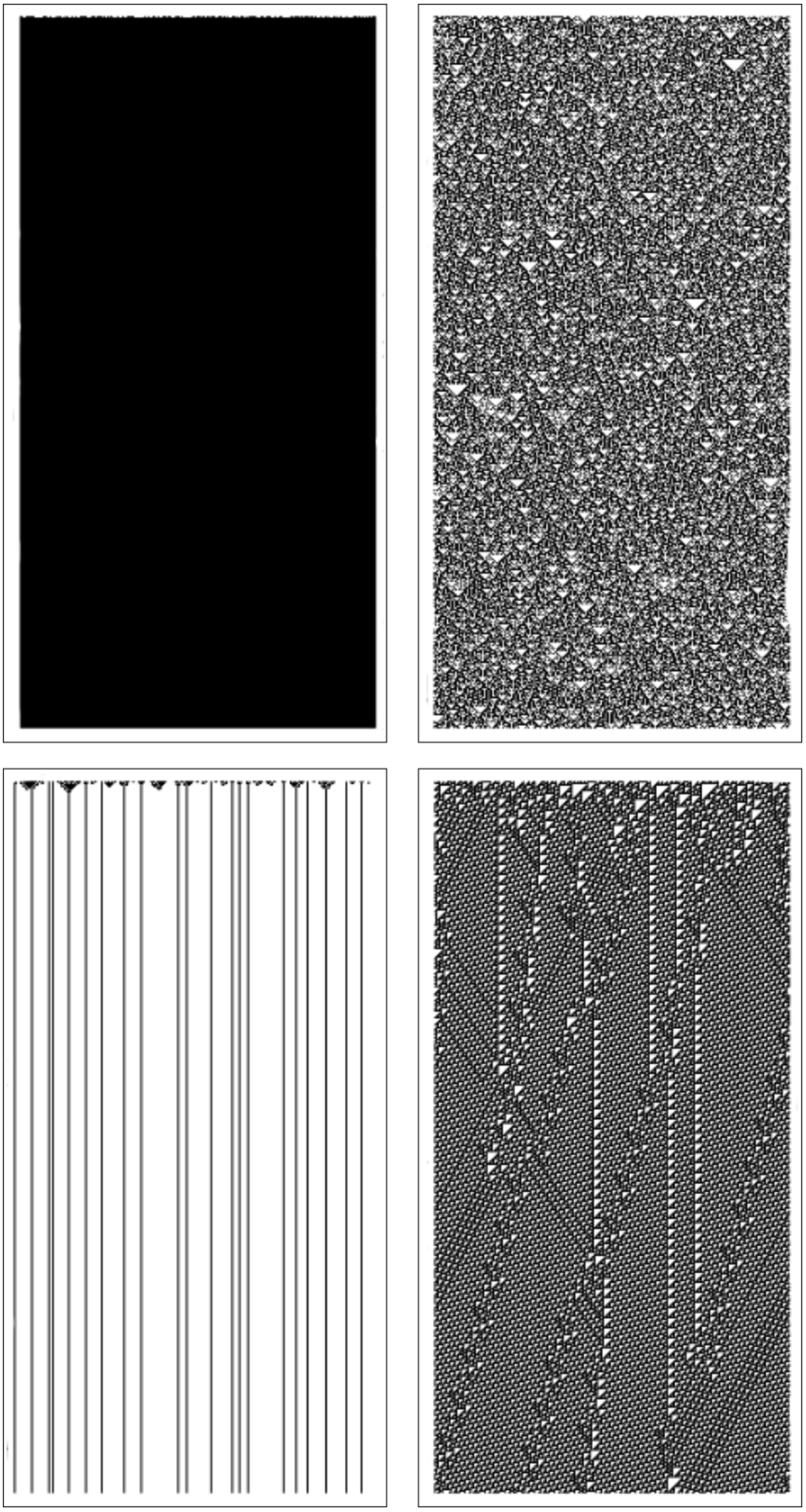
slučajnih inicijalnih uslova, ovi automati brzo konvergiraju u monotonu evoluciju u kojoj su sve ćelije ili aktivne ili ne. Pravilo 164, prikazano dole levo na slikama, jeste pravilo klase II: posle slučajnih inicijalnih uslova, automate ove klase odlikuje vremenska evolucija u periodičnim formama koje se egzaktno reprodukuju sa svakim korakom. Pravilo 90, prikazano gore desno na slikama, koje pripada Volframovoj klasi III, odlikuje evolucija u slučajnost. Posle slučajnih inicijalnih uslova, vremenska evolucija ovakvih celularnih automata ne pokazuje praktično nikakve predvidljive odlike. Konačno, pravilo 110 pripada izuzetno interesantnoj klasi IV celularnih automata. Posle slučajnih inicijalnih uslova, ovi automati pokazuju evoluciju koja je mešavina klase II i III, odlikovana prožimanjem slučajnosti i regularnosti, pojavom lokalizovanih struktura koje je donekle moguće pratiti, koje međusobno ulaze u složene interakcije u kojima neke nestaju, od nekih nastaju nove, drugačije strukture, a neke ostaju skoro nepromenjene.

Tokom rada na teoriji celularnih automata, Stiven Volfram je još veoma rano posumnjao da su celularni automati klase IV *Tjuring-ekvivalentni* (Wolfram, 2002). To znači da ih je, zadajući im različite inicijalne uslove, moguće programirati da izvode sva izračunavanja koja može da izvede Tjuringova mašina, što dalje znači da automati klase IV jesu sistemi univerzalnog izračunavanja. Programirati direktno celularne automate zadavanjem različitih inicijalnih stanja je praktično nezamislivo, pa je tvrdnja o Tjuring-ekvivalentnosti klase IV zbog toga dugi niz godina predstavljala samo hipotezu. Konačno, posle godina rada na tom problemu, Volframov asistent Metju Kuk je uspeo da dokaže da automat klase IV sa pravilom 110 ima kapacitet za univerzalno izračunavanje[24] . Danas pravilo 110 za elementarne celularne automate predstavlja najjednostavniji Tjuring-ekvivalentan sistem koji nam je uopšte poznat.

Posledica ovog dokaza omogućila je Volframu da izađe pred naučnu javnost sa jednom od najhrabrijih tvrdnji od početka proučavanja kompjutacionih sistema uopšte. Ukoliko su sistemi tolike jednostavnosti kao elementarni, jednodimenzionalni celularnih automati, sposobni za univerzalna izračunavanja, onda smo verovatno okruženi ogromnim brojem sistema sposobnih za univerzalno izračunavanje. Na osnovu ove plauzibilne pretpostavke, Volfram uvodi *princip kompjutacione ekvivalencije*: postoji, verovatno veoma nizak, prag kompleksnosti bilo kog sistema u univerzumu, preko koga on postaje sposoban za univerzalna izračunavanja. Ako odbacimo samo one sisteme čija su ponašanja očigledno trivijalno jednostavna, od preostalih sistema koji su kompleksni, ogroman broj njih bi morali da budu

sposobni da iznesu univerzalna izračunavanja. Razlike u kompleksnosti različitih izračunavanja su moguće samo do određenog praga kompleksnosti: sve preko njega su izračunavanja koja mogu da izvedu samo univerzalni sistemi. Drugim rečima, ogromna većina procesa koje opažamo u fizičkom, biološkom i socijalnom okruženju su *kompjutaciono ekvivalentni*: oni su posledica vremenske evolucije različito fizički realizovanih sistema univerzalnog izračunavanja koji se prema svojim kompjutacionim moćima ne razlikuju (Wolfram, 2002). Volframov princip kompjutacione ekvivalencije nije formulisan matematički, kao što ni njegova klasifikacija sistema u četiri klase nije formulisana matematički, već na osnovu fenomenoloških osobina celularnih automata. Prava diskusija ovog principa, ako se pokaže da ga je moguće postaviti dovoljno egzaktno da bude prihvaćen kao pravi naučni princip, tek predstoji.

Dva bitna zaključka o kompjutacionizmu u psihologiji slede iz rasprave o emergentnim sistemima. Prvi je da ljudski um, ako je uopšte kompjutacioni sistem, ne mora biti sistem baziran na eksplicitnim simboličkim reprezentacijama. Ona nastaje kao posledica direktne analogije uma sa teorijom formalnih sistema i arhitekturom digitalnih kompjutera koji su proizvod ljudskog dizajna. Međutim, čini se da biološka evolucija „dizajnira“ svoje sisteme na drugi način, za koji celularni automati i neuronske mreže predstavljaju bolje analogije. Drugi zaključak je da mesto ljudskog uma među drugim prirodnim sistemima uopšte ne mora da bude posebno, što je intuitivna pretpostavka svih rasprava o umu u istoriji psihologije i filozofije. Ako je izračunavanje, u suštini, sve što um čini, onda on čini ono što izgleda čini ogromna većina prirodnih sistema koji nas okružuju. Neuronske mreže, sa svojom arhitekturom, algoritmima učenja i vremenskom evolucijom strukture deluju kao superkompjuteri u odnosu na jednostavnost pravila elementarnih celularnih automata. Nasuprot intuiciji, ispostavlja se da mnogo jednostavniji sistemi od konekcionističkih neuronskih mreža, sistemi koje odlikuje izuzetna jednostavnost, jesu sposobni da iznesu izračunavanja ma koje kompleksnosti, tj. da dinamički razviju ponašanja ma koje složenosti.



Slika 9b. *Vremenska evolucija četiri elementarna jednodimenzionalna celularna automata.* Inicijalni uslovi za sva četiri automata na ovoj slici su slučajni. Poštujući Volframovu notaciju, gore levo: pravilo 255 (klasa I), dole levo: pravilo 164 (klasa II), gore desno: pravilo 90 (klasa III), dole desno: pravilo 110 (klasa IV).

Naravno, primena određenih modela - kao što su neuronske mreže - je izbor zasnovan na relativnoj jednostavnosti njihove analize. Niko ne bi odabrao da modelira bilo koji proces celularnim automatom klase IV čije programiranje predstavlja skoro pa prometejski podvig. Ipak, značaj teorijskih uvida koji proističu iz proučavanja emergentnih struktura je ogroman. U poređenju sa simbolicističkom paradigmom, emergentizam, bar *prima facie*, više odgovara opisu kognitivnih sistema koje odlikuju izuzetna fleksibilnost u prilagođavanju, masovna interaktivnost i visoka kompleksnost.

### 5.2.3 Kognitivni konstruktivizam i paradigma utelotvorene kognicije

Epistemološka pretpostavka prema kojoj ljudski um ne može da ima drugačiji odnos prema realnosti do interpretativnog zajedničkа je svim konstruktivističkim pravcima unutar savremene psihologije. Prema ovoj pretpostavci, um nije sredstvo pasivne recepcije sveta koji objektivno postoji, već *svojom aktivnošću nameće značenja i red informacijama* iz okoline (Raskin, 2002). Genetička epistemologija Žana Pijažea (Piaget, 1973/1994) i Kelijeva psihologija personalnih konstrukata (Kelly, 1955/1991) svakako su najpoznatije konstruktivističke teorije. Konstruktivizmu pripadaju i shvatanja socijalnih konstrukcionista (Gergen, 1999). Razgranata metateorijska diskusija koja počiva na istoj epistemološkoj pretpostavci povezuje *teoriju autopoietičkih sistema* biologa Maturane i Varele (Maturana & Varela, 1973, 1987), *radikalni konstruktivizam* u kibernetici fon Glasersfelda (Glasersfeld, 1984, 1989, 1995, 1999, 2001), *kibernetiku drugog reda* fon Ferstera (Foerster, 2002), *enaktivističku paradigmu* Varele, Tompsona i Rošove (Varela, Thompson & Rosch, 1992), ideje *utelotvorene kognicije* (engl. *Embodied Cognition*) u kognitivnoj psihologiji (Clark, 1998, Shapiro, 2011) i robotici i veštačkoj inteligenciji (Brooks, 1991, Anderson, 2003) u konceptualnu celinu koju mi nazivamo *kognitivnim konstruktivizmom*. Dva su razloga zašto se odlučujemo na uvođenje ovog novog termina. Prvi se odnosi na činjenicu da postoji veći broj teorijskih pristupa za koje je zajednička osnovna pretpostavka o tome da tek subjekat kroz svoje kognitivne činove konstituiše ono što je svet za njega. Drugi razlog je taj što verujemo da termin „*konstruktivizam*“ dobro ukazuje na suštinu ovakvih pristupa, prema kojima je, kao što ćemo videti, teško analitički razdvojiti ono što je „*svet*“ nekog saznajnog subjekta od onoga što su „*kognitivni činovi*“ tog subjekta: okolina i nosilac kognitivnih funkcija se uzajamno definišu, i bilo koje od ova dva bi moralo da promeni svoje suštinske odlike ako bi se suštinske odlike onog drugog promenile.

Neke bitne ideje koje preuzimaju savremeni konstruktivisti prethodno su razvijene u okviru *ekološke škole* pod uticajem Gibsonove teorije percepcije (npr. Shaw & McIntyre, 1974). Karakteristično je da svi navedeni teorijski pravci unutar kognitivnog konstruktivizma prihvataju emergentističku paradigmu i grade svoja shvatanja na njenim osnovama; nije nam poznat nijedan kognitivni konstruktivista koji razvija svoje ideje na osnovama simbolicističke paradigme.

Kognitivni konstruktivizam zahteva potpunu promenu eksplanatorne sheme koja je zajednička simbolicistima i konekcionistima. Dok teorije simbolicista i konekcionista suštinski počivaju na eksplanatornoj moći koncepta *reprezentacije*, u konstruktivizmu reprezentacije nemaju eksplanatornu moć da objasne saznajne procese i složena ponašanja. Klasični eksplanatorni model kognitivne psihologije izgleda ovako: (1) kognitivni sistemi preko senzornog aparata primaju informacije iz svoje sredine i na osnovu njih razvijaju reprezentaciju te okoline; (2) sve reakcije organizama čije kognitivne sisteme proučavamo jesu motorne radnje planirane na osnovu informacija sadržanih u reprezentaciji sredine. I simbolicistički i konekcionistički pristup koriste ovaj model objašnjenja, razlikujući se u shvatanju prirode internih reprezentacija, koje su eksplicitni simboli u formalnim sistemima simbolicista, a distribuirane reprezentacije kod konekcionista. Suština kognitivnog konstruktivizma nije u stavu prema prirodi reprezentacija, već u odbacivanju samog eksplanatornog modela koji na njima počiva.

Konceptualnu osnovu za kognitivne konstruktiviste predstavlja teorija autopoietičkih sistema čileanskih biologa Humberta Maturane i Fransiska Varele. Ova teorija je konzistentno izložena u knjizi „*Autopoiesis and Cognition: the Realization of the Living*“ (Maturana & Varela, 1973, 1987). *Autopoieza* (engl. *autopoiesis*) je termin koji su Maturana i Varela uveli da bi njime označili suštinsku osobinu živih sistema uopšte, kombinujući grčke reči da označe sisteme sposobne da „*sebe stvaraju*“. Maturana i Varela su predložili potpunu promenu epistemološke analize procesa adaptacije organizma sredini. Njihova epistemološka analiza počiva na shvatanju posmatrača (subjekta) kao suštinskog u definisanju sistema koji se posmatra i bazira sa na elementarnoj ontologiji u kojoj ključne uloge igraju *jedinstva* (engl. *unity*) i *distinkcije* (engl. *distinction*). Posmatrač je onaj koji određuje šta u svom okruženju definiše kao *jedinstvo* (engl. *unity*). Jedinstvo je elementarna ontološka kategorija autopoietičke teorije i posmatrač ga određuje kao „drugo“ (od sebe) koje može biti iskorišćeno u manipulacijama, deskripcijama i interakcijama sa drugim posmatračima. Inherentno relativna operacija posmatrača je operacija

*distinkcije*, koja predstavlja elementaran epistemički čin izdvajanja onoga što će smatrati određenim jedinstvom iz fundamentalnog fenomenološkog kontinuuma koji ga okružuje. Operacijama distinkcije izdvajamo sebe kao posebnu ontološku kategoriju od ostatka sveta, ljude i druge žive organizme koji nas okružuju, posebne fizičke objekte, konstelacije organizama i objekata. Postuliranjem jedne ovakve ontologije „*stvari*" pre epistemološke analize, autopoietička teorija se delom izlaže Drajfusove kritici (kritici ontološke pretpostavke). Kroz naredne korake razvoja autopoietičke teorije videćemo kako se ova pozicija relativizuje tako da se ne može više govoriti o univerzumu činjenica određenih nezavisno od konteksta, što je osnovna karakteristika ontologije simbolicizma koju kritikuje Drajfus.

Svako jedinstvo karakteriše specifična *organizacija* procesa koji su fundamentalni za njega: ono menja svoju strukturu, *zadržavajući svoj identitet sve dok održava iste organizacione procese* - procese koji su za njega karakteristični. Razbijanje čaše koju posmatramo na stolu je transformacija koja fundamentalno narušava organizaciju fizičkih procesa koji su uopšte omogućili da čaša bude distinkcijom određena kao jedinstvo u fenomenološkom kontinuumu. Veoma značajan za razumevanje teorije autopoietičkih sistema i kognitivnog konstruktivizma jeste fokus na to da su neki procesi u sredini prethodno omogućavali kognitivne akte koji bez njih ne bi bili mogući, ili ne bi imali smisla. Zakoni fizike koji uređuju našu sredinu omogućavaju tip materijalne ravnoteže zahvaljujući kome je nešto kao čaša uopšte moguće: naša distinkcija neke čaše kao jedinstva u fenomenološkom kontinuumu je operacija koja ne bi bila moguća da fizički zakoni kakvi jesu prethodno nisu omogućili stabilnu organizaciju fizičkih procesa koji čine jednu čašu. Naše saznanje sveta je neraskidivo povezano sa prirodom tog sveta: naše distinkcije izdvajaju jedinstva iz fenomenološkog kontinuuma na način koji otkriva duboku simetriju osobina naše percepcije i kognitivnih procesa s jedne, i zakonitosti organizacije materije oko nas, s druge strane. Bilo koje jedinstvo, rekli smo, počiva na specifičnoj vrsti organizacije za njega fundamentalnih procesa. U interakcijama, jedinstva prolaze kroz transformacije, od kojih su destruktivne one kojima se ukida njihova specifična organizacija. Interakcije dva jedinstva u svetu od kojih ni jedno ne trpi destruktivne transformacije nazivaju se *strukturalnim povezivanjem* (engl. *structural coupling*). Kroz strukturalno povezivanje, dva sistema izazivaju *perturbacije* jedan u drugome, promene kojima svaki može da se prilagodi interakcijama koje ne narušavaju njegovu fundamentalnu organizaciju. U strukturalnom povezivanju, teško je govoriti o tome da je jedno jedinstvo subjekat, a drugo objekat, kao u klasičnim analizama

u kojima organizmi svojim ponašanjem modifikuju sredinu, ili sredina svojim promenama vrši selekciju organizama kroz proces prirodne evolucije. Strukturalna povezivanja između različitih jedinstava (dva organizma, čoveka i alata, čoveka i kompjutera, dve grupe ljudi) koje ne uključuju destruktivne transformacije počivaju na kompatibilnim osobinama organizacije jednog i drugog jedinstva. Upotreba nekog alata u klasičnoj intencionalnoj analizi počiva na podeli situacije na subjekat i objekat: čovek preko impulsa svom senzomotornom sistemu upotrebljava fizički objekat da bi ostvario neku promenu u sredini. To je potpuno očigledno, ali je takođe potpuno očigledno i da čovek može da upotrebljava neki objekat kao alat samo ako je taj objekat prethodno dizajniran ergonomski tako da biće sa senzomotornim aparatom čoveka može da ga upotrebi, uzimajući u obzir sva njegova ograničenja. Nazad u cirkularan odnos organizma i sredine, neki objekat ne bi ni bio prepoznat kao alat da nema te osobine, niti bi čovek dizajnirao nešto kao alat tako da on ne zadovoljava ograničenja ljudskog organizma. Tipična interakcija čoveka i alata je u analizi zamagljena našom intuicijom o čoveku kao nosiocu „volje“ i „namere“, „izvoru energije“ koji upotrebljava pasivan fizički objekat: analiza Maturane i Varele upućuje na drugačije posmatranje takve interakcije kroz koncept strukturalnog povezivanja, u kome su čovek i alat koji koristi strukturalno povezani na način koji jesu zahvaljući čitavom spletu kompatibilnosti organizacione strukture jednog i drugog jedinstva koji određuje kako ta interakcija može, a kako ne može, da izgleda. Suštinski, „subjekat“ i „objekat“ su teorijski duhovi, koncepti koji su neophodni samo našoj prirodnoj, svakodnevnoj intuiciji.

U autopoietičkoj teoriji, *autonomni* su svi sistemi koji su definisani kao kompozitna jedinstva - jedinstva čija komponente čine mrežu u međusobnoj interakciji - koji zadovoljavaju dva uslova: (*i*) da kroz svoje interakcije neprestano regenerišu mrežu interakcija koja ih je stvorila, i (*ii*) da realizuju mrežu sopstvenih interakcija konzistentno u vremenu i prostoru, tako specifikujući svoj topološki domen, granicu, obezbeđujući da mogu da budu distinkcijama izdvojeni kao jedinstva (Varela, 1981). Živi, biološki sistemi, prema Maturani i Vareli, spadaju u višu kategoriju sistema koje oni nazivaju autopoietičkim sistemima. *Autopoietički* je onaj autonoman sistem koji kroz svoje interakcije i transformacije *proizvodi* sopstvene komponente, pored održavanja mreže procesa koja garantuje njegov identitet.

Ovakva epistemološka analiza živih sistema upućuje na bitne promene u našem razumevanju bihejvioralnih pojmova. Pojam ponašanja, centralni pojam biologije

i psihologije, u klasičnim analizama viših organizama poput čoveka, razume se kao *proizvod* centralnog nervnog sistema. Uloga centralnog nervnog sistema je da proizvodi ponašanje u funkciji prilagođavanja sredini. Analiza Maturane i Varele ukazuje na to da je ponašanje pogrešno razumeti kao proizvod organizma. Organizam, kao autopoietički sistem, baziran je na skupu interakcija između svojih komponenti, i kroz moguća strukturalna povezivanja sa sredinom on i sredina *šire domen svojih interakcija*: organizam i njegova fizička, biološka, psihološka i socijalna sredina se međusobno prožimaju u strukturalnom povezivanju. Njihove interakcije, pre povezivanja svojstvene pojedinačnim sistemima, počinju da se protežu kroz domen više sistema u strukturalnom povezivanju. Intencionalna analiza, koja počiva na intuiciji o čoveku i drugim biološkim sistemima kao izvorima aktivnosti i inicijatorima interakcija, ne otkriva da je ponašanje samo proširenje domena interakcija svojstvenih različitim sistemima, i da ponašanje uzima oblik koji uzima zahvaljujući ograničenjima i sistema koji posmatramo (pa govorimo o „njegovom“ ponašanju) i sredine koja ga okružuje. Obostrana ograničenja definišu formu strukturalnog povezivanja organizma i sredine, pa prema tome i formu ponašanja, koje predstavlja našu izolaciju parcijalnih interakcija koje onda pripisujemo samo jednom od sistema koji su strukturalno povezani - onome koji smo izabrali da proučavamo. Tako Maturana i Varela definišu ponašanje kao *deskripciju koji pravi posmatrač o promenama sistema u odnosu na sredinu sa kojom se taj sistem nalazi u interakciji* (Maturana & Varela, 1987).

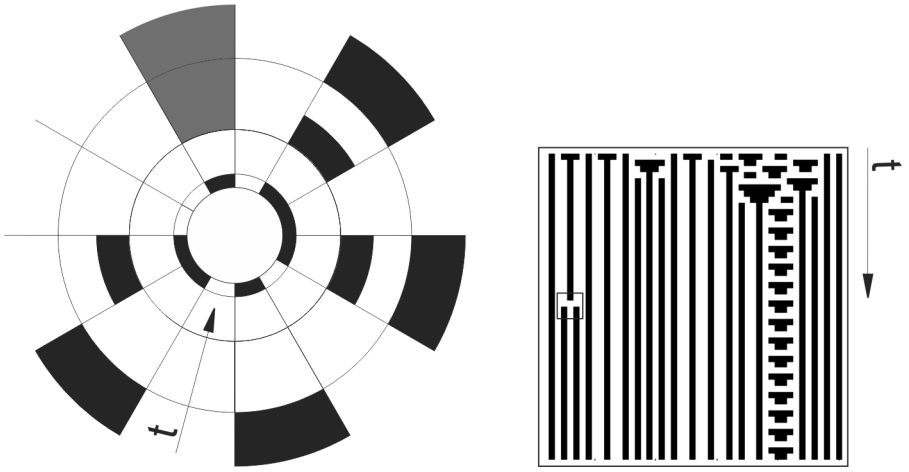
Autopoietička analiza živih sistema ukazuje da nikada, suštinski, ne možemo da govorimo o ponašanju nekog sistema koristeći samo predikate kojima opisujemo *taj* sistem: sredina i drugi sistemi koji su strukturalno povezani sa sistemom koji opisujemo takođe moraju da čine deo naše deskripcije. Razlog za ovo, kao što ćemo videti, nije samo u tome što je tek kroz analizu strukturalnog povezivanja moguć *potpun* opis nekog kognitivnog sistema. Elaboracija konstruktivističke analize pokazuje da tek analizom strukturalnog povezivanja opis nekog sistema postaje *funkcionalan*, u tom smislu što predikate kojima opisujemo neki kognitivni sistem možemo da interpretiramo tek kada u analizu uključimo i strukturu povezanosti tog sistema sa sredinom. Ovakvu elaboraciju nalazimo u radovima enaktivista, koji praktično predstavljaju manifest savremenog kognitivnog konstruktivizma.

Teorija *enaktivizma*, koju uvode Varela, Tompson i Rošova (1991), predstavlja elaboraciju i proširenje autopoietičke teorije na domen kognitivnih fenomena i danas je svakako najreprezentativniji vid kognitivnog konstruktivizma. Enaktivisti

direktno kritikuju reprezentacionističke teorije - simbolicizam i konekcionizam - sa pozicija koje smo već diskutovali. Interne reprezentacije, prema njihovom shvatanju, same po sebi nemaju eksplanatornu moć. Suština shvatanja ljudskih saznajnih fenomena postaje jasna tek kada se proučavaju *otelotvorenja uma*: činjenica da je on sistem koji preko tačno određenog senzomotornog sistema razmenjuje informacije sa okolinom, kao i činjenica da se on neprestano nalazi u kontinuumu sa dinamikom okoline. Na osnovu prethodne autopoietičke analize pojma ponašanja, koja je pokazala da definicija ponašanja nema smisla bez reference na strukturu okoline u kojoj posmatramo neko ponašanje, isto tako shvatamo da kognitivni fenomeni i reprezentacije nemaju smisla ukoliko nije specifikovana sredina u odnosu na koju oni uopšte imaju neku funkciju. Ono što prepoznajemo kao strukture okoline koje su relevantne u nekom činu (ponašanju) nekog organizma neraskidivo je povezano sa našim pokušajem da razumemo koje kognitivne strukture razvija taj organizam da bi mogao na okolinu da reaguje takvim činom.

Prema enaktivistima, svet za neki organizam nije „datost“, kako to pretpostavljaju simbolicisti i konekcionisti čiji modeli grade reprezentacije tog sveta. Organizam sa određenim kognitivnim sistemom *ustanovljava*[25] svet kroz evoluciju mogućih strukturalnih povezivanja sa njim: struktura okoline, struktura senzomotornog sistema i struktura kognitivnog sistema određuju domen takvih mogućih interakcija. Tako saznanje nikada nije pasivna obrada primljenih informacija, već je ono uvek *kognitivni čin*: aktivan proces koji je moguć samo kroz interakciju.

U ilustraciji enaktivističke ideje Varela, Tompson i Rošova (1991) koriste jednostavan primer elementarnog jednodimenzionalnog celularnog automata. Svaki jednodimenzionalni celularni automat je povezan u prsten, kao što smo primetili: prva i poslednja ćelija u nizu su susedi, iako ih češće posmatramo „razmotane“. Varela, Tompson i Roš u misaonom eksperimentu „bacaju“ jedan ovakav prsten jedinica i nula koje predstavljaju inicijalno stanje za automat sa pravilom 133[26] u sredinu koja se sastoji od slučajno generisanih nula i jedinica. Svaki put kad nula iz okoline dodirne jedinicu iz automata, ćelija označena kao aktivna tom jedinicom kopira nulu iz okoline i postaje neaktivna. Proces je ilustrovan na Slici 10.




Slika 10. *Slučajna okolina izaziva perturbacije u elementarnom jednodimenzionalnom celularnom automatu* (prema Varela, Thompson, & Rosch, 1991). Na levom panelu, celularni automat je prsten ćelija u sredini. Slučajna okolina je predstavljena aktivnim i neaktivnim ćelijama generisanim u „talasima“ koji se približavaju automatu u smeru strelice. Dva siva polja predstavljaju dva uzastupna talasa aktivacije koji će pogoditi celularni automat u *t+1* i *t+2*. Na desnom panelu, vremenska evolucija celularnog automata: označen je momenat u kome dve perturbacije za redom menjaju spaciotemporalnu strukturu vremenske evolucije.

Na Slici 10, celularni automat (zatvoren prsten ćelija u sredini na levom panelu) se nalazi u slučajnoj okolini koja u svakom diskretnom trenutku vremena proizvodi nasumično broj aktivnih i neaktivnih ćelija. Kada ćelije iz okoline „dodirnu“ ćelije celularnog automata, ovaj kopira njihovo stanje bez obzira na interna pravila sopstvene vremenske evolucije. Na desnom panelu Slike 10. prikazana je vremenska evolucija celularnog automata po pravilu 133. Startovan od slučajnih inicijalnih uslova, ovaj automat evoluira u repetitivnu strukturu klase II. U jednom trenutku, označenom na dijagramu vremenske evolucije automata, jedna od stabilnijih linija prima dve perturbacije za redom: prethodno aktivna ćelija automata dva puta iz okoline prima signal „1“, ond. dva puta za redom biva „pogođena“ aktivnom ćelijom. Ovo je jedina perturbacija koja menja spaciotemporalnu strukturu vremenske evolucije ovog automata: vidimo da posle duple perturbacije jedna prethodno stabilna linija biva zamenjena dvema stabilnim linijama. Celularni automat počinje da pokazuje elementarnu kognitivnu funkciju selektivnog reagovanja na okolinu, pošto se ponaša kao da „bira“ duple perturbacije kao jedino na šta će „reagovati“ promenom strukture svoje evolucije. Suština enaktivističke teorije je u sledećem: *niti struktura pravila koja generišu ovaj celularni automat eksplicitno određuje da on može da se ponaša kao da prepoznaje duple perturbacije iz okoline*, *niti*

*bilo šta u strukturi okoline to može da objasni.* Samo *interakcija* između ova dva kompjutaciona sistema - automata i slučajne okoline - otkriva mogućnost da celularni automat interpretiramo kao sistem sposoban za elementarno prepoznavanje određenih promena u okolini. Tek u strukturalnom povezivanju (*i*) pravila za generisanje celularnog automata preko (*ii*) „senzomotornog sistema“ automata tj. njegove osobine da kopira stanja okoline, sa (*iii*) stanjem okoline, pojavljuje se prethodno nepoznata (i otud, emergentna) odlika automata koju mi, kao posmatrači, interpretiramo kao elementarno selektivno ponašanje. Kada ne bismo imali informaciju o tome da je dupla perturbacija ono što je promenilo spaciotemporalnu strukturu evolucije ovog automata, pojava nove spaciotemporalne strukture bi za nas bila neinterpretabilna. Isto tako, kada ne bismo poznavali pravila koja generišu ovaj automat, ne bismo razumeli zašto on tako reaguje na impulse iz okoline sa kojom je povezan. „Sredina“ i „kognitivni sistem“ u konstruktivističkoj teoriji nose dva aspekta jedne iste informacije koja omogućava interpretaciju nečega kao kognitivnog čina uopšte. Ako bi sam celularni automat bio saznajno biće i koristio dijagram svoje evolucije kao neku vrstu rudimentarnog memorijskog sistema, promeni u tom memorijskom sistemu koja nije u skladu sa pravilima njegovog generisanja on ne bi mogao da dodeli smislenu interpretaciju kada ne bi bio u stanju da je dovede u vezu sa tačno određenom promenom u sredini. Tek u strukturalnom povezivanju sa sredinom, predikati za opis sredine, poput „dve neaktivne ćelije jedna za drugom“, i predikati za opis sistema, poput „zamena jedne dvema stabilnim linijama“, postaju semantički interpretabilni. Kroz strukturalno povezivanje sa sredinom ovaj jednostavan kompjutacioni sistem *ustanovljava* šta je svet za njega tako što razvija (a) osnovu za prepoznavanje određenih regularnosti u sredini i (b) njihovu interpretabilnost.

Kognitivni konstruktivizam, dakle, zahteva radikalnu promenu pozicije u odnosu na simbolicizam i konekcionizam: koncept reprezentacije, od centralnog značaja u kognitivnoj psihologiji od kada ova disciplina postoji, prepušta eksplanatornu moć interakcijama u strukturalnom povezivanju uma i okoline. U kontekstu filozofije uma, predlozi slični konstruktivističkim poznati su od ranije (up. poređenje paradigmi „enkodingizma“ i „interaktivizma“, Bickhard, 1996). Veliki problem sa razvoj enaktivističkih teorija predstavlja složenost interakcija koje je neophodno analizirati da bi se sa enaktivističkog stanovišta govorilo o realnom kognitivnom sistemu. Videli smo kako jednostavni kompjutacioni sistemi omogućavaju jasne primere enaktivističke analize, ali nam ostaje samo da se pitamo kako bi takva

analiza bila sprovedena da je umesto automata klase II diskutovan neki od automata klase IV, čije smo zanimljive osobine imali prilike da upoznamo. Konstruktivističke teorije zadržavaju kontinuitet sa emergentističkim pristupom: osobine sistema koje dobijaju smisao tek u njegovim interakcijama sa svetom jesu emergentne odlike tih interakcija. Ideje emergentizma i konstruktivizma danas okupljaju skoro sve pripadnike naučne zajednice koja zauzima kritički odnos prema standardnoj simbolicističkoj paradigmi KKP. Osnovna kritika konstruktivističke paradigme je što se ona skoro isključivo zadržava na nivou apstraktne teorijske analize. Sigurno, to je posledica izuzetne složenosti koja se javlja pri pokušaju konstruktivističke analize realnog kognitivnog funkcionisanja čoveka.

Diskusijom kognitivnog konstruktivizma završavamo pregled kompjutacionističke paradigme u savremenoj kognitivnoj psihologiji: sada su nam poznate i standardna paradigma (simbolicizam) i njene kritike (emergentizam i konstruktivizam). Bez obzira na razlike u teorijskim pozicijama, pretpostavka savremene kognitivne psihologije jeste da su saznajni procesi u nekom smislu izračunavanja, ili da se mogu opisati kao izračunavanja. Ono što karakteriše odnos tri diskutovane kompjutacione paradigme, pored razlika u prirodi shvatanja kognitivnih funkcija, jeste i forma naučnog objašnjenja koju one pružaju.

## 5.3 Naučno objašnjenje u tri teorijske paradigme KKP i njihov odnos

Standardna simbolicistička paradigma formuliše naučna objašnjenja kognitivnih funkcija u formi *objašnjenja putem mehanizma* (Wright & Bechtel, 2007). Algoritmi koji izračunavaju parametre potrebne za kontrolu ponašanja u odnosu na probleme koje kognitivnom sistemu nameće okolina predstavljaju *mehanizme interne kognitivne obrade informacija* i jesu naučna objašnjenja simbolicističkih teorija. Mehanizmi su paradigmatičan oblik naučnog objašnjenja ne samo u psihologiji, a u simbolicističkoj paradigmi KKP oni su sasvim specifičnog tipa *mehanizama za obradu informacija.* Metafora digitalnog kompjutera, čiji *hardver* (u analogiji kojom se gradi objašnjenje kognicije, neurofiziološki aparat) kontroliše čisto simbolička forma *softvera* (u analogiji, kognitivni procesi kao procesi izračunavanja u okviru određenog formalnog sistema), predstavlja osnovu za izvođenje ovog tipa mehanicističkog objašnjenja u simbolicističkoj paradigmi KKP. Poreklo savremenog mehanicističkog pogleda na svet, odn. shvatanje

da je naučno razumevanje Univerzuma uopšte moguće ako ga razumemo kao vrstu mehanizma čije funkcionisanje treba egzaktno opisati, nalazi se u filozofiji Rene Dekarta; zato se često u teorijskim diskusijama u KKP karakterističan pristup objašnjenju preko mehanizama za obradu informacija naziva *kartezijanskim*. Kompjutacionistička revolucija XX veka u velikoj meri se ogleda u izumu nove vrste eksplanatornog mehanizma, upravo kompjutacionog mehanizma, odn. mogućnosti da se simbolička aktivnost, karakteristična za čoveka, obavlja potpuno mehanički u odgovarajućem fizičkom sistemu: digitalnom kompjuteru. Savremena filozofija nauke karakteriše *mehanizam* kao objašnjenje za određene fenomene na sledeći način: „*Mehanizam je struktura koja izvodi određenu funkciju kroz svoje komponentne delove, komponentne operacije i njihovu organizaciju. Orkestrirano funkcionisanje mehanizma odgovorno je za jedan ili više fenomena*“ (citirano prema Bechtel & Abrahamsen, 2005, str. 423, naš prevod). Organizacija komponenata mehanizma može biti čisto temporalna, tako da komponente operišu sekvencijalno, proizvodeći proces odgovoran za određeni fenomen pod opservacijom, dok živi sistemi pokazuju kompleksnije forme organizacije koje često nisu sekvencijalne. Neuronske mreže, eksplanatorni mehanizmi konekcionista, podrazumevaju visoko paralelan oblik organizacije koji, iako sekvencijalan u procesu učenja, tek zahvaljujući toj osobini paralelizma pruža eksplanatornu moć koju konekcionisti očekuju od svojih modela. Simbolicistički pristup, odn. klasična kognitivna obrada informacija, razmatra i paralelne i serijalne oblike organizacije kognitivnih mehanizama (up. Townsend & Ashby, 1983 za matematičko uvođenje paralelnih i serijskih modela), ali za nju nije karakterističan opis putem mehanizama visokog paralelizma karakterističnih za konekcionistički, emergenistički pristup.

Karakteristična razlika simbolicističkog i konekcionističkog pristupa je, dakle, u *formi* mehanizama koje biraju u mehanicističkom objašnjenju kognitivnih fenomena. Međutim, po nama je značajnija jedna druga razlika između ova dva pristupa, razlika koja se javlja kad ih razmotrimo kao objašnjenja putem naučnih zakona. I simbolicistička i konekcionistička paradigma su, osim postuliranja eksplanatornih mehanizama, u stanju da objašnjavaju kognitivne fenomene i putem *deskripcije odgovarajućih naučnih zakona.* Kognitivni mehanizmi, u formi eksplicitnih simboličkih reprezentacija ili u formi neuronskih mreža, karakterišu se preko skupova parametara - najčešće realnih brojeva - čije vrednosti, zajedno sa strukturom postuliranog mehanizma, određuju autput funkcije kognitivnog sistema. Autput funkcije sistema determinišu njegov odgovor na uticaj sredine, dakle ponašanje koje

mi osmatramo u eksperimentalnim studijama. Tek kao funkcija opisa (a) sredine, tj. eksperimentalnog dizajna, karakteristika stimulusa koji se koriste, (b) postuliranog kognitivnog mehanizma i (c) odgovarajućih parametara za njegovo funkcionisanje, razvija se naučno objašnjenje u KKP. Pravilnosti u odnosu karakteristika stimulusa i odgovora opisuju se matematički, odgovarajućim funkcijama čije su forme posledica strukture postuliranog mehanizma i odgovarajućih vrednosti parametara koje karakterišu njegov rad. Naučni zakoni se u KKP javljaju na Marovom kompjutacionom nivou 3. kao posledica algoritama koji se analiziraju na nivou 2; na nivou 3, teorije su *kompjutacione*, dok su na nivou 2. one *procesne*[27]. Uzimajući primer konekcionističkih modela, naučno objašnjenje u tom slučaju počiva na (a) arhitekturi neuronske mreže koja je postulirana, dakle složaju veza, broju jedinica i slojeva jedinica, (b) funkciji aktivacije jedinica u mreži, i konačno (c) vrednostima težina veza koje rezultiraju kao stabilno stanje algoritma učenja mreže. Tek u konjukciji sva tri (arhitekture, aktivacionih funkcija i vrednosti parametara) dobijamo konekcionističku teoriju određenog fenomena. Odgovor, odn. rezultat autput funkcije nekog kognitivnog mehanizma, onda mora da ima osobine matematičke funkcije koja objašnjava vezu između variranja situacionih parametara, odn. opisa stimulusa u određenom eksperimentalnom setingu, i variranja odgovora koje ljudski ispitanici daju sa variranjem tih situacionih parametara. Na taj način, simbolicističke i konekcionističke teorije formulišu i drugi tip naučnog objašnjenja kognitivnih fenomena, *objašnjenja putem zakona*. Eksplanatorni mehanizmi ovih teorija, u konjukciji sa odgovarajućim vrednostima njihovih parametara i sklopom matematičkih funkcija koje izračunavaju onda pružaju objašnjenja za zakone koji povezuju dva tipa varijabli koje karakteristično proučavamo u naučnoj psihologiji: S tipa, odn. opisa stimulacije, i R tipa, odn. opisa odgovora. Eksplanatorni mehanizmi, funkcije koje ih karakterišu i vrednosti odgovarajućih parametara tako su u pravom smislu reči hipotetski konstrukti kako ih je formulisala filozofija nauke prve polovine XX veka: neopservabilni entiteti koji omogućavaju objašnjenje opservabilnih fenomena (MacCorquodale & Meehl, 1948). Više puta do sada mi smo naglasili da je karika koja često nedostaje u lancu teorijske analize KKP ona koja objašnjava da nije moguće postulirati bilo koje hipotetske konstrukte u objašnjenju bilo kog kognitivnog fenomena. Struktura naučnih podataka koje pružaju ispitanici u eksperimentalnoj psihologiji (ili eksperimentalnoj ekonomiji, ili ma kojoj drugoj društvenoj nauci koja koristi eksperimentalni metod) ograničava skup matematičkih inferencija kojima se utvrđuje egzistencija mogućih neopservabilnih konstrukata

u objašnjenju tih podataka - npr. funkcija korisnosti, ili funkcija ponderisanja verovatnoća u teoriji odlučivanja. Nažalost, svest o ovom ograničenju najčešće nije eksplicitna u empirijskim studijama u okviru KKP, sa izuzetkom teorije odlučivanja u kojoj se veoma vodi računa o odnosu struktura opservabilnih podataka i hipotetskih konstrukata koji učestvuju u njihovom objašnjenju.

Pošto smo ovako rasvetlili odnos između postuliranih kognitivnih mehanizama i naučnih zakona u KKP, sada je potrebno da razumemo razliku između emergentističkih - konekcionističkih i dinamičkih teorija - s jedne, i simbolicističkih teorija, s druge strane, u odnosu prema *objašnjenju putem zakona* (v. Psillos, 2007, za savremene diskusije tipova naučnih objašnjenja i odnosa prema naučnim zakonima). Ta suštinska razlika između teorija koje pripadaju ovim dvema paradigmama je u formi *pravila korespondencije* koja povezuju formalne, hipotetske konstrukte teorije, odn. njene dispozicione i teorijske pojmove, sa opservabilnim fenomenima (Carnap, 1956)[28]. Prema klasičnoj, sintaksičkoj koncepciji strukture naučnih teorija, svi teorijski termini neke naučne teorije moraju da budu povezani sa opservacionim terminima preko pravila korespondencije, koja su jedina pravila naučne teorije koja sadrže i teorijske i opservacione termine. Na taj način naučna teorija dobija fundamente u opservacionom domenu (a veći deo napora pozitivista poput Karnapa u periodu posle Drugog svetskog rata svodio se na to da se pokaže kako tim putem nauka obezbeđuje empirijski sadržaj, odn. smislenost za svoje teorijske koncepte; uspeh tog pokušaja, možda najbolje oličenog u Karnapovom metodu remzifikacije naučnih teorija, tumači se kao delimičan, up. Carnap, 1966, Psillos, 2000, Melia & Saasti, 2006, Friedman, 2009). Podsećajući se sada prirode kognitivnih teorija karakterističnih za emergentističke teorije, shvatamo da pravila korespondencije u njihovom slučaju moraju da budu veoma složena, kao i da njihova priroda nije očigledna. U emergentističkim teorijama, struktura koja objašnjava kognitivne fenomene posledica je sinhrone i asinhrone, visoko paralelne interakcije obrade informacija na mikro-nivou deskripcije sistema: ona je uvek neko stabilno stanje u faznom prostoru odgovarajućeg dinamičkog sistema, prostoru koji predstavlja domen deskripcije sistema na makro-nivou. Primer su stabilna stanja parametara (težina veza) u neuronskoj mreži posle dovoljnog broja iteracija odgovarajućeg algoritma učenja. Sa vrednostima parametera koje postižu u svojim stabilnim stanjima, autput funkcije dinamičkih sistema poput neuronskih mreža postaju „prave funkcije“ u odnosu na situacione parametre tj. stimuluse: mreža daje odgovore koji korespondiraju očekivanom ponašanju sistema koji uči da

optimalno („tačno“) reaguje na strukturu svoje okoline. Dakle, teorijski pojmovi neke konekcionističke teorije uopšte nisu očigledni: oni predstavljaju određene skupove vrednosti velikog broja varijabli u opisu neke neuronske mreže kao dinamičkog sistema, i tek struktura svih odnosa među njima jeste teorijski koncept koji pravila korespondencije preslikavaju na opservacione termine. Kompletno mehanicističko objašnjenje u ovakvim sistemima je praktično neupotrebljivo, jer bi ono zahtevalo iznošenje kompletne dinamičke istorije sistema koji objašnjava neko ponašanje (konekcionistički autori nekad i pružaju uvid u takve dinamičke istorije, up. Rogers & McClelland, 2004). Ovaj uvid nas vraća na pitanje same definicije emergentnih odlika, koje su po našem mišljenju efikasno rešili još Hempel i Openhajm 1965: neka je $Pt$ oznaka za relaciju „biti deo“, koja se odnosi na neki sistem $w$ i njegove delove; odlika $W$ sistema $w$ je emergentna odlika u odnosu na neke atribute $G$ u deskripciji sistema $W$, relacije $Pt$ i teoriju $T$, ukoliko nije moguće izvesti inferenciju ka $W$ iz teorije $T$ u odnosu na atribute $G$ i sve relevantne $Pt$ relacije (Hempel & Oppenheim, 1965/2008). Ponovimo već rečeno: u emergentističkom objašnjenju kognitivnih fenomena, npr. postuliranjem određenje neuronske mreže, osobine dinamičkog sistema koji karakteriše tu neuronsku mrežu na makro-nivou analize nisu direktno interpretabilne na nivou mikro-analize gde nalazimo jedinice mreže - neurone - i njihove funkcije aktivacija. Međutim, za našu trenutnu raspravu bitan element Hempelove i Openhajmove definicije emergentne odlike je onaj koji ukazuje na to da „emergentnost“ nekog svojstva ne može biti shvaćena bez odnosa prema nekoj teoriji $T$ koja važi na mikro-nivou analize. Ono što sledi je upravo naš zaključak o prirodi pravila korespondencije koja bi povezivala neka emergentna svojstva sa opservacionim varijablama: njih nije moguće spustiti na nivo mehanizma u smislu deskripcije na mikro-nivou analize. Međutim, za neke sisteme ni povlačenje korespondencije između stabilnih makro-stanja ka opservabilnim strukturama nije jednostavno: ovo je razlog zbog kojeg smo tvrdili da neki emergentistički sistemi, poput celularnih automata, uopšte ne predstavljaju pogodne modele za naučno objašnjenje saznanja. Zbog ogromne kompleksnosti interakcija u takvim sistemima, teško je steći čak i hipotetički uvid u pravila korespondencije koja bi povezivala sve relevantne teorijske koncepte sa opservacionim terminima. Samo teorijski koncepti sa makro-nivoa analize mogu da učestvuju u smislenim pravilima korespodencije u mehanicističkim objašnjenjima koja uključuju emergentističke sisteme. Međutim, u slučaju sistema kao što su celularni automati, i te stabilne, kvalitativne odlike (poput onih koje analizira Wolfram) na makro-nivou nije lako

prepoznati. Celularni automati treba pre posmatrati kao pomoćna, didaktička sredstva kojima se demonstriraju određeni teorijski koncepti, nego kao kandidate za realna naučna objašnjenja kognitivnih fenomena. Za razliku od upravo diskutovane osobine emergentističkih teorija, pravila korespondencije koja povezuju teorijske koncepte teorija u simbolicističkom ruhu sa opservacionim terminima su svakako daleko jednostavnija. Iako simbolicistički sistemi po sebi uopšte ne moraju da budu jednostavni, komponente njihovih eksplanatornih mehanizama moguće je razdvojiti tako da može da se identifikuje doprinos svake od njih u formiranju odgovarajuće autput funkcije kognitivnog sistema.

Koja je karakteristična forma konstruktivističkog naučnog objašnjenja, u paradigmi enaktivizma odn. utelotvorene kognicije? Objašnjenje u kome eksplanatornu ulogu igra neopservabilan teorijski koncept kognitivne reprezentacije to sigurno nije. Analiza koja se dosledno izvodi pod konstruktivističkom paradigmom ne omogućava - osim kao aproksimaciju - ni jasno razdvajanje značenja opservacionih termina S (sredinskih parametara, „stimulusa“) i R („reakcija“, osmotrenog ponašanja): podsetimo se konstruktivističke definicije ponašanja kao deskripcije koji koju pravi posmatrač o promenama sistema *u odnosu na sredinu sa kojom se taj sistem nalazi u interakciji* (Maturana & Varela, 1987). Ukoliko R-termini u uspostavljanju dispozicionih pojmova, pre uvođenja pravih teorijskih pojmova (u Karnapovom smislu reči) koji bi artikulisali teorije na apstraktnijem nivou, uvek nose referencu prema S-terminima, koncepcija strukture naučne teorije - bez obzira da li je izražena u sintaksičkoj (Carnap, 1956, 1966) ili semantičkoj verziji (Suppes, 1960, 2002) - mora da pretrpi izvesne izmene da bi se omogućila analiza takvih odnosa. Ovaj problem u određenju za psihologiju i druge društvene nauke veoma važne klase dispozicionih pojmova biće u fokusu naše rasprave u VI poglavlju, kada ćemo pokušati da pokažemo kako je konstruktivističko shvatanje u raspravi o racionalnosti saznanja praktično neizbežno.

# 6 Racionalnost saznanja u kompjutacionizmu

Promena u istoriji psihologije koja je suštinski označila napuštanje bihejviorizma u korist kognitivne psihologije, u prvim decenijama druge polovine XX veka, jeste promena metodološkog stava prema proučavanju internih, neopservabilnih psiholoških stanja i mehanizama. Dok je klasični bihejviorizam u potpunosti isključio mogućnost naučnog govora o neopservabilnim psihološkim konceptima

poput pažnje i pamćenja, svodeći psihologiju na jednostavnu S-R formulu, a neobihejviorizam i neo-neobihejviorizam postepeno popuštali ovu strogu metodološku ogradu (Radonjić, 1967/94), kognitivna psihologija je potpuno otvorila vrata za proučavanje internih kognitivnih mehanizama i stanja kao hipotetskih konstrukata. Kompjutacionizam, kao njena teorija uma, obezbedio je podršku za ovakav potez. Sa razvojem shvatanja da je veštačka implementacija formalnih sistema u digitalnim kompjuterima moguća, kao i da fizički sistemi simbola pokazuju iste apstraktne osobine koje važe za simboličke funkcije čoveka - ideja o psihološkom proučavanju internih kognitivnih mehanizama čoveka dobila je krila na kojima se razvila u savremenu KKP. Međutim, veza sa bihejviorističkim načelima nije u potpunosti prekinuta. Ponašanje je prestalo da bude jedini, ekskluzivan predmet psihološke nauke, ali je ostalo *jedini izvor opservabilnih podataka* u naučnoj psihologiji[29], pošto kognitivna psihologija nije promenila stav prema kome introspekcija ne može biti izvor naučnih podataka. Ova situacija upućuje na centralni značaj *egzaktne formulacije odnosa između neopservabilnih hipotetskih konstrukata i metodologije koja povezuje tvrdnje o njihovoj egzistenciji i osobinama sa strukturama naučnih podataka koji o toj egzistenciji ili osobinama svedoče.*

U prvom poglavlju rada, kroz diskusiju problema racionalnog izbora, videli smo kako se aksiomatski gradi naučna teorija koja povezuje ograničenja u strukturi naučnih podataka sa mogućnošću inferencija o neopservabilnim konstruktima. Fon Nojman i Morgenšternova teorija očekivane korisnosti aksiomatizuje jednostavne intuicije o izboru u uslovima rizika, i pokazuje da ukoliko te intuicije važe, ponašanje posmatranog sistema može alternativno da se opiše pomoću internih, neopservabilnih funkcija korisnosti, a njegove odluke dovedu u korespondenciju sa izračunavanjem očekivane korisnosti bilo koje distribucije verovatnoća nad određenim vrednostima. Činjenica da ponašanje nekog subjekta zadovoljava aksiome teorije očekivane korisnosti ekvivalentna je sa tvrdnjom o tome da taj subjekat svoje odluke donosi na osnovu interne funkcije korisnosti i izračunavanja očekivane korisnosti po jednačini (2) do koje je došao još Bernuli. U ovom delu naše rasprave upoznali smo izvor ovakve strategije konstrukcije naučne teorije kroz rad Remzija. Tek ukoliko empirijski podaci koje prikupljamo zadovoljavaju određene pretpostavke, moguća je inferencija o neopservabilnim konstruktima koji postaju osnovne gradivne jedinice odgovarajuće kognitivne teorije. Činjenica da veliki broj ljudi u empirijskim istraživanjima krši aksiomatiku fon Nojmana i Morgnešterna, dajući psiholozima i ekonomistima na uvid strukture podataka koje ne zadovoljavaju

ograničenja koja slede iz njihove teorije, motivisala je razvoj alternativnih shvatanja poput teorije izgleda. Matematička inferencija kojom se dokazuje postojanje funkcije korisnosti u teoriji fon Nojmana i Morgenšterna moguća je samo ako je aksiom nezavisnosti (uz druge aksiome) zadovoljen; ukoliko ga ispitanici u eksperimentalnim situacijama krše, nije oboren samo aksiom, već i cela inferencija koja tvrdi da postoje neopservabilni konstrukti karakteristični za teoriju očekivane korisnosti. Različite strukture empirijskih podataka ukazuju na mogućnost inferencija ka različitim neopservabilnim konstruktima. Aksiomatika teorije izgleda, koju još nismo diskutovali, tako, omogućava inferencije o sasvim drugačijim teorijskim konceptima od onih na koje upućuje aksiomatika teorije očekivane korisnosti, na primer o funkciji ponderisanja verovatnoće, ili averziji prema gubicima. U svakom slučaju, aksiomatska ograničenja se odnose na strukturu opservabilnih podataka, što za psihologiju znači: na *strukturu ponašanja*. Kibernetičke analize uma kao crne kutije, „*inženjering unazad*“ uma, te metodološke okosnice KKP, mogući su, dakle, samo pod uslovom fiksnog, egzaktnog opisa strukture ponašanja koju „proizvodi“ um. U kojoj meri i na koji specifičan način su takva ograničenja osvešćena u radu na nekim od najvažnijih problema u savremenoj KKP diskutovaćemo u III delu. Ono što nas sada interesuje su fundamentalna ograničenja analize racionalnosti saznanja u odnosu na ovaj princip povezivanja strukture ponašanja sa neopservabilnim konstruktima. Vodilja naše diskusije ostaje više puta ponovljeno shvatanje da je *racionalan subjekt S čije ponašanje konzistentno otkriva da on dela u skladu sa svojim verovanjima B kako bi ostvario svoje ciljeve G u sredini E*. Videli smo da se ovakvo shvatanje implicitno nalazi u svim aktualnim strategijama analize ljudskog saznanja kao što su Andersonova racionalna analiza ili Marovo seciranje analize problema saznanja kroz kompjutacioni, algoritamski i implementacioni nivo. Takođe, videli smo da je Remzijeva analiza - kao i sve koje slede njenu formu - takođe moguća samo ako se pretpostavi motivisan subjekt, subjekt koji ima cilj, koji pokušava da stekne što više vrednosti može u situaciji koja uključuje ograničene resurse. Međutim, ispunjavanje samo ovog uslova ne garantuje siguran razvoj shvatanja o internim kognitivnim stanjima, funkcijama i mehanizmima: u narednim redovima diskutujemo dodatna ograničenja koja onemogućavaju razvoj takvog shvatanja u potpunosti.

## 6.1 Realna kompleksnost ponašanja, prostor hipotetskih konstrukata i problem selekcije modela

Prvi problem sa kojim se suočavamo je problem *realne kompleksnosti ponašanja* i njegov odnos prema *kompleksnosti hipotetskih konstrukata* koji, po pretpostavci, reprezentuju funkcije i mehanizme koji omogućavaju, „proizvode“ to ponašanje. Pod „realnom kompleksnošću“ ovde mislimo na kompleksnost ponašanja kakvu srećemo u ekološki relevantnim (vanlaboratorijskim) uslovima, podjednako kao i u diskusiji posmatranja veoma složenih ponašanja pod eksperimentalnim uslovima - ako su i kada takva posmatranja moguća. Ograničenja ponašanja koja propisuje aksiomatika teorije očekivane korisnosti odnose se na ekstremno jednostavan podskup ljudskog ponašanja u problemima odlučivanja. U prvom poglavlju smo videli koliko komplikovane naučne problematike je sledilo iz razmatranja već tako jednostavnih struktura ponašanja. Ne samo usložnjavanje problema odlučivanja, npr. proučavanjem problema multiatributivnog odlučivanja, u kome su alternative opisane na više od jedne dimenzije vrednosti, ili proučavanjem složenih interakcije aktera u teoriji igara, već i svest o tome da se kognitivni procesi koje proučavamo u realnom okruženju uvek odvijaju u kontekstu ogromnog broja drugih kognitivnih procesa, upućuju nas na staru istinu o tome da je naše proučavanje mentalnih procesa uvek fundamentalno ograničeno. Klasična dela kibernetike ne propuštaju da nas opomenu na ovaj problem (Ashby, 1956): kognitivni model, baš kao i model svakog drugog kompleksnog sistema koji nauka analizira, nužno je *sveden na onaj broj varijabli* (stepeni slobode) *koji je moguće artikulisati u eksperimentalnom setingu i teorijskoj analizi*. Analiza kognitivnog funkcionisanja čoveka i viših organizama je *an generale* ograničena na današnjem stupnju razvoja nauka; fundamentalno je pitanje da li u totalitetu ona uopšte, principijelno, ikada može da bude dosegnuta.

Iz ovog ograničenja slede dve suštinske posledice po konstrukciju teorija kognitivne psihologije. Prva, na neki način principijelno manje opasna posledica, odnosi se na kompleksnost naučne - matematičke i konceptualne - analize problema koje proučavamo. Pretpostavimo da smo u mogućnosti da formulišemo aksiomatski sistem koji bi plauzibilno, blisko intuiciji, ograničavao strukture nekog veoma složenog ponašanja, omogućavajući inferencije ka velikom broju složenih, neopservabilnih kognitivnih funkcija i mehanizama. Čak i kada bi u principu eksperimentalno proučavanje veoma složenog ponašanja bilo moguće, sasvim je izvesno da bi nas i tek mala usložnjavanja matematičkog sistema koji bismo morali

da proučavamo suočila sa sistemima jednačina koje bi bile izuzetno teške, ako ne i nerešive, analitičkim metodama, a moguće i van domašaja naših metoda numeričke aproksimacije. Uzimajući u obzir rastuću snagu izračunavanja u savremenim kompjuterskim sistemima koje koristimo za simulacije i analize naučnih podataka, recimo da se ovde suočavamo sa problemom koji je možda „u principu rešiv"; ipak, ovaj stav ne treba prihvatiti bez opreza.

Druga posledica nas suočava sa daleko ozbiljnijim problemom, problemom za koji verujemo da u velikom broju situacija ni u principu nije rešiv. Aksiomatike naših teorija, shvatili smo, zapravo definišu ograničenja struktura ponašanja koje ustanovljavamo naučnim posmatranjem. Ukoliko je neka određena aksiomatika zadovoljena, ona obezbeđuje inferencije ka neopservabilnim konstruktima, koje omogućavaju opis sistema ekvivalentan aksiomatskom. Kognitivna psihologija zatim postulira te konstrukte - mehanizme i funkcije - kao interne („u glavi", „u kognitivnom sistemu") i na njih prenosi eksplanatornu moć svojih teorija. Sa porastom u kompleksnosti ponašanja koje se posmatra, a koji je nužan ukoliko pokušavamo da razvijemo teorije o složenijim kognitivnim procesima, aksiomatike koje postuliramo da bismo mogli da povezujemo opservacije sa hipotetičkim konstruktima postajaće sve složenije. Posledica ovog usložnjavanja struktura ponašanja je ne samo u tome da neopservabilni eksplanatorni mehanizmi takođe postaju složeniji, već da se otvara mogućnost da *više alternativnih opisa (mehanizama, funkcija, procesa) istog sistema zadovoljava jedinstvenu aksiomatiku.* Primer koji ilustruje ovu situaciju odnosi se na već diskutovane funkcije korisnosti. Objasnili smo u I delu da iz fon Nojman-Morgenšternove teorije očekivane korisnosti sledi da se donosilac odluka ponaša kao da izračunava funkciju korisnosti, ali da iz iste teorije ne slede sve osobine funkcije korisnosti, već samo neka ograničenja njenih osobina. Razlika koju diskutujemo je razlika između modela teorije očekivane korisnosti datog u *neparametarskoj formi*, gde funkcija korisnosti nije precizno definisana veću su dokazane samo njena egzistencija i neke osobine, i modela teorije u *parametarskoj formi*, gde je funkcija korisnosti definitino određena, na primer kao stepena funkcija korisnosti $u(x) = x^{\rho}$, ili kao eksponencijalna funkcija korisnosti, $u(x) = 1 - e^{-\theta x}, \theta > 0$ (up. Wakker, 2010, za diskusiju eksponencijalne korisnosti). Specifikujući precizno matematičku funkciju u modelu neke teorije, ta teorija dobija dopunske osobine koje su takođe predmet potencijalne empirijske falsifikacije. Tako svako preciziranje modela dodatno komplikuje (ionako ne jednostavnu) teorijsku analizu. Obe funkcije iz našeg primera, i stepena, i eksponencijalna, zadovoljavaju

osobine funkcije korisnosti koje slede iz vNM aksiomatike. Kada shvatimo da zapravo velika klasa matematičkih funkcija zadovoljava te osobine, razumemo i kolika kompleksnost analize je neophodna za selekciju modela teorije odlučivanja; podsećamo da govorimo o bazičnom kognitivnom procesu odlučivanja u uslovima rizika koji toliko prožima sve ljudske kognitivne aktivnosti da je objašnjenje nečega složenijeg od ovog procesa praktično nezamislivo ukoliko prethodno ne možemo da objasnimo njega. Dakle, posledica po konstrukciju kognitivne teorije je sledeća: sa progresivnim usložnjavanjem struktura ponašanja koje posmatramo, pod uslovom da možemo da ih posmatramo u eksperimentalnim uslovima, broj potencijalnih modela koje testiramo veoma brzo raste, što netrivijalno usložnjava falsifikaciju teorija odn. problem selekcije modela. Dodajmo tome neophodnost da se svi modeli analiziraju uzimajući u obzir neke razumne prepostavke o greškama merenja, što je npr. odlika najnovijih eksperimentalnih testova u oblasti odlučivanja, i počećemo da stičemo pravu sliku o složenosti problema selekcije formalnih modela u kognitivnoj psihologiji. Dalje, ovaj zaključak vodi ka tome da će više različitih, međusobno ekvivalentnih kognitivnih mehanizama, ili skupova funkcija, moći da proizvede ponašanje strukture ekvivalentne onoj koju propisuje neka složena aksiomatika. Ovakav rezultat ne mora da bude nužan, ali mi verujemo da je praktično neizbežan sa porastom kompleksnosti ponašanja koje se proučava. Štaviše, ovakva situacija u potpunosti odgovara trenutnoj slici rezultata eksperimentalnih testova više različitih modela odlučivanja u uslovima rizika i neizvesnosti, slici koju ćemo detaljno diskutovati u III delu ove rasprave. Strategija kojom bi se ovaj problem izbegao postuliranjem izuzetno preciznih aksiomatskih sistema, sistema koje bi zadovoljavale samo usko ograničene klase modela, ili u idealnom slučaju, jedan model, nije prihvatljiva zbog cene koja bi morala da se plati kompleksnošću hipoteza, odn. kompleksnošću strukture samog aksiomatskog sistema sa takvim osobinama; na stranu pitanje da li bi ljudsko ponašanje, u suštini stohastičko, ikada moglo da proizvede strukture podataka koje bi zadovoljavale tako strogo ograničene aksiomatske sisteme.

Nažalost, problemi povezani sa kompleksnošću struktura naučnih podataka i teorijskih koncepata ne prestaju ovde. Suštinska posledica po analizu racionalnosti saznanja u okviru KKP koja sledi iz ove diskusije je sledeća. Zahvaljujući prethodno diskutovanim ograničenjima, očigledno je da lako možemo da se nađemo u situaciji u kojoj (a) struktura osmotrenog ponašanja odgovara nekim racionalnim kriterijumima, odn. zadovoljava neku intuitivno prihvatljivu aksiomatiku pogodnu

za oblast koju proučavamo, dok je (b) ponašanje koje posmatramo proizvod neopservabilnih kognitivnih procesa, ili kompleksnih interakcija procesa, koji, uzeti u celini, ne odgovaraju nužno kriterijumima racionalnosti. Ovo je posledica toga što je, u prostoru hipotetskih, teorijskih konstrukata, moguće kreirati mehanizme i skupove funkcija koji kao *podskup svog autputa sadrže i ponašanje koje je očekivano u nekom aksiomatskom okviru*, iako sami ti mehanizmi i funkcije nisu ustanovljeni kao posledice inferencija iz aksiomatskog sistema koji testiramo. Ova činjenica nam govori da se neprestano nalazimo u nezavidnoj situaciji u kojoj nikada ne znamo da li test određenog aksiomatskog ovkira može da se sprovede tako da nam *bar garantuje da znamo zašto određeni aksiomatski okvir pada, ako pada, i zašto opstaje kroz empirijske testove, ako opstaje.* Nažalost, praksa savremene KKP u empirijskim istraživanjima ne obraća mnogo pažnje na ovakva ograničenja. Skoro je paradoksalno da istorijski prevaziđen bihejvioristički pokret u ovom smislu metodološke strogosti može da predstavlja samo uzor savremenoj KKP. *Vice versa*: u praksi naučne analize možemo da se nađemo u situaciji u kojoj ponašanje koje posmatramo ne zadovoljava normativne kriterijume racionalnosti, dok je ono proizvod latentno racionalnih kognitivnih procesa koji, zahvaljujući kompleksnim interakcijama sa drugim procesima, ili grešci merenja, ili nivou šuma koji je inherentan u radu sistema složenosti ljudskog kognitivnog sistema, ne dozvoljavaju opservaciju tog ponašanja kao normativnog. Raspravljajući dublje problem odlučivanja u III i V delu ove rasprave videćemo da najnoviji radovi u ovoj oblasti prepoznaju upravo ovakvu mogućnost.

Problemi diskutovani u ovoj sekciji mogu da se posmatraju kao instance *generalizacije Bernštajnovog problema*: sistem sa prevelikim brojem stepeni slobode uvek omogućava višestruke matematičke deskripcije, tako čineći problem selekcije modela nerešivim. Aksiomatizovanje naučnih teorija u kognitivnoj psihologiji onda možemo da shvatimo kao strategiju *redukcije broja stepeni slobode*, ali smo već pokazali da takva strategija dovodi do novih problema: neaksiomatizovani deo procesa i funkcija može da doprinese strukturi opservabilnih rezultata na način koji zamagljuje analizu pod odabranim aksiomatskim pretpostavkama. Zahvaljujući svemu navedenom, granice analize racionalnosti saznanja prirodno se svode u diskusije lokalnih problema, pod metodološkim imperativom izbora onih metoda posmatranja i deskripcije kojima bi takvi lokalni problemi - poput „jednostavnog“ odlučivanja u uslovima rizika - bar u nekoj meri mogli eksperimentalno da se izoluju, a teorije koje testiramo ostale onog nivoa složenosti koji omogućuje bar

nekakvu selekciju njihovih modela. Svi koji smatraju da eventualnim uspehom takve eksperimentalne izolacije problem selekcije odgovarajućeg modela postaje rešiv bar u domenu lokalizovanih ponašanja i podskupova kognitivnih funkcija i procesa, pokazaćemo u VI delu naše rasprave, greše.

## 6.2 Falsifikabilnost kompjutacionističkog programa

Naš prethodni kritički pregled kompjutacionističke paradigme omogućava nam da sada postavimo interesantno pitanje statusa naučnog programa KKP kao *empirijske hipoteze*. Diskusija racionalnosti saznanja u ovoj tezi odnosi se na pitanje naučnog statusa ovog pojma u okvirima nauke, dakle, samo u okvirima onih naučnih programa koji predstavljaju empirijske, falsifikabilne hipoteze. Vodeće pristalice CTM u klasičnoj simbolicističkoj paradigmi ne propuštaju da naglase da je CTM empirijska hipoteza; često se u diskursu filozofije uma ograđuju tokom rasprave, tvrdeći da CTM ne iznosi (isključivo) nužne, apriorne istine (up. Fodor, 2008, za odnos prema pitanju da li je CTM „prava filozofija“). Za nas je odnos CTM prema drugim argumentima u filozofiji uma manje značajan: ukoliko CTM jeste domen teorijske psihologije, za nas je bitno da ustanovimo da li je ona u celini podložna empirijskom testiranju i tako falsifikabilna. CTM predstavlja prvu ideju u istoriji nauke koja je pružila makar perspektivu za naturalizaciju svih kognitivnih procesa; opet, čak i najveći zagovornici ove teorije sumnjaju da je ona u sadašnjoj formulaciji u stanju da odgovori na takav zahtev u potpunosti (up. Fodor, 2000, za detaljnu analizu ograničenja CTM). Ipak, ove kritike se odnose na kognitivne funkcije koje je teško (i možda nemoguće) shvatiti kao kompjutacione u smislu u kom danas razumemo koncept izračunavanja. Nas interesuje da li tvrdnje CTM koje se odnose na one kognitivne funkcije koje bar u principu podležu kompjutacionoj deskripciji predstavljaju empirijske hipoteze u tom smislu da je njihova falsifikacija moguća.

Raspravljaćemo ovo pitanje na primeru veoma složenog problema u oblasti psihološke semantike. U najnovijoj kompletnoj ekspoziciji CTM i LOTH, Fodor ovom problemu posvećuje posebnu pažnju, pokušavajući da pokaže kako upravo zahvaljujući suštinskim osobinama CTM on može biti rešen (Fodor, 2008). Radi se o klasičnom problemu za raselijanske koncepcije značenja (Speaks, 2011), koje obuhvataju Fodorovu LOTH hipotezu: problemu *Fregeovih slučajeva* (Zalta, 2012, Speaks, 2011). Gotlib Frege će ostati čuven u istoriji semantike po uvođenju distinkcije između *smisla* (u nemačkom originalu „*Sinn*“, engl. "*sense*") i *reference* (u nemačkom originalu „*Bedeutung*“, engl. *reference*) u raspravu o značenju

nekog termina (Frege, 1892/1960). Razlika između termina VENERA i ZVEZDA VEČERNJAČA je u tome što oba imaju istu referencu, odn. realni, postojeći objekat: planetu *Veneru*, ali različit smisao. Ovo se ogleda u tome što je iskaz o identitu 'VENERA = VENERA' potpuno neinformativan, za razliku od iskaza 'VENERA = ZVEZDA VEČERNJAČA', koji kao da uvećava naše znanje. Ukoliko drugi iskaz nije neinformativan, a intuicija nam jasno govori da on to nije, mora da postoji neka relacija, neki atribut F tako da (*i*) F($b$) važi, (*ii*) F($a$) ne važi, i ujedno (*iii*) $a = b$. Frege je ovaj problem rešio upravo uvođenjem distinkcije između smisla i reference, tvrdeći da $a$ i $b$ imaju istu referencu, ali različit smisao. Međutim, ovo rešenje nije moguće primeniti u kompjutacionoj paradigmi: kompjutaciona paradigma trpi samo egzaktna određenja entiteta kojima operiše, tako da je *referencijalni sadržaj* jedina vrsta sadržaja koju simboli mogu da imaju, pa je samim time referencijalni sadržaj i jedini mogući *mentalni sadržaj* u kompjutacionizmu. U teorijskom razvoju CTM izuzetno se obraća pažnja na ovaj „detalj" u teoriji značenja: CTM i Fodorova LOTH predstavljaju osnovu za naturalističku teoriju uma samo ukoliko je sav mentalni sadržaj moguće shvatiti kao referencijalan. Ovo očigledno zahteva redukciju onoga što je Frege nazvao smislom na koncept reference, i Fodor sprovodi upravo takvu redukciju. Naša tvrdnja, pošto iznesemo Fodorovo rešenje, biće da upravo to rešenje predstavlja deo CTM i LOTH koji nije moguće empirijski falsifikovati; ako smo u pravu, posledica je da ovaj značajan deo savremenog kompjutacionističkog programa nije empirijska hipoteza.

Pokušaj redukcije smisla na referencu sproveden je u potpunosti u najnovijoj formulaciji LOTH, u Fodorovoj knjizi „*LOT 2: The Language of Thought Revisited*" iz 2008: sve što sledi odnosi se na Fodorovo rešenje izneto u 3. poglavlju navedenog dela: „*LOT Meets Frege's Problem (Among Others)*". U našem prikazu simbolicističke paradigme KKP već smo konstatovali da ona zahteva da svi kompleksni koncepti budu funkcije značenja jednostavnih koncepata koji su njihovi konstituenti. Fodor prvo otklanja problem smisla za kompleksne koncepte. Diskutujući moguće razlike u značenjima termina ZVEZDA VEČERNJAČA (engl. MORNING STAR) i ZVEZDA VEČERNJAČA (engl. EVENING STAR), koja imaju istu referencu, on razvija sledeći argument: pošto su značenja ova dva koncepta pod simbolicističkom KKP proizvod operacija kombinatorne sintakse jezika uma nad značenjima njihovih konstituenata, sasvim je moguće da neko ima koncept ZVEZDE VEČERNJAČE bez toga da ima koncept ZORE, bez kog ne može da ima koncept ZVEZDE ZORNJAČE, i obrnuto za koncept VEČERI. Dakle, zahvaljujući

tome što je značenje kompleksnih koncepata u potpunosti svodivo na značenje njihovih konstituenata, moguće je pod CTM i LOTH da neko F važi za neko $a$, ali ne važi za neko $b$, iako a i b imaju istu referencu. Drugim rečima, u jeziku uma sa odgovarajućom kombinatornom sintaksom moguće je razviti dve različite interne deskripcije sa istim referencijalnim sadržajem, takve da kad nad njima operišu procesi izračunavanja koji su, podsetimo se, *osetljivi na sintaksu, a ne na sadržaj reprezentacija nad kojima operišu*, ne moraju sve inferencije koje su moguće nad jednom i nad drugom deskripcijom da budu identične. Fodoru preostaje da reši problem Fregeovih slučajeva za jednostavne koncepte. Jednostavi su oni koncepti koji nemaju nikakvu konstituentnu strukturu: dok pojam Šareni vrabac ima konstituetnu strukturu, tj. proizvod je operacije kombinatorne sintakse uma nad drugim jednostavnim ili kompleksnim konceptima Šareno i Vrabac, koncepti ličnih imena poput Aleksandar ili Justinijan *prima facie* nemaju nikakvu konstituentnu strukturu koja bi bila proizvod operacije kombinatorne sintakse u jeziku uma. Fodor rešava problem Fregeovih slučajeva donekle neuobičajenim potezom koji u suštini „svodi“ jednostavne koncepte na kompleksne koncepte. Još jednom, suštinska osobina CTM je da kompjutacioni kognitivni procesi operišu nad kognitivnim reprezentacijama, vršeći izračunavanja po pravilima određenog formalnog sistema, ali tako da su - kao i u svim formalnim sistemima - oni osetljivi samo na sintaksičke osobine reprezentacija, ne i na njihove semantičke osobine. Kognitivni sistem u svom funkcionisanju „zna“ koji kognitivni proces može da se primeni na neku reprezentaciju tako što samo određeni procesi mogu da prepoznaju (da *parsiraju*, rečnikom kompjutacione lingvistike) sintaksičku strukturu te reprezentacije; to ne mogu svi procesi. Ukoliko Fregeovi slučajevi pokazuju da je moguća koegzistencija verovanja $\mathrm{F}(a)$ i $\neg\mathrm{F}(b)$ za $a$ i $b$ koji imaju isti *referencijalni sadržaj*, onda je dovoljno omogućiti da se reprezentacije $a$ i $b$ *sintaksički razlikuju*. Sintaksičko rešenje je u CTM potpuno jednostavno: npr, neka kognitivni sistem u jeziku uma obeleži jednu reprezentaciju neke osobe terminom „Osoba$_1$“, a neku drugu terminom „Osoba$_2$“, gde su termini jezika uma „Osoba$_1$“ i „Osoba$_2$“ dve sintaksički različite reprezentacije sa istim referencijalnim sadržajem. Onda iz same postavke CTM, koja nije problematična, sledi da mogu da postoje kompjutacioni procesi koji su osetljivi na razliku. Problem je rešen ako skup procesa nad sintaksičkom strukturom „Osoba$_1$“ vodi u skup inferencija (rezultata kognitivne obrade informacija) koji se ne poklapaju u potpunosti sa skupom inferencija u koji vode procesi nad reprezentacijom „Osoba$_2$“. Kao što smo rekli, Fodorovo

rešenje „svodi jednostavno na kompleksno“: jednostavni koncepti mogu da budu *koreferentni* (da imaju istu referencu) a da za njih važe različiti predikati ukoliko se pretpostavi da su i oni u jeziku uma reprezentovani kao kompleksni koncepti, makar tek toliko da različiti kompjutacioni procesi prepoznaju sintaksu njihovih reprezentacija kao različitu i tako izvode različite inferencije iz jednog i iz drugog internog opisa. Primetimo da Fodorovo rešenje krši princip kognitivne ekonomije koji se često smatra karakteristikom racionalnosti kognitivnih funkcija (Rescher, 1989): njegovo rešenje Fregeovih slučaja umnožava broj internih reprezentacija koje kognitivni sistem koristi da bi reprezentovao samo jedan koncept.

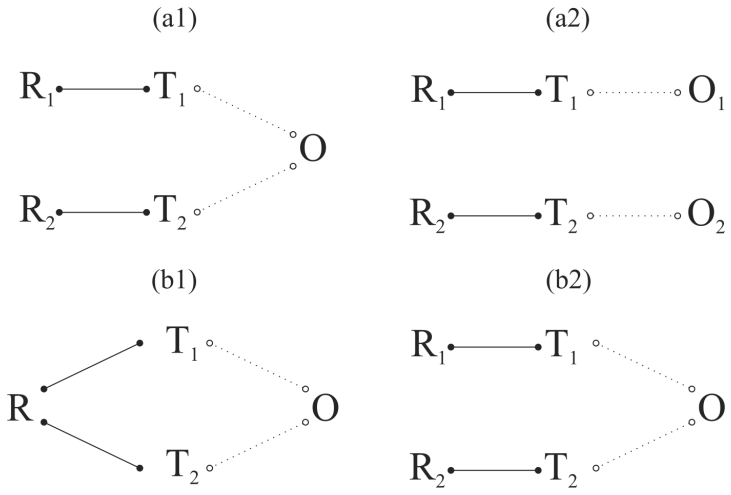
Filozofiji ostavljamo diskusiju da li je na ovaj način moguće rešiti sve Fregeove slučajeve i tako potpuno redukovati smisao na referencu, a mi postavljamo pitanje: da li je Fodorova koncepcija značenja u najnovijoj formulaciji LOTH empirijska hipoteza (kao što on tvrdi da jeste, up. Fodor, 2008)? Problem je veoma značajan u kontekstu psihološkog objašnjenja: Fregeovi slučajevi se javljaju u situacijama kada osoba poseduje (najmanje) dve reprezentacije nekog objekta, ali *nije svesna* njihove koreferencijalnosti; tada je moguće da ponašanje te osobe interpretiramo kao da ono nije u saglasnosti sa njenim verovanjima, željama i ciljevima, što predstavlja direktno kršenje osnovnog postulata intencionalne psihologije, iako ta osoba može da veruje da su njeni činovi u saglasnosti sa njenim verovanjima, željama i ciljevima (Schneider, 2005). Dakle, da bismo diskutovali Fodorovu teoriju značenja, prvo moramo da se uverimo da se kognitivni sistem nalazi pred problemom Fregeovih slučajeva, a to je moguće samo ako osoba čije ponašanje diskutujemo nije svesna koreferencijalnosti nekih reprezentacija koje ima. Slika 11. ilustruje moguće odnose opservabilnih i neopservabilnih elemenata testa korefencijalnosti koji će nam pomoći da pokažemo da Fodorova koncepcija značenja nije empirijska hipoteza. Vratimo se primerima iz Fodorove knjige: Fodor diskutuje Fregeove slučajeve koji bi mogli da se jave u nedoumicama o ličnostima, služeći se primerom istorijske ličnosti, Ignaci Jana Paderevskog (1860 - 1941), poljskog pijaniste, kompozitora, diplomate, političara i premijera. Da li je Paderevski pijanista ista osoba kao i Padarevski političar? Dilema je sasvim moguća. Pretpostavimo da smo, negde početkom XX veka, odlučili da prisustvujemo klavirskom koncertu koji izvodi Paderevski. Neko može da nas zapita da li bismo se kladili da će Padarevskom tokom pauze između dva stava pasti na pamet da održi politički govor: koliko novca bismo uložili na takvu opkladu? Ako verujemo da su termini „Padarevski pijanista“ i „Padarevski političar“ koreferencijalni, mogli bismo da ponudimo neku sumu. Ako verujemo da nisu, ne

bismo ponudili ništa (već bismo bili začuđeni ponudom opklade). Naša metodologija za ustanovljavanje koreferencijalnosti je potpuno remzijanska: koreferencijalnost dva termina, ili nečija sumnja u to da bi oni mogli biti koreferencijalni, može da se izmeri relativno u odnosu na cenu informacije koja nedostaje za odluku: nula, ukoliko smo sigurni da termini nisu koreferencijalni i da postoje dva Paderevska, pijanista i političar, i neko $x$, srazmerno tome koliko verujemo da termini jesu koreferencijalni, te da PADAREVSKI PIJANISTA jeste PADAREVSKI POLITIČAR.

Recimo da sumnjamo da su termini koreferencijalni samo ukoliko je naša ponuda za opkladu neki iznos veći od nule; naravno, postoje i drugi razlozi zbog kojih Paderevski može da odluči da ne održi politički govor usred koncerta čak i ako PADEREVSKI PIJANISTA = PADEREVSKI POLITIČAR. Shema (a1) na Slici 11. ilustruje ovu situaciju: postoje dve reprezentacije u jeziku uma, PADEREVSKI PIJANISTA i PADEREVSKI POLITIČAR, svaka od njih vezana za odgovarajući termin u prirodnom jeziku, međutim, termini su koreferencijalni i reference vezane za oba termina i obe reprezentacije konvergiraju ka jednom realnom objektu. Ukoliko je naša ponuda za opkladu nula, smatramo da termini nisu koreferentni: ostajemo sa dve reprezentacije, dva termina, i dve relacije denotacije ka dva denotata (panel (a2) na Slici 11). Kako se situacija menja ukoliko smo saznali da dva termina, „Paderevski pijanista“ i „Paderevski političar“ jesu koreferentna? Postoje dva moguća rešenja: shema (b1) ilustruje rešenje u kome kognitivni sistem gradi jednu generičku reprezentaciju koju vezuje za dva koreferentna termina, dok shema (b2) ilustruje rešenje u kome kognitivni sistem i dalje koristi dve reprezentacije, vezane za dva koreferentna termina, sa jedinstvenim denotatom - i zato predstavlja isto reprezentaciono stanje kao shema (a1).

Prema rešenju koje nudi Fodor, i koje podrazumeva uvođenje sintaksički kompleksnih reprezentacija za jednostavne koncepte kao što su lična imena, nema nikakve *opažljive*, *bihejvioralne razlike* u funkcionisanju sistema posle ustanovljavanja koreferencijalnosti dva termina između situacija prikazanih shemama (b1) i (b2) na Slici 11. Ako se posle koncerta zateknemo sa gospodinom Paderevskim, neko koga bismo zamolili da mu dospe još čaja bi svejedno mogao da tu akciju reprezentuje preko reprezentacije PADEREVSKI PIJANISTA kao i preko reprezentacije PADEREVSKI POLITIČAR. Trebalo bi da je potpuno jasno da nam *bihejvioralni test u Fregeovim slučajevima omogućava da odredimo da li neka osoba veruje da su dva termina koreferencijalna, ali ne i to da li se interne reprezentacije vezane za koreferencijalne termine koriste selektivno ili ne.* Ako je Fodorovo

rešenje za Fregeove slučaje to da se za dva koncepta uvode dve sintaksički složene reprezentacije, prevodeći tako kontekste Paderevskog *kao pijaniste* i Paderevskog *kao političara* iz dva smisla u dva referencijalno ista ali sintaksički različita mentalna sadržaja, ne postoji bihejvioralni test kojim možemo da utvrdimo da li je to rešenje primenjeno ili nije. Fodorova referencijalna teorija značenja je empirijska hipoteza samo u smislu reči „da bi moglo biti tako“: ona nije empirijska hipoteza u strogom naučnom smislu reči u kome su empirijske hipoteze podložne falsifikaciji odgovarajućim eksperimentom.




Slika 11. *Opservabilni i neopservabilni entiteti u bihejvioralnom testu koreferencijalnosti.* Shema (a1): test ustanovljava koreferencijalnost dva termina (dve reprezentacije); shema (a2): test ne ustanovljava koreferencijalnost; shema (b1): situacija u kojoj posle saznanja o koreferencijalnosti opstaje samo jedna reprezentacija, shema (b2): situacija u kojoj posle saznanja o koreferencijalnosti opstaju dve reprezentacije. Pune linije označavaju relacije između reprezentacija (R, u jeziku uma) i jezičkih termina (T, u prirodnom jeziku), isprekidane linije označavaju relacije denotacija (između termina T i denotata O). Objašnjenje testa koreferencijalnosti je u dato u tekstu.

Može da se postavi pitanje da li je Fodorova teorija značenja u najnovijoj verziji LOTH iz 2008. u principu falsifikabilna, i to je mesto gde ćemo verovatno doći do razlike u značenju koncepta „empirijska hipoteza“ koje imaju na umu filozofi, s jedne, i empirijski naučnici, s druge strane. Test Fodorove teorije podrazumevao bi da smo u (*i*) u nekoj kontrolisanoj situaciji osigurali da neka osoba koristi isti termin u dva smisla, (*ii*) za oba od ta dva smisla nekom eksperimentalnom tehnikom izolovali jedinstvene interne reprezentacije koje učestvuju u njihovoj upotrebi, (*iii*) dokazali da su te reprezentacije koreferencijalne, i (*iv*) upoznali sintaksu jezika uma do te mere da možemo da kažemo da je jedna od te dve reprezentacije samo sintaksička varijacija druge. Drugim rečima, ne preostaje nam ništa drugo do da

pokušamo *direktnu opservaciju* rada internog reprezentacionog sistema da bismo testirali ovu teoriju. Sa stanovišta filozofije uma, teorija značenja u LOTH je možda „u principu“ falsifikabilna hipoteza; sa stanovišta naučne psihologije, da je direktni test upravo opisan kroz (*i*) do (*iv*) uopšte moguć, ne bismo uopšte ni imali probleme koji motivišu ovu raspravu! Koje su posledice po empirijski status KKP i analizu racionalnosti u njegovom okviru? S jedne strane, mi moramo da prihvatimo da ograničenja metoda kojima raspolažemo u empirijskom proučavanju uma omogućavaju ispitivanje samo hipoteza određenog nivoa složenosti. S druge strane, jedini predlog za kompletnu naturalizaciju kognitivnih fenomena koji je na stolu sadrži tvrdnje koje su verovatno i „u principu“ van domašaja eksperimentalne metodologije kojom raspolažemo. Nažalost, deo tog predloga za koji smo pokazali da nije podložan eksperimentalnoj falsifikaciji nalazi se u raspravi oko jednog od najvažnijih problema ljudske i veštačke inteligencije u kognitivnim naukama: kako prirodni procesi mogu da označavaju iste objekte dok njima upravljaju zakoni neke (naturalizovane) semantike koji omogućavaju različite interpretacije tih istih objekata?

Opšti zaključak iz dve prethodne diskusije racionalnosti saznanja u okviru KKP je da je pitanje racionalnosti saznanja neraskidivo povezano sa prirodom metodologije koja se koristi da bi se na to pitanje odgovorilo. Ovaj zaključak predstavlja imperativ svih diskusija koje slede: sve one će biti vođene u odnosu na to kako metodološka ograničenja (a) određuju našu predstavu o kognitivnom sistemu uopšte, i (b) kako ona određuju naše odgovore na pitanja o tome da li su, i u kojoj meri, kognitivne funkcije i procesi racionalni. Tradicija kognitivne psihologije, započeta u „kognitivnoj revoluciji“ kasnih 50-ih i 60-ih godina XX veka, sa izuzetkom tek nekoliko oblasti poput odlučivanja i teorija sličnosti, grubo je zapostavila analizu odnosa teorijskih koncepata, s jedne, i operacija merenja i strukture eskperimentalnih procedura, s druge strane. Metodološki stav koji zauzimamo svakako ne predstavlja nekakav prolog povratku u bihejviorizam, ali predstavlja pokušaj obnavljanje prakse koja za bihejvioriste jeste bila svakodnevnica: prakse da se o odnosu metodologije i merenja prema teorijskim konceptima *misli pre* nego što se teorijski koncepti konstituišu. U nekim oblastima, poput teorije odlučivanja, ovakav stav nikada nije napustio prvi plan. Kao što se nadamo da će naše diskusije u narednim poglavljima pokazati, savremena KKP se, zahvaljujući svim ovde diskutovanim problemma, suočava pre sa *problemom redukcije broja modela i teorija* nego sa problemom generisanja novih modela i teorija.

## 6.3 Metateorijski okvir za analizu racionalnosti saznanja

Već u prvom poglavlju ove teze uveli smo razliku između normativnih i deskriptivnih teorija. Pristalice normativnih teorija po pravilu vide empirijska odstupanja od racionalnih rešenja kao (a) posledice primene specifičnih metodoloških procedura, ili kao (b) posledice pogrešno postavljenog problema, npr. kada eksperimentalna situacija kojom se dobijaju odstupanja od racionalnosti ne predstavlja uzorak realne strukture sredine u kojoj se proučavani problem rešava, ili kao (c) posledice grešaka merenja, šuma inherentnog radu kogntivnog sistema. Pristalice deskriptivnih teorija, s druge strane, vide empirijska odstupanja od racionalnih, normativnih standarda kao suštinske pokazatelje da normativne teorije treba korigovati. Njihova strategija objašnjenja empirijskih odstupanja od racionalnosti se najčešće bazira na (a) ukazivanju na ograničenu procesnu moć kognitivnog sistema, koja kao posledicu ima odstupanje od racionalnosti, (b) ukazivanju na dejstvo određenih faktora koje normativne teorije nisu uključile u svoje analize, ili (c) na mogućnost da kognitivni sistem ne reprezentuje problem koji rešava na način na koji su to pretpostavile normativne teorije. U odnosu na shvatanja o racionalnosti saznanja, ove dve grupe teorija tvore osnovnu dimenziju suprostavljenih gledišta u debati o racionalnosti. Stanovič, u skladu sa prirodom dve pozicije u debati o racionalnosti, naziva zagovornike shvatanja o kognitivnom sistemu kao ograničeno racionalnom *melioristima*, a zagovornike shvatanja o tome da su odstupanja od racionalnosti posledica faktora koji nemaju veze sa inače racionalnim kompetencijama kognitivnog sistema *panglosijancima* (Stanovich, 1999, prema Stanovich & West, 2000). Duh debate o racionalnosti odlično ilustruju sledeće reči Danijela Kanemana, vodećeg melioriste u dihotomiji Stanoviča, kojima komentariše panglosijanske predloge Koena (Cohen, 1981/2008), objašnjavajući da oni uvek nude „*...zgodnu zbirku odbrana koje mogu da se iskoriste ako su ispitanici okrivljeni za greške: privremeno ludilo, teško detinjstvo, da su žrtve nameštaljke ili loše sudijske procene - jedno od ovih će sigurno proraditi, i obnoviti pretpostavku racionalnosti.*“ (citirano prema Kahneman, 1981, naš prevod).

Pre nego što se i sami uključimo u debatu o racionalnosti u III delu naše rasprave, analizirajući odnos normativnih i dekriptivnih shvatanja u nekoliko oblasti viših i simboličkih kognitivnih funkcija, preciznije ćemo odrediti značenje nekoliko *metateorijskih i metodoloških koncepata* razvijenih tokom ove debate. Ovi koncepti su od ključnog značaja za karakterisanje različitih teorija o kognitivnim funkcijama na mapi debate o racionalnosti saznanja.

*Sajmonov pojam zadovoljavanja* (engl. *satisficing*). U radu koji je praktično otvorio debatu o racionalnosti, ekonomista i nobelovac Herbert Sajmon, 1955. godine, piše:

*„Tradicionalna ekonomska teorija postulira „ekonomskog čoveka" koji je, u sklopu toga što je „ekonomski", ujedno i „racionalan". Za ovog čoveka se pretpostavlja da poseduje znanje o relevantnim aspektima svog okruženja koje je, ako ne apsolutno kompletno, makar impresivno čisto i izdašno. Za njega se pretpostavlja da ima dobro organizovan i stabilan sistem preferencija, kao i kompjutacione sposobnosti koje mu omogućavaju da izračuna koja će mu od alternativnih akcija koje su mu na raspolaganju dopustiti da dostigne najvišu moguću tačku na njegovoj skali preferencija.*" (citirano prema Simon, 1955a, naš prevod).

Sajmon nastavlja sa formalnim određenjem jedne po jedne sposobnosti „racionalnog" čoveka i za svaku od njih pokazuje da, pod pretpostavkama da govorimo o „psihološkom", dakle realnom, čoveku, malo koja od njih može da mu se pripiše bez oklevanja. Stabilnost dobro organizovanih preferencija, oličena u funkciji korisnosti neke osobe, može da bude idealizacija koja mahom ne odlikuje ljude. Nasuprot funkciji korisnosti koju je ekonomska misao razvijala od Bernulija do fon Nojmana, Morgenšterna i Sevidža, Sajmon postavlja diskretnu funkciju koja moguće ishode neke neizvesne situacije deli na one koji „zadovoljavaju" i one koji „ne zadovoljavaju" donosioca odluka. Tako je donošenje odluka za pojedinca olakšano, a pretpostavke o njegovim idealizovanim sposobnostima minimizovane. Posledica je, naravno, da pojedinac sa takvom „funkcijom korisnosti" više ne donosi odluke koje su racionalne u smislu maksimalne moguće (optimalne) saglasnosti sa normativnim kriterijumima, već odluke koje *zadovoljavaju* njegove aspiracije. Prema teoriji očekivane korisnosti, racionalnog čoveka mora da odlikuje sposobnost da sve korisnosti alternativa koje su pred njim ponderiše odgovarajućim verovatnoćama da bi doneo odluku. Međutim, nema prostora za takvu pretpostavku u slučaju „psihološkog", realnog čoveka: čak i ako on na raspologanju uvek ima sve verovatnoće koje su potrebne za ocenu neke riskantne situacije, ništa ne garantuje da je on u stanju da integriše sve te informacije sa svojim preferencijama na način na koji normativna teorija to od njega očekuje. Kompjutaciona ograničenja ljudskih kognitivnih sposobnosti postoje: ono o čemu matematički ekonomista ne mora da vodi računa dok razvija model tržišta baziran na idealizovanom akteru u ekonomskim interakcijama je upravo ono što je osnovni naučni problem psihologa koji proučava saznajne funkcije i inteligenciju. Jednom kada oceni koja od rizičnih opcija mu

donosi najveću očekivanu korisnost, racionalni čovek normativne teorije ne ulaže nikakav napor u selekciji mogućih akcija koje bi dovele do njenog ostvarenja; s druge strane, realni čovek ne mora uvek da ima na raspolaganju internu reprezentaciju svih alternativnih akcija, tako da određenu cenu u njegovom odlučivanju može da nosi i kognitivni rad vezan za pretragu skupa - ili čak konstrukciju skupa - mogućih akcija. Praktično nijedna intuicija koja učestvuje u konstrukciji onog racionalnog donosioca odluka kojeg su kao aktera teorije igara zamislili fon Nojman i Morgenštern ne odlikuje realnog, „psihološkog“ aktera ekološki validnih, realnih interakcija.

Pojam *zadovoljavanja*, shvaćen kao karakteristika kognitivnih funkcija i procesa, predstavlja osnovni pojam koji razlikuje odgovarajuće normativne i deskriptivne teorije, teorije racionalnosti i teorije ograničene racionalnosti. Herbert Sajmon nikada nije upotrebio reč iracionalnost da bi okarakterisao pokušaj da se u kognitivnim naukama i ekonomiji modelira ponašanje realnog čoveka. U njegovom razvoju shvatanja izbora koji zadovoljavaju donosioca odluka eksplicitno je prisutna ideja da strategije koje zadovoljavaju mogu da budu racionalne, odn. da je i njih moguće shvatiti kao *optimalna rešenja*, ali tek kada se u analizu uključe faktori kao što su cena kognitivnog izračunavanja odn. mentalnog napora koji osoba ulaže u rešavanje problema, mogućnost da sve alternativne akcije za rešavanje problema nisu uvek dostupne sistemu koji se analizira, ili mogućnost da kognitivni sistem jednostavno nema uvek i u svakoj situaciji na raspolaganju stabilan, uređen skup preferencija.

Zbog ovakve karakterizacije pojma zadovoljavanja, u debati o racionalnosti se ne trudimo da povučemo razliku između racionalnog i *iracionalnog* ponašanja ili racionalnih i *iracionalnih* kognitivnih fenomena, funkcija i procesa. Uzmimo za primer tipično odstupanje od normativnih kriterijuma: fenomen ponderisanja verovatnoća u već diskutovanoj teoriji izgleda Kanemana i Tverskog. Empirijski nalaz da ljudi opažaju niske verovatnoće kao više nego što jesu i obrnuto za visoke verovatnoće je jedan od najstabilnijih empirijskih nalaza o suđenju i donošenju odluka uopšte (Griffin & Brenner, 2004); takav nalaz direktno ukazuje na odstupanje od normativnih standarda prema kojima donosilac odluka idealno informisan. Pokazuje se, međutim, da je u teoriji odlučivanja takav nalaz moguće objasniti sa stanovišta koje je i blisko psihološkoj intuiciji i zasnovano na prethodno utemeljenim pojmovima optimizma i pesimizma (Diecidue & Wakker, 2001, Wakker, 2010). S druge strane, Viskuzi pokazuje da je taj nalaz lako protumačiti kao prirodnu posledicu u suštini racionalne bejzijanske inferencije donosioca odluka koji integriše

svoja prethodna verovanja o verovatnoćama sa verovatnoćama koje su mu prikazane u eksperimentalnoj situaciji (Viscusi, 1989). Fenomeni koji bi svedočili o nečemu što bismo kolokvijalno mogli da nazovemo iracionalnim, kao što bi to npr. bila ekstremna sklonost ka riziku, su retki kako u teoriji tako i u empirijskim nalazima. U debati o racionalnosti nikad ne nalazimo teorijske koncepte koji bi izmicali *ma kakvoj* logici i tako predstavljali čisto deskriptivne konstrukte sa jednom jedinom ulogom: da objasne empirijske devijacije od normativnih kriterijuma. Zadovoljavanje je zato koncept koji na odgovarajući način karakteriše rešenja u okviru teorija ograničene racionalnosti: ta rešenja su i adaptivna i optimalna *kada se problemska situacija reformuliše tako da više odgovara prirodnoj deskripciji i prirodnim ograničenjima sistema čije se ponašanje analizira*, a suštinski deo debate o racionalnosti zato i jeste metateorijsko pitanje o odgovarajućoj deskripciji problema i sistema koji ga rešava.

*Normativna adekvatnost.* Pitanje adekvatnosti predloženog normativnog okvira za neku kognitivnu teoriju je ozbiljno filozofsko i logičko pitanje. Na prvi pogled, deluje da je za svaku kognitivnu funkciju koju možemo da zamislimo normativni okvir *a priori* određen odgovarajućom oblašću logike ili matematike. Međutim, kako Koen pokazuje (Cohen, 1981), to ne mora da bude tačno. Koen se zalaže za (za nas prihvatljivu) tvrdnju da normativni okvir treba odabrati tako da odgovara najpribližnije našem empirijskim znanju o *zajedničkim ljudskim intuicijama* u domenu na koji se određena kognitivna funkcija odnosi. Argument koji vezuje normativni okvir za formalne nauke se onda ponovo javlja: zar nisu upravo teorije logike, verovatnoće i geometrije, one teorije koje počivaju na *očiglednim* aksiomima, aksiomima koji opisuju naše najdublje intuicije o stvarnosti? Ovo pitanje zahetva da mu posvetimo posebnu pažnju.

(I) Prva kritika ovakvog shvatanja jeste sledeća: *intuitivnost aksiomatskog okvira nije njegova nužna karakteristika.* Neki aksiomatski okviri, na primer aksiomi teorije očekivane korisnosti, jesu intuitivni. Kolmogorevljevi aksiomi teorije verovatnoće takođe počivaju na intuicijama koje bi morale biti bliske svakoj osobi. Međutim, u istorijskom razvoju aksiomatskog metoda, odnos prema ljudskoj intuiciji se menjao, dopuštajući i razvoj aksiomatskih okvira koji grubo odstupaju od naše intuicije. Primer jer aksiomatski okvir neeuklidskih geometrija, u kojima se aksiom paralelnosti (Euklidov V postulat) menja u forme koje nisu bliske ljudskoj intuiciji o geometrijskom prostoru (dok se u slučaju tzv. apsolutne geometrije čak ni ne koristi). Činjenica da ovakvi aksiomatski sistemi nalaze primenu u razvoju

teorija koje se poklapaju sa našim empirijskim uvidima o prirodi (kao npr. u Ajnštajnovoj teoriji opšte relativnosti) je jedna od najfascinantnijih osobina naučnog saznanja uopšte. Drugi primer bi mogla da bude Sevidžova aksiomatika subjektivne očekivane korisnosti. Struktura njenih aksioma je takva da bismo (verovatno), posle odgovarajuće diskusije koja motiviše način razmišljanja u toj teoriji, svačiju intuiciju uspeli da ubedimo u opravdanost njenih aksioma. Međutim, njena složenost izvesno isključuje to da bi osoba koja prethodno nije motivisana da o odlučivanju razmišlja na način koji predlaže Sevidž tvrdila da prepoznaje baš njegov skup principa kao svoje intuicije o odlučivanju u uslovima neizvesnosti i rizika.

(II) Druga kritika ovog argumenta je Koenova i glasi ovako: *svakodnevne ljudske intuicije u određenim domenima stvarnosti se ne poklapaju nužno sa aksiomatskim okvirima tih domena, čak i kada aksiomi neke teorije mogu da budu opravdani ukazivanjem na njihovo intuitivno značenje.* Smatramo da su dve interpretacije Koenovog stava moguće.

(IIa) Prema prvoj, koje se drži sam Koen, suštinski je razumeti da između (*i*) intuicija o validnosti nekog aksiomatskog sistema *sa stanovišta inferencija koje on obezbeđuje*, i (*ii*) intuicija o validinosti nekog aksiomatskog sistema *u odnosu na njegovu empirijsku primenu* postoji bitna razlika: ove dve grupe intuicija ne moraju nužno da se poklapaju. Primeri koje navodi Koen odnose se na logičku dedukciju, za koju je dobro poznato da predstavlja intuitivno razumljivu teoriju samo ako razlikujemo *formalno validne inferencije* od *semantičke istinitih argumenta.* Iskoristićemo primer čuvenog logičara Beta (Beth, 1955/1987). Da li iz premisa (*i*) „*Neki panteri nisu sisari*“, i (*ii*) „*Neki sisari nisu labudovi*“, logički sledi konkluzija (*iii*) „*Neki panteri nisu labudovi*“? Premisa argumenta nije tačna, međutim, upotreba logičke implikacije nam omogućava da iz netačne premise izvedemo tačan zaključak (i onemogućava obrnuto, da iz tačne premise izvedemo netačnu konkluziju). Bet nas savetuje da logičku vezu u ovom argumentu proverimo tako što ćemo zameniti termine „*panter*“, „*labud*“ i „*sisar*“ redom terminima „*svinja*“, „*prasac*“ i „*mamut*“. Posle zamene, argument uzima sledeći oblik: iz (*i*) „*Neke svinje nisu mamuti*“ i (*ii*) „*Neki mamuti nisu prasci*“, sledi konkluzija (*iii*) „*Neke svinje nisu prasci*“. Nova konkluzija je *lažna*, dok je nova premisa *tačna*: pošto znamo da ovakva inferencija logički nije validna, jasno je da forma argumenta ni sa prvobitno datim terminima nije mogla da obezbedi da konkluzija logički sledi iz premisa. Validnost inferencija je predmet logike; istinitost, predmet semantike (za elaboraciju ove razlike v. Beth, 1955/1987). Ipak, ostaje psihološki utisak da bi inferencija sa

prvobitno datim terminima „*panter*“, „*labud*“ i „*sisar*“ nekako mogla da bude validna. Ljudski um nepripremljen na bavljenje logikom kao normativnom disciplinom nema obavezu da poznaje distinkciju između validnosti u logici i istinitosti u semantici. Koen zato tvrdi da normativni okviri kognitivnih teorija moraju da budu odabrani na osnovu njihovog poklapanja sa ljudskom intuicijom *u odnosu na empirijsku primenu određenog formalnog sistema*, a ne na osnovu njihovog poklapanja sa intuicijom o validnosti inferencija koje njihovi aksiomi omogućavaju.

(IIb) Mi predlažemo još jednu interpretaciju Koenovog gledišta da svakodnevne intuicije ne moraju da se poklapaju sa aksiomatskim okvirma čak i kad oni počivaju na intuitivno jasnim argumentima: *ljudska intuicija može i da bude zadovoljena aksiomatskim postulatima neke teorije, pa čak i da se poklapa sa njima u empirijskoj primeni, ali to ne znači da je zato ona automatski zadovoljena i u slučaju svih inferencija koje ti aksiomatski postulati omogućavaju.* Jedino pravilo inferencije logičkog računa prvog reda je *Modus Ponens*: $[(Q \Rightarrow P), Q] \Rightarrow P$, i ne čini se da nešto nije u redu sa jednostavnom intuicijom koje ono simbolički kodira. Pogledajmo sada formulu računa prvog reda koju koristi Koen u jednom primeru u istom radu (1981) koji diskutujemo: $[(A \Rightarrow B) \wedge (C \Rightarrow D)] \Rightarrow [(A \Rightarrow D) \vee (C \Rightarrow B)]$. Ova inferencija jeste zasnovana na aksiomima o upotrebi logičkih veznika i pravilu *Modus Ponens*, ali nema razloga se pretpostavi da ova inferencija ima odgovarajuću sliku u ljudskoj intuiciji. Ako se pozovemo ovde na diskusiju prethodno uvedenog koncepta *zadovoljavanja*, jasno je da ljudskom kognitivnom sistemu ne moraju uvek da budu na raspolaganju *sve inferencije* koje omogućava neki intuitivan aksiomatski sistem, čak i ako psihološka intuicija korespondira logičkoj i matematičkoj intuiciji sadržanoj u tom aksiomatskom sistemu, pa čak i ako obe odgovaraju Koenovom kriterijumu poklapanja sa ljudskom intuicijom u odnosu na empirijsku primenu.

Već smo uveli razliku između personalnog i subpersonalnog nivoa psihološkog objašnjenja u raspravi o folk-psihologiji. Smatramo da prethodna diskusija normativne adekvatnosti motiviše uvođenje još jednog nivoa psihološkog objašnjenja koji može biti od velike koristi u debati o racionalnosti. Naša intrepretacija (IIb) Koenovog argumenta nas vodi u sledeće razmišljanje. Pretpostavimo da postoji poklapanje između ljudske intuicije i formalne aksiomatizacije određenog domena stvarnosti. Na kom nivou psihološkog objašnjenja ustanovljavamo to poklapanje? Ukoliko govorimo o objašnjenju na personalnom nivou, na kome je validan opis u predikatima folk-psihologije, onda mi očekujemo da ljudi budu svesni da se njihove intuicije poklapaju sa formalnim aksiomima *kada im značenje tih aksioma ponudimo*

*u jeziku koji je za njih prihvatljiv* (jezik formalne logike neće biti prihvatljiv za veliki broj ljudi). Da li je moguće govoriti o korespondenciji intuicije sa formalnim aksiomima na subpersonalnom nivou opisa? Naravno da jeste. Pretpostavimo da uopšte ne postoji korespondencija između intuicije donosioca odluka i aksiomatike teorije očekivane korisnosti, i to u smislu da npr. taj donosilac odluka uopšte svesno ne smatra da je obavezan da poštuje aksiom nezavisnosti ove teorije. Zamislimo eksperiment u kome, umesto da probleme odlučivanja formulišemo verbalno, mi njih formulišemo kao zadatke u domenu senzomotorne koordinacije. Takvi eksperimenti su zaista izvedeni (Wu, Delgado & Maloney, 2009). Ukoliko odlučivanje ispitanika u ovakvom zadatku, gde su verbalni opisi stimulusa zamenjeni odgovarajućim senzomotornim zadacima, odlikuje maksimizacija očekivane korisnosti, mi možemo da tvrdimo da je teorija očekivane korisnosti tačna u domenu senzomotornog odlučivanja, i to psihološko objašnjenje se sada nalazi na subpersonalnom nivou. Međutim, postoji još jedna mogućnost. Pretpostavimo da smo ustanovili da se intuicije neke osobe poklapaju sa formalnim aksiomima teorije očekivane korisnosti na personalnom nivou: osoba je svesna postojanja aksioma i razume njihovo značenje dobro, te tvrdi da se oni poklapaju sa njenom intuicijom o tome kako treba donositi odluke u uslovima rizika. Međutim, u eksperimentu izbora između određenih rizičnih lozova ponašanje ove osobe ne odgovara ponašanju donosioca odluka koji bi izbor vršio maksimizujući očekivanu korisnost. Očigledno ograničenje eksperimentalnih procedura u ovakvim istraživanjima (sve te procedure su međusobno veoma slične, ako ne istovetne) je to što se od ispitanika traži da odluku ne donese uz pomoć papira i olovke, ili kalkulatora, ili bilo kako *eksplicitno primenjujući principe odgovarajuće teorije u određenom simboličkom kodu* (npr. simboličkom kodu elementarne aritmetike). Ispitujući normativnost ljudskih odluka na ovaj način, zaista je moguće da i eksperti u teoriji odlučivanja, ako raspolažu ograničenim simboličkim sredstvima, odstupaju od normativno racionalnih odluka. Predlažemo da se nivoi psihološkog objašnjenja prošire za još jedan nivo, koji ćemo nazvati *suprapersonalnim*, i koji odlikuje psihološko objašnjenje nekog kognitivne funkcije *u uslovima u kojima čovek može da se osloni na eksteriorizovane, socijalno standardizovane, simboličke sisteme i druga sredstva koja mu omogućavaju da svoje intuicije podrži tim sistemima i sredstvima kako bi proširio svoje kognitivne kapacitete*. Odlika ljudskog ekološkog okruženja jeste i to da je ono simboličko i tehnološko u upravo opisanom smislu. Ako je evolucija ljudske vrste čoveku omogućila da proširi svoje kognitivne kapacitete oslanjajući se na izum pisma,

računanja, abakusa, tekst procesora i radne stanice, mi smatramo da je neophodno za celokupnu analizu ljudske racionalnosti uzeti u obzir i mogućnost da čovek koristi kapacitete svog kognitivnog sistema unapred pripremljen za funkcionisanje *u takvoj simboličko-tehnološkoj sredini*[30]. Termini RACIONALNOST$_1$ i RACIONALNOST$_2$, kako smo ih koristili u I delu ove rasprave, lako se razvrstavaju po ovako organizovanim nivoima analize u psihologiji: RACIONALNOST$_1$ se očigledno nalazi samo na suprapersonalnom nivou analize, dok RACIONALNOST$_2$ pripada isključivo personalnom i subpersonalnom nivou.

Pitanje normativne adekvatnosti teorijskog okvira je očigledno izuzetno složeno. Mi smatramo da ga je moguće postaviti na sva tri nivoa psihološkog objašnjenja: subpersonalnom, personalnom i suprapersonalnom koji smo upravo definisali, i da je na svakom od tih nivoa potrebno ustanoviti koji, i zašto, teorijski okvir za određeno kognitivno funkcionisanje možemo da opravdamo kao normativan. Mesto za pitanje o poklapanju intuicije sa normativnim okvirom otvoreno je na svakom od tri nivoa objašnjenja. Na subpersonalnom nivou, nivou koji je najzastupljeniji u diskusijama i empirijskim istraživanjima KKP, te intuicije ne mogu da budu svesne. Onda se postavlja pitanje šta su one u stvari, a kao da se nameće odgovor da su one zapravo neke suštinske odrednice kognitivnog funkcionisanja čoveka, odn. da ih u nauci možemo shvatiti kao simbolički zapis matematičkih osnova ljudskih kognitivnih funkcija - bile one u skladu sa nekim normativnim okvirom ili ne. Na personalnom nivou, intuicije su te prirode da zahtevaju svest subjekta o njihovom poklapanju sa normativnim standardima, ako su one normativne prirode. Na suprapersonalnom nivou, intuicije su te prirode da omogućavaju inferencije o složenijim problemima zahvaljujući oslanjanju na simboličko-tehnološko okruženje čoveka, i na tom nivou tek smatramo da je moguće testirati i odnos intuicije prema validnosti u inferencijalnom smislu koju odbacuje Koen (Cohen, 1981/2008). U daljim diskusijama ćemo videti da postoje domeni ljudskog kognitivnog funkcionisanja za koje je otvoreno pitanje da li uopšte mogu da se okarakterišu ma kakvim normativnim okvirom, na ma kom nivou analize.

*Kompetencija i performansa.* Distinkcija između kompetencije i performanse u kognitivnoj psihologiji poreklom je iz lingvistike; prvi je ovu distinkciju uveo Čomski (Chomsky, 1965), omogućivši da se jasno govori o (a) internalizovanim, ne nužno svesnim pravilima upotrebe nekog jezika koja dele svi njegovi govornici i koja suštinski omogućavaju komunikaciju među njima, i (b) u govornoj i pisanoj praksi opservabilnim jezičkim formama koje ne moraju da budu savršeno formirane

u skladu sa internalizovanim pravilima. Distinkcija kompetencije i performanse je, u metodologiji lingvistike, poslužila da se idealizovane jezičke strukture odrede kao predmet proučavanja: sintaksa nekog jezika može da se proučava ne uzimajući u obzir greške u redosledu reči u rečenici, upotrebi padeške gramatike ili jednostavno greške u govoru koje *nisu sistematske prirode*. Generalizovana u kognitivnoj psihologiji, ova distinkcija igra istu ulogu u debati o racionalnosti. Dok odstupanja od normativnih standarda nisu sistematska, uvek je moguće govoriti o racionalnoj kompetenciji kognitivnog sistema u nekom domenu kognitivnih funkcija. Jedna od strategija odbrane normativne pozicije u debati o racionalnosti sastoji se upravo u pokušaju da se odstupanja od normativnih standarda posmatraju kao domen performanse, odn. da se greške u odstupanju od normativnosti posmatraju kao greške u izvođenju akcija, greške koje su posledica praktične i ne uvek savršene *primene* inače racionalne kompetencije. Veoma uticajan Koenov rad koji smo diskutovali u sekciji o normativnoj adekvatnosti iznosi upravo ovakav argument (Cohen, 1981). U Koenovom shvatanju pojma intuicije prepoznaje se potpuno slaganje sa načinom na koji pojam kompetencije razume Čomski: „*Njena najbliža analogija je ona sa intuicijom o dobroj gramatičkoj formiranosti.*“ (citirano prema Cohen, 1981/2008).

Problem sa ovim oblikom objašnjenja odstupanja od normativnih standarda predstavlja, naravno, veliki broj empirijskih nalaza koji pokazuju da su ta odstupanja sistematske prirode: da se greške ponavljaju, i da to ponavljanje grešaka uzima formu pravilnih struktura bihejvioralnih podataka. Uprkos tome, postoje teorijski radovi koji pokazuju da pretpostavljene racionalne kompetencije u određenom domenu, kombinovane sa određenim idejama o distribuciji slučajnih grešaka u ponašanju, mogu da objasne i neka od sistematskih odstupanja od racionalnosti (up. Blavatskyy, 2006, za stohastičko proširenje teorije očekivane korisnosti u formu čija je eksplanatorna moć uporediva sa onom kumulativne teorije izgleda).

Deluje razumno usvojiti metodološki princip da ako kognitivni sistem raspolaže racionalnom kompetencijom u nekom domenu, on jeste racionalan u tom domenu, bez obzira ne greške u performansi. Kao dobru ilustraciju, u kontekstu rasprave o principu maksimalne očekivane korisnosti, Pavličić citira fon Mizesa: „*Greška, neefikasnost i promašaj se ne smeju mešati sa neracionalnošću. Osoba koja gađa želi, po pravilu, da pogodi metu. Ako promaši, ona nije "neracionalna", već je slab strelac.*“ (Ludwig von Mises, citat preuzet iz Pavličić, 1997).

*Optimalnost i kognitivna ekonomija.* Princip kognitivne ekonomije prvi put je u istoriji kognitivne psihologije upotrebljen u klasičnom radu Kolinsa i Kilijana o hijerarhijskoj organizaciji semantičke memorije (Collins & Quillian, 1969), gde je kao argument za hijerarhijsku, taksonomsku strukture reprezentacije semantičkih informacija uzeta činjenica da svaki koncept u okviru taksonomije zahteva samo jednu reprezentaciju koja u potpunosti definiše sve relacije inkluzije u koje on ulazi (kojih je kategorija on član, i koje su kategorije njegovi članovi). Princip kognitivne ekonomije ovde koristimo u generalizovanoj formi da označimo teorijsko uverenje da kognitivni sistem pokušava da reši probleme adaptacije koje mu sredina nameće uz *minimalnu moguću upotrebu resursa* (informacija, procesa, funkcija, energije) koji su mu na raspolaganju (Savion & Morado, 2002). Neki autori vide kognitivnu ekonomiju kao fenomen širi od psihološkog, uviđajući da je optimalna organizacija znanja uopšte, u smislu odnosa cene i efikasnosti upotrebe tog znanja, neraskidivo povezana sa našim shvatanjem racionalnosti (Rescher, 1989). U svakom slučaju, ideja da će kognitivni sistem pokušati da reši neki adaptivni problem optimalno ne samo sa stanovišta strukture informacija koje definišu problem u okruženju, već i uz optimalnu upotrebu resursa koji su mu na raspolaganju, predstavlja teorijski stav sa suštinskim posledicama u debati o racionalnosti. Prvo, ne možemo da ne smatramo racionalnim kognitivni sistem koji optimizuje upotrebu svojih resursa. Drugo, optimizacija resursa može da kao svoju posledicu ima bihejvioralnu performansu koja odstupa od normativnih standarda: Sajmonov koncept zadovoljavanja uključuje tu ideju, da uz svako rešavanje problema ide i cena izračunavanja koje mora da se izvrši da bi problem bio rešen. Pitanje racionalnosti saznanja u celini je ovime veoma zakomplikovano: da li ćemo racionalnim nazvati kognitivni sistem čija bihejvioralna performansa ne dostiže normativne standarde, ali ne kao (a) posledice toga što sistem ne raspolaže racionalnom kompetencijom, već kao (b) posledice toga što sistem suboptimalno alocira raspoložive resurse?

*Proceduralna i deskriptivna invarijantnost.* Oba koncepta smo već koristili tokom prethodnih diskusija tako da ih ovde samo preciznije određujemo. Pod *deskriptivnom invarijantnošću* podrazumevamo normativan zahtev da se rešavanje nekog adaptivnog problema od strane kognitivnog sistema ne menja sa promenom deskripcije tog problema. Ukoliko kognitivni sistem pokazuje racionalnu performansu u rešavanju određenog problema datog u deskripciji $D_1$, ali ne pokazuje takvu performansu kada je isti problem dat u deskripciji $D_2$, nastupa kršenje deskriptivne invarijantnosti. Pod deskripcijom ovde ne

podrazumevamo isključivo jezičku, verbalnu deskripciju. Kršenja deskriptivne invarijantnosti su osnova za objašnjenja u paradigmi ograničene racionalnosti koja počivaju na ideji da kognitivni sistem ne reprezentuje problemsku situaciju na način koji je izomorfan reprezentaciji koju sadrži eksperimentalni nacrt. Proceduralna invarijantnost je normativan zahtev da racionalna performansa kognitivnog sistema za neki problem, ukoliko je osmotrena empirijski pod metodološkom procedurom $M_1$, mora da bude osmotrena i pod nekom drugom metodološkom procedurom $M_2$. Za proceduralnu invarijantnost je vezan dodatni problem koji ne propušta da konstatuje savremena filozofija nauke (Suppes, 2002), a to je da u analizi naučnog rada retko kada operacionalizujemo metodološke procedure; operacionalizacija se svodi na precizan opis eksperimentalnih nacrta, dok detalji same procedure (eksperimentalne instrukcije, ograničenja ponašanja u eksperimentalnoj situaciji i konkretan način prikupljanja podataka) najčešće nisu deo operacionalizacije. Veliki broj empirijskih odstupanja od racionalnosti spada u kršenja deskriptivne i proceduralne invarijantnosti; to što proceduralna invarijantnost nije operacionalizovana u teorijskim i metodološkim analizama predstavlja, kao što ćemo videti, problem daleko veći od pažnje koja mu je posvećena.

*Evoluciona, ekološka i environmentalna racionalnost.* Razmotrimo distinkciju u shvatanju racionalnosti kognitivnih funkcija koju uvodi evolucioni robotičar Harvi (Harvey, 2005). Prema Harviju, kognitivne funkcije možemo da posmatramo kao racionalne sa stanovišta (a) da su one proizvodi evolucionog procesa tokom kojeg su organizmi prilagođavani specifičnim uslovima sredine u kojoj žive, i sa stanovišta (b) da su one racionalne na način na koji to kolokvijalno kažemo za više kognitivne funkcije čoveka, koje Harvi vidi kao specifične prvenstveno zahvaljujući distinktivnoj odlici čoveka da koristi jezik. Racionalnost u drugom Harvijevom smislu bi verovatno odgovarala racionalnosti na našem suprapersonalnom nivou analize, na kome možemo da tvrdimo da je čovek „racionalna životinja“ jer se u svojoj adaptaciji sredini oslanja na eksteriorizovane simboličke funkcije u tehnološkim okruženjima koja mu omogućavaju adaptivnu dobit koju druge vrste ne mogu da razviju. Sa razvojem teorije evolucije, smatra Harvi, zahvaljujući ingenioznom uvidu Čarlsa Darvina da su sistemi koji *prima facie* izgledaju kao da su proizvod inteligentnog dizajna mogući kao posledice nenamernih, nenavođenih i neplaniranih procesa evolucije, otvoren je horizont za razumevanje racionalnosti u prvom od dva ponuđena smisla. Racionalnost viših kognitivnih i simboličkih funkcija čoveka proizvod je racionalnog procesa evolucije u prvom smislu te reči, dok je naša

mogućnost da shvatimo taj odnos posledica naše racionalnosti u drugom smislu reči: pozicija koja odlikuje naše epistemološke, naučne i sve druge refleksivne aktivnosti.

Razumevajući racionalnost kao proizvod procesa evolucije, postavlja se pitanje kako je uopšte moguća ograničena racionalnost ljudskog kognitivnog sistema. Prema našim naučnim shvatanjima, proces evolucije uobličuje funkcije nekog organizma, uključujući i kognitivne funkcije, na način koji u potpunosti odgovara logici optimizacije i zadovoljavanja ograničenja (engl. *constraint satisfaction*) koja nameće okolina u kojoj se organizam razvija. Neuronske mreže razvijaju optimalne odgovore sistema tako što optimizuju njegovo funkcionisanje u odnosu na sredinske promene. Na kraju procesa, odgovor sistema je optimalan u smislu reči da on postiže maksimalnu adaptivnu vrednost u odnosu na kompjutacione resurse kojima sistem raspolaže i složenosti izračunavanja koju nameću odgovarajući problemi učenja u odgovarajućoj sredini. Pod uslovom da je evolucija na raspolaganju imala dovoljno resursa da kognitivnom sistemu obezbedi dovoljno kompjutacione moći, svaki kognitini sistem je onda savršeno prilagođen da rešava odgovarajuće probleme adaptacije. Dakle, ako kognitivna performansa ne odgovara normativno racionalnoj slici u potpunosti, argument iz teorije evolucije glasi da bi kognitivna kompetencija na neki način *morala* da odgovara normativnoj slici.

Ni iz perspektive teorije evolucije nije uopšte jednostavno tvrditi da je racionalnost saznanja, makar samo u smislu kompetencije, nužna. Argument koji pruža američki psiholog Geri Markus predstavlja odlično polazno objašnjenje zašto je ograničena racionalnost moguća i kao proizvod optimalnog evolucionog procesa (Marcus, 2006). Markusov argument je u suštini veoma jednostavan: još Čarls Darvin, primećuje on, je shvatao da su postojeće forme života nasledenici nekih formi života koje su postojale pre njih. Komparativna analiza vrsta, zajedno sa analizom fosilnog materijala, nas empirijski uverava da je to tako, i danas mi to smatramo jednom od osnovnih empirijskih odlika procesa evolucije. Markus navodi veliki broj empirijskih argumenata kojima zastupa tezu da su različite kognitivne funkcije koje mi proučavamo danas u prošlosti imale zajedničke „kognitivne pretke“. Ono što je značajno za našu analizu racionalnosti saznanja je da se Markusov argument[31] direktno primenjuje u njoj na sledeći način: ako su kognitivne funkcije kakve mi proučavamo danas rezultati evolucije polazeći od nekih drugih, evoluciono prevaziđenih kognitivnih funkcija, onda njihova kompetencija i performansa koju mi opažamo u bihejvioralnim testovima mora biti ograničena. Evolucija, kako objašnjava Markus, „nema vremena“ da rešava svaki novi kognitivni

problem adaptacije uvođenjem novog rešenja koje bi predstavljalo optimalni adaptivni odgovor kognitivnog sistema na taj problem. U nuždi, evolucija rešava takve probleme evoluirajući nove kogntivne sisteme „prepravkom“ već postojećih kognitivnih sistema, tako zapravo dolazeći do rešenja do kojeg može da dođe, pre nego do idealnog rešenja. Drugim rečima, nisu sredinska ograničenja jedina koja utiču na oblik optimizacije kognitivnog odgovora organizma, već su i *ograničenja data prethodnom dinamičkom evolucijom sistema* uključena u optimizaciju neke kognitivne funkcije koju mi posmatramo. Na bihejvioralnom planu, gde donosimo zaključke o uspešnosti u optimalnoj adaptaciji, ponašanje tako otkriva formu ograničene racionalnosti za svaku onu kognitivnu funkciju koja u procesu evolucije nije potpuno optimizovana - zahvaljujući ograničenjima koja su posledica njene *prethodne* evolucione dinamike.

Koncept *ekološke racionalnosti*, iako povezan sa shvatanjem da je proces evolucije neizostavna komponenta objašnjenja racionalnosti ili odstupanja od nje, ima specifično značenje u debati o racionalnosti. Koncept ekološke racionalnosti dugujemo nemačkom psihologu Gerdu Gigerenceru i njegovim kolegama iz istraživačke grupe ABC (engl. *Adaptive Behavior and Cognition*). Ovaj koncept iznet je i diskutovan u mnogobrojnim publikacijama ove istraživačke grupe (npr. Todd & Gigerenzer, 2000, 2007, Gigerenzer, 2008, za novije ekspozicije ovog istraživačkog programa). Prema Gigerenceru, normativnu osnovu za analizu kognitivnog funkcionisanja čoveka ne čini logika, niti je čine druge formalne teorije, već nju predstavlja *struktura realnog prirodnog okruženja* u kome se adaptacija organizma sa određenim kognitivnim sistemom odvija. U strukturi koja je složena poput našeg realnog okruženja, kognitivni sistem najčešće ni ne može da razvije optimalna rešenja, jer su ona preskupa u smislu kompjutacionih resursa koji su njemu na raspolaganju. Igra šaha je, kako primećuje Gigernecer, samo jedna ilustracija za problem koji je realno nemoguće rešiti primenom najbolje, optimalne strategije izbora poteza. Zato je evoluciono rešenje koje primenjuje naš kognitivni sistem skup kognitivnih funkcija i procesa koje Gigernezer naziva *adaptivnim alatkama* (engl. *adaptive toolbox*), a te alatke predstavljaju specifičan skup strategija rešavanja problema koje se nazivaju *heuristikama*. Heuristike su koncept koji je zastupljen u kognitivnim naukama još od prvih istraživanja veštačke inteligencije. One suštinski predstavljaju algoritme koji na pojednostavljen način, brzo i jeftino sa stanovišta potrošnje kompjutacionih resursa, rešavaju problem na zadovoljavajuć način. Kao primer, posebno popularna u najnovijim istraživanjima

odlučivanja u uslovima rizika je tzv. *heuristika prioriteta* (Brandstätter, Gigerenzer & Hertwig, 2006), za koju se pokazuje da je u stanju da primenom samo elemenatrno jednostavnih kognitivnih operacija poređenja vrednosti i verovatnoća na rizičnim lozovima objasni veliki broj robustnih empirijskih nalaza koji se inače objašnjavaju komplikovanim i kompjutaciono zahtevnim teorijama poput teorije izgleda. U istoriji psihološke misli su verovatno najpoznatije heuristike *reprezentativnosti* i *dostupnosti* Tverskog i Kanemana, čijom se primenom objašnjavaju odstupanja od normativnih okvira u donošenju sudova o verovatnoćama (Tversky & Kahneman, 1973). Heuristike, prema Gigerenzeru, predstavljaju brz, jeftin i pametan način da se reši veliki broj adaptivnih problema. One su „brze i jeftine“ zato što počivaju na evoluciono određenim funkcijama koje su prilagodile kognitivni sistem organizma sredini onoliko koliko je to bilo moguće, tako da zahvaljujući tome što su „gotova rešenja“ heuristike ne zahtevaju veliku kompjutacionu snagu da bi funkcionisale; „pametne su“, kako se izražava Gigerenzer, jer koriste ekološki relevantnu strukturu podataka koja karakteriše stabilnu sredinu u kojoj se odvija adaptacija organizma, pa nema potrebe da troše kompjutacione resurse na prikupljanje i izračunavanje novih informacija.

Koncept *environmentalne racionalnosti* ćemo u narednim diskusijama koristiti u usko specikovanom značenju koje upravo iznosimo. Kao što ćemo videti, jedna od strategija odgovora na empirijske nalaze koji ustanovljavaju ograničenu racionalnost u nekom domenu kognitivnog funkcionisanja bazira se na sledećoj ideji:

(*i*) empirijski nalazi koji govore o ograničenoj racionalnosti posledica su racionalnih, normativnih procesa, *ako se interpretacija tih nalaza vrši pod pretpostavkom da*

(*iia*) kognitivni sistem *integriše podatke koji su mu predstavljeni* u određenoj problemskoj (eksperimentalnoj) situaciji sa

(*iib*) *drugim podacima relevantnim za strukturu problema koji rešava, a koji su uskladišteni kao prethodne reprezentacije tog kognitivnog sistema.*

Svaki put kada se te prethodne informacije koje kognitivni sistem koristi, integrišući ih sa informacijama koje sadrži struktura problema adaptacije kojoj je izložen, odnose na neku *deskripciju ekološki relevantnog okruženja u kom kognitivni sistem inače rešava taj problem adaptacije*, govorićemo o tom objašnjenju kao pozivanju na *environmentalnu racionalnost.* Ilustrovaćemo ideju o environmentalnoj racionalnosti jednim primerom. Pretpostavimo da u realnom okruženju u kome

kognitivni sistem donosi sudove o verovatnoćama nekih neizvesnih događaja postoje određene stabilne distribucije verovatnoća tih događaja. Na primer, možemo da pitamo naše ispitanike koja je verovatnoća da će sutra padati kiša u eksperimentu koji organizujemo u Beogradu. Ukoliko je jedan od naših ispitanika proveo neke godine života u Londonu, moguće je da će on dati drugačiji sud o ovoj verovatnoći (precenjujući je kao višu od realne verovatnoće da će sutra padati kiša u Beogradu), jer je svoje kratkotrajno poznavanje vremenskih prilika u Beogradu podredio svom mnogo boljem poznavanju vremenskih prilika u Londonu. U odnosu na događaj „sutra će padati kiša“, naš imaginarni ispitanik ima već prethodno razvijenu reprezentaciju verovatnoće ovog događaja. Ta reprezentacija je razvijena u sredini *drugačije strukture* od one u kojoj mi od njega tražimo da donese svoj sud. Ukoliko takav ispitanik koristi svoje prethodno znanje o nekim događajima zajedno sa znanjem koje je stekao u sredini u kojoj mi istražujemo njegove kognitivne funkcije, moguće je da će naše opservacije upućivati na ograničenu racionalnost njegovih kognitivnih funkcija. Međutim, *ukoliko bismo mi uzeli u obzir* da on koristi i te prethodne informacije koje nisu eksplicitne u empirijskom testu njegovih kognitivnih funkcija koji sprovodimo, moguće je da bismo želeli da promenimo naš stav o tome da su te funkcije ograničeno racionalne. Ovo pozivanje na upotrebu prethodno reprezentovanih, *apriornih* struktura podataka u objašnjenju suđenja, odlučivanja, rezonovanja i drugih kognitivnih funkcija karakteristično je za trenutno veoma popularne modele bejzijanske racionalnosti koje diskutujemo sledeće.

*Racionalne bejzijanske teorije.* Termin *bejzijanske teorije* potiče od matematičke forme koju sve teorije ove grupe uzimaju, forme dobro poznate *Bejzove teoreme* u teoriji verovatnoće. Označimo sa $A$ određeni *događaj*, a sa $H_1$, $H_2$, ..., $H_n$ *hipoteze*, odn. neke moguće okolnosti pod kojima očekujemo da taj događaj može da nastupi (ili da je mogao da nastupi, ako analizu sprovodimo retrospektivno). Zavisnu verovatnoću da se događaj $A$ ostvario pod hipotezom $H_i$ označavamo kao $\mathrm{P}(A|H_i)$. Interesuje nas verovatnoća da se događaj $A$ odigrao pod nekom od alternativnih hipoteza $H_1$, $H_2$, ..., $H_n$; očigledno, ono što nas interesuje je zavisna verovatnoća $\mathrm{P}(H_i|A)$. Jednostavna inferencija na osnovu aksioma teorije verovatnoće pokazuje da ova zavisna verovatnoća može da se izračuna kao:

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_j P(A|H_j)P(H_j)} \tag{11}$$

Termin $\mathrm{P}(A|H_i)$ označava obrnutnu zavisnu verovatnoću od one koja nas

interesuje, $\mathrm{P}(H_i|A)$. Zavisna verovatnoća $\mathrm{P}(A|H_i)$ zapravo predstavlja koncept matematičke statistike poznat kao *verodostojnost* (engl. *likelihood*). Zamislimo da su $H_1$, $H_2$, ..., $H_n$ neke različite eksplanatorne hipoteze za objašnjenje osmotrenih podataka $A$. Verodostojnost, zavisna verovatnoća $\mathrm{P}(A|H_i)$, nam govori o tome koliko je *verovatno da osmotrimo podatke A ako je tačna hipoteza* $H_i$. Međutim, nas u naučnoj praksi po pravilu interesuje kolika je verovatnoća da je *neka eksplanatorna hipoteza tačna za date podatke*, a ne obrnuto. Ako nam je na raspologanju više hipoteza, za koje možemo da izračunamo verodostojnosti u obliku $P(A|H_i)$, Bejzova teorema (11) nam omogućava da izračunamo i verovatnoću hipoteze koja nas interesuje. Termin $P(H_i)$ označava *a priori* verovatnoću hipoteze: verovatnoću sa kojom smatramo da je neka $H_i$ tačna pre nego što dobijemo podatke koji o njenoj tačnosti svedoče[32]. Dakle, u bejzijanskoj analizi verovatnoće hipoteza, mi moramo da poznajemo verodostojnosti hipoteza, odn. verovatnoće da su osmotreni podaci dobijeni pod pretpostavkom da je svaka od njih tačna, i da imamo na raspolaganju neka prethodna uverenja o tačnosti hipoteza, izraženih u formi *a priori* verovatnoća. U literaturi se konceptualno Bejzova teorema često predstavlja u obliku (termini su na engleskom jeziku): $Posterior = Likelihood \times Prior$; verovatnoće *a posteriori* proizvod su verodostojnosti i verovatnoća *a priori*.

Bejzova teorema se smatra *normativnim teorijskim okvirom* za procese *revizije verovanja* (engl. *belief updating*), koji predstavljaju neke od najvažnijih kognitivnih procesa uopšte. Učenje je, na primer, kao proces nezamislivo bez mogućnosti revizije verovanja o rizičnim ili neizvesnim aspektima okoline. Ogroman broj drugih kognitivnih procesa podrazumeva neku mogućnost revizije verovanja, odn. korekcije verovanja o određenim događajima u svetu pošto kognitivni sistem prikupi nove evidencione podatke o tim događajima. U principu, ne postoji teorijsko ograničenje koje bi sprečilo da svi kognitivni procesi budu izraženi u formi odgovarajućih bejzijanskih modela; od teoreme teorije verovatnoće, bejzijanska inferencija se do danas razvila u kompletnu filozofiju o tome kako treba modelirati kognitivne procese u psihologiji (Oaksford & Chater, 2009, up. Jones & Love, 2011, za elaboriranu diskusiju bejzijanskog pristupa). Proces bejzijanske inferencije podrazumeva primenu Bejzove teoreme u obnavljanju verovanja: prethodna verovanja su reprezentovana verovatnoćama *a priori*, podaci se izražavaju u formi verodostojnosti, da bi se izračunale verovatnoće *a posteriori*. Jedan suštinski deo debate o racionalnosti posvećen je pitanju da li je ljudski kognitivni sistem sposoban za racionalnu bejzijansku inferenciju, odn. da li procesi obnove verovanja slede formu

Bejzove teoreme (Kahneman & Tversky, 1972, Cosmides & Tooby, 1996). Literatura o bejzijanskim kognitivnim modelima je danas bespregledna; njihova popularnost značajno je porasla od 80-ih i 90-ih godina XX veka, kada su kompjuterski sistemi sposobni da ocene modele komplikovanih bejzijanskih inferencija postali široko dostupni na univerzitetima. Nadovezujući se na primer koji smo predstavili u prethodnom uvođenju pojma enviromentalne racionalnosti, ovde možemo da se pozovemo na karakteristične rezultate bejzijanske paradigme koje predstavljaju Grifits i Tenenbaum (Griffiths & Tenenbaum, 2006). Grifits i Tenenbaum su velikom broju ispitanika postavili jednostavna pitanja o proceni trajanja nekog intervala ukoliko im se prethodno da informacija o tome koliko dugo proces koji se posmatra *već traje*. Na primer, ispitanicima bi postavili pitanje o tome koliko očekuju da će živeti osoba koju su sreli i koja trenutno ima 39 godina, ili osoba koju su sreli i koja trenutno ima 96 godina. Kognitivni sistem ovaj problem može da reši ukoliko poznaje distribuciju trajanja ljudskog života u relevantnoj populaciji uopšte, tj. distribuciju verovatnoće da nečiji život traje 18, 19, 20, 21, ..., 96, 97, 98, itd. godina. Očigledno, distribucija o kojoj govorimo daje *a priori* verovatnoće neophodne za bejzijansku inferenciju. Ukoliko poznaje ovu distribuciju, kognitivni sistem sada mora da formuliše relevantnu *verodostojnost*, odn. zavisnu verovatnoću da neko već ima npr. 39 godina ako pripada populaciji sa datom *a priori* distribucijom (Grifits i Tenenbaum pretpostavljaju da je verodostojnost uniformna u ovom slučaju, što odslikava verovanje da smo osobu čiji se životni vek procenjuje mogli da sretnemo bilo kada tokom njenog života). Na osnovu ove dve informacije, bejzijanska inferencija omogućava da se izvede *a posteriori* verovatnoća da će život neke osobe trajati određen broj godina *ako znamo da ona pripada datoj populaciji* i *ako znamo da ona već ima određen broj godina*. Upotrebom tzv. bejzijanske funkcije predikcije, Grifits i Tenenbaum pokazuju da postoji izuzetno visoko slaganje ljudskih procena o trajanju života individua sa procenama koje daje bejzijanski model na osnovu javno dostupnih statističkih podatka (Griffiths & Tenenbaum, 2006). Ovaj nalaz snažno protivreči nalazima o tome da ljudi generalno ne koriste mehanizam bejzijanske inferencije u reviziji relevantnih verovanja.

Objašnjenje pozivanjem ne princip normativne bejzijanske inferencije danas je ubedljivo najpopularnija *strategija racionalizacije ograničene racionalnosti*. Ukoliko je neko osmotreno ponašanje ograničeno racionalno, bejzijanski pristup pruža objašnjenje da ono zapravo zadovoljava normativne standarde *ako se u obzir uzmu prethodna verovanja* (*a priori* verovatnoće) *subjekta čije se ponašanje analizira*.

Drugim rečima, inkorporacija prethodnih verovanja koja se kroz Bejzovu teoremu integrišu sa objektivnim podacima - onima koje pruža odgovarajući eksperimentalni nacrt - može da koriguje osmotrena odstupanja od normativnog rešenja za problem adaptacije koji se analizira. Andersonova metodologija racionalne analize, kao i njegovi radovi u kojima iznosi racionalne analize pojedinih kognitivnih funkcija, potpuno je formulisana u bejzijanskom duhu. Bejzijanski pristup oličava skoro savršeno formu standardnog psihološkog objašnjenja u kome je ponašanje neke osobe proizvod aktualne problemske situacije u kojoj se ona nalazi i prethodnih znanja koja ona o rešavanju takve problemske situacije ima. Ne smemo da propustimo da primetimo da forma bejzijanskog objašnjenja savršeno odgovara formi objašnjenja pozivanjem na environmentalnu racionalnost koju smo prethodno diskutovali. Kakvi problemi se javljaju u ovoj strategiji objašnjenja ograničene racionalnosti videćemo u VI delu rada, pošto prethodno u V delu temeljno prodiskutujemo osobine jedne racionalne bejzijanske teorije odlučivanja.

# Deo III

# Debata o racionalnosti

Sadržaj naučne periodike u oblasti viših i simboličkih kognitivnih funkcija danas je bespregledan. Ogroman broj istraživača svakodnevno doprinosi novim empirijskim rezultatima, a u praktično svim oblastima proučavanja viših kognitivnih procesa nalazimo po nekoliko suprostavljenih teorija ili hipoteza. Naredne redove zato ne treba posmatrati kao pregled debate o racionalnosti u ovom domenu. Kritičko razmatranje nekoliko oblasti koje ovde predstavljamo vođeno je izborom onih naučnih radova, hipoteza, teorija i empirijskih istraživanja koja su blisko konceptualno povezana sa suštinom debate o racionalnosti saznanja. Već prethodne diskusije su dovoljno ukazale na apstraktnost i širinu ove problematike: činjenica da se u okviru nje, zbog tih osobina, može diskutovati ma koji naučni rad u oblasti viših kognitivnih funkcija, čini svaki pokušaj pregleda unapred ograničenim. Naše je mišljenje da je bolje kritičkom oštricom seći kroz oblasti, pronalazeći mesta koja markantno ukazuju na način izgradnje naučnih teorija o višim kognitivnim funkcijama i povezujući ih sa metateorijskim analitičkim okvirom koji je predložen na prethodnim stranicama, nego pokušavati da se debata o racionalnosti predstavi u celini.

# 7 Racionalnost viših i simboličkih funkcija

Podela na „niže“ i „više“ kognitivne funkcije, uobičajena u udžbenicima, organizaciji nastavnih kurikuluma i pregledima kognitivne psihologije, nije odgovarajuća klasifikacija kognitivnih procesa. Prema uobičajenoj klasifikaciji, donošenje odluka spada u više kognitivne funkcije. To je potpuno netačno; odlučivanje je toliko prevalentna kognitivna funkcija da je nju moguće proučavati u jedinstvenom teorijskom okviru na svim nivoima kognitivnog funkcionisanja čoveka od psihofizičkog i senzomotornog do simboličkog. Sporno je da li procese suđenja možemo nedvosmisleno da shvatimo kao „više“ kognitivne procese. Ova podela je, dakle, više posledica sociolingvističke podele rada među psiholozima nego što je rezultat primene fundamentalnih principa. Ipak, zbog očigledne potrebe da se obim diskusije o racionalnosti saznanja ograniči, mi je se ovde pridržavamo. Sve analize koje slede odnose se na one funkcije koje kolokvijalno, dakle, nazivamo „višim“ i simboličkim kognitivnim funkcijama. Pod simboličkim kognitivim funkcijama u diskusijama koje slede mislimo na sve one funkcije koje se odnose na značenje i upotrebu znakova. Kao što je već rečeno, psiholingvistika, odn. njen deo koji se odnosi na proučavanje kognitivnog statusa morfologije i sintakse, te fonologije i fonetike, isključena je iz naše rasprave zbog velikog broja specifičnosti koje odlikuju ovu oblast. Jedina oblast u kojoj se naša rasprava graniči sa psiholingvističkom je oblast psihološke semantike. Diskusije u oblasti percepcije takođe nisu uključene u našu analizu racionalnosti saznanja; ipak neće izostati naše pozivanje na one argumente iz ove oblasti koji su dovoljno opšti da ih je moguće primeniti i na više i simboličke funkcije.

Redom razvijamo kritičku diskusiju normativnih i deskriptivnih objašnjenja u (7.1) oblasti odlučivanja u uslovima rizika i neizvesnosti, (7.2) kauzalnog učenja, (7.3) epizodičke memorije, (7.4) rezonovanja i suđenja, (7.5) konceptualne organizacije u funkciji kategorizacije, i (7.6) konceptualne organizacije u funkciji otkrivanja novih koncepata. Svaku sekciju završavamo kraćim kritičkim komentarom o statusu debate o racionalnosti u odgovarajućoj oblasti. Zaključke kritičkih diskusija u svakoj od navedenih oblasti izvodimo u sekciji 8 gde pokušavamo da odredimo status istraživanja u pobrojanim oblastima unutar metateorijskog okvira za analizu racionalnosti saznanja.

## 7.1 Odlučivanje u uslovima rizika i neizvesnosti

Konceptualnu osnovu za diskusiju odlučivanja u uslovima rizika i neizvesnosti već smo razvili. Bitna dihotomija koja se poštuje u ovoj oblasti podrazumeva razliku između odlučivanja u uslovima *rizika* i u uslovima *neizvesnosti* (Knight, 1921). U odlučivanju u uslovima rizika, donosiocu odluka su poznate (date) objektivne verovatnoće mogućih ishoda, kao na primer u rizičnom lozu (50 *EUR*,$\frac{1}{2}$; 25 *EUR*,$\frac{1}{2}$), lozu koji sa verovatnoćom od $\frac{1}{2}$ donosi igraču 50 ili 25 evra. Odlučivanje u uslovima *neizvesnosti* podrazumeva da donosilac odluka mora da se osloni na svoju subjektivnu ocenu verovatnoća, na osnovu bilo kojih informacija koje može da iskoristi da do takve ocene dođe. Na primer, u izboru između lozova (50 *EUR*, „*Sutra pada kiša u Parizu*“; 25 *EUR*, „*Sutra ne pada kiša u Parizu*“) i (25 *EUR*, „*Sutra pada sneg u Beogradu*“; 75 *EUR*, „*Sutra ne pada sneg u Beogradu*“), donosilac odluka mora na osnovu subjektivnih informacija da proceni verovatnoće događaja „*Sutra pada kiša u Parizu*“ i „*Sutra pada sneg u Beogradu*“. Normativni okvir je potpuno jasan: njega predstavljaju teorija očekivane korisnosti fon Nojmana i Morgenšterna (EU), za uslove rizika, i Sevidžova teorija subjektivne očekivane korisnosti (SEU), za uslove neizvesnosti i rizika. Podsetimo se da je Sevidž razvio aksiomatski okvir (teoriju SEU) za odlučivanje u uslovima neizvesnosti čiji aksiomi tzv. komparativne (ili kvalitativne) verovatnoće garantuju da je neizvesnost moguće matematički reprezentovati merom koja potpuno odgovara meri verovatnoće. To znači da je na osnovu izbora nekog donosioca odluka moguće kvantifikovati njegova subjektivna verovanja o neizvesnim ishodima tako da ta kvantifikacija bude izražena kao verovatnoća. Ne treba zaboraviti da se u formalnom, matematičkom razvoju teorija odlučivanja slučaj rizika pokazuje kao specijalan slučaj neizvesnosti, tako da je rešenje problema za slučaj neizvesnosti generalnije od rešenja za slučaj rizika. Odlučivanje u uslovima rizika i neizvesnosti je jedina oblast koju razmatramo u kojoj je normativni okvir - bar koncenzusom u stručnim krugovima - utvrđen, aksiomatizovan, i kroz odgovarajuće parametarske matematičke modele povezan sa empirijskim podacima. Problem odlučivanja u uslovima rizika i neizvesnosti ima jedinstven pedigre u analizama racionalnosti saznanja: istorijski, to je problem kojim je rasprava započeta, i oko koga se vode teorijski i matematički najsofisticiranije diskusije u debati o racionalnosti.

Usvajamo sledeću notaciju za slučaj rizika. *Loz* predstavlja distribuciju verovatnoća nad skupom ishoda koji mogu biti pozitivni, negativni (gubici) ili neutralni (nula), i označava se: $(x_1,p_1;x_2,p_2;..;x_n,p_n)$, za loz koji sa verovatnoćom

$p_1$ donosi ishod $x_1$, sa verovatnoćom $p_2$ donosi ishod $x_2$, ..., sa verovatnoćom $p_n$ donosi ishod $x_n$. Diskusiju organizujemo na sledeći način. Prvo razmatramo status normativnih i deskriptivnih objašnjenja za tri komponente rizika koje prepoznaje savremena teorija odlučivanja (Wakker, 2010), redom: komponetne korisnosti, ponderisanja verovatnoća i averzije prema gubicima; zatim diskutujemo status objašnjenja koja se oslanjaju na heuristike i mogućnosti objašnjenja nekih bitnih problema proceduralne i deskriptivne invarijantnosti u odlučivanju. Konačno, diskutujemo status stohastičkih teorija korisnosti koje predstavljaju neka savremena proširenja EU i RDU.

*Funkcija korisnosti.* Nikad ne treba gubiti iz vida da se govoreći o funkciji korisnosti u teoriji odlučivanja može misliti na dva različita objekta. Funkcija korisnosti EU, koja izračunava korisnost nekog rizičnog *loza*, uvedena jednačinom (2) još na početku ove teze, razlikuje se od funkcije korisnosti koja preslikava vrednosti u korisnosti u *uslovima odsustva rizika*, $u : x \to u(x)$, i koja se najčešće naziva *Bernulijevom funkcijom korisnosti* da bi se razlikovala od prve. Strategija izgradnje svake teorije odlučivanja počiva na sledećim koracima. Prvo se konstruiše skup aksioma koje donosilac odluka mora da poštuje. Taj skup aksioma opisuje osobine binarne relacije preferencije, $\succsim$, i određuje striktno interpretaciju iskaza poput $P \succsim Q$ - da donosilac odluka preferira loz $P$ nad lozom $Q$ ili je indiferentan između $P$ i $Q$. U sledećem koraku, polazeći od aksioma gradi se najbitnija matematička inferencija teorije, a to je dokaz *reprezentacione teoreme*. Reprezentaciona teorema pokazuje da je skup aksioma teorije, koji predstavljaju osobinu relacije binarne preferencije *za lozove*, $\succsim$, *ekvivalentan* funkciji korisnosti *za lozove*, tako da ako za dva loza $P$ i $Q$ važi $P \succsim Q$, onda $P \succeq Q \iff U(P) \geqslant U(Q)$. Drugim rečima, dokaz reprezentacione teoreme pokazuje da $(i)$ ako strukura ponašanja tj. opaženi izbori neke osobe zadovoljavaju aksiome teorije koju diskutujemo, onda se $(ii)$ ista ta struktura ponašanja dobija ako se pretpostavi da ta osoba koristi internu, neopservabilnu funkciju korisnosti za lozove $U(\cdot)$. Dokazi reprezentacionih teorema u teorijama odlučivanja vode i ka ustanovljavanju osobina funkcije korisnosti za lozove U$(\cdot)$, ali te osobine još uvek ne određuju specifičnu formu funkcije korisnosti. Funkcija korisnosti predstavlja *meru* relacije preferencije: ona preslikava svaki rizični loz u određeni realni broj, na taj način da uređenje među tim realnim brojevima odgovara uređenju preferencija donosioca odluka. Funkcije korisnosti nad lozovima su skoro bez izuzetka mere na intervalnoj skali; jedino je funkcija korisnosti teorije izgleda mera na racio skali, ali je ta osobina obezbeđena shvatanjem da je korisnost

neutralnog ishoda $u(0) = 0$ po definiciji. Bernulijeva funkcija korisnosti se izgrađuje u trećem koraku, pokazivanjem da za tzv. *svedene lozove*[33], lozove kod kojih se sva verovatnoća nalazi na samo jednom ishodu (na primer, loz koji sa verovatnoćom 1 donosi 50 evra), tako da oni sada predstavljaju jedan određen siguran ishod, mora da postoji funkcija $u : x \to u(x)$, koja vrednosti sigurnih ishoda preslikava u subjektivne korisnosti. Razlika u notaciji je ovde od fundamentalnog značaja: funkcija $U(\cdot)$ je funkcija korisnosti nad lozovima, a funkcija $u(\cdot)$ nad sigurnim ishodima; termin funkcija korisnosti koji je poznat psiholozima i koji predstavlja neku vrstu paralele psihofizičkim funkcijama u domenu odlučivanja odnosi se na $u(\cdot)$.

Kao što smo već izneli, u matematičkom razvoju neke teorije odlučivanja, inferencije u sklopu reprezentacione teoreme omogućavaju da se ustanove samo neke opšte osobine funkcije korisnosti. To znači da više matematičkih funkcija mogu da zadovolje te osobine. Kada je model teorije odlučivanja dat sa tek uopštenim osobinama funkcija, bez *post hoc* izbora konkretne funkcionalne forme koja će biti korišćena kao Bernulijeva funkcija korisnosti te teorije, kažemo da je model dat u *neparametarskoj formi* (jer će parametre dobiti tek kad odaberemo konkretnu funkciju); kada je dat sa tačno određenom, npr. stepenom funkcijom korisnosti, on je dat u *parametarskoj formi* (kod stepene funkcije taj parametar je njen eksponent i on, kao što smo već diskutovali, određuje stepen averzije prema riziku). Odnos između parametarskog i neparametarskog oblika neke teorije odlučivanja dodatno komplikuje matematičke inferencije u određenoj teoriji. Za garancije da određeni parametarski oblik postoji i odgovara osobinama funkcije korisnosti neke teorije odlučivanja najčešće su neophodne dodatne pretpostavke, kao npr. pretpostavka o *homogenosti preferencija* za teoriju izgleda u formi datoj jednačinama (3) - (5) (Tversky & Kahneman, 1992, up. Tversky, 1967 za prvo uvođenje stepene funkcije korisnosti na osnovu homogenosti preferencija). Već smo, diskutujući Sajmonov pojam zadovoljenja, videli da postoje teorijska i empirijska opravdanja za drugačija shvatanja o funkciji korisnosti. Sajmon predlaže (Bernulijevu) funkciju korisnosti koja je diskretna (Simon, 1955a), deleći sve ishode na one koji donosioca odluka zadovoljavaju i one koje ga ne zadovoljavaju. Funkcije korisnosti su ipak najčešće date u parametarskoj formi i kontinuirane; najpopularnija među njima svakako je stepena funkcija, $u(x) = x^\rho$, koja za $\rho < 1$ opisuje averziju prema riziku, koncept koji smo već diskutovali; sklonost prema riziku ne smatra se racionalnom osobinom. Sve konkavne Bernulijeve funkcije korisnosti imaju osobinu averzije prema riziku; sve konveksne odgovaraju sklonosti ka riziku.

Fenomen da opažena vrednost tj. korisnost nije linearna funkcija objektivne vrednosti, te da postoje određena psihofizička preslikavanja koja upravljaju ljudskim odlukama u domenu rizika, jedan je od fundamentalnih nalaza u ovoj oblasti. Previše bitnih empirijskih činjenica i teorijskih konstrukata, poput averzije prema riziku i marginalne opadajuće vrednosti novca, vezani su za ovu ideju. Tako funkcija korisnosti ostaje centralni konstrukt svake teorije odlučivanja u uslovima rizika. U cilju razvoja deskriptivne teorije odlučivanja koja bi mogla da objasni empirijska odstupanja od normativne EU razvijen je veći broj različitih formalizacija odlučivanja. Sve one koriste istu strategiju razvoja odgovarajuće reprezentacione teoreme i funkcije korisnosti nad lozovima polazaći od odgovarajućeg skupa aksioma. Najpoznatija takva teorija je, svakako, kumulativna teorija izgleda Tverskog i Kanemana (Chateauneuf & Wakker, 1999, Tversky & Kahneman, 1992, Wakker, 2010, Wakker & Tversky, 1993). Ova teorija počiva na prethodnim razvojima RDU - *modela zavisnih od ranga* (Quiggin, 1982, Schmeidler, 1989, up. Abdellaoui, 2009, za diskusiju osnovnih osobina RDU). Razvoj teorija odlučivanja posle empirijskih kritika EU je komplikovan i nije ga moguće prikazati ovde (up. Schmidt, 2004, za detaljan pregled). Međutim, neke opšte karakteristike tog razvoja možemo da diskutujemo. Jedna koja je od suštinskog značaja za analizu racionalnosti saznanja odnosi se na *intuitivnost aksiomatskih sistema* razvijenih posle teorije očekivane korisnosti. Sa razvojem alternativnih teorija odlučivanja, razvijane su sve „pametnije“ tehnike aksiomatizacije relacije preferencije, koje su omogućavale razvoj deskriptivno moćnijih modela i ujedno omogućavale pojednostavljenje matematičkih inferencija i dokaznih postupaka u ovoj tehnički zahtevnoj oblasti. Na osnovu pregleda više deskriptivnih teorija, stiče se utisak kao da su sa tim razvojem teorija odlučivanja istraživači u oblasti sve manje i manje obraćali pažnje na to koliko su njihove aksiomatske osnove intuitivno prihvatljive (Milovanović, 2011), što je za neke autore, kao što smo videli, osnovni uslov normativne adekvatnosti modela. Uvek je moguć odgovor meliorista da deskriptivne teorije ni ne moraju da odgovaraju na taj zahtev, pošto ni sami ispitanici ne odgovaraju normativnim standardima racionalnosti, ali je teško izbeći pitanje o tome da li je moguće da ljudi svakodnevno donose odluke bazirajući se na principima koji im intuitivno nisu bliski. Postoje i teorijski pokušaji da se osnove složenih modela deskriptivnih teorija odlučivanja kao što je RDU povežu sa psihološkim intuicijama (Diecidue & Wakker, 2001). U sledećim redovima ulazimo dublje u aksiomatske osnove kumulativne teorije izgleda Kanemana i Tverskog, u nastojanju da izgradimo tle za precizniju diskusiju odnosa

procesa formalizacije odlučivanja, deskriptivne validnosti teorije koja rezultira u tom procesu, intuitivnih osnova predloženog aksiomatskog okvira i njegove normativne adekvatnosti.

Da bismo ispitali detaljnije odnos (a) aksiomatizacije deskriptivne teorije odlučivanja i (b) intuitivnosti normativnog okvira koji je takvom aksiomatizacijom dat, diskutujemo sledeću aksiomatizaciju koja je deo najvažnijeg razvoja u oblasti - kumulativne teorije izgleda Kanemana i Tverskog. Vaker i Tverski pokazuju (Wakker & Tversky, 1993, Wakker, 2010) da je uvođenjem matematičke operacije *otkupljivanja (*engl. *tradeoff)*, koju ćemo sada predstaviti, i njenim odgovarajućim prilagođavanjem za aksiomatizaciju nekog modela odlučivanja, moguće aksiomatizovati redom teoriju EU, teoriju RDU, i teoriju izgleda za slučajeve neizvesnosti (Chateauneuf & Wakker, 1999, pokazuju da je rešavanjem slučaja neizvesnosti za teoriju izgleda slučaj rizika rešiv istom metodom na nešto jednostavniji način). Šta je otkupljivanje u kontekstu rizičnih lozova? Neka su $\alpha, \beta, \gamma, \delta$ neka četiri sigurna ishoda, i neka su dati lozovi $P = (x_1,\ p_1;\ x_2,\ p_2;\ ...;\ x_n,\ p_n)$, i $Q = (y_1,\ p_1;\ y_2,\ p_2;\ ...;\ y_n,\ p_n)$, sa *istim* distribucijama verovatnoće ($p_1$, $p_2$, ..., $p_n$). Neka je loz $\alpha_j P$ loz koji nastaje tako što se $j$-ti ishod $x_j$ na lozu $P$ zameni ishodom $\alpha$; $\beta_j Q$, $\gamma_j P$ i $\delta_j Q$ imaju analogno značenje zamene $j$-tih ishoda u odgovarajućim lozovima. Onda, kažemo da ishodi $\alpha, \beta$ *otkupljuju* $\gamma, \delta$, u notaciji $\alpha, \beta \overset{*}{\succ} \gamma, \delta$ ako

$$\alpha_j P \succsim \beta_j Q \tag{12}$$

$$\gamma_j P \prec \delta_j Q \tag{13}$$

Drugim rečima: ako ishodi $\alpha, \gamma$ zamene isti ishod u lozu $P$, a ishodi $\beta, \delta$ zamene odgovarajući ishod u lozu $Q$, gde lozovi $P$ i $Q$ imaju istu distribuciju verovatnoće nad ishodima, i ako posle te operacije zamene dobijamo preferencije $\alpha_j P \succsim \beta_j Q$ i $\gamma_j P \prec \delta_j Q$, onda ishodi $\alpha, \beta$ otkupljuju ishode $\gamma, \delta$. Izraz $\alpha, \beta \overset{*}{\succsim} \gamma, \delta$ koristi relaciju $\overset{*}{\succsim}$ koja označava samo to da je odnos preferencija takav da u jednačini (13) umesto $\prec$ nalazimo $\precsim$; relacije $\overset{*}{\prec}$ i $\overset{*}{\precsim}$ samo obrću preferencije u jednačinama (12) i (13) na očekivan način. Vaker i Tverski pokazuju da je sledeći uslov dovoljan za aksiomatizaciju teorije očekivane korisnosti:

*Konzistencija otkupa*: konzistencija otkupa je zadovoljena ako ne postoje ishodi

$\alpha, \beta, \gamma, \delta$ tako da u $\alpha, \beta \overset{*}{\succ} \gamma, \delta$ i $\alpha, \beta \overset{*}{\precsim} \gamma, \delta$ važe.

Na prvi pogled, konzistencija otkupa predstavlja čist formalizam zakomplikovan uvođenjem novih, ekstravagantnih relacija u teoriju odlučivanja; tek posledice po odnose korisnosti sigurnih ishoda koje nastaju uvođenjem uslova konzistencije otkupa omogućavaju da se jasno razume njegov značaj. Vaker i Tverski pokazuju da iz diskutovanog uslova konzistencije zapravo sledi *izostanak kontradikcije u uređenju parova razlika korisnosti ishoda* (na Bernulijevoj funkciji korisnosti), odn. da ako je konzistencija otkupljivanja zadovoljena, *ne mogu da se pojave takvi ishodi* $\alpha, \beta, \gamma, \delta$ tako da istovremeno $u(\alpha)-u(\beta) \geqslant u(\gamma)-u(\delta)$ i $u(\alpha)-u(\beta) < u(\gamma)-u(\delta)$. Vidimo da je tek u ovom narednom koraku formalnog izvođenja aksiomatizacija koju predlažu Vaker i Tverski približena intuiciji. Pretpostavimo da su $\alpha, \beta, \gamma, \delta$ monetarni dobici u iznosima 3, 5, 1 i 9 evra, redom. Neka su $u(3)$, $u(5)$, $u(1)$ i $u(9)$ odgovarajuće korisnosti koje tim dobicima dodeljuje kognitivni sistem nekog donosioca odluka. Pošto je funkcija korisnosti monotona transformacija vrednosti, imamo $u(9) > u(5) > u(3) > u(1)$. Jasno da intuiciji nikako ne može da odgovara situacija u kojoj *istovremeno* važi $u(3) - u(5) \geqslant u(1) - u(9)$ i $u(3) - u(5) < u(1) - u(9)$. Vaker i Tverski matematički pokazuju da uslov konzistencije otkupa, ako je zadovoljen, onemogućava da se upravo opisana situacija pojavi u odnosima između korisnosti koje neki donosilac odluka pridaje vrednostima o kojima donosi odluke.

Pošto se u skup aksioma uvede konzistencija otkupa, potrebni su još aksiom *slabog uređenja* (koji zajednički uvodi već diskutovane osobine simetriju i tranzitivnosti relacije preferencije), *aksiom kontinuiteta* (takođe već diskutovan) i jedan standardan uslov koji se naziva *monotonošću* da bi se aksiomatizovala fon Nojman-Morgnešternova EU u celini. Svi pobrojani uslovi su standardni i manje problematični od aksioma nezavisnosti koji više nije neophodan za aksiomatizaciju EU posle uvođenja relacije otkupljivanja. Ovo ne znači da preferencije ne zadovoljavaju nezavisnost, koju je sada moguće izvesti kao teoremu EU, već samo to da nezavisnost ne mora da bude u skupu polaznih aksioma. Uopšte, velika većina deskriptivnih teorija odlučivanja posle EU je bila bazirana na ideji slabljenja uslova koji navodi aksiom nezavisnosti; ideja Vakera i Tverskog, koja danas predstavlja standardan način aksiomatizacije teorija odlučivanja, je prva (po našem znanju) koja u potpunosti napušta strategiju modifikacije aksioma nezavisnosti.

U razvoju deskriptivnih modela, kao što su RDU i nešto složenija kumulativna teorija izgleda, Vaker i Tverski pokazuju sledeće: uvođenjem postepenih modifikacija odgovarajućih uslova konzistencije, naime, *konzsitencije komonotonog otkupljivanja*

(engl. *comonotonic tradeoff consistency*) i *konzistencije saoznačenog otkupljivanja* (engl. *co-signed tradeoff consistency*), moguće je aksiomatizovati kumulativnu teoriju izgleda (uz već navedene standardne uslove). Komonotono otkupljivanje vezano je za, kako samo ime govori, *komonotone lozove*. U formalnim analizama koncept komonotonosti počiva na komplikovanijim konstrukcijama teorije merenja koje težimo da izbegnemo u ovom izlaganju; zato ćemo ga objasniti na primeru. Prepostavimo da ponovo imamo dva loza $P$ i $Q$ koji predstavljaju istu distribuciju verovatnoća nad određenim ishodima. Neka je, na primer, loz $P$: (50 *EUR*, 15%; 25 *EUR*, 35%; 5 *EUR*, 50%), a loz $Q$: (10 *EUR*, 15%; 5 *EUR*, 35%; 1 *EUR*, 50%). Fokusirajmo se za trenutak na samo neke od opcija na lozu $P$: vidimo da ovaj loz, na primer, donosi 50 evra sa verovatnoćom 15%, kao i da donosi 25 evra sa verovatnoćom 35%. Dakle, dobitak vezan za verovatnoću od 15% na lozu $P$ (50 evra) je veći od dobitka vezanog za verovatnoću od 35% (25 evra) na tom lozu. Posmatrajmo sada loz $Q$ i uočimo istu osobinu: on daje 10 evra sa 15%, i 5 evra sa 35%; baš kao i na lozu $P$, dobitak vezan za verovatnoću od 15% na lozu $Q$ (10 evra) je veći od dobitka vezanog za verovatnoću od 35% (5 evra) na istom tom lozu. Ako za dva loza važi da je za sve dobitke (ili gubitke) ponuđene sa korespondetnim (istim) verovatnoćama na dva loza pravac razlika između njih *uvek isti* na način upravo ilustrovan primerom, onda kažemo da su ta dva loza *komonotona*. Tako lozovi $P$: (50 *EUR*, 15%; 25 *EUR*, 35%; 5 *EUR*, 50%), i loz $R$: (5 *EUR*, 15%; 10 *EUR*, 35%; 1 *EUR*, 50%) nisu komonotoni: dok na lozu $P$ sa 15% dobijamo 50 evra a sa 35% 25 evra, na lozu $R$ sa 15% dobijamo 5 evra a sa 35% 10 evra, i vidimo da je pravac razlike između dobitaka koji su ponuđeni sa 15% i 35% na lozu $P$ obrnut od pravca razlike između dobitaka ponuđenih sa istim verovatnoćama na lozu $R$[34].

Ako *konzistentnost otkupljivanja važi samo za komonotone lozove*, prema Vakeru i Tverskom, zadovoljen je uslov *konzistencije komonotonog otkupljivanja*. Dalje, zahtevom da svi ishodi $\alpha, \beta, \gamma, \delta$ u jednačinama (12) i (13) *budu istog znaka*, odn. ili su $\alpha, \beta, \gamma, \delta$ svi dobici ili su svi gubici, uvodimo uslov *konzistencije saoznačenog otkupljivanja*. U normativnoj teoriji očekivane korisnosti (EU) konzistencija otkupljivanja važi *za sve lozove*; u klasičnim RDU modelima koji prethode teoriji izgleda Kanemana i Tverskog ona je zadovoljena samo na komonotonim lozovima, dok u kumulativnoj teoriji izgleda ona mora da bude zadovoljena na komonotonim i saoznačenim lozovima da bi teorija mogla da se aksiomatizuje i razvije do forme koju smo predstavili još u I delu naše rasprave.[35] Vidimo da koncept konzistencije otkupa - posle netrivijalnih formalnih poteza u strukturi izgradnje teorija odlučivanja -

obezbeđuje elegantan poredak teorija, poredak u kome jednostavniji RDU modeli predstavljaju specijalan slučaj kumulativne teorije izgleda, dok je normativna EU specijalan slučaj RDU modela (pa tako i specijalan slučaj kumulativne teorije izgleda).

Osobine aksiomatskog okvira kumulativne teorije izgleda motivišu veoma važno pitanje u diskusiji odnosa intuicije i normativne adekvatnosti u debati o racionalnosti. Ako se posle prethodne diskusije aksiomatike teorije izgleda sada podsetimo fon Nojman-Morgnšternove aksiomatike teorije očekivane korisnosti, uočavamo sledeće: (a) osobina aksiomatike fon Nojmana i Morgenšterna je ta da su ključni aksiomi teorije očekivane korisnosti *direktno intuitivno prihvatljivi*, ili bar podležu direktnom testu naše intuicije u meri u kojoj možemo da tvrdimo za njih da li su očigledni ili ne, dok (b) je osobina aksiomatike teorije izgleda ta da njene aksiome igraju *čisto formalnu ulogu*, obezbeđujući inferencije koje će tek u određenom koraku razvoja teorije uzeti oblik koji može da ima sliku u ljudskoj intuiciji. Aksiom nezavisnosti, najproblematičniji deo teorije očekivane korisnosti, predstavlja tvrdnju sa kojom bi se intuitivno i lako složila većina ljudi: suočeni sa njegovim sadržajem, kao da imamo unutrašnje osećanje slaganja sa tvrdnjom da ograničenje koje on opisuje zaista odlikuje racionalnog donosioca odluka. U krajnjoj liniji, ništa što aksiom tvrdi ne predstavlja nešto što bi ljudski um odbio kao nelogično ili iracionalno na prvi pogled. Međutim, konzistentnost otkupljivanja - i njene razrade u formi komonotone i saoznačene konzistencije otkupljivanja, suštinske za razvoj teorije izgleda - nema tu osobinu *očiglednosti* koju ima aksiom nezavisnosti. Konzistentnost otkupljivanja, posle niza formalnih inferencija, vodi ka očiglednoj tvrdnji da razlike u korisnostima ishoda moraju da budu konzistentne, odn. da ne mogu da se pojave takva četiri ishoda $\alpha, \beta, \gamma, \delta$ da istovremeno $u(\alpha) - u(\beta) \geqslant u(\gamma) - u(\delta)$ i $u(\alpha) - u(\beta) < u(\gamma) - u(\delta)$ važe. Možda ovo tvrđenje predstavlja momenat blizak intuiciji u diskusiji osnova teorije izgleda, ali taj momenat očigledno *pripada rezultatima inferencija* koje su udaljene posledice aksiomatskih osnova teorije, *ne samim aksiomatskim osnovama*. Same poteškoće u uvođenju aksioma konzistentnosti otkupljivanja - počevši od potrebe da se definiše nova relacija otkupljivanja, preko neophodnosti da se definiše komonotonost lozova (koju smo, podsećamo, izneli u formalno pojednostavljenom vidu) ukazuju na to da aksiomatika teorije izgleda nije mesto gde treba tražiti korespondenciju između intuicije osobe (koja po pretpostavci nije ekspert u teoriji odlučivanja) i aksiomatskog okvira teorije.

U ovoj diskusiji odnosa intuitivne prihvatljivosti aksiomatika dve teorije

odlučivanja prepoznajemo upozorenje da je neophodno razlikovati intuitivnu prihvatljivosti aksioma teorije od intuitivne prihvatljivosti inferencija u koje te aksiome vode. Podsetimo se sada kratkog pregleda empirijskih istraživanja koja su motivisala razvoj deskriptivinih teorija - poput Aleovih rezultata i rezultata Kanemana i Tverskog - koji smo pružili u I delu naše rasprave. Kao što znamo, kritika normativne teorije očekivane korisnosti počiva na *direktnom testu njenih fundamentalnih tvrđenja* i u suštini se razvija oko neodrživosti aksioma nezavisnosti. Teoriju izgleda Kanemana i Tverskog, zbog čisto formalne uloge koju igra njen centralni aksiom komonotone i kooznačene konzistencije otkupljivanja, uopšte ne možemo da testiramo na ovaj način. Tvrdnja na prvi pogled deluje metodološki neodrživom: naravno da je takav test moguć, jer je uvek moguće dizajnirati eksperimentalnu situaciju u kojoj bismo ispitali da li strukture ponašanja, strukture opservabilnih odluka ispitanika, zadovoljavaju (ma koliko složene) formalne deskripcije koje su sadržaj tog aksioma teorije izgleda. Međutim, postavimo sebi sledeće pitanje: ukoliko bi se pokazalo da u takvom testu održivosti aksiomatike teorije izgleda ispitanici ne pokazuju ponašanje konzistentno sa njenim centralnim aksiomom, da li bi ostali iznenađeni kada bismo im ukazali na to koji uslov odlučivanja krše kao što su to sigurno bili ispitanici Morisa Alea kada su shvatili da njihove odluke ne poštuju aksiom nezavisnosti? Svakako ne. Zašto? Jednostavno, zato što aksiom konzistencije otkupljivanja ni na koji način nije formulisan, niti može da bude formulisan, jezikom koji bi se približio intuiciji bilo koje osobe koja bez prethodnog detaljnog proučavanja savremene teorije odlučivanja svakodnevno donosi odluke u uslovima rizika i neizvesnosti. Aksiom konzistencije otkupljivanja je veoma složena formalna konstrukcija teorije izgleda. Za veoma složene formalne konstrukcije mi (makar u principu) možemo da konstruišemo bihejvioralne testove, naravno: problem je u tome što možemo *uvek* da ih konstruišemo, za ma koju formalnu konstrukciju. Svaki bihejvioralni test testiru *neki* skup formalnih konstrukcija; rezultati svakog takvog testa odbacuju jedne i potvrđuju neke druge takve konstrukcije. Za trenutak bismo mogli da sliku razvoja teorija odlučivanja izokrenemo i pitamo se: zašto odgovarajuću teoriju odlučivanja ne bismo izgradili čisto induktivnim putem, planirajući seriju sistematskih bihejvioralnih eksperimenata, i postepeno odbacujući moguće formalne konstrukcije koje testovi eliminišu a zadržavajući one koje su konzistentne sa rezultatima posmatranja? Jednostavno, zato što mogućih formalnih konstrukcija koje se na neki način odnose na ljudsko donošenje odluka ima beskonačno. Pravo

pitanje je: zašto smo u razvoju deskriptivnih teorija odlučivanja posvetili toliko pažnje upravo testovima aksioma nezavisnosti? Odgovor je, naravno, zato što je upravo *intuitivna prihvatljivost* tog aksioma motivisala istraživače da se posvete proučavanju razloga zbog kojih on nije konzistentan sa rezultatima eksperimentalnih istraživanja. Ako bismo tvrdili da su status aksioma nezavisnosti teorije očekivane korisnosti i status aksioma konzistencije otkupljivanja u odnosu na postupak izgradnje naučnih teorija isti, direktno bismo odbacili upečatljivu evidenciju koju predstavlja sama istorija razvoja teorija odlučivanja u drugoj polovini XX veka.

Matematičar i logičar mogu da pokažu beskonačnu kreativnost u stvaranju novih formalnih sistema. Istorija matematike svedoči o tome da je razvoj aksiomatskih sistema koji su daleki od ma kakve ljudske intuicije pokazao da inferencije koje slede iz takvih aksiomatskih sistema mogu da budu od zadivljujuće koristi i značaja kako u matematici, tako i u empirijskim naukama. Pretpostavimo sada da ljudske empirijske odluke koje možemo da zabeležimo u nekoj eksperimentalnoj situaciji odgovaraju aksiomima neke teorije odlučivanja čija intuicija nije očigledna. Pošto se u postupku izgradnje svake teorije odlučivanja aksiomatika odnosi na strukturu ponašanja, kao što smo to već razumeli, direktno sledi tvrđenje *da ljudski kognitivni sistem proizvodi ponašanje na osnovama koje mu nisu intuitivno prihvatljive ili jasne.* Na subpersonalnom nivou analize, nivou koji odlikuje ogromnu većinu teorija i modela kompjutacione kognitivne psihologije, prethodna tvrdnja možda ne predstavlja problem jer sa na tom nivou analize ni ne zahteva to da je osoba čije kognitivne funkcije analiziramo svesna principa na kojima one počivaju. Ipak, smatramo bi ovakav stav predstavljao samo guranje problema pod tepih: u oblasti odlučivanja, svi mi (i naučnici i ispitanici) smo dobro svesni toga da je odlučivanje proces koji je svakako prisutan i na personalnom nivou analize. Na suprapersonalnom nivou analize koji smo mi predložili, aksiomatika teorije izgleda može da postane intuitivno prihvatljiva, sigurno: neophodan je samo višemesečni trening u formalno ni najmanje naivnoj oblasti savremene teorije odlučivanja da biste od nekog napravili osobu za koju je komonotona i saoznačena konzistentnost otkupljivanja intuitivno prihvatljiva; posle dovoljno vežbe i potrošenog papira, ljudski um najčešće automatizuje formalne inferencije i prestane da se pita o tome šta su one značile u početku. Ravnodušnost prema očiglednosti koja odlikuje neki aksiomatski sistem može da bude sasvim prihvatljiva za matematičara, logičara ili fizičara: psihologu ona očigledno može da napravi krupne probleme.

Tverski i Kaneman, uvodeći kumulativnu teoriju odlučivanja 1992. godine,

konstatuju da su „*Teorije izbora u najboljem slučaju aproksimativne i nepotpune.*“ (citirano prema Tversky & Kahneman, 1992, naš prevod). Ako je složenost ovog domena kognitivnih funkcija tolika, postavlja se pitanje šta sa stanovišta analize racionalnosti više opravdano: težiti formalnim, striktno aksiomatizovanim teorijama, žrtvujući pri tom intuitivne osnove teorije zarad elegantnih matematičkih inferencija, ili proučavati fenomen sa čisto deskriptivnog stanovišta bez strogih normativnih osnova, zadovoljavajući se ciljem da pružimo empirijski validan opis fenomena, a o odnosu prema normativnim zahtevima diskutovati *post hoc*? Podsetimo se, kumulativna teorija izgleda sadrži EU kao svoj specijalan slučaj, a njena aksiomatizacija, ukoliko se ma ko poziva na intuitivnost uslova konzistencije otkupljivanja, predstavlja normativnu osnovu (o čijoj se *normativnoj adekvatnosti* očigledno da raspravljati). Da li je ovo nužno za deskriptivnu teoriju odlučivanja? Da li postoje jednostavniji načini da se plati dug normativnim osnovama, ukoliko smo uopšte postigli koncenzus da teorija mora da ima određen normativni okvir?

*Ponderisanje verovatnoća.* Upotreba pondera odluke u odlučivanju smatra se jednom od empirijski najproverenijih činjenica u ovoj oblasti. Tipična slika ograničene racionalnosti izgleda ovako: ispitanici se odnose prema niskim verovatnoćama kao da su nešto više, a prema visokim verovatnoćama kao da su nešto više, pri tom tretirajući nemogućnost kao verovatnoću 0, a sigurne ishode kao da su dati sa verovatnoćom 1. Empirijska generalizacija Kanemana i Tverskog koja je omogućila deskriptivno rešenje Aleovih paradoksa i paradoksa zajedničke proporcije u prvoj verziji teorije izgleda (Kahneman & Tversky, 1979) - da ako je izbor između $(y, pq)$ i $(x, p)$ indiferentan, onda je $(y, pqr)$ preferirano u odnosu na $(x, pr)$, $0 < p, q, r < 1$ - omogućila je otkriće kvalitativnog empirijskog oblika funkcije ponderisanja verovatnoća, u oznaci $w(\cdot)$ (v. Prelec, 1998, za sistematsku analizu konstrukcije ove funkcije); ovaj oblik je prikazan na Slici 4b. još u našim prvim diskusijama odlučivanja. Parametarski oblik funkcije ponderisanja verovatnoće *nije osobina* koja sledi iz aksiomatskog okvira kumulativne teorije izgleda. U literaturi je predloženo nekoliko funkcionalnih, parametarskih formi, od kojih je najpopularnija Prelecova jednoparametarska forma (poznata kao *compound-invariace form*, zbog jedne formalne osobine na kojoj se zasniva i koju nećemo ovde diskutovati). Što se samog oblika funkcije tiče, „inverznog-S“ oblika u kome konveksni region smenjuje konkavni na višim verovatnoćama, sa $w(0) = 0$ i $w(1) = 1$, već smo primetili koliko je neuobičajen i zahteva da bude opravdan jakim empirijskim razlozima. Aksiomatski okvir kumulativne teorije izgleda omogućava samo inferenciju ka nekoj funkciji $w(\cdot)$

koja transformiše verovatnoće, koja je striktno rastuća na domenu (0, 1) i za koju važi $w(0) = 0$ i $w(1) = 1$; lako je uočiti da funkcija na Slici 4b. zadovoljava aksiomatska ograničenja teorije izgleda. U odnosu na prethodnu diskusiju korisnosti kao centralnog koncepta teorije odlučivanja, sada je potrebno da dodamo da u teoriji izgleda Tverskog i Kanemana ona ulazi u komplikovane interakcije sa funkcijom ponderisanja verovatnoće, interakcije koje objašnjavaju bihejvioralna odstupanja od normativnog okvira teorije očekivane korisnosti.

Empirijska istraživanja metodom ocene monetarnih ekvivalenata rizičnih lozova, u kojoj od ispitanika tražimo da ponudi minimalnu cenu po kojoj bi prodao loz oblika ($x$,$p$;$y$,1-$p$) potvrdila su oblik funkcije ponderisanja verovatnoće. Iako ne bez kritika, koje ćemo izneti u V delu ove teze, ova istraživanja upućuju na to da je ovaj čudan oblik funkcije ponderisanja verovatnoće dobra slika empirijske realnosti (Tversky & Kahneman, 1992, Gonzales & Wu, 1999). Posebno snažan empirijski dokaz za ovo dala je studija Gonzalesa i Vua iz 1999. godine u kojoj su funkcija korisnosti i funkcija ponderisanja verovatnoća izvedene za 10 ispitanika neparametrijskom metodom, dakle ne pretpostavljajući određene funkcionalne forme za ove funkcije. Kvalitativno, Gonzales i Vu su pokazali da ispitanici u oceni monetarnih ekvivalenata koriste konkavnu funkciju korisnosti i inverznu-S funkciju ponderisanja verovatnoće koje dobro odgovaraju poznatim funkcionalnim formama. Funkciju sa kvalitativno istim osobinama empirijski izvode Tverski i Foks u odlučivanju u uslovima neizvesnosti (Tversky & Fox, 1995).

Ponderisanje verovatnoća se smatra za jedan od osnovnih izvora odstupanja od normativnih standarda u teoriji odlučivanja. Međutim, pokazuje se da postoji mehanizam koji omogućava da se ponderisanje verovatnoća tumači kao proizvod racionalnih principa, i to principa bejzijanske inferencije. Taj mehanizam bejzijanske inferencije deo je Viskuzijeve teorije *teorije perspektivne reference* (engl. *Prospective Reference Theory*, Viscusi, 1989). Viskuzi je došao do zaključka da ukoliko pretpostavimo da u odlučivanju u uslovima rizika ljudi polaze od nekih *a priori* verovantoća koje se odnose na njihovo iskustvo u osvajanju ishoda koji se nalaze na lozu, onda se ponderisanje verovatnoća, posle primene odgovarajućeg bejzijanskog modela, prirodno javlja u evaluaciji lozova. Na primer, suočeni sa 15% šansi da osvoje 30 evra i 85% šansi da osvoje 5 evra, ljudi mogu da iskoriste svoja prethodna verovanja o tome kakve su mogućnosti za osvajanje 30 ili 5 evra, i da kroz proces bejzijanske inferencije modifikuju verovatnoće date na lozu u *a posteriori* verovatnoće koje će stvarno iskoristiti u evaluaciji vrednosti ovog loza.

Ponderisanje verovatnoća se u ovoj teoriji otkriva na empirijski poznat način: niske verovantoće su precenje, a visoke potcenjene. Viskuzijev model odlučivanja je u stanju da objasni ponderisanje verovatnoća, empirijsku generalizaciju Kanemana i Tverskog, pa tako i Aleov paradoks, čak i ako se pretpostavi da su sve *a priori* verovatnoće jednake i iznose $\frac{1}{n}$, gde je $n$ broj ishoda na lozu; u slučaju loza sa dva ishoda, dakle, ispitanik bi pretpostavio da su oni *a priori* podjednako verovatni, da bi onda kroz bejzijansku inferenciju modifikovao verovatnoće na lozu donekle podižući niže i ponešto spuštajući više. Adekvatan bejzijanski model u slučaju lozova, koje posmatramo kao diskretne distribucije verovatnoće koje su univarijante (u slučaju loza sa samo dva ishoda, gde su verovatnoće $p$ i 1-$p$) ili multivarijantne (u slučaju lozova sa više od dva ishoda) jeste osnovni i najjednostavniji model bejzijanske inferencije koji koristi binomijalne i multinomijalne (u multivarijantnom slučaju) distribucije za opis verodostojnosti, a njima konjugovane Beta i Dirišleove (u multivarijantnom slučaju) distribucije za opis verovatnoća *a priori* i *a posteriori*. Funkcije ponderisanja verovatnoća koje slede iz ovog modela su *linearne*, precenjuju niske a potcenjuju visoke verovatnoće, a definicijom je neophodno uvesti da je $w(0) = 0$ i $w(1) = 1$. Viskuzijev model nećemo sada formalno predstavljati jer će on biti osnova za razvoj jednog novog modela odlučivanja u V delu ove teze; tada ćemo ga proučiti u detaljima. Ova teorija je primila dve snažne empirijske kritike: jedan je rad Dejvida Harlesa iz 1993. u kome je direktno testirana Viskuzijeva teorija (Harless, 1993), a drugi je istraživanje više modela odlučivanja u eksperimentu izbora Blavackog (Blavatskyy, 2011) u kome je teorija perspektivne reference uvek bila dominirana od strane drugih modela. Ipak, obe empirijske kritike odnose se na formu teorije u kojoj sve a priori verovatnoće imaju vrednost $\frac{1}{n}$ za lozove sa $n$ ishoda, pretpostavke za koju ćemo u V delu pokazati da nije realna.

Konačno, neki najnoviji radovi pokazuju da postoji i još jedna mogućnost zbog koje se javlja ponderisanje verovanoća u odlučivanju. Stjuart, Čater i Braun u razvoju teorije odluke putem semplovanja (engl. *Decision By Sampling Theory*, Stewart, Chater & Brown, 2006, 2009), pokazuju da forma funkcije ponderisanja verovatnoće može da bude posledica distribucije odgovarajućih vrednosti i verovatnoća u ljudskom pamćenju, koje po pretpostavci odslikava realnu strukturu okoline. U objašnjenju za razvoj inverzne-S funkcije ponderisanja verovatnoća oni pokazuju da distribucija jezičkih formi koje se koriste za karakterizaciju različitih verovatnoća ima tu osobinu da najmanje ima izraza za verovatnoće u srednjem rasponu, a najviše za niske i visoke verovatnoće. Kada se izračuna relativni rang

verovatnoća sa ovakve distribucije dobija se funkcija koja u odnosu na objektivnu verovatnoću ima nešto više relativne rangove od očekivanih za male i nešto niže relativne rangove od očekivanih za visoke verovatnoće. Za našu diskusiju je značajno to što ovaj rad predstavlja još jednu ideju o tome da je ponderisanje verovatnoća proizvod racionalne inferencije koja polazi od podataka o realnoj strukturi okoline. Viskuzijeva teorija ne uključuje nikakve pretpostavke o distribuciji verovatnoća ili ishoda u ekonomskom okruženju donosica odluka; možda je kombinacija intuicije o ponderisanju verovatnoća kao posledici realne strukture okruženja i racionalne bejzijanske inferencije u stanju da objasni ograničenu racionalnost u ovom domenu sa normativnog stanovišta? Na ovo pitanje ćemo odgovor dati u V delu ove teze.

*Averzija prema gubicima.* Prema Vakeru, averzija prema gubicima, kao jedan od fenomena zavisnosti odluka od referentne tačke, predstavlja najgrublje odstupanje od normativne racionalnosti u odlučivanju (Wakker, 2010). U odnosu na *status quo* koji u većini primena teorije izgleda definišemo kao 0, svi ishodi ispod referentne tačke su gubici, dok su svi izgledi iznad referentne tačke dobici. U racionalnoj teoriji, kao što smo videli, nema osnove za tvrdnju da se odnos prema ishodima sa jedne i druge strane referentne tačke razlikuje; empirijski, Kaneman i Tverski su pokazali da indeks averzije prema gubicima ima srednju vrednost *oko dva*: gubici „bole“ oko dva puta više nego što dobici „raduju“ (Tversky & Kahneman, 1992). Već smo diskutovali promenu oblika funkcije korisnosti (Slika 5.) koju ovaj empirijski nalaz diskutuje: ona je konkavna za dobitke, konveksna za gubitke, i *strmija* za gubitke nego za dobitke. Ovo poslednje je posledica averzije prema gubicima; konkavnost u domenu dobitaka i konveksnost u domenu gubitaka su posledica diskutovanih efekata refleksije. Prvu teorijsku analizu averzije prema gubicima pružili su Tverski i Kaneman (Tversky & Kahneman, 1991); fenomen je inkorporiran u aksiomatsku strukturu kumulativne teorije izgleda (Wakker & Tversky, 1993), a od tada je teorijski diskutovana više puta, u odnosu na probleme merenja (npr. Köbberling & Wakker, 2005) kao i u odnosu prema određenim parametarskim formama modela teorije izgleda (Bradley, al-Nowaihi & Dhami, 2008).

Da li je averzija prema gubicima empirijski stabilan fenomen? Vaker na osnovu rezultata pregleda većeg broja empirijskih radova ustanovljava da vrednost indeksa averzije prema gubicima od oko dva nije stabilna, a da se u nekim studijama javlja čak obrnut fenomen od averzije prema gubicima, *sklonost ka dobicima* - u kome je vrednost indeksa „averzije prema gubicima“ između 0 i 1 (Wakker, 2010). Poznato je da se averzija prema gubicima javlja i u uslovima odlučivanja bez rizika, što najbolje

ilustruje *efekat zaduživanja* (engl. *endowment effect*). Efekat zaduživanja se ogleda u činjenici da se cena koju su ljudi spremni da ponude kao minimalnu po kojoj bi prodali nešto *što je već u njihovom vlasništvu* razlikuje od iznosa koji su spremni da ponude kao maksimalan za koji bi kupili tu istu stvar: minimalna cena prodaje je veća od maksimalnog iznosa koji bi ponudili za istu stvar (Kahneman, Knetsch, & Thaler, 1990). Experimental Test of the endowment effect and the Coase Theorem. Journal of Political Economy 98(6), 1325-1348.). U empirijskoj studiji poređenja averzije prema gubicima u kontekstu odlučivanje bez rizika i kontekstu odlučivanja u uslovima rizika, Gahter, Džonson i Herman su našli pozitivnu i značajnu korelaciju između ocenjene averzije prema gubicima u dva konteksta od .635 (Gachter, Johnson & Herrmann, 2007). Međutim, vrednosti ocenjene averzije prema gubicima kroz ispitanike su značajno varirale: prosek je u ovoj studiji iznosio 2.62, sa standardnom devijacijom od 2.28 i u aproksimativnom rasponu od 1 do 4 (!). Studija Sokol-Hesnera i saradnika takođe nalazi veliki raspon averzije prema gubicima: od njihovih 30 ispitanika, 9 pokazuju sklonost ka dobicima (averziju prema gubicima između 0 i 1), 7 su neutralni u odnosu na dobitke i gubitke, a 14 pokazuju tipičnu averziju prema gubicima. Međutim, još zanimljiviji nalaz ove studije je da je na stepen averzije prema gubicima moguće uticati eksperimentalnim instrukcijama da se u odlučivanju koriste različite kognitivne strategije. U metodologiji ove studije, rizični lozovi se uvek porede sa sigurnim ishodima - ispitanik bira da li želi da odigra rizičan loz, ili uzme ponuđen siguran iznos. Jedna strategija koju Sokol-Hesner i saradnici instrukcijama indukuju kod svojih ispitanika upućuje ih na to da svaki loz posmatraju kao izolovan događaj, kao da je samo on važan (tzv. „*attend*“ strategija); druga instrukcija upućuje ispitanike na to da razmišljaju o svim lozovima koje vide kao delovima portfolija, odn. da se ponašaju kao da je bitno koliko zarađuju ukupno na celoj seriji lozova, a ne na svakom posebno (tzv. „*regulate*“ strategija). Sokol-Hesner i saradnici pokazuju da je u drugoj strategiji, gde ispitanici posmatraju lozove kao seriju izbora koji čine portfolio, indeks averzije prema gubicima značajno niži u odnosu na prvu strategiju, gde se svaki loz posmatra posebno, kao da je odluka na njemu najznačajnija koja može da se donese u eksperimentu (Sokol-Hessner et al, 2009).

Studija Plota i Zilerove koristi metodologiju određivanja spremnosti da se cena prihvati i spremnosti da se cena plati baziranu na dobro poznatoj BDM proceduri u eksperimentalnoj ekonomiji (Becker, De Groot & Marschak, 1964) i uključuje trening ispitanika o efektu zaduživanja (Plott & Zeiler, 2005). Pod ovakvom

metodologijom, Plot i Zilerova nisu našli nikakve dokaze za postojanje averzije prema gubicima. Trening ispitanika uveden je upravo zbog sumnje autora da averzija prema gubicima nije fundamentalan empirijski fenomen, već posledica miskoncepcija o odlučivanju koje bi ispitanici mogli da imaju u manje kontrolisanim eksperimentalnim procedurama. Status ovog empirijskog nalaza je nesiguran: Ajsoni, Lums i Sagden su samo parcijalno uspeli da repliciraju rezultate Plota i Zilerove po istom nacrtu, a u reanalizi njihovog nacrta pokazali da Plot i Zilerova nisu analizirali sve situacije u kojima je do diskrepance između spremnosti da se cena plati i da se cena prihvati moglo da dođe (Isoni, Loomes & Sugden, u štampi).

Osim empirijske nestabilnosti, averzija prema gubicima sa sobom unosi još neke probleme u teoriju izgleda. Koberlingova i Vaker su razvili jednostavnu demonstraciju paradoksa koji se javlja kada se teorija izgleda postavi u njenoj najpopularnijoj formi (i ujedno jedinoj formi u kojoj je empirijski testirana, isključujući neparametrijske testove). Ta forma podrazumeva jednačine (3) - (8). Sa stepenom funkcijom korisnosti, za koju Tverski i Kaneman dopuštaju da ima različite eksponente za dobitke i gubitke (Tversky & Kahneman, 1992), dobijamo posledicu da se dve funkcije korisnosti, za dobitke i gubitke, uvek seku u tački (1,1), ako ih posmatramo u istom (I) kvadrantu, odn. ako diskutujemo samo apsolutne vrednosti dobitaka i gubitaka. Ovo je direktna posledica prirode stepene funkcije koja, bez obzira na vrednost eksponenta, uvek ima vrednost 1 za $1^{\rho}$. Iz ovoga sledi da za dve stepene funkcije korisnosti uvek postoji region u kome je prva iznad druge, i region u kome je druga iznad prve, što dalje znači da postoji region u kome apsolutne vrednosti dobitaka dominiraju nad apsolutnim vrednostima gubitaka, što je protivno ideji o averziji prema gubicima na celom domenu funkcije korisnosti. Uvođenje indeksa averzije prema gubicima, kao u jednačini (8), ne rešava problem, jer indeks sada jednostavno reskalira funkciju korisnosti za gubitke tako da se samo tačka preseka dve funkcije menja kada se one posmatraju u I kvadrantu (Köbberling & Wakker, 2005). U V delu ove teze pokazaćemo da u konjukciji sa novim teorijskim uvidima u strukturu teorije izgleda (Bradley, al-Nowaihi & Dhami, 2008) iz ovog problema koji su primetili Koberlingova i Vaker sledi još dublji paradoks u strukturi ove teorije.

Konačno, da li je averzija prema gubicima, koliko god bila empirijski nestabilan i teorijski nezgodan konstrukt, svedočanstvo o odstupanju od racionalnosti u odlučivanju u uslovima rizika i neizvesnosti? Razmišljamo na sledeći način: neka je naša referentna tačka neko $S_0$ - to je, recimo, količina novca koju sada posedujemo.

Ukoliko u rizičnim ekonomskim interakcijama u sledećem koraku steknemo vrednost $S_1$, naša konačna pozicija je neko $S_2$=$S_0$+$S_1$; ukoliko u toj interakciji izgubimo vrednost $S_1$, naša konačna pozicija je neko $S'_2$=$S_0$-$S_1$. Ako smo izgubili novac kroz ekonomsku interakciju u kojoj smo mogli i da zaradimo i da izgubimo, naša nova pozicija $S'_2$ diktira da sada moramo da zaradimo novac da bismo se vratili tek na $S_0$ - iako smo imali priliku da budemo na finalnoj poziciji $S_2$. Da bismo stigli do, po pretpostavci željene, pozicije $S_2$, sada je potrebno da *uložimo nešto u nove ekonomske interakcije*, ili, *da podnesemo dodatan rizik*, kako bismo poboljšali svoju poziciju. Iz ove jednostavne logike koja uzima u obzir to da ni ekonomske interakcije nisu besplatne u realnom okruženju, već zahtevaju izvesno ulaganje da bi se obavljale, ne sledi da je averzija prema gubicima iracionalan fenomen. Kognitivni sistem lavice koja u savani donosi odluku o tome za kojom zebrom treba da se da u trk sigurno uzima u obzir i procenu toga koliko će energije potrošiti na lov sa neizvesnim ishodom - jer, ukoliko ne ulovi zebru na koju je usresredila pažnju, moraće da lovi drugu, iako je već izgubila investeciju u energiji loveći prvu[36]. Ova pretpostavka se smatra potpuno samorazumevajuća u analizama bihejvioralne ekologije (Mangel & Clark, 1988). Ako je neka zebra veoma brza, i u sistemu odlučivanja lavice reprezentovana kao potencijalan gubitak, sledi da je za lavicu racionalno da njen sistem odlučivanja uključi parametar koji će je dodatno odvraćati od ulaganja. Za donosioca odluka koji zna da mora da uloži određene resurse da bi popravljao svoju poziciju na skali korisnosti, averzija prema gubicima je racionalan izraz pokušaja da se izbegne porast investicija koje neće moći da doprinesu pozitivnom priraštaju ukupne korisnosti. Naše mišljenje je da je averzija prema gubicima teorijski konstrukt deskriptivnih teorija koji svedoči o odstupanju od racionalnosti samo ako se za normativnu racionalnost proglasi isključivo matematička konstrukcija teorije očekivane korisnosti. Da li je averzija prema gubicima racionalna ili ne, zavisiće, dakle, od teorijskog, interpretativnog okvira koji odaberemo. Ako su sve ekonomske interakcije uvek besplatne, ona nije racionalna; ukoliko one imaju cenu, a teško je zamisliti realnu ekonomsku interakciju koja bi bila besplatna, averzija prema gubicima onda može da odlikuje i racionalnog donosioca odluka.

*Heuristike u odlučivanju i pitanje proceduralne invarijantnosti.* Heuristike smo već definisali u inventaru naših metateorijskih koncepata za analizu racionalnosti saznanja. Najnovija literatura u oblasti odlučivanja u uslovima rizika pokazuje da je moguće konstruisati heuristike zapanjujuće jednostavnosti koje objašnjavaju veliki

broj robusnih empirijskih nalaza odstupanja od normativnih standarda. Heuristike, za razliku od ideje bejzijanske inferencije, ili odlučivanja na osnovu prethodno uskladištenih informacija o relevantnoj strukturi okoline, ne predstavljaju strategije racionalizacije ograničene racionalnosti. Heuristike, same po sebi, nisu normativno racionalne; u Gigerencerovom shvatanju, koje smo već izložili, njih odlikuje ekološka racionalnost, ali je ova tvrdnja karakteristična samo za Gigerencerov istraživački program (Gigerenzer, 2008). Razmotrićemo jednu heuristiku čija je popularnost danas naročita, a koji je poreklom upravo iz Gigerencerovog skupa adaptivnih alatki kognitivnog sistema: *heuristiku prioriteta* (Brandstätter, Gigerenzer & Hertwig, 2006).

Heuristika prioriteta se odnosi na lozove sa dva ishoda oblika ($x$,$p$;$y$,1-$p$). Ime nosi po strategiji koju ilustruje, a koja *prioritet stavlja na to da se izbegne dobitak manjeg od dva ishoda.* Heuristika prioriteta uvodi redosled prioriteta i operacija poređenja karakterističnih za odluke na svakom prioritetu koji se analizira dok se ne donese konačna odluka. Taj redosled prioriteta izgleda ovako: (1) minimalni dobitak, (2) verovatnoća minimalnog dobitka, (3) maksimalni dobitak. Redosled poređenja i odgovarajuća odluka na svakom prioritetu (1) - (3) izgleda ovako: (1) ako se minimalni dobitak razlikuje za više od 1/10 od maksimalnog dobitka na lozu, odaberi loz sa maksimalnim dobitkom; inače, (2), ako se verovatnoće minimalnog i maksimalnog dobitka razlikuju za više od 1/10 skale (skala verovatnoće je od 0 do 1), odaberi loz sa povoljnijom verovatnoćom; inače, (3) odaberi loz sa maksimalnim dobitkom. Gigerencer i saradnici naglašavaju „heurističku prirodu“ ove heuristike: on se koristi samo u „lakim“ problemima odlučivanja, samo u slučajevima jednostavnih lozova tipa ($x$,$p$;$y$,1-$p$) , oslanja se na jednostavne i brze operacije poređenja, i u suštini predstavlja jednu logičnu strategiju: pokušaj analize da li je moguće izbeći zaradu manjeg od dva iznosa na lozu. Fascinantna je empirijska performansa ovako jednostavne teorije. Heuristika prioriteta može da objasni sledeća robusna empirijska odstupanja od normativne racionalnosti: Aleov paradoks, četvoročanu strukturu stavova prema riziku (koja uključuje averziju prema riziku za srednje i visoke verovatnoće gubitaka i sklonost ka riziku za niske verovatnoće dobitaka; sklonost ka riziku za srednje i visoke verovatnoće gubitaka i averziju prema riziku za niske verovatnoće gubitaka, Tversky & Kahneman, 1992), efekte izvesnosti i sigurnosti (Kahneman & Tversky, 1979) i pojave intranzitivnih preferencija među lozovima. Drugim rečima, praktično svi empirijski nalazi koji su motivisali razvoj komplikovanih deskriptivnih teorija odlučivanja u drugoj polovini XX veka mogu da

budu (makar na domenu jednostavnih lozova, koji su uostalom najčešće korišćeni u eksperimentalnoj praksi) objašnjeni primenom jednostavnih pravila odlučivanja koje obuhvata heuristika prioriteta.

U jednoj novijoj studiji, Blavacki je vršio selekciju modela teorija odlučivanja koristeći specijalno dizajniran eksperiment izbora i poseban ekonometrijski model kojim je kontrolisao određene probleme u selekciji modela koje prethodne studije nisu uspele da kontrolišu (Blavatskyy, 2011). U eksperimentima izbora (engl. *choice experiments*) od ispitanika tražimo da izaberu jedan od dva ponuđena loza koji bi pre odigrali; ovo je verovatno najjednostavnija eksperimentalna procedura u celoj oblasti (ali je zato veoma zahtevna sa stanovišta statističke selekcije adekvatnog modela odlučivanja). Između većeg broja teorija odlučivanja koje je uključio u studiju, Blavacki je modelirao i jednu jednostavnu heuristiku (heuristika prioriteta u punoj formi, nažalost, nije mogla da se primeni na njegove stimuluse). Ta heuristika koristi sledeća pravila za donošenje odluke: (1) izaberi loz koji nosi manju verovatnoću najmanjeg ponuđenog dobitka; ako je ta verovatnoća ista na oba loza, (2) izaberi loz koji nosi veću verovatnoću najvećeg ponuđenog dobitka. Svi lozovi Blavackog su pozitivni ili ne-negativni i sadrže samo ishode {0 *EUR*, 5 *EUR*, 20 *EUR*, 25 *EUR*, 40 *EUR*}. Rezultati Blavackog pokazuju da su izbori nešto preko 25% od njegovih 38 ispitanika najbolje objašnjeni primenom ove jednostavne heuristike - heuristike koju je Blavacki predložio za potrebe ove empirijske studije, bez prethodne teorijske elaboracije kakvu pružaju Gigerencer i saradnici za heuristik prioriteta (Brandstätter, Gigerenzer & Hertwig, 2006).

*Pitanje proceduralne invarijatnosti.* Jedno od robustnih odstupanja od normativne teorije predstavlja dobro potkrepljen empirijski nalaz *zamene preferencija* (engl. *preference reversals*). Ako za neka dva loza, $P$ i $Q$, tražimo od ispitanika da odrede minimalne cene po kojima bi ih prodali, odn. da odrede njihove monetarne ekvivalente, a zatim postavimo pred njih izbor između ta dva loza $P$ i $Q$, često se dešava da izbori nisu konzistentni sa odnosom monetarnih ekvivalenata kakvi su dobijeni od istih ispitanika (Cox, 2008, Tversky, Slovic & Kahneman, 1990). Koks daje opis paradigmatične eksperimentalne situacije: postoje dva loza, P-loz sa relativno visokom verovatnoćom malog dobitka, i V-loz sa relativno niskom verovatnoćom velikog dobitka. Tipična slika ograničene racionalnosti u ovakvoj paradigmi se sastoji u (a) određivanju višeg monetarnog ekvivalenta za V-loz, i (b) češćem izboru P-loza kada se u paru poredi sa V-lozom (Cox, 2008). Racionalna teorija izbora ne pravi razliku između cene loza i njegove

vrednosti u izboru: one moraju da budu konzistentne. Veoma značajna činjenica u analizi racionalnosti saznanja u ovoj oblasti je ta da *nije razvijena nijedna teorija odlučivanja koja u jedinstvenom okviru može da objasni ovaj fenomen* (ali ćemo uskoro videti da neke verzije stohastičke EU mogu da reše ovaj problem, Blavatskyy, 2006). Standardno psihološko objašnjenje za ovaj fenomen počiva na *hipotezi o kompatibilnosti* (Slovic, Griffin & Tversky, 1990, Tversky & Thaler, 1990): loz sa visokom vrednošću u kognitivnoj reprezentaciji problema korespondira sa skalom vrednosti, odn. skalom na kojoj se izražavaju monetarni ekvivalenti, dok loz sa visokom verovatnoćom u zadatku izbora korespondira sa skalom verovatnoće koja je relevantna u izboru (gde se onda, po pretpostavci, postavlja pitanje o tome koji će od dva loza *verovatnije* doneti zaradu). Stimulus koji ostvari višu korespondenciju na odgovarajućoj skali, po ovoj hipotezi, dobija viši ponder u odlučivanju; ovo je za sada jedino teorijsko objašnjenje fenomena zamene preferencija koji predstavlja očigledno kršenje normativnog standarda.

*Stohatističke teorije odlučivanja.* Sve ovde do sada razmatrane teorije odlučivanja su *determinističke*. Teorija očekivane korisnosti, podjednako kao i kumulativna teorija izgleda, računaju očekivane korisnosti rizičnih lozova. Kada su te očekivane korisnosti izračunate, izbor se donosi na osnovu toga koji loz ima višu očekivanu korisnost. Jasno je da je ljudsko ponašanje retko kada determinističko, a empirijske analize u odlučivanju u uslovima rizika samo potvrđuju tu intuiciju: Hej i Orm nalaze da je samo oko 3/4 odluka u eksperimentima izbora konstantno pri ponovljenim merenjima kroz ispite ispitanike (Hey & Orme, 1994). Ovakav nalaz jasno motiviše pokušaj da se u determinističke teorije odlučivanja unese slučajan odn. stohastički element; teorijsko pitanje je, naravno, kako i na kom nivou teorije odlučivanja?

Blavacki (Blavatskyy, 2006) i Lums, Mofat i Sagden (Loomes, Moffatt & Sugden, 2002) diskutuju ovaj problem i nalaze da su moguće tri strategije njegovog rešavanja. Prva strategija je da se pretpostavi da su preferencije same po sebi stohastičke prirode, odn. da postoji određena distribucija verovatnoća koja utiče na izbor između lozova čija se očekivana korisnost izračunava sledeći neku teoriju odlučivanja. Ovo je pokušaj da se stohastički element ugradi veoma duboko u strukturu teorije odlučivanja: ako neki model odlučivanja odlikuje određenu osobu nekim skupom parametara, npr. $[\rho, \gamma, \lambda]$ za (redom) stepen averzije prema riziku, parametar ponderisanja verovatnoća i averziju prema gubicima, onda ova ideja podrazumeva postojanje distribucija verovatnoća nad ovim parametrima za svakog

ispitanika, tako da je izračunavanje očekivane korisnosti *inherentno* stohastički proces. Ovaj pristup su u novijim istraživanjima aktualizovali Lums i Sagden (Loomes & Sugden, 1995, prema Loomes, Moffatt & Sugden, 2002), i on se naziva *modelom slučajnih preferencija*. Drugi pristup inkorporaciji stohastičkog elementa u odlučivanje je da se pretpostavi da se element slučajnosti javlja tek u fazi izračunavanja odluke, tj. da neka individua ima stabilne, „prave" preferencije između lozova $P$ i $Q$, ali u izračunavanju razlike između njihovih očekivanih korisnosti neki stohastički proces unosi grešku, tako da je razlika između U($P$) i U($Q$) dopunjena nekim terminom greške: U($P$)-U($Q$)+$\varepsilon$. Ovaj model se još naziva i *fehnerijanskim modelom odlučivanja*. Pretpostavke o distribuciji slučajne greške $\varepsilon$ su onda od ključnog značaja za empirijske performanse ovakve teorije. Konačno, treća mogućnost je da se slučajna greška javlja tek u fazi akcije, odn. davanja odgovora, tj. da nema suštinske veze sa odlučivanjem. Ovakav pristup se razvija inkorporiranjem konstantne slučajne greške u sve odluke ispitanika čiji se izbori analiziraju, karakterističan je za neke od prvih studija stohastičkih elemenata u odlučivanju (npr. Camerer & Harless, 1994), i danas se praktično više ne razmatra.

Ono što je od ključnog značaja za analizu racionalnosti saznanja je sledeće: inkorporacijom stohastičkih elemenata u normativnu teoriju odlučivanja ona dobija empirijske osobine koje korespondiraju nekim od nalaza o ograničenoj racionalnosti (Blavatskyy, 2006). Blavacki, npr. sistematski proširuje teoriju očekivane korisnosti sa samo dve (relativno intuitivno prihvatljive i teorijski zasnovane) stohastičke pretpostavke i pokazuje da ona u tom obliku može da objasni Aleov paradoks, četvoročlanu strukturu stavova prema riziku i *čak efekte zamene preferencija*, što smo prethodno diskutovali kao veoma težak problem za klasične (determinističke) teorije odlučivanja. Ne treba gubiti iz vida bitnu sledeću metodološku prepreku kada se diskutuju empirijske performanse teorija odlučivanja: nijedna teorija odlučivanja ne može da bude ocenjena u eksperimentu izbora između lozova *ako se prethodno ne proširi stohastičkim elementom*. Uopšte, ocena parametara neke teorije odlučivanja iz binarnih odluka ispitanika izvodi se isključivo na taj način: teorija odlučivanja koja se ocenjuje se tretira kao *centralni model* oko kojeg se razvija *stohastički model*. Tek razvoj stohastičkog modela onda objašnjava kako (centralna) deterministička teorija može da odgovara i podacima ispitanika koji nije idealno konzistentan u svojim odlukama. Dakle, u analizi eksperimenata izbora, neizbežno srećemo problem *konfundacije teorije odlučivanja sa teorijom greške* koja joj se pridaje da bi ona mogla da objasni realne podatke koji sadrže šum (pri čemu je „šum" zaista šum samo

za determinističke teorije; u stohastičkim teorijama, on je inherentan deo procesa odlučivanja).

Uvođenjem stohastičkog elementa u determinističke teorije odlučivanja otvara se sledeći, nimalo jednostavan problem: kako statistički, na osnovu bihejvioralnih podataka, razdvojiti efekte „jezgra teorije“ od efekata koji su posledica njenog stohastičkog elementa? Teorije odlučivanja se baziraju na teoriji merenja koja nema nikakve veze sa tzv. „klasičnim modelom“ na kome počiva uobičajena psihometrija; u oblasti odlučivanja, jedina teorija merenja koja uopšte omogućava smislenu analizu je tzv. *združeno aditivno merenj*e (engl. *Additive Conjoint Measurement,* Luce & Tukey, 1964), koju zbog njene složenosti izbegavamo da uključimo u diskusije u ovom radu. Preostaje samo da nam se veruje na reč da ni u okviru ove teorije merenja, koja se smatra za pravu revoluciju u teoriji merenja još od Euklida (Michell, 1999), nije moguće dati generalno rešenje za problem koji postavljaju kombinacije determinističkih i stohastičkih teorija odlučivanja. Stohastičke teorije odlučivanja, kao i „pakovanje“ determinističkih teorija u stohastičke modele, u stvari predstavlja kombinovanje „dve filozofije“, jedne iz združene aditivne teorije merenja na kojoj počivaju inferencije o funkcijama korisnosti i drugim konstruktima samih teorija odlučivanja, i druga iz „klasične teorije“ merenja (iz koje potiču psiholozima poznatiji pojmovi distribucije greške ili jedinstvenog skora). Zato je problem koji postavlja razvoj stohastičkih teorija odlučivanja veoma težak, a analiza posledica njihovog razvoja još uvek nejasna.

*Problem selekcije modela odlučivanja.* Selekcija modela odlučivanja odgovarajućim statističkim procedurama takođe je odlika novijih razvoja u oblasti odlučivanja. Prve studije koje su pokušale da odrede koji je „pravi“ bihejvioralni model odlučivanja pojavile su se 90-ih godina; i danas su ovakvi pokušaji malobrojni. Generalno, opšti zaključak svih studija bio bi da čak ni dominacija modela korisnosti zavisnih od ranga (RDU), poput teorije izgleda, nad teorijom očekivane korisnosti, nije siguran nalaz, uprkos poznatoj osobi RDU modela da lako inkorporiraju robustna odstupanja od normativnog okvira. Studija Heja i Orma poredi deset modela odlučivanja u uslovima rizika, uključujući teoriju očekivane korisnosti i RDU model (koji na pozitivnim i ne-negativnim lozovima ima istu formu kao teorija izgleda), pod fehnerijanskim modelom inkorporacije stohastičkog elementa u odlučivanje, i dolazi do zaključka da je nemoguće ustanoviti dominaciju alternativnih teorija odlučivanja nad teorijom očekivane korisnosti u objašnjenju empirijskih izbora njihovih 40 ispitanika (Hey & Orme, 1994). Studija Lumsa, Mofeta i Sagdena

dolazi do zaključka da RDU model, koji je kombinovan sa (*i*) modelom slučajnih preferencija (dakle, sadrži inherentno stohastički element u odlučivanju) i (*ii*) sa modelom konstatne greške, pokazuje bolje empirijske performanse od očekivane korisnosti u stohastičkom ruhu u dva eksperimenta izbora sa po 90 parova lozova (Loomes, Moffatt & Sugden, 2002). Nećemo na ovom mestu ulaziti u diskusiju šta je to tačna interpretacija modela odlučivanja koji kombinuje RDU sa dva različita mehanizma inkorporacije stohastičkog elementa u odlučivanje. Konačno, najnovija u ovom pregledu je studija Blavackog, u kojoj je iskorišćen originalan ekonometrijski model inkorporacije stohastičkog elementa u odlučivanje, koji najbolje možemo da opišemo kao kombinaciju fehnerijanskog modela sa dopunskim pretpostavkama o „razumnoj" distribuciji grešaka. Blavacki dolazi do sledećih rezultata: 10 od 38 njegovih ispitanika najbolje opisuje jednostavna heuristika koji smo već diskutovali; očekivana korisnost i RDU (teorija izgleda) najbolje opisuju odluke oko 1/4 ispitanika svaka; ostali ispitanici se raspodeljuju na druge modele odlučivanja (u veoma niskim procentima) ili su podjednako dobro objašnjeni modelima *više teorija istovremeno*. Ipak, do ubedljivo najfascinantnijeg rezultata u selekciji modela odlučivanja dolaze Harison i Rutstromova (Harrison & Rutström, 2008). Ovi istraživači kritikuju pretpostavku koja je zajednička svim pokušajima selekcije modela odlučivanja uopšte, a to je da je neki ispitanik u eksperimentu izbora ujedno i *reprezentativni subjekat neke teorije odlučivanja*. Različiti ispitanici mogu da se ponašaju različito u izboru: neki mogu da ponderišu verovatnoće više ili manje, u kom slučaju su oni reprezentativni subjekti teorije izgleda, dok neki mogu da preslikaju verovatnoće 1:1 u pondere odluke, u kom slučaju su oni reprezentativni subjekti teorije očekivane korisnosti. Harison i Rutstromova razvijaju statističku metodologiju koja im omogućava da *simultano* modeliraju odgovore ispitanike pod obe teorije i odrede proporciju odgovora koje bolje objašnjava jedna ili druga teorija. Njihov zaključak je ono što bi svako u procesu selekcije modela najmanje poželeo da čuje: *ne postoji reprezentativni subjekat teorije očekivane korisnosti ili teorije izgleda*. Empirijski izbori ispitanika se najbolje objašnjavaju statističkom mešavinom (engl. *mixture model*) dve teorije, i teško je govoriti o ispitaniku čije bi odgovore bolje modelirala jedna ili druga teorija.

*Komentari o debati o racionalnosti odlučivanja u uslovima rizika.* Znajući da je debata o racionalnosti ponikla upravo u oblasti odlučivanja u uslovima rizika i neizvesnosti ne čudi to što su u njoj prepoznatljive sve pozicije koje slede iz metateorijskog okvira za analizu racionalnosti saznanja koji smo predstavili. Pitanje

koje je grubo zapostavljeno u novijim diskusijama problema odlučivanja odnosi se na status normativne adekvatnosti odgovarajućih teorija. Naša analiza aksiomatskog okvira teorije izgleda Kanemana i Tverskog upućuje na dominantno formalistički stav koji se zauzima u strategiji njene izgradnje. Intuitivne osnove ove teorije, ako postoje, nalaze se „usred“ njene formalne konstrukcije; centralna aksioma na kojoj počivaju njena deskriptivna validnost i generalnost u odnosu na druge teorije ne predstavlja tvrdnju koja bi mogla da ima korespondenciju sa uobičajenom intuicijom donosioca odluka. Prema našem mišljenju, stohastička proširenja teorije očekivane korisnosti, ili pokušaji inkorporacije principa očekivane korisnosti u šire teorije koje bi je povezale sa bejzijanskim principima (poput onih Viskuzijeve teorije) i principima environmentalne racionalnosti imaju više šansi da u budućnosti dovedu do teorije odlučivanja koji bi zadovoljavala i intuicije donosioca odluka i predstavljala normativno adekvatnu teoriju. Pokušaj ovakve sinteze predstavljamo u V delu ove teze.

Shvatanje stohastičkog elementa u odlučivanju kao inherentnog samom procesu (a ne kao puke posledice greške merenja) možda može da rezultira razvojem teorija koje bi izbegavale problem konfundacije ocene modela odlučivanja i modela greške. Trenutno, u pogledu problema selekcije adekvatnog modela odlučivanja u odnosu na empirijske podatke, čini se da postoje samo razlozi za zabrinutost. Studije koje smo predstavili pokazuju da skoro ravnopravno - sa malom prednošću RDU modela poput teorije izgleda - *različiti* modeli odlučivanja najbolje opisuju ponašanje *različitih* ispitanika. „Najboljeg modela odlučivanja“ nema ni na horizontu ovih složenih studija. Još više brinu nalazi studije Harisona i Rutstrumove koji pokazuju da je ideja o *reprezentativnom subjektu* (tj. subjektu čije sve izbore najbolje karakteriše jedna teorija odlučivanja) suštinski pogrešna. Do sličnih zaključaka dolazi Blavacki (2011) za veći broj svojih ispitanika. Podsetimo se da Hej i Orm nalaze da tek oko 3/4 odluka u eksperimentima izbora ostaje isto u odgovorima istog ispitanika pri ponovljenim merenjima i suočićemo se sa veoma teškim problemom: *da li je uopšte moguće ljudsko odlučivanje karakterisati jedinstvenim modelom odlučivanja*? Ukoliko prihvatimo stav da ne postoje reprezentativni subjekti neke određene teorije odlučivanja, stav koji sledi iz studije Harisona i Rutstrumove, otvara se još složenije pitanje: *da li ljudski kognitivni sistem možda varira algoritamske (procesne) strategije (na drugom Marovom nivou opisa) rešavanja jednog istog kompjutacionog problema (na Marovom trećem nivou opisa)*? Dalje, postoji li mogućnost da određeni kompjutacioni problem, koji mi kao

istraživači definišemo određenim eksperimentalnim nacrtom i dajemo mu određenu matematičku formu (smatrajući ga tako konstantnim problemom), *kognitivni sistem neprestano reinterpretira kroz skup donekle različitih problema*? U oba slučaja besmisleno je govoriti o reprezentativnom subjektu *jedne* teorije odlučivanja. Zašto bi to bilo tako? Detaljna diskusija ovih pitanja, u VI delu naše rasprave, motiviše jedan od naših osnovnih argumenata u ovoj raspravi o racionalnosti saznanja.

## 7.2 Kauzalno učenje

Problematika odnosa normativnih i deskriptivnih objašnjenja u oblasti kauzalnosti - otkrivanja, učenja, razumevanja i primene uzročno-posledičnih odnosa - predstavlja možda formalno najsloženiju oblast naše diskusije. Zbog toga će naš formalni tretman ovog problema biti manje strog, a obim diskusije donekle ograničen. Prvo ćemo posvetiti pažnju problemu kauzalne indukcije, odn. učenja kauzalnih odnosa; u sekciji 7.4 vratićemo se kauzalnosti kroz diskusiju primene kauzalnog znanja u rezonovanju. Naš prvi korak je upoznavanje sa savremenom formalizacijom problema probabilističkih kauzalnih odnosa.

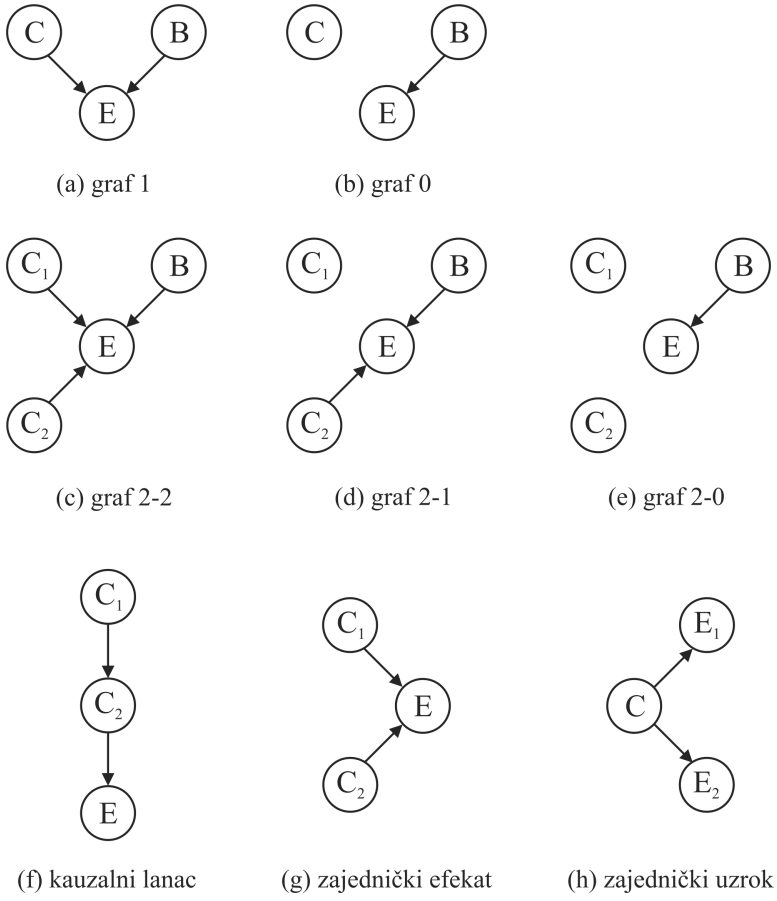
*Formalne i normativne pretpostavke teorije probabilističkih kauzalnih odnosa.* Šta bi moglo biti jednostavnije od zaključivanja o kauzalnim odnosima u fizičkoj i socijalnoj okolini potpuno prožetoj kauzalnim mehanizmima? Britanski filozof, u istoriji filozofije čuveni autor „*Rasprave o ljudskoj prirodi*“ (1739/1983), Škot Dejvid Hjum, čudio sa nad ovim procesom temeljnije od drugih u dugačkoj tradiciji rasprave o kauzalnosti. Svaka bilijarska lopta, udarena određenom snagom, izvršiće određeni rad i promeniti svoju poziciju na stolu. Kroz ponavljanje velikog broja zajedničkog pojavljivanja dva fenomena, um stvara asocijacije između impresija, ideja *intervencije* - udaranja lopte - i *efekta*, njene promene položaja. Hjum konstatuje: tri su znaka na osnovu kojih ljudski um razvija unutrašnje razumevanje kauzalnog odnosa, tri relacije od kojih je on komponovan: (i) *spacio-temporalna kontingencija* uzroka i posledice, činjenica da se oni nalaze blisko jedno drugom u prostoru i vremenu, da su koherentni u tom smislu, zatim, (ii) *sukcesija*, odn. vremenski sled (to da uzroci prethode svojim posledicama), i konačno (iii) *nužnost veze* između uzroka i posledice. Kauzalni odnos nije isto što i korelacija, to je mantra koju studenti psihologije dobro nauče (ili bi trebalo dobro da nauče) još u I semestru svojih studija. Kauzalni odnos odlikuje unutrašnja nužnost, suštinska veza između prirode uzroka i njegovih posledica. Nije sve što je u Univerzumu

povezano - povezano kauzalno. Međutim, ljudskom umu su na raspolaganju samo informacije o tome u kojoj meri dva fenomena *kovariraju*. O tim informacijama svedoči kontingencija, zajedničko pojavljivanje uzroka i posledica u prihvatljivom, koherentnom spacio-temporalnom okviru posmatranja. Pored vremenskog sleda, koji ne dopušta da efekti prethode svojim uzrocima, kontingencija je jedina opservabilna informacija koju mi imamo: kako, onda, razlikujemo kauzalne odnose od pukih (engl. *spurious*) korelacija? Kako um zaključuje o toj unutrašnjoj nužnosti u povezanosti uzroka i posledica? Ovaj problem, jedan od najtežih koje je ljudski um uopšte, ikad postavio, naziva se *Hjumovim problemom kauzalne indukcije*. Pod pretpostavkom da je vremenski sled fenomena određen (tj. da nije narušen sled po kome potencijalni efekti slede a ne prethode potencijalnim uzrocima), opservabilne informacije su samo one o zajedničkom pojavljivanju uzroka i posledica. Njih je moguće opisati statistički, naravno. Hjumov problem kauzalne indukcije onda glasi: kako na osnovu opservabilnih statističkih informacija o kontingenciji fenomena ljudski um donosi zaključak o tome koja kontingencija ukazuje na kauzalni odnos, a koja ne? Kako se otkriva „unutrašnja nužnost“ u povezanosti stvari koja neke odnose čini kauzalnim, a neke ne? Ogroman priraštaj naučne periodike posvećene ovom problemu tokom XX veka svedoči o neprestanom interesovanju filozofa, matematičara, psihologa, ekonomista, biologa i drugih naučnika za odgovor na ovo pitanje. Pošto je pozitivistička filozofija, pre svega zahvaljujući uticaju Bertrana Rasela, praktično „proterala“ raspravu o kauzalnosti iz domena naučnih pitanja, bile su potrebne decenije da se ono vrati u fokus rasprave filozofije nauke i nauka o saznanju. Krajem XX veka, 80-ih i 90-ih godina, prisustvujemo pravoj naučnoj revoluciji u razumevanu ovog problema. Bez ograde se može reći da živimo u jednoj od najinteresantnijih epoha matematike, filozofije i psihologije saznanja u odnosu na rešavanje ovog, s vremena na vreme metafizičkog, s vremena na vreme matematičkog pitanja.

Već i površna diskusija kauzalne indukcije upućuje na složenost problema. Prvo, možemo da govorimo o problemu zaključivanja o *postojanju ili odsustvu kauzalnog odnosa između neke dve varijable*: jednog pretpostavljenog uzroka i njegove pretpostavljene posledice. Da li pušenje *uopšte* izaziva rak pluća? Ovako postavljen problem naziva se *problemom učenja kauzalne strukture* (Griffiths & Tenenbaum, 2005). Drugo, možemo da postavimo pitanje o tome kako se donosi *sud o snazi nekog pretpostavljenog ili ustanovljenog kauzalnog odnosa između dve varijable*, od kojih je jedna pretpostavljeni uzrok (ili *kauzalni faktor*) najčešće

označena kao $C$ (engl. *Cause*), a druga pretpostavljeni efekat, najčešće označena kao $E$ (engl. *Effect*). U kojoj meri će pušenje doprineti verovatnoći da neka osoba oboli? Ovako postavljeno pitanje odnosi se na *problem ocene snage kauzalnog odnosa* (Griffiths & Tenenbaum, 2005, up. Lu et al, 2008 za diskusiju odnosa ova dva pitanja). Da pružimo još jedan primer, $C$ može da bude određeni medicinski tretman, a $E$ neki njegov očekivani efekat, recimo promena stanja indikatora na nekom standardizovanom testu. Pitanje učenja kauzalne strukture odnosi se na donošenje suda o tome da li između medicinskog tretmana $C$ i promene stanja indikatora $E$ uopšte postoji uzročno-posledični odnos. Drugo pitanje se odnosi na to koliko je taj kauzalni odnos snažan, ukoliko on postoji. Uzroke generalno delimo u dve kategorije: *generativne*, koji doprinose pojavi efekta $E$, i *preventivne*, koji doprinose sprečavanju pojave efekta $E$. Jasno je da ovde sve vreme govorimo o *probabilističkoj kauzalnosti*: ako su kauzalni odnosi koje razmatramo *deterministički*, onda svaka pojava generativnog uzroka dovodi do pojave odgovarajućeg efekta, a svaka pojava preventivnog uzroka onemogućava pojavu odgovarajućeg efekta. U diskusiji determinističke kauzalnosti tako nema smisla govoriti o snazi kauzalnog odnosa. Ako uzroci samo sa određenom verovatnoćom doprinose pojavi svojih posledica, ili samo sa određenom verovatnoćom blokiraju njihovu pojavu (u slučaju preventivnih uzroka), govorimo o probabilističkoj kauzalnosti. Očigledno, upravo ta verovatnoća sa kojom uzroci utiču na pojavu ili odsustvo svojih efekata treba da korespondira verovanju o snazi određenog kauzalnog odnosa. U svetu kojim su okružena ljudska bića nalazimo i determinističke i probabilističke kauzalne odnose, ali ovi drugi suštinski dominiraju, unoseći element neizvesnosti i rizika u naše okruženje. Otkrivanje probabilističkih kauzalnih odnosa i procena snage kauzalne veze između različitih parametara našeg okruženja zato predstavlja adaptivni problem od fundamentalnog značaja za ljudska bića. Konačno, problem kauzalnog učenja se prirodno usložnjava na dva načina. Prvi je uvođenje većeg broja potencijalnih kauzalnih ili pretpostavljenih efekata. Ako jednu istu posledicu može da izazove više različitih uzroka, osim osnovnog kauzalnog dejstva svakog od njih pojedinačno govorićemo i o *kauzalnim interakcijama* među njima - ako one postoje, pošto je i njih potrebno prethodno ustanoviti, a tek zatim oceniti nihovu snagu. Slično, više različitih efekata može da ima jedan isti zajednički uzrok. Drugi put u usložnjavanje problema kauzalne indukcije jeste slobodno usložnjavanje strukture kauzalnih odnosa između više varijabli koja može da uzme oblik najrazličitijih *kauzalnih mreža* (Pearl, 2000, Glymour, 2001, 2003). Slika 12. predstavlja nekoliko

grafova koji ilustruju upravo diskutovanu problematiku.




Slika 12. *Osnovni grafovi u analizi kauzalne indukcije i primeri jednostavnih kauzalnih mreža* (objašnjenje u tekstu).

Svi pretpostavljeni kauzalni odnosi na grafovima na Slici 12. su generativni. Svi grafovi kojima se služimo u matematičkom tretmanu kauzalnosti imaju određene osobine. Čvorovi ovih grafova prikazuju varijable (uzroke i posledice), a linije predstavljaju veze među njima. Grafovi kauzalnih modela su uvek *usmereni aciklični grafovi* (engl. *Directed Acyclic Graph*, skr. DAG).: (a) veze u njima su usmerene tako da odslikavaju smer kauzalnih uticaja *od* uzroka *ka* efektima *i ne obrnuto*, te (b) u ovakvom grafu ne postoji način da se prateći smer kauzalnog uticaja od nekih potonjih efekata („potomaka“ u grafu) stigne do njihovih prethodnih uzroka („predaka“ u grafu). Graf (a) predstavlja situaciju u kojoj postoji jedan uzrok, $C$, kauzalno povezan sa jednim efektom, $E$; $B$ predstavlja konglomerat svih drugih mogućih uzroka koji mogu da dovedu do $E$. Uskoro ćemo objasniti značaj reprezentacije ovog konglomerata $B$ drugih mogućih uzroka. Graf (b) predstavlja iste varijable osim što sada između varijable $C$ i varijable $E$ ne postoji kauzalni

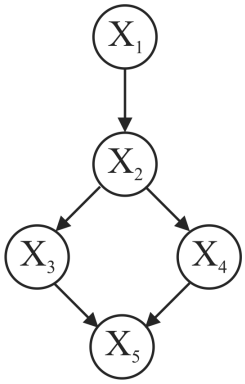
odnos. Zbog svog značaja u analizi elementarne kauzalne indukcije ova dva grafa imaju ustaljena imena: graf (a) je „graf 1“, dok se graf (b) naziva „graf 0“. Graf (c) predstavlja dva uzroka, $C_1$ i $C_2$, kauzalno povezana sa istom posledicom $E$, i konglomerat ostalih mogućih uzroka iste varijable $E$ označen sa $B$. Graf (d) predstavlja iste varijable kao graf (c) samo što je u njemu uklonjen jedan kauzalni odnos; u grafu (e) uklonjena su oba kauzalna odnosa. Primetimo da i u grafu (c) i u grafu (d) konglomerat drugih mogućih uzroka varijable $E$, ponovo označen sa $B$, uvek ostaje kauzalno povezan sa njom. Takođe, pošto grafovi (c), (d) i (e) obuhvataju više od jednog mogućeg uzroka pojave $E$, moguće su i kauzalne interakcije između uzroka koji su na njima predstavljeni. Te interakcije nisu uključene u reprezentacije ovih grafova. Grafove (c), (d) i (e) smo, po analogiji sa grafovima (a) i (b), nazvali „grafom 2-2“, „grafom 2-1“ i „grafom 2-0“, gde prvi broj u imenu grafa označava koliko potencijalnih uzroka (kauzalnih faktora) se uopšte razmatra, a drugi koliko uzroka su stvarno i kauzalno povezani sa efektom. Treći red grafova takođe prikazuje jednostavne kauzalne mreže. Graf (f) prikazuje tzv. *kauzalni lanac*: varijabla $C_1$ je kauzalni faktor varijable $C_2$ koja je kauzalni faktor varijable $E$. Graf (g) prikazuje *mrežu zajedničkog efekta*: dva uzroka, $C_1$ i $C_2$, imaju zajedničku posledicu $E$. Primetimo da je ovaj graf formalno identičan grafu (c), osim što ne prikazuje eksplicitno konglomerat drugih mogućih uzroka $B$. Konačno, graf (h) prikazuje *mrežu zajedničkog uzroka*, u kojoj jedan isti kauzalni faktor $C$ ima dve posledice $E_1$ i $E_2$. Mreža zajedničkog uzroka se naziva još i *kauzalnom viljuškom*.

U svim narednim analizama pretpostavićemo da su varijable o čijim kauzalnim odnosima diskutujemo binarne: one mogu da uzmu samo jednu od dve vrednosti, npr. „prisutna“ ili „odsutna“. Tako ćemo diskutovati samo odnose između varijabli kao što su „kiša pada/kiša ne pada“ i „nebo je oblačno/nebo nije oblačno“[37]. Grafovi poput onih na Slici 12. predstavljaju suštinski deo svake kauzalne mreže. Kauzalna mreža je formalni, matematički konstrukt koji koristimo da bismo egzaktno govorili o probabilističkim kauzalnima odnosima između nekih varijabli. Ponekad se kauzalne mreže nazivaju još i *bejzijanskim mrežama* (engl. *Bayesian Networks*)[38]. Svaku kauzalnu mrežu definiše (*i*) *jedan tačno određen usmereni aciklični graf* i (*ii*) *zajednička distribucija verovatnoće javljanja vrednosti varijabli koje su deo tog grafa*. Posvetimo se za kratkom drugom delu definicije kauzalne mreže. U grafu neke kauzalne mreže, sve varijable se javljaju sa određenim verovatnoćama. Te verovatnoće su ili *egzogene*, u kom slučaju su one definisane nekim kauzalnim odnosima koji se nalaze *van* same kauzalne mreže koju posmatramo[39], ili su

određene kauzalnim odnosima između varijabli i verovatnoćama onih varijabli koje predstavljaju uzroke. Pogledajmo ponovo grafove (f), (g) i (h) na Slici 12. Verovatnoća pojave varijable $E$ u kauzalnom lancu na grafu (f) određena je snagom kauzalne veze između $C_2$ i $E$, kao i verovatnoćom pojave $C_2$; verovatnoća pojave $C_2$ određena je snagom kauzalne veze između $C_1$ i $C_2$, kao i verovatnoćom pojave $C_1$; ali verovatnoća pojave $C_1$ je egzogena - ona nije definisana kauzalnim odnosima u kauzalnom lancu na grafu (f). Svakako, i egzogena varijabla mora imati neke uzroke - ali ti uzroci *nisu u fokusu analize* koju predstavlja određena mreža. Isto tako, na grafu (g), oba uzroka su egzogene varijable, kao što je to i jedini uzrok u kauzalnoj viljušci na grafu (h). Očigledno, ako su date verovatnoće egzogenih varijabli i snage kauzalnih odnosa (uskoro ćemo uvideti da su i te snage verovatnoće), moguće je odrediti verovatnoću pojave svih varijabli u mreži. Zajednička distribucija verovatnoće svih varijabli u mreži ($X_1$, $X_2$, .., $X_n$), u oznaci $P(X_1, X_2, .., X_n)$, tako predstavlja suštinsku odliku svake kauzalne mreže. Centralni normativni uslov u analizi probabilističkih kauzalnih odnosa odnosi se na bitnu osobinu ove zajedničke distribucije verovatnoće i naziva se *kauzalnim Markovljevim uslovom* (Pearl 2000, 2010). Posmatrajmo neku varijablu $X$ u određenoj kauzalnoj mreži. Sve varijable koje su neposredni uzroci $X$, odn. koje su povezane direktno linijama sa $X$ tako da smer kauzalnog uticaja ide od tih varijabli ka $X$, nazivaju se *roditeljima* te varijable. Sve varijable koje prethode $X$ a nalaze se u kauzalnim lancima koje prethode njoj i njenim roditeljima nazivaju se *precima* te varijable. Slično, sve varijable koje su pod direktnim kauzalnim uticajem $X$, odn. koje povezuju direktno linije kauzalnog uticaja od $X$ ka njima, nazivaju se *decom* te varijable. Sve varijable koje se, polazeći od početnog kauzalnog uticaja $X$, nalaze u kauzalnim lancima posle nje u istoj kauzalnoj mreži, nazivaju se *potomcima* te varijable. Sada možemo da definišemo centralni normativni uslov za analizu probabilističke kauzalnosti.

Kauzalni Markovljev uslov: *verovatnoća javljanja svake varijable u kauzalnoj mreži nezavisna je od stanja ne-potomaka te varijable, uslovno u odnosu na roditelje te varijable.*

Ovako određen kauzalni Markovljev uslov objasnićemo kroz primer na Slici 13. Posmatrajmo varijablu $X_3$ u kauzalnoj mreži na Slici 13. Njen roditelj je varijabla $X_2$, koja je takođe roditelj varijable $X_4$. Varijable $X_3$ i $X_4$ imaju zajedničkog roditelja i još jednog zajedničkog pretka koji nije roditelj: to je varijabla $X_1$. Varijabla $X_5$ je dete varijable $X_3$, kao što je i dete varijable $X_4$; $X_1$, $X_2$, $X_3$ i $X_4$ su sve njeni preci.




Slika 13. *Jednostavna kauzalna mreža za diskusiju kauzalnog Markovljevog uslova* (objašnjenje u tekstu).

*Ne-potomci* varijable $X_3$ koju posmatramo su varijable: $X_1$, $X_2$ i $X_4$, pri čemu joj je $X_2$ roditelj; varijabla $X_5$, njeno dete, jeste njen potomak. Pretpostavimo da ne znamo ništa o tome da li je $X_2$, koja je roditelj $X_3$ koju posmatramo, prisutna ili nije. Ako znamo da je prisutna varijabla $X_1$, koja je predak (ali ne i roditelj) $X_3$, da li možemo da donesemo neki sud o verovatnoći javljanja $X_3$? Sigurno da možemo: ako je $X_1$ prisutna, onda postoji neka verovatnoća da ona kauzalno utiče na $X_2$, koja je roditelj $X_3$ i može kauzalno da utiče na nju; dakle, znanje o tome da je prisutna $X_1$ utiče na naš sud o verovatnoći javljanja $X_3$ - ona je veća, ako je $X_1$ prisutna. Tako, naše znanje o verovatnoći da se $X_1$ javi utiče na naš sud o verovatnoći da će se javiti $X_3$. Pretpostavimo sada da mi *intervenišemo* u kauzalnoj mreži fiksirajući vrednost binarne varijable $X_2$ tako da je ona uvek prisutna. Ovakve intervencije su ključne u analizi kauzalnih mreža. Da li sada naše znanje o verovatnoći javljanja $X_1$ uvećava naše znanje o verovatnoći javljanja $X_3$? Odgovor je, naravno, ne: pošto je posle naše intervencije roditelj varijable $X_3$, varijabla $X_2$, uvek prisutna, to da li će se $X_3$ javiti ili ne zavisi isključivo od snage kauzalnog odnosa između njenog roditelja i nje. Mi smo ovom intervencijom efektivno „odsekli“ uticaj kauzalnog lanca koji povezuje $X_1$ sa $X_3$. Verovatnoća javljanja $X_1$ više ne utiče na naš sud o verovatnoći javljanja $X_3$, koja je sada određena isključivo verovatnoćom da kauzalna veza između njenog roditelja $X_2$ i nje „proradi“ i dovede do njene pojave. Pretpostavimo sada je varijabla $X_4$ prisutna. Šta sada možemo da kažemo o verovatnoći javljanja varijable $X_3$? Pošto je prisutna $X_4$, koja je takođe ne-potomak varijable $X_3$, a njih dve imaju zajedničkog roditelja $X_2$, znanje o tome da se $X_4$ javlja nam sugeriše da je $X_2$ takođe prisutna, pa je tako veća i verovatnoća da će se javiti $X_3$ koju posmatramo. Međutim, ukoliko mi opet intervenišemo u kauzalnoj mreži tako da fiksiramo vrednost $X_2$ i ona postane uvek prisutna, onda su verovatnoća pojave $X_4$ i verovatnoća pojave

$X_3$ *nezavisne*: obe su, naime, određene isključivo verovatnoćom da će ih proizvesti njihov roditelj $X_2$, koja je data. Ovakva izolacija potencijalnih kauzalnih uticaja na određene varijable naziva se *zaklanjanjem* (engl. *screening-off*), a prvi je diskutuje Hans Rajhenbah (Reichenbach, 1956, prema Glymour & Eberhardt, 2011), u epohi u kojoj je pozitivistička filozofija čvrsto drži do stava da je kauzalnost metafizičko pitanje koje kao takvo ne može da dobije odgovorajući pozitivan tretman.

Kauzalni Markovljev uslov konceptualno i matematički rezimira našu prethodnu diskusiju na sledeći način: *pod uslovom da su dati roditelji neke varijable, odn. da znamo da je njihovo dejstvo prisutno, verovatnoća javljanja varijable koju posmatramo je nezavisna od njenih ne-potomaka.* Kauzalni Markovljev uslov garantuje to da je verovatnoća pojave neke posledice u mreži kauzalnih odnosa određena samo verovatnoćom pojave njenih roditelja i snagom direktnih kauzalnih odnosa koji mogu da dovedu do nje (isto važi i za analizu preventivnih kauzalnih odnosa). Da li ljudi u rezonovanju u kauzalnim odnosima poštuju ovaj normativni uslov videćemo u sekciji 7.4. Matematički izražen, ovaj uslov glasi:

$$P(X_1, X_2, ...X_n) = \prod_{i=1}^{n} P(X_i|Par(X_i)) \tag{14}$$

gde skup $Par(X_i)$ označava skup roditelja varijable $X_i$. Kauzalni Markovljev uslov iskazuje to da je zajednička distribucija verovatnoće svih varijabli u kauzalnoj mreži jednaka proizvodu zavisnih verovatnoća da se one jave *ako su dati njihovi roditelji.* Pošto za data prisustva roditelja zavisne verovatnoće javljanja varijabli u mreži postaju međusobno nezavisne, zajednička verovatnoća javljanja varijabli je jednostavno proizvod odgovarajućih zavisnih verovatnoća varijabli - baš kao što je u teoriji verovatnoće to uvek slučaj sa nezavisnim varijablama.

Fundamentalan doprinos razvoju teorije kauzalnih mreža - ili kauzalnih modela - koja danas predstavlja osnovni skup formalnih sredstava za tretman problema probabilističkih kauzalnih odnosa dao je američki matematičar i kompjuterski naučnik Džudia Perl (Pearl, 2000, 2010). Upravo njegov rad, 80-ih godina XX veka, omogućio je formalizaciju „*logike intervencije*“ u kauzalnim mrežama, pokazujući da, nezavisno od jezika matematičke statistike, rasprava o problemu kauzalnosti zahteva razvoj preciznog *jezika metodologije*, koji formalizacijom osnovnih operacija *eksperimentisanja sa kauzalnim faktorima* otkriva suštinske mogućnosti i ograničenja našeg zaključivanja u oblasti probabilističkih kauzalnih odnosa (Pearl, 2000). Suštinska poruka teorije kauzalnih modela je da samo

posmatranjem kovarijacije između fenomena ne možemo saznati mnogo o prirodi probabilističkih kauzalnih odnosa; tek aktivno učestvujući u svetu, utičući na potencijalne kauzalne faktore, realno ili hipotetički, mi počinjemo da koristimo naše (ograničene) mogućnosti kauzalnog rezonovanja. Uporedo sa Perlovom sintezom u teoriji kauzalnih mreža, filozofija nauke se oporavlja od pozitivističke zabrane analize kauzalnosti kroz radove više istaknutih mislilaca, od kojih se posebno ističe doprinos Nensi Kartvrajt, britanske filozofkinje čija knjiga „*Kapaciteti prirode i njihovo merenje*“ predstavlja pravu prekretnicu u razumevanju analize kauzalnih odnosa u naučnom, eksperimentalnom zaključivanju (Cartwright, 1989).

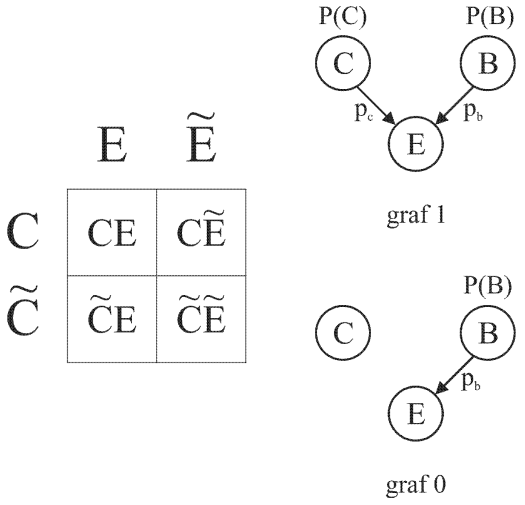
*Pristup problemu kauzalnog učenja kroz teoriju kauzalnih mreža.* Raspravu smo počeli razlikovanjem problema učenja kauzalne strukture od problema učenja o snazi kauzalnog odnosa, tj. donošenja suda o tome koliki je intenzitet kauzalnog odnosa između dve varijable ukoliko on postoji. Posvetićemo se sada drugom od ova problema, koji u užem smislu specifikuje problem kauzalne indukcije. Na osnovu čega neka osoba može da donese sud o tome koliki je intenzitet kauzalne veze između nekog uzroka $C$ i njegove posledice $E$? U svetu probabilističke kauzalnosti, određene verovatnoće, naravno, predstavljaju jedine informacije na osnovu kojih takav sud može da se donese. Dakle, pretpostavimo da imamo pred sobom jedan binarni uzrok $C$, i jednu binarnu posledicu $E$. Za potrebe primera, neka $C$ uzima vrednosti „novi medicinski tretman je primenjen“, u oznaci: $C$, i „novi medicinski tretman nije primenjen“, u oznaci $\widetilde{C}$; $E$ neka uzima vrednosti „došlo je do poboljšanja u stanju pacijenata“, u oznaci $E$, i „nije došlo do poboljšanja u stanju pacijenata“, u oznaci $\widetilde{E}$. U svetu binarnih uzroka i posledica, moguće su samo sledeće četiri situacije koje nastaju kombinacijom dva stanja uzroka i dva stanja posledice: $CE$ (prisutni i uzrok i efekat), $\widetilde{C}E$ (uzrok nije prisutan ali jeste posledica), $C\widetilde{E}$ (uzrok je prisutan ali ne i posledica) i $\widetilde{C}\widetilde{E}$ (nisu prisutni ni uzrok ni posledica). Kada posmatramo ovakav problem, u kome je naš zadatak da procenimo kolika je snaga kauzalne veze između uzroka i efekta, na raspologanju su nam po pravilu samo posmatranja ove četiri situacije. Intuicija nam govori da frekvencije njihovog javljanja u nekom uzorku posmatranja mogu da ukažu na stepen njihove povezanosti; problem kauzalne indukcije traži odgovor na pitanje *kako* se one koriste u tu svrhu. Frekvencije kombinacija dva stanja potencijalnog uzroka i njegove posledice daje nam punu informaciju o *kovarijaciji* (ili *kontingenciji*) dve diskretne promenljive. Kovarijacija, naravno, još nije kauzalnost, a standardni način da se ona predstavi jeste *kontingencijska tabela*:

Tabela 2. *Kontingencijska tabela elementarne kauzalne indukcije.*

| | $E$ | $\widetilde{E}$ |
|---|---|---|
| $C$ | $CE_{(\mathrm{a})}$ | $C\widetilde{E}_{(\mathrm{b})}$ |
| $\widetilde{C}$ | $\widetilde{C}E_{(\mathrm{c})}$ | $\widetilde{C}\widetilde{E}_{(\mathrm{d})}$ |

Tabela 2. daje sumarni pregled frekvencija događaja $CE$ (prisustvo $C$ i $E$, ćelija $a$), $C\widetilde{E}$ (prisustvo $C$ i odsustvo $E$, ćelija $b$), $\widetilde{C}E$ (odsustvo $C$ i prisustvo $E$, ćelija $c$) i $\widetilde{C}\widetilde{E}$ (odsustvo i $C$ i $E$, ćelija $d$). Ukupan broj posmatranja u ćeliji je zbir četiri frekvencije koje ona sadrži, u oznaci: N. Zadatak kognitivnog sistema u procesu kauzalne indukcije je da na osnovu poznavanja samo ovih frekvencija (ili verovatnoća, ako se posmatranja u svim ćelijama podele sa N), zaključi koliki je stepen uzročno-posledične veze između $C$ i $E$. Kao primer za podatke u kontingencijskoj tabeli možemo da uzmemo (a) broj pacijenata kod kojih je novi tretman primenjen i kod kojih je došlo do poboljšanja, (b) broj pacijenata kod kojih je novi tretman primenjen ali kod kojih nije došlo do poboljšanja, (c) broj pacijenata kod kojih nije primenjen novi tretman ali je došlo do poboljšanja i (d) broj pacijenata kod kojih niti je primenjen novi tretman niti je došlo do poboljšanja. Fundamentalni problem sa kojim se suočava kognitivni sistem u kauzalnoj indukciji formulišemo na sledeći način. Opservabilne varijable su frekvencije u kontingencijskoj tabeli. Kognitivni sistem je izložen frekvencijama zajedničke pojave uzroka i posledice kao i njihovog zajedničkog odsustvovanja, te situacijama u kojima se uzrok javlja bez posledice i posledica bez uzroka. Pošto govorimo o probabilističkoj kauzalnosti, sasvim je moguće da pojava uzroka neće svaki put dovesti do javljanja posledice, kao i to da će se posledica nekada javljati u odsustvu uzroka - jer oba učestvuju u složenoj mreži kauzalnih odnosa sa drugim varijablama. Stanje nekih pacijenata u našem primeru može „spontano“ da se promeni i ako nisu izloženi novom tretmanu koji ispitujemo. U kauzalnim analizama, „spontano“ nikada nije stvarno spontano; sve mora da ima neki uzrok, a u situacijama u kojima se posledica javlja u odsustvu uzroka koji mi analiziramo kažemo da je do nje dovelo nešto od *pozadinskih uzroka* koji nisu eksplicitni u kauzalnoj mreži. Konglomerat takvih pozadinskih uzroka predstavlja čvor $B$ na grafovima kauzalnih mreža u prva dva reda Slike 12. Posmatrajmo sada grafove kauzalnih mreža na Slici 14.

Na grafu 1 vidimo da do efekta $E$ mogu da dovedu dve varijable. Jedna je uzrok $C$ čiju kauzalnu snagu je potrebno da ocenimo. Druga je konglomerat svih drugih mogućih uzroka posledice $E$, koji označavamo sa $B$. Uzrok $C$ se javlja sa verovatnoćom $P(C)$, dok se svi pozadinski uzroci $B$ javljaju sa verovatnoćom $P(B)$. Ako je prisutan, uzrok $C$ generiše efekat $E$ sa verovatnoćom $p_c$. Ova verovatnoća predstavlja *kauzalnu moć* (engl. *Causal Power*) uzroka $C$. Kauzalnu moć ne treba mešati sa zavisnom verovatnoćom da će se efekat javiti ako je dat uzrok, $P(E|C)$. Kauzalna moć je konceptualno, i videćemo, matematički, precizno određena: *to je verovatnoća da kauzalna veza između C i E, mehanizam koji prenosi kauzalni uticaj sa jedne na drugu varijablu, postane aktivna u situacijama u kojima je uzrok* C *prisutan.* Ne zaboravimo: $C$ je samo probabilistički uzrok efekta $E$. Nije nužno da pri svakoj njegovoj pojavi on ispolji svoju kauzalnu moć da generiše posledicu $E$. On to čini sa verovatnoćom $p_c$. Isto tako, konglomerat pozadinskih uzroka koji mogu da dovedu do $E$, označen sa $B$, ima kauzalnu moć $p_b$. Svi podaci koji su uopšte dostupni kognitivnom sistemu koji pokušava da oceni kolika je kauzalna moć uzroka dati su kontingencijskom tabelom. Na osnovu frekvencija u kontingencijskoj tabeli moguće je izračunati nekoliko verovatnoća na osnovu kojih je, videćemo, moguće doneti sud o tome kolika je $p_c$ za određeni uzrok $C$. Primetimo da dve kauzalne moći na grafu 1 *nisu direktno opservabilne.*




Slika 14. *Grafovi neophodni za rešenje problema elementarne kauzalne indukcije i kontingencijska tabela sa odgovarajućim frekvencijama (objašnjenje u tekstu).*

Verovatnoću javljanja uzroka, $P(C)$, sasvim je lako izračunati iz kontingencijske tabele: ona je suma frekvencija onih situacija u kojima je uzrok prisutan, $CE$ i $C\widetilde{E}$, podeljena sa brojem posmatranja N:.

$$P(C) = \frac{CE + C\widetilde{E}}{N} = \frac{a+b}{a+b+c+d} \tag{15}$$

Zavisna verovatnoća da je efekat prisutan ako je uzrok prisutan, $P(E|C)$, takođe se računa direktno iz kontingencijske tabele, kao:

$$P(E|C) = \frac{P(CE)}{P(C)} = \frac{CE}{CE + C\widetilde{E}} = \frac{a}{a+b} \tag{16}$$

Slično, zavisna verovatnoća da je efekat prisutan ako uzrok nije prisutan, $P(E|\widetilde{C})$, računa se kao:

$$P(E|\widetilde{C}) = \frac{P(\widetilde{C}E)}{P(\widetilde{C})} = \frac{\widetilde{C}E}{\widetilde{C}E + \widetilde{C}\widetilde{E}} = \frac{c}{c+d} \tag{17}$$

Vratimo se sada grafu 1 sa Slike 14. Formulisaćemo zavisnu verovatnoću da je efekat $E$ prisutan ako je uzrok $C$ prisutan, $P(E|C)$, upotrebom parametara koje opisuju relacije na grafu 1, bez obzira na to što neki od njih nisu opservabilni za kognitivni sistemi koji pokušava da izračuna kauzalnu moć uzroka $C$. Zavisna verovatnoća $P(E|C)$ može da se izrazi kao:

$$P(E|C) = p_c + P(B) \cdot p_b - p_c \cdot P(B) \cdot p_b \tag{18}$$

Pošto smo pretpostavili da je $C$ prisutan(tj. $P(C) = 1$), njegov doprinos verovatnoći pojavi $E$ zavisi samo od kauzalne moći $p_c$. Ovome treba dodati doprinos konglomerata svih drugih mogućih uzroka, koji je proizvod verovatnoće njihovog javljanja, $P(B)$ i njihove kauzalne moći, $p_b$. Konačno, pošto smo sada već uračunali oba kauzalna uticaja na grafu 1, moramo od svega da oduzmemo doprinos mogućnosti da $C$ i $B$ *istovremeno* proizvedu $E$. Sada smo izrazili $P(E|C)$ kroz teorijske varijable grafa 1; podsećamo da nijedna od dve kauzalne moći, kao ni verovatnoća javljanja konglomerata drugih uzroka koji nisu deo analize na grafu 1, nisu opservabilni. Zapitajmo se sada kako da teorijski izrazimo $\mathrm{P}(E|\widetilde{C})$, zavisnu verovatnoću da se uzrok javlja u odsustvu efekta. Ta situacija prikazana je na grafu 0, na kome je uzrok $C$ diskonektovan od efekta $E$, označavajući to da samo konglomerat mogućih drugih uzroka $B$ može da dovede do eventualne pojave $E$. Očigledno je da

$$P(E|\widetilde{C}) = P(B) \cdot p_b \tag{19}$$

odn. da je zavisna verovatnoća $\mathrm{P}(E|\widetilde{C})$ da se efekat $E$ javi u odsustvu uzroka $C$ jednaka proizvodu kauzalne moći pozadinskog uzroka $B$, $p_\mathrm{b}$, i verovatnoće da se taj pozadinski uzrok javi, $\mathrm{P}(B)$. Interesantno: iz jednačine (19) saznajemo da opservabilna zavisna verovatnoća $\mathrm{P}(E|\widetilde{C})$, koju možemo da izračunamo direktno iz frekvencija u kontingencijskoj tabeli, predstavlja ocenu proizvoda dve neopservabilne varijable na grafu 1 i grafu 0, $P(B) \cdot p_b$. Ako ponovo pogledamo jednačinu (18), videćemo da ovaj proizvod igra ulogu i u njoj. Zamenjujući izraz $P(B) \cdot p_b$ sa $\mathrm{P}(E|\widetilde{C})$ u jednačinama (18) i (19), i rešavajući po $p_c$, kauzalnoj moći uzroka $C$ koju pokušavamo da ocenimo, dobijamo

$$p_c = \frac{P(E|C) - P(E|\widetilde{C})}{1 - P(E|\widetilde{C})} \tag{20}$$

i vidimo da je na osnovu opservabilnih verovatnoća, verovatnoća koje je moguće izračunati direktno iz kontingencijske tabele, moguće oceniti neopservabilnu $p_c$ - kauzalnu moć uzroka $C$: verovatnoću sa kojom on, kada se javlja, prenosi svoj kauzalni uticaj na efekat $E$. Poznavaoci teorije učenja će prepoznati u brojiocu jednačine (20) vrednost $\Delta P$, probabilističkog kontrasta, granične vrednosti Rescorla-Wagner modela učenja kovarijacije i klasičnog uslovljavanja (Allan, 1980, Chapman & Robins, 1990, Cheng & Novick, 1990, Cheng, 1997, Danks, 2003). Pošto je $\Delta P$ definisana kroz $\mathrm{P}(E|C)$ i $\mathrm{P}(E|\widetilde{C})$ kao

$$\Delta P = P(E|C) - P(E|\widetilde{C}) = \frac{a}{a+b} - \frac{c}{c+d} \tag{21}$$

u teoriji kauzalnog učenja izraz za kauzalnu moć uobičajeno pišemo kao

$$p_c = \frac{\Delta P}{1 - P(E|\widetilde{C})} \tag{22}$$

Inferencija za preventivne uzroke je podjenako jednostavna i vodi ka izrazu

$$q_c = -\frac{\Delta P}{P(E|\widetilde{C})} \tag{23}$$

Vrednost $\Delta P$, kao mera kontingencije, može da se izračuna direktno iz frekvencija događaja u kontingencijskoj tabeli. Vrednost kauzalne moći, $p_c$, može da se oceni na osnovu nje i zavisne verovatnoće $\mathrm{P}(E|\widetilde{C})$. Između probabilističkog kontrasta i kauzalne moći postoji duboka veza i suštinska razlika. Probabilistički kontrast je *asocijativna mera* intenziteta kauzalne veze koja govori o stepenu intenziteta kauzalnog odnosa u situaciji u kojoj uticaj nekog potencijalnog uzroka ocenjujemo *u prisustvu drugih mogućih kauzalnih uticaja*. Kauzalna moć, nasuprot tome, govori o intenzitetu kauzalnog uticaja izraženog kroz verovatnoću da će kauzalni mehanizam biti aktivan i dovesti do realizacije u vidu očekivanog efekta, *pod pretpostavkom da se potencijalni uzrok posmatra u izolaciji*, odn. bez istovremenog uticaja drugih mogućih uzroka $B$. U odnosu na to kako se razume problem kauzalne indukcije, i $\Delta P$ i $p_c$ mogu da se interpretiraju kao *normativne ocene* snage kauzalnog odnosa[40].

Upravo predstavljen pristup problemu kauzalne indukcije karakterističan je tek za najnoviju epohu njegovog proučavanja. 90-ih godina u psihologiji dolazi do uvida u mogućnost primene teorije kauzalnih modela u objašnjenju kauzalnog učenja. Primena teorije kauzalnih modela u kognitivnoj psihologiji razvija se istovremeno u radovima nekoliko autora (Waldmann, & Holyoak, 1992, Waldmann, 1996, Cheng, Park, Yarlas & Holyoak, 1996, Cheng, 1997), ali najznačajniji doprinos je sigurno doprinos američke naučnice Patriše Čeng iz 1997 godine. U njenom radu „*From covariation to causation: A causal power theory*“, objavljenom u časopisu *Psychological Review*, 1997, Čengova je dala derivaciju kauzalne moći $p_c$ kako smo je ovde predstavili i diskutovala njen odnos sa probabilističkim kontrastom (u čijem razvoju kao teorije ljudske ocene kauzalnog odnosa je prethodno i sama učestvovala, Cheng & Novick, 1990, 1992). Rad Čengove iz 1997. je danas sigurno jedan od najuticajnijih radova u istoriji eksperimentalne psihologije uopšte. Njen najveći doprinos je verovatno u jasnom prepoznavanju uslova pod kojima je inferencija kauzalne moći (jednačine (17 -23)) moguća. Ovi uslovi predstavljaju pretpostavke jedne lokalne *kantijanske metafizike* (Holyoak & Cheng, 2010) pod kojom je ocena neopservabilne kauzalne moći moguća. Pretpostavke pod kojima je moguća inferencija kauzalne moći probabilističkog uzroka na osnovu jednačina (22) i (23) su sledeće (oznake prate notaciju sa grafova 1 i 0 na Slici 14) :

($i$) Kada se $C$ pojavljuje, on uzrokuje $E$ sa verovatnoćom $p_c$; kada se $B$ pojavljuje, on uzrokuje $E$ sa verovatnoćom $p_b$; i ništa drugo ne uzrokuje $E$;

($ii$) $C$ i $B$ uzrokuju $E$ nezavisno jedno od drugog (preciznije: ne postoji kauzalna interakcija između $C$ i $B$);

(*iii*) $C$ i $B$ uzrokuju pojavu $E$ kauzalnim moćima $p_c$ i $p_b$ koje su nezavisne od verovatnoće javljanja $C$ i $B$; i

(iv) $E$ se pojavljuje samo kada je uzrokovan (Cheng, 1997).

Uslove za derivaciju kauzalne moći koje specifikuje Čengova treba posmatrati kao *normativne uslove* teorije kauzalne moći. Uslovi koje specifikuje Čengova predstavljaju pojednostavljenje kompleksnosti realnog kauzalnog Univerzuma - ali pojednostavljenje one vrste koje nam omoguća da donesemo bar neke zaključke u ekstremno komplikovanoj situaciji zaključivanja i učenja u uslovima probabilističke kauzalnosti. Nije teško pretpostaviti da se probabilistički uzroci nalaze u kauzalnim interakcijama sa većim brojem drugih probabilističkih uzroka koji se u derivaciji kauzalne moći vode pod konglomeratom pozadinskih uzroka $B$. Ipak, Čengova diskutuje i situacije u kojima neki od normativnih uslova za ocenu kauzalne moći nisu zadovoljeni, i zaključuje da u situaciji kada $C$ i $B$ nisu nezavisni, izračunavanje kauzalne moći vodi ka konzervativnoj oceni snage kauzalnog odnosa, što je povoljna karakteristika za ovakvu meru (Cheng, 2000). Čengova je, u saradnji sa Lorom Novik, ponudila i proširenje teorije kauzalne moći ka izračunavanju *kauzalnih moći interakcija dva uzroka* (Novick & Cheng, 2004). Objašnjenje kauzalnih interakcija u teoriji kauzalnih modela je dosta složenije od derivacije kauzalne moći jednog uzroka van interakcije. Uprkos tome što su empirijski dobro dokumentovani efekti interakcije uzroka u kauzalnom učenju (Spellman, 1996, Novick & Cheng, 2004, Rehder & Milovanović, 2007), u našoj diskusiji ne možemo da posvetimo više prostora ovom važnom problemu. Teorija kauzalne moći, njeno proširenje na problem ocene snage kauzalnih interakcija (Novick & Cheng, 2004) i njena bejzijanska verzija (Lu et al, 2008) svakako predstavljaju jednu od najelegantnijih matematičkih teorija u psihologiji uopšte.

*Učenje kovarijacija, kauzalno učenje i asocijacionističke teorije.* Teorija kauzalne moći ni iz daleka ne predstavlja jedinog kandidata za psihološku teoriju o tome kako kognitivni sistem donosi sud o intenzitetu kauzalne povezanosti fenomena. Već smo pokazali da kauzalna moć stoji u jasnom odnosu prema jednoj asocijativnoj meri snage kauzalnog odnosa - probabilističkom kontrastu, $\Delta P$: u jednačinama (22) i (23) vidimo da probabilistički kontrast učestvuje u izračunavanju kauzalne moći. Dakle, svaki kognitivni sistem koji izračunava kauzalnu moć kao svoju ocenu snage kauzalnog odnosa, istovremeno izračunava i probabilistički kontrast -$\Delta P$ - kao meru kontingencije. Međutim, $\Delta P$ kao mera kontingencije se posmatra i kao teorija ljudske ocene intenziteta kauzalnog odnosa - bez obzira na njen odnos prema

kauzalnim moćima. U psihološka istraživanja uvode je Vord i Dženkins (Ward & Jenkins, 1965a, Jenkins & Ward, 1965b, prema Allan, 1980), i od kako je uvedena $\Delta P$ ne prestaje da igra jednu od glavnih uloga u matematičkim teorijama učenja. Model probabilističkog kontrasta duboko je povezan sa asocijacionističkim teorijama učenja i temeljno proučen u tradiciji istraživanja *učenja kovarijacija* kod ljudi (Shanks, 2004). *Asocijacionističke teorije* ocene intenziteta kauzalnog odnosa, kojima sada posvećujemo pažnju, poreklom su iz te istorijski ranije, klasične epohe teorije učenja - preciznije, iz epohe razvoja matematičkih teorija klasičnog i operantnog uslovljavanja, teorija koje predstavljaju analogiju ljudskoj oceni kovarijacije i kauzalnosti u proučavanju ponašanja životinja.

Frekvencije broja slučajeva $a$, $b$, $c$ i $d$ - u kojima je prisustvo jedne varijable praćeno prisustvom ili odsustvom druge i obrnuto - u svakoj novoj kontingencijskoj tabeli definišu jedan osnovni *nacrt eksperimenta klasičnog* (respodentnog, Pavlovljevog) *uslovljavanja*. Za našu diskusiju neće biti značajno da li govorimo o jednom ili drugom osnovnom obliku učenja. U terminologiji asocijacionističkih teorija učenja, varijabla koju smo do sada nazivali uzrokom se češće naziva *znakom* (engl. *Cue*), a varijabla koju smo tretirali kao posledicu - ishodom (engl. *Outcome*). Ova razlika u terminologiji odslikava suštinsku razliku između asocijacionističkog tretmana kauzalnosti i tretmana koji ona dobija u teoriji kauzalnih modela. U teoriji kauzalnih modela, kao što smo videli, intenzitet kauzalnog odnosa je verovatnoća da će kauzalni mehanizam između dva fenomena biti aktiviran i dovesti do realizacije efekta. Bez reprezentacije putem kauzalnih mreža ovakvu tvrdnju nije ni moguće formalno izraziti. Nasuprot ovoj teoriji, asocijacionističke teorije ne tvrde ništa drugo do toga da se kauzalni odnosi uče i njihovi intenziteti procenjuju istim mehanizmima koji učestvuju u stvaranju S-S i S-R asocijacija u klasičnim asocijacionističkim teorijama učenja.

Model asocijativnog učenja koji su (prvo nezavisno jedan od drugog, a zatim zajednički) razvili Alan Reskorla i Ričard Vagner, poznat kao *Reskorla-Vagner* (skr. *RW*) *model*, predstavlja sigurno najčešće diskutovanu asocijacionističku teoriju (Rescorla & Wagner, 1972). U početku razvijen da objasni fenomene poput osenčavanja i blokiranja u eksperimentima klasičnom uslovljavanja, ovaj model je zahvaljujući dobrom odnosu jednostavnosti i eksplanatorne moći postao neka vrsta standardne asocijacionističke teorije učenja. Svi kasniji asocijacionistički modeli su, na jedan ili drugi način, počivali na logici RW modela. Formalno, RW model je dat na sledeći način:

$$\Delta V_x = \alpha_x \beta (\lambda - \sum_i V_i) \tag{24}$$

gde su $\alpha_x, \beta$ parametri rate učenja (engl. *learning rate*) koji su određeni suštinskim osobinama uslovnog ($\alpha$) i bezuslovnog ($\beta$) stimulusa, a parametar $\lambda$ određuje maksimalan stepen asocijativne moći koje uslovni stimulus može da stekne (odn. koji određeni bezuslovni stimulus podržava). Parametri $\alpha_x, \beta$ uzimaju vrednosti između 0 i 1 i suštinski modeliraju one osobine stimulusa, koje - poput intenziteta - mogu da utiču na brzinu razvoja asocijativne veze. Parametar $\lambda$ određuje maksimalnu asocijativnu moć koju podržava određeni bezuslovni stimulus. Na primer, veoma bolan elektrošok bi podržavao veću asocijativnu moć koja bi mogla da bude raspodeljena između potencijalno velikog broja uslovnih stimulusa (znakova) koji bi signalizirali njegovu pojavu od nekog tek veoma slabog elektrošoka. Tako $\lambda$ određuje *asimptotu učenja*, pojam od velikog značaja u RW modelu. U psihologiji učenja dobro je poznato da ova osobina RW modela omogućava objašnjenje pojave blokiranja (Kamin, 1969, prema Rescorla & Wagner, 1972) koju nijedna teorija pre RW nije uspela da objasni na zadovoljavajuć način. RW model izračunava $\Delta V_x$, promenu u intenzitetu asocijativne veze između znaka (uslovnog stimulusa) $x$ i bezuslovnog stimulusa, posle određenog broja situacija u kojima je znak praćen ili nije praćen ishodom (bezuslovnim stimulusom) i situacija u kojima se ishod pojavljuje bez znaka, ili oba odsustvuju. To su upravo situacije koje predstavlja svaka kontingencijska tabela. Učenje se odvija tako što se pri svakom sledećem pokušaju (engl. *trial*) koji može da spada u jednu od četiri moguće situacije ustanovljava kolika je razlika između maksimalne moguće asocijativne moći (asimptote) $\lambda$ i sume asocijativne moći svih znakova (uslovnih stimulusa) koji učestvuju u eksperimentu, u oznaci $\sum_i V_i$. Termin ove razlike predstavlja *grešku predikcije* (engl. *prediction error*) i ona je osnovni termin u svim modelima učenja koji se baziraju na R-W. Kada se dostigne asimptotska vrednost, greška predikcije postaje nula, i učenje se završava. RW model je tipičan primer dinamičkog modela koji menja svoju vrednost posle svakog diskretnog koraka u eksperimentu (svaki diskretni korak obuhvata ekspoziciju jedne od četiri moguće situacije odn. kombinacije prisustva i odsustva znaka i ishoda). Njegova stabilna, asimptotska stanja bila su predmet mnogih studija zahvaljujući kojima su danas modeliranje RW modelom i simulacije odgovarajućih eksperimenata bitno olakšani. U primenama RW modela uobičajeno je da parametar $\lambda$ uzima vrednost 1 u situacijama kada je

ishod prisutan i 0 kada nije prisutan. Međutim, parametar $\beta$ se takođe odnosi na karakteristike bezuslovnog stimulusa odn. ishoda. Ukoliko se RW model ograniči tako da parametar $\beta$ uzima istu vrednost kada je ishod prisutan i kada je odsutan (situacije $a$ i $c$, odn. $b$ i $d$ u kontingencijskog tabeli), asimptotska vrednost jednačine (24) postaje $\Delta P$ (Chapman & Robins, 1990). Radovi Čengove (Cheng, 1997) i Denksa (Danks, 2003) pokazali su da RW model ima asimptotsku vrednost u $\Delta P$ za široku klasu različitih nacrta učenja kovarijacija ili klasičnog uslovljavanja, a poznato je i da interakcije više različitih uslovnih stimulusa - efekte za čije je objašnjenje RW model prvobitno razvijen - vode ka tome da ovaj model ima asimptotske vrednosti u vrednostima *uslovnih* $\Delta P$ (v. Spellman, 1996, za diskusiju uslovnih $\Delta P$, v. Tangen & Allan, 2003, za dokaz ove tvrdnje). Za široku klasu mogućih nacrta od kojih svaki definiše određena kontingencijska tabela, dakle, RW model koincidira u svojim predikcijama sa modelom probabilističkog kontrasta. Ipak, njih ne treba poistovećivati.

Pristup asocijacionističkih teorija problemu kauzalnog učenja je taj da se ono posmatra kao slučaj asocijativnog učenja u kome uzrok (ili uzroci) igraju ulogu uslovnih, a posledice ulogu bezuslovnih stimulusa. Predikcija ovih modela je da snaga asocijativne veze, koju svaki od njih izračunava na različit način, koincidira sa procenama o snazi kauzalnog odnosa ljudskih ispitanika. Do sada je bilo malo reči o različitim metodama merenja intenziteta kauzalnog odnosa; pokazaće se da razlike među njima mogu da igraju ključnu ulogu u diskusijama odnosa normativnih i deskriptivnih teorija. Za sada, očigledno je da ono što se u ljudskoj psihologiji naziva učenjem kovarijacija jeste metodološka analogija klasičnom i operantnom uslovljavanju u psihologiji životinja (mada se obe paradigme primenjuju i u ljudskoj psihologiji). U psihologiji životinja, tipični indikatori učenja su frekvencija i intenzitet razvijene (uslovne) reakcije, dok od ljudi u eksperimentalnim uslovima zahtevamo da na određenoj skali, pod određenom procedurom, donesu svoju procenu o intenzitetu povezanosti između neka dva fenomena, u slučaju paradigme učenja kovarijacija, ili o intenzitetu kauzalnog odnosa, u slučaju kauzalnog učenja. Prema asocijacionističkim teorijama, ne postoji suštinska razlika između ove dve paradigme u ljudskoj psihologiji. Pristalice teorije kauzalnih modela i kauzalne moći, naravno, baziraju svoje stavove upravo na tvrdnji da kauzalno učenje nije asocijativni proces i da ne može da se svede na učenje asocijacija između dve varijable.

Činjenica je da postoje ubedljivi empirijski argumenti u prilog pristalica teorije kauzalnih modela. U danas već klasičnom radu, Voldman i Holiouk su pokazali

da fenomeni koje predviđaju asocijacionističke teorije ne uzimaju istu formu kada se znaci i ishodi u učenju interpretiraju kao pravi uzroci i posledice (Waldmann & Holyoak, 1992). Koristeći kauzalne mreže zajedničke posledice, u kojima je definisano više uzroka jedne iste posledice, i kauzalnih viljuški, u kojima je definisan jedan isti uzrok za više posledica, Voldman i Holiuk uvode eksperimentalne paradigme *dijagnostičkog* i *prediktivnog učenja*. U dijagnostičkom učenju, jedan uzrok ima više posledica; u prediktivnom, više uzroka ima jednu zajedničku posledicu. U obe paradigme koriste se iste varijable, a asocijativne veze među njima date su potpuno istovetnim kontingencijskim tabelama. Voldman i Holiouk pokazuju da se poznat fenomen (sličan osenčavanju i blokiranju u klasičnom uslovljavanju) da asocijativna snaga jednog znaka (uslovnog stimulsa) opada uvođenjem dodatnog znaka koji je prediktivan za isti ishod (a) javlja u situaciji prediktivnog učenja, kada više uzroka predviđa pojavu iste posledice, ali (b) ne javlja u situaciji dijagnostičkog učenja, kada jedan uzrok ima više posledica. Drugim rečima, u predviđanju zajedničke posledice, više njenih uzroka međusobno se blokiraju, što rezultira u nižim procenama intenziteta veze između svakog pojedinačno i zajedničke posledice. S druge strane, više efekata jednog istog uzroka se međusobno ne blokiraju, odn. svaki od njih sa tim istim uzrokom razvija u potpunosti snagu odgovarajućeg kauzalnog odnosa. Pošto su u obe eksperimentalne paradigme korišćene potpuno istovetne varijable sa istovetnim nacrtom učenja kovarijacija, a paradigme dijagnostičkog i prediktivnog učenja razvijene tako što su jednostavno obrnute uloge uzroka i posledica (uzrok u dijagnostičkom učenju postaje posledica u prediktivnom, a posledice u dijagnostičkom učenju postaju uzroci u prediktivnom), Voldman i Holiouk zaključuju da se ljudski kognitivni sistem u učenju oslanja na razumevanje strukture odgovarajućeg kauzalnog modela i ne zasniva svoje sudove samo na kovarijaciji između promenljivih. Suštinska razlika između teorije kauzalnih modela i asocijacionističkih teorija je u *usmernosti* veze između dva fenomena: dok je asocijativna veza između nekog $C$ i $E$ u asocijacionističkim teorijama potpuno simetrična, u teoriji kauzalnog modela ona je usmerena: C$\mapsto$E, i ne E$\mapsto$C; ovo je suštinska asimetrija na kojoj počiva bilo koja teorija koja kauzalne odnose posmatra kao fundamentalne, a ne kao posledice interpretacije pukih kovarijacija između varijabli u okolini.

Vratimo se asocijacionističkim teorijama učenja. RW model doživeo je nekoliko korekcija u pokušajima da se njegova eksplanatorna moć proširi na neke efekte koje osnovna verzija nije mogla da objasni; poznatije modifikacije su Van-Hame i

Vasermanova (Van-Hamme & Wasserman, 1994) i Pirsova (Pearce, 1987), bazirana na nekada popularnoj teoriji Pirsa i Hola (Pearce & Hall, 1980) i još starijem konfiguralnom modelu Etkinsona i Estesa (Atkinson & Estes, 1963, prema Pearce, 1987). Pirsova modifikacija RW modela je formalno suviše složena da bismo je ovde diskutovali detaljno, ali treba napomenuti da je u pitanju više nego temeljna revizija pretpostavki na kojima počiva RW model. Iako Pirsov model konačno razvija intenzitete asocijativnih veza kroz modifikaciju jednačine RW modela (24), pretpostavke na kojima počiva i mehanizmi učenja koje obuhvata ga zapravo čine jedinstvenom i specifičnom teorijom učenja. Razvijen je i metod asimptotske ocene (inače u primeni ne tako jednostavne) Pirsove teorije asocijativnog učenja (Perales & Shanks, 2007). Suštinski doprinos Pirsovog modela ogleda se u operaciji izračunavanja sličnosti između stimulusa u eksperimentu uslovljavanja ili učenja kovarijacija. Ovo izračunavanje sličnosti bazira se na (*i*) pravilima multplikativne prirode (koja ćemo sresti u diskusiji tzv. modela primeraka u kategorizaciji, v. 7.5) i (*ii*) jasno određenim odnosima između *konfiguracija* (ekspozicija više znakova tj. uslovnih stimulusa istovremeno) i pojedinačnih znakova u učenju.

Pored asocijacionističkih modela diskutovanih do sada, oblast učenja kovarijacija ponudila je i više jednostavnih heuristika koje se sve baziraju na istoj logici upotrebe informacija iz kontingencijske tabele. Sve ove heuristike uslovno možemo da podvedemo pod kategoriju asocijacionističkih teorija, jer svakako stoje u antagonističkom odnosu prema teoriji kauzalnih modela, a u krajnjem slučaju svaka od njih u praksi samo predstavlja način kako da se drugačije opiše kovariranje dve binarne promenljive. Još Inhelderova i Pijaže (Piagett & Inhelder, 1958, prema Allan, 1980), predlažu heuristiku za ocenu snage kauzalnog odnosa poznat kao $\Delta D$ :

$$\Delta D = (a + b) - (c + d) \tag{25}$$

koji očigledno predstavlja najjednostavniji način da se iskoriste konfirmatorne ($a$, $d$) i diskonfirmatorne ($c$, $b$) informacije iz kontingencijske tabele. Perales i Šenks predlažu nešto komplikovaniju verziju ove heuristike, za koju nalaze da ju je prvi primenio Bjuzmejer (Busemeyer, 1991, prema Perales & Shanks, 2007). Nazivaju je *integracijom evidencije* (engl. *Evidence Integration*):

$$EI = \frac{(w_a a + w_d d) - (w_c c + w_b b)}{w_a a + w_b b + w_c c + w_d d} \tag{26}$$

Iz forme ovog asocijacionističkog modela jasno je da se radi o ponderisanoj verziji $\Delta D$: četiri slobodna parametra, $w_a$, $w_b$, $w_c$ i $w_d$ određuju pondere (značaj) frekvencija iz svake od četiri ćelije kontingencijske tabele; razlika između konfirmatorne i diskonfirmatorne informacije se normalizuje sumom ponderisanih frekvencija u imeniocu. Pored navedenih, postoji još načina da se izračunavaju zavisnosti binarnih varijabli iz kontingencijskih tabela; praktično svaki od njih je testiran bilo u oblasti učenja kovarijacija, bilo kao model kauzalnog učenja (Allan, 1980; up. Hattori & Oaksford, 2007, za iscrpan pregled pravila izračunavanja stepena kovarijacije iz 2×2 kontingencijskih tabela). Značaj pojedinih informacija iz kontingencijske tabele u učenju kovarijacija i kauzalnih odnosa kod ljudi nije jednak, i to je ono što motiviše ideje poput upravo uvedenog modela integracije evidencije. Naime, još su Šustak i Sternberg u poznatom radu iz 1981. godine primenom jednostavnog linearnog multipla-regresionog modela utvrdili stepen značaja koji ljudi pridaju frekvencijama ćelijama kontingencijske tabele u oceni snage kauzalnog odnosa: $a > b > c > d$; ovaj redosled se pokazuje se kao jedan od najkonzistentnijih nalaza u oblasti kauzalnog učenja uopšte (Schustack & Sternberg, 1981, up. Hattori & Oaksford, 2007 - nalaz o odnosu $a > b > c > d$ poznat je od najranijih eksperimentalnih istraživanja kauzalnog učenja i učenja kovarijacija). Međutim, ni informacije iz kontingencijske tabele izražene kao zavisne verovatnoće, poput onih koje koristimo u izračunavanju probabilističkog kontrasta $\Delta P$, ne doprinose ljudskim procenama intenziteta kauzalnih odnosa podjednako: u ljudskim procenama, očigledno je da veću težinu dobija vrednost P($E|C$) nego P($E|\widetilde{C}$) (up. Perales & Shanks, 2007, za izuzetnu ilustraciju značaja ovog nalaza; Lober & Shanks, 2000, za eksperimentalnu demonstraciju). Ovaj empirijski nalaz motiviše uvođenje još jednog ne-normativnog, asocijacionističkog modela, modela ponderisanog $\Delta P$:

$$w\Delta P = w_1 P(E|C) - w_2 P(E|\widetilde{C}) \tag{27}$$

Slobodni parametri ovog modela su $w_1$ i $w_2$, ponderi na dve odgovarajuće zavisne verovatnoće za koje se očekuje odnos $w_1 > w_2$; ako su njihove vrednosti 1, model je očigledno ekvivalentan probabilističkom kontrastu.

*Empirijske performanse modela kauzalnog učenja.* Svi prethodno diskutovani modeli ocene intenziteta kauzalnog odnosa između dve varijable povdrgavaju se strogim eksperimentalnim testovima već decenijama. I posle velikog broja eksperimentalnog studija i detaljnog rada na razvoju formalnih modela ove

kognitivne funkcije još uvek se nalazimo daleko od zaključka o tome kako ljudski kognitivni sistem uči o probabilističkim kauzalnim odnosima. Meta-analize velikog broja studija pokazuju da jednostavne heuristike (poput EI definisane jednačinom (26)), koji linearnim ponderisanjem kombinuju evidenciju o kauzalnom odnosu datu kontingencijskom tabelom, pokazuju bolje empirijske performanse od teorijski utemeljenih i složenih kauzalnih i asocijacionističkih modela (Perales & Shanks, 2007; vidi još Hattori & Oaksford, 2007). S druge strane, najnovija istraživanja pokazuju da bejzijanske verzije modela zasnovanih na teoriji kauzalnih mreža pokazuju empirijske performanse koje su u najmanju ruku uporedive sa performansama takvih heuristika (Lu et al, 2008).

U svakom eksperimentu kauzalnog učenja, ispitanicima se na neki način predstavljaju informacije koje sadrži određena kontingencijska tabela, a od njih se zahteva da kroz određenu proceduru daju svoju procenu o snazi odgovarajućeg kauzalnog odnosa. Prezentacijom različitih kontingencijskih tabela prikuplja se više procena intenziteta kauzalnih odnosa, a prosečne procene više ispitanika se koriste kao eksperimentalni podaci kojima se testiraju odgovarajući formalni modeli. Svi modeli koje smo diskutovali koriste samo frekvencije iz kontingencijske tabele (i ponekad slobodne parametre) u izračunavanju snage kauzalnog odnosa; regresiona analiza se koristi da bi se evaluirale predikcije tih modela. Ispitanicima se u tipičnom istraživanju objašnjava određena situacija koja bi mogla da odgovara nekom realnom problemu, na primer da nauka istražuje dejstvo određenog hemijskog jedinjenja na stepen mutacija kod određenog tipa bakterija, i da će im biti prikazani rezultati određenog broj testova u kojima su bakterije izložene dejstvu tog jedinjenja ili nisu, i u kojima je došlo ili nije došlo do mutacija. Na osnovu ovako ili slično formulisanog scenarija, očekuje se da ispitanici razviju reprezentaciju potencijalnog uzroka i njegove posledice, i da na osnovu informacija o njihovog kovarijaciji mogu da donesu sud o intenzitetu kauzalnog odnosa među njima. Većina novijih eksperimentalnih studija u oblasti kauzalnog učenja fokusira se na testove teorije kauzalne moći (Cheng, 1997) i pokušaje dizajniranja krucijalnih eksperimenata koji bi obezbedili istovremenu falsifikaciju ove teorije i podršku za model probabilističkog kontrasta ili obrnuto. Analize rezultata ovakvih studija pokazuju da postoje nacrti u kojima, pri istim vrednostima kauzalne moći, ispitanici daju različite procene intenziteta kauzalnih odnosa, najčešće pod uticajem vrednosti zavisnih verovatnoća $\mathrm{P}(E|C)$ i $\mathrm{P}(E|\widetilde{C})$ (npr. Wasserman, Elek, Chatlosh, & Baker, 1993, prema naknadnim analizama koje diskutuje Shanks, 2004). U takvim studijama tipičan rezultat

pokazuje da je $\Delta P$ nešto bolji prediktor bihejvioralnih procena snage kauzalnih odnosa od kauzalne moći.

Postoje dva osnovna pristupa deskripciji informacija iz kontingencijske tabele i nekoliko različitih eksperimentalnih procedura kroz koje variraju načini davanja odgovora i prezentacija samih stimulusa ispitanicima. Podaci iz kontingencijske tabele mogu da se daju slučaj po slučaj (engl. *trial-by-trial*). U ovakvoj proceduri, ispitanici vide npr. rezultate jednog testa u kome hemijsko jedinjenje jeste primenjeno na uzorak bakterija i dobijaju informaciju o tome da li su bakterije iz tog uzorka mutirale ili nisu. U sledećem slučaju, npr. prikazuju im se rezultati testa koji svedoče o tome da se mutacija spontano javila u uzorku bakterija koje nisu bile izložene dejstvu odgovarajuće hemikalije, itd. Više studija koristi ovaj način prezentacije informacija (npr. eksperiment 3 u Buehner, Cheng & Clifford, 2003; veliki broj studija koje koriste ovu metodu prezentacije uključene su u noviju meta-analizu Peralesa i Šenksa, Perales & Shanks, 2007). Smatra se da prikazivanje informacija slučaj po slučaj dovodi do konfundacije nekontrolisanih varijabli koje se odnose na memorijske kapacitete i procesa kauzalnog učenja[41] (Buehner, Cheng & Clifford, 2003). Taj problem može da reši *sumarno prikazivanje informacija* (engl. *summary presentation format*). Studije koje koriste ovaj pristup prikazivanju informacija o kovarijaciji (npr. eksperiment 2 u Buehner, Cheng & Clifford, 2003) najčešće na jednom panelu (odštampanom, ili na ekranu kompjutera) prikazuju vizuelnim sredstvima slučajeve koji odgovaraju $a$, $b$, $c$ i $d$ ćelijama kontingencijske tabele; u nekim slučajevima, materijal je vizuelno organizovan tako da jasno razdvaja slučajeve koji odgovaraju istoj ćeliji tabele u vizuelnom polju ispitanika. Ovako organizovan vizuelni materijal verovatno olakšava kognitivim procesima da ocene proporcije različitih slučajeva koje odgovaraju zavisnim verovatnoćama $\mathrm{P}(E|C)$ i $\mathrm{P}(E|\widetilde{C})$, favorizujući tako klasu modela poput raznih verzija $\Delta P$ i kauzalne moći (eksperimenti 3a i 3b u studiji Perales & Shanks, 2008, sistematski ispituju razliku između slučajnog rasporeda slučajeva u sumarnoj prezentaciji i rasporedu koji grupiše slučajeve koji odgovaraju određenim ćelijama kontingencijske tabele). S druge strane, asocijacionistički modeli učenja poput RW modela ili Pirsovog modela mogu da se primene samo u slučaju sekvencijalnog, slučaj po slučaj prikazivanja informacija; jednačine ovih modela koriguju snage odgovarajućih asocijativnih veza korak po korak, upravo odgovarajući ovakvoj prezentaciji podataka[42]. Ređe se koristi mešoviti format prezentacije, u kome se ispitanicima na ekranu kompjutera prikazuje slučaj po slučaj, ali tako da svaki slučaj posle prikazivanja ostaje na

ekranu, što konačno vodi ka sumarnoj prezentaciji do koje se stiže kroz slučaj po slučaj proceduru prikazivanja (npr. eksperiment 1 u studiji Buehner, Cheng & Clifford, 2003; nije nam poznato da je ovakav format prezentacije korišćen i u drugim studijama).

U eksperimentima kauzalnog učenja posebna pažnja se poklanja načinu na koji je formulisano pitanje koje ispitanika treba da vodi ka proceni snage kauzalnog odnosa. Pretpostavimo da posle prezentacije kontingencijskih informacija ispitanicima postavimo pitanje „*U kojoj meri uzrok C dovodi do pojave efekta E?*“, ili neku sličnu varijantu ovog pitanja. Buener i saradnici problematizuju ovakav pristup postavljanju tzv. kauzalnog pitanja (Buehner, Cheng & Clifford, 2003). Ukoliko ispitanici interpretiraju ovo pitanje kao pitanje o tome u kojoj meri $C$ izaziva $E$ u prisustvu drugih mogućih uzroka, normativno adekvatan odgovor na njega predstavlja probabilistički kontrast $\Delta P$. Ukoliko određeni efekat $E$ ima verovatnoću spontanog javljanja od 10% u nekom uzorku (gde pod „spontanim javljanjem“ podrazumevamo da do njega dovodi konglomerat pozadinskih uzroka $B$), vrednost od $\Delta P = .20$ između nekog potencijalnog uzroka $C$ i tog efekta $E$ doslovce znači da bi uvođenjem dejstva $C$ stopa javljanja $E$ skočila na 30%. Za razliku od ovog efekta koji $\Delta P$ ima na kauzalno učenje, dejstvo uzroka kauzalne moći određenog intenziteta na verovatnoću pojave efekta u prisustvu pozadinskih uzroka $B$ je nelinearno i ne podleže prethodnom opisu. Tek ako se postavi pitanje o tome u kojoj meri uzrok $C$ izaziva efekat $E$ *u izolaciji od drugih potencijalnih uzroka*, normativno adekvatan odgovor predstavlja mera kauzalne moći. Buener i saradnici zato predlažu da se kauzalno pitanje za test teorije kauzalne moći formuliše kao: „*U koliko od 100 slučajeva u kojima efekat E nije prisutan bi se on pojavio kada bi smo uveli dejstvo uzroka C?*“. Ovako postavljeno pitanje naziva se *kontrafaktualnim kauzalnim pitanjem*. Razlika između ova dva pitanja, bar *prima facie*, nije trivijalna u odnosu na test teorije kauzalne moći. Primetimo da oba pitanja podrazumevaju da su ispitanici sigurni da *C jeste uzrok E*: ona se ne odnose na situaciju u kojoj ispitanici mogu da budu više ili manje sigurni u to da li kauzalni odnos uopšte postoji. Ovo pitanje smo na početku rasprave o kauzalnom učenju predstavili kao problem učenja kauzalne strukture. Grifits i Tenebaum će u razvoju njihovog bejzijanskog modela kauzalne podrške dodatno zakomplikovati problem upravo pretpostavljajući da kognitivni sistem zapravo sud o kauzalnosti donosi kao sud o snazi evidencije da kauzalni odnos uopšte postoji (Griffiths & Tenenbaum, 2005). Eksperimentalne studije Buenera i saradnika (koja predstavlja

najjaču evidenciju u prilog teoriji kauzalne moći do sada), kao i Kolinsa i Šenksa (Collins & Shanks, 2006) zaista su potvrdile da kauzalna moć daje tačne predikcije bihejvioralne ocene intenziteta kauzalnog odnosa kada se kauzalno pitanje postavi u kontrafaktualnoj formi. Eksperimentalni testovi alternativnih teorija kauzalnog učenja baziraju se na situacijama u kojima istraživači pokušavaju da različitim kontingencijskim tabelama koje karakterišu iste vrednosti jedne ili druge teorije (npr. iste vrednosti probabilističkog kontrasta ili kauzalne moći) isprovociraju različite bihejvioralne ocene intenziteta kauzalne veze: ovakav nalaz definitivno falsifikuje jednu ili drugu teoriju koja se testira. Upravo ovakav nacrt koristili su Kolins i Šenks kada su pokazali da za različite kontingencijse tabele sa istom vrednošću kauzalne moći ispitanici daju različite procene intenziteta kauzalne veze kada se postavi standardno kauzalno pitanje, ali iste procene kada se pitanje da u kontrafaktualnoj formi (koja favorizuje teoriju kauzalne moći, Collins & Shanks, 2006). Ipak, u studiji Peralesa i Šenksa iz 2008, ovi nalazi su dovedeni u pitanje. Potonja studija koristi nešto drugačije kontingencijske tabele od onih koje je koristila studija Kolinsa i Šenksa iz 2006 (razlike između relevantnih zavisnih verovatnoća za izračunavanje$\Delta P$ i kauzalne moći su veće u studiji iz 2008). Perales i Šenks koriste sekvencijalnu, slučaj po slučaj prezentaciju kontingencije, u kojoj ispitanici za svaki prikazani slučaj prvo pokušavaju da predvide da li će se efekat pojaviti ili neće, da bi potom saznali da li je u tom slučaju efekat bio ili nije bio prisutan, i dobili fidbek o svom predviđanju kao tačnom ili netačnom. Kao što vidimo, i eksperimentalna procedura koja se koristi u novoj studiji Peralesa i Šenksa iz 2008. je drugačija od prethodno diskutovanih (prisustvo predikcije i fidbeka u učenju, mada njihova primena nije retkost u studijama ovog tipa; Šenks veruje da je prezentacija kontingencije slučaj po slučaj ekološki relevantnija od sumarnih prezentacija, Shanks, 2004). Dvema grupama ispitanika su postavljena različita pitanja: standardno kauzalno pitanje i kontrafaktualno kauzalno pitanje. Perales i Šenks su kroz eksperimentalnu proceduru u eksperimentu 1 uveli više olakšavajućih okolnosti koje bi trebalo da favorizuju donošenje suda na osnovu kauzalne moći, ali njihovi rezultati svejedno pokazuju da ispitanici daju različite ocene intenziteta kauzalnih veza za različite kontingencijske tabele sa *istim* vrednostima kauzalne moći (up. eksperiment 1, Perales & Shanks, 2008). Rezultati ove važne eksperimentalne studije pokazuju da ni posle dopunskih manipulacija oblicima kauzalnog pitanja dizajniranih da obezbede procenu intenziteta kauzalne veze uzroka C u izolaciji dobijene efekte nije moguće eliminisati (čak ni kroz modifikaciju$\Delta P$ i kauzalne

moći uključivanjem slobodnih parametara poput onih u ponderisanom $\Delta P$ modelu, jednačina (27)). Čak ni posle prilagođavanja eksperimentalne procedure i formata prikazivanja informacija teoriji kauzalne moći (upotrebom sumarne prezentacije u eksperimentima 3a i 3b, Perales & Shanks, 2008) u ovoj studiji nije obezbeđena empirijska podrška za ovaj model[43]. Prethodno predstavljen model integracije evidencije (jednačina (26)) omogućava najbolje objašnjenje bihejvioralnih procena intenziteta kauzalnih odnosa u zajedničkom skupu eksperimentalnih podataka koji čine upravo diskutovana studija Peralesa i Šenksa, studija Kolinsa i Šenksa (Collins & Shanks, 2006) i Buenera i saradnika (Buehner, Cheng & Clifford, 2003) - iako je u sve tri studije korišćeno kontrafaktualno pitanje koje, prema pretpostavci, favorizuje ocenu kauzalne moći (Perales & Shanks, 2007).

Najkompletniji zaključak o empirijskim performansama modela kauzalnog učenja kod ljudi donosi meta-analiza koju su sproveli Perales i Šenks (Perales & Shanks, 2007). Autori su prikupili rezultate velikog broja studija u kojima su dosledno korišćeni (a) sekvencijalni, slučaj po slučaj pristup prezentaciji kontingencijske tabele, i (b) standardno kauzalno pitanje, dakle pitanje koje nije kontrafaktualno i koje ne specifikuje kontekst u kome treba doneti sud o kauzalnom odnosu. Meta-analiza Perales i Šenksa je sprovedena na uzorku od 114 različitih kontingencijski tabela korišćenih u više studija; za svaku tabelu, prosečna procena intenziteta kauzalnog odnosa između dve binarne varijable merena je na skali od 0 do 100 (tj. - 100.. 0.. 100, pošto su obuhvaćeni i preventivni uzroci). Testirano je više modela procene intenziteta kauzalnog odnosa (od kojih je većina specifikovana u jednačinama (22)-(27)), uključujući i jedan hibridni model koji autori nazivaju *hibridnim pseudonormativnim modelom* (skr. HPN), forme

$$HPN = \gamma \Delta P' + (1 - \gamma) p'_c \tag{28}$$

gde su $\Delta P'$ i $p'_c$ ponderisane verzije probabilističkog kontrasta i kauzalne moći: $\Delta P'$ je izračunat uključujući dva pondera na $\mathrm{P}(E|C)$ i $\mathrm{P}(E|\widetilde{C})$ kao u ponderisanom modelu (27), a zatim u tom obliku iskorišćen da bi se izračunala vrednost ponderisane kauzalne moći $p'_c$. Parametar $\gamma$ igra ulogu ponderisanja *kombinacije* dva modela koji zajedno čine HPN. Model je, očigledno, pokušaj da se iskoristi sva prediktivna moć normativnih modela, bez obzira na tip kauzalnog pitanja koje favorizuje jedan ili drugi model, i to u generalizovanom obliku koji uključuje pondere na odgovarajućim zavisnim verovatnoćama. Perales i Šenks su ocenili parametre[44]

različitih modela kauzalnog učenja na polovini podataka iz 114 kontingencijskih tabela; zatim su linearnom regresionom analizom ocenili prediktivnu moć svih modela sa dobijenim vrednostima parametara na podacima iz druge polovine od 114 tabela u studiji. Ovaj postupak u poređenju različitih matematičkih modela efektivno otklanja prednost modela sa više slobodnih parametara.

Rezultati Peralesa i Šenksa pokazuju da teorijski dobro utemeljeni, normativni modeli probabilističkog kontrasta i kauzalne moći ne dostižu prediktivnu moć ($R^2$) jednostavnijih modela poput integracije evidencije (EI - jednačina (26)), asocijacionističke heuristike koji sa četiri slobodna parametra kontroliše doprinos svake od četiri frekvencije u kontingencijskoj tabeli procenama intenziteta kauzalne moći. Njihovi rezultati pokazuju da kroz veliki broj slučajeva EI model ima vrednost $R^2 = .89$, veću od $R^2 = .86$ koliko postiže kauzalna moć ili $R^2 = .80$ koliko postiže probabilistički kontrast $\Delta P$. Za njom zaostaju i složeni asocijacionistički modeli poput RW ($R^2 = .81$) i Pirsovog modela ($R^2 = .85$). Jedini model uporediv sa empirijskim performansama ove jednostavne heuristike jeste hibridni model HPN dat jednačinom (28), sa $R^2 = .88$.

Rezultati još jedne meta-analize ukazuju na to da jednostavne heuristike objašnjavaju ljudske procene intenziteta kauzalnih odnosa bolje od složenih kauzalnih ili asocijacionističkih modela - ili njihovih asimptota poput $\Delta P$. Hatori i Oaksford predlažu heuristiku pod imenom *heuristike dva faktora* (engl. *dual factor heuristic*) u sledećoj formi:

$$H = \sqrt{P(E|C)P(C|E)} \tag{29}$$

odn. geometrijske sredine zavisnih verovatnoća $\mathrm{P}(E|C)$ i $\mathrm{P}(C|E)$ (Hattori & Oaksford, 2007). Heuristika H je *motivisana racionalno*: ona se javlja kao granična vrednost dobro poznatog koeficijenta korelacije $\phi$ pod pretpostavkom da frekvencija ćelije $d$ u kontingencijskoj tabeli *teži beskonačnosti*. Objasnimo ovo. Rezonovanje koje vodi ovoj ideji je sledeće: u realnom životu, mi smo retko kada izloženi kontingenciji nekog konkretnog uzroka i neke konkretne posledice tako da su nam sve informacije predstavljene odjednom, kao u kontingencijskoj tabeli. Uzroke i posledice mi upoznajemo sekvencijalno, kroz vreme, odn. srećemo ponekad poneku situaciju u kojoj su prisutni i uzrok i posledica, poneku u kojoj je prisutan samo uzrok, poneku u kojoj je prisutna samo posledica, a ogroman deo vremena su iz našeg

perceptivnog polja odsutni i uzroci i posledice o kojima govorimo. Ovo rezonovanje opravdava zašto u analizi kauzalnog učenja ima smisla pretpostaviti da kognitivni sistem zanemaruje ćeliju $d$, koja nosi frekvencije slučajeva u kojima su odsutni i $C$ i $E$. Ta frekvencija je apriori ogromna: veći deo života mi nemamo nikakav dodir ni sa bakterijama koje mutiraju ili ne, ni sa radijacijom koja je prisutna ili nije, o kojima eksperimentator u laboratoriji od nas zahteva da rezonujemo. Hatori i Oaksford se tako oslanjaju na *hipotezu o retkosti uzroka i posledica* (koju Oaksford koristi u probabilističkim objašnjenjima procesa rezonovanja, up. 7.4): ako je frekvencia $d$ izuzetno visoka, onda su vrednosti $\mathrm{P}(C)$ i $\mathrm{P}(E)$ nužno veoma male. Jednačina (29) se dobija kao rezultat jednostavne matematičke derivacije u kojoj se koeficijent korelacije $\phi$ izračunava pod pretpostavkom da $d \to \infty$ (v. Hattori & Oaksford, str. 770). Hatori i Oaksford testiraju heuristiku H uz veliki broj drugih modela kovarijacije i kauzalnog učenja, na podacima iz većeg broja eksperimentalnih studija, i pokazuju da H postiže srednju vrednost koeficijenta determinacije $R^2 = .90$. Heuristika H (i drugi modeli kovarijacije) u ovoj studiji po empirijskim performansama ostavljaju daleko iza sebe normativne modele poput $\Delta P$ ili kauzalne moći $p_c$.

*Bejzijanski pristup problemu kauzalnog učenja.* Najnovije doprinose formalnim modelima kauzalnog učenja predstavlja grupa bejzijanskih teorija koje sve dele zajednički konceptualni i matematički okvir[45] (Tenenbaum & Griffiths, 2001, Griffiths & Tenenbaum, 2005, 2009, Lu et al, 2007, 2008). Pogledajmo ponovo grafove 1 i 0 na Slici 14. Na grafu 1, uzrok $C$ je povezan sa efektom $E$, dok ja na grafu 0 sa efektom $E$ povezan samo konglomerat pozadinskih uzroka $B$. I dalje se držimo fundamentalne pretpostavke cele oblasti kauzalnog učenja da kognitivni sistem sud o snazi veze između $C$ i $E$ donosi na osnovu opservabilnih frekvencija odn. na osnovu kovarijacije dve binarne varijable. Način na koji su pitanje kauzalnog učenja tj. ocene snage kauzalnog odnosa konceptualno formulisali Tenenbaum i Grifits u razvoju svog modela *kauzalne podrške* (engl. *Causal Support*, Tenenbaum & Griffiths, 2001, Griffiths & Tenenbaum, 2005) je sledeći: za date frekvencije koje definišu kontingenciju $C$ i $E$, kako kognitivni sistem donosi odluku o tome da li *uopšte postoji kauzalna veza* između njih? Ovo pitanje smo još na početku rasprave definisali kao pitanje učenja kauzalne strukture: efektivno, kognitivni sistem koji rešava ovaj problem treba da donese odluku da li data evidencija (kontingencijska tabela) svedoči u prilog tome da je realnost opisana grafom 1 (u kome je $C$ stvarni uzrok *efekta* $E$) ili grafom 0 (u kome $C$ nije uopšte povezan sa efektom

$E$). Bejzijanski pristup rešavanju ovog problema se u modelu kauzalne podrške formalizuje na sledeći način:

$$\frac{logP(graf\,1|D)}{logP(graf\,0|D)} = \frac{logP(D|graf\,1)}{logP(D|graf\,0)} + \frac{logP(graf\,1)}{logP(graf\,0)} \tag{30}$$

i odmah interpretiramo ovaj na prvi pogled neuobičajen način da se izrazi bejzijanski model: model je dat u logaritamskoj formi i otud je forma Bejzove teoreme ovde aditivna a ne multiplikativna, kako je uobičajeno. Simbol $D$ označava podatke koje nosi kontingencijska tabela, odnosno frekvencije slučajeva $a$, $b$, $c$ i $d$. Model dat u formi logaritamskih odnosa formalizuje, sa leve strane, odnos *a posteriori* verovatnoća (a) da je graf 1 (uzrok $C$ je stvarno povezan sa efektom $E$) tačan opis realnosti ako su dati podaci $D$ u kontigencijskoj tabeli i (b) da je graf 0 (uzrok $C$ nije povezan sa efektom $E$) tačan opis realnosti ako su dati isti ti podaci $D$. Prvi član sa desne strane jednačine (29) jeste odnos funkcija verodostojnosti, odn. verovatnoća da podaci kakve nosi $D$ jesu generisani strukturom grafa 1 ili grafa 0. Konačno, drugi logaritamski odnos sa desne strane jednačine (29) prikazuje odnose apriori verovatnoća grafa 1 i grafa 0: Tenenbaum i Grifits pretpostavljaju da ljudski kognitivni sistem ne pravi nikakve pretpostavke o tome da li je struktura grafa 1 u okolini više ili manje verovatna od strukture grafa 0, te pretpostavljaju da su ove apriori verovatnoće uniformne. Pod ovom pretpostavkom, izračunavanje *a posteriori* verovatnoća nije ni potrebno: kauzalna podrška, koja u ovoj teoriji predstavlja meru snage kauzalnog odnosa, jeste član $\frac{logP(D|graf\,1)}{logP(D|graf\,0)}$, poznat u teoriji bejzijanske statistike još i kao Bejzov faktor (Griffiths & Tenenbaum, 2005). Ovaj izraz nije ništa drugo do logaritam odnosa dve verodostojnosti, odn. odnosa verovatnoće da je datu kontingencijsku tabelu generisao graf 1 (prema kome kauzalni odnos zaista postoji) i da je tu tabelu generisao graf 0 (prema kome kauzalnog odnosa između $C$ i $E$ zapravo nema). Primetimo da ovakvo rezonovanje u potpunosti prati formu *kontrafaktualnog rezonovanja* o kauzalnim odnosima: poređenjem hipotetičkog sveta grafa 1, u kome $C$ jeste uzrok $E$ sa, hipotetičkim svetom grafa 0, u kome $C$ nije uzrok $E$, kognitivni sistem „simulira“ eksperiment u kome se utvrđuje doprinos uvođenja varijable $C$ pojavi varijable $E$ pošto je već ocenjena njena verovatnoća u odsustvu $C$. Kognitivni sistem, dakle, donosi odluku o strukturi kauzalnih odnosa, odn. sud o tome da li data kontingencija svedoči o tome da $C$ zaista jeste uzrok $E$; kauzalna podrška je mera stepena poverenja u taj sud.

Model kauzalne podrške dat u formi jednačine (29) odslikava samo opštu formu

ove teorije. Da bi se izračunala verovatnoća da su podaci u tabeli $D$ generisani grafom 1 ili grafom 0, potrebne su dopunske pretpostavke. Te pretpostavke su u ovoj teoriji iste kao pretpostavke koje pravi Čengova u razvoju teorije kauzalne moći (Cheng, 1997), tako da Tenenbaum i Grifits podrazumevaju da su kauzalne moći mere intenziteta dejstva uzroka na posledice. Pod ovakvom pretpostavkom, pokazuje se da je moguće izračunati verovatnoće P($D|graf\ 1$) i P($D|graf\ 0$) u jednačini (29) na sledeći način:

$$P(D|graf\,1) = \int_0^1 \int_0^1 P(D|b, c, graf\,1) P(b, c|graf\,1)\, db\, dc \tag{31}$$

$$P(D|graf\,0) = \int_0^1 P(D|b, graf\,0) P(b|graf\,0)\, db \tag{32}$$

gde je moguće izračunati zavisne verovatnoće (a) da su dobijene frekvencije u kontingencijskoj tabeli $D$ pod grafom 1 sa kauzalnim moćima $b$ i $c$ (up. Sliku 14), odn. $P(D|b, c, graf\,1)$ i (b) da su dobijene frekvencije u kontingencijskoj tabeli $D$ pod grafom 0 sa kauzalnom moći $b$, odn. $P(D|b, graf\,0)$, zahvaljujući vezi između kauzalne moći i frekvencija u kontingencijskoj tabeli koje je ustanovila još Čengova 1997 (jednačine (18-22)). Verovatnoće $P(b, c|graf\,1)$ u jednačini (30) i $P(b|graf\,0)$ u jednačini (31) su apriori verovatnoće da kauzalne moći $c$ i $b$ uzimaju određenu vrednost u zavisnosti od toga da li realnost odgovara grafu 1 ili grafu 0; kao što smo rekli, Tenenbaum i Grifits pretpostavljaju da su ove verovatnoće uniformno distribuirane, tj. da su sve moguće vrednosti ovih kauzalnih vrednosti apriori jednake. Dok je jednačinu (31) moguće izračunati analitički, jednačina (31) zahteva numeričku aproksimaciju; Grifits i Tenenbaum predlažu i algoritam za aproksimaciju njene vrednosti (up. dodatak A.3., Griffiths & Tenenbaum, 2005).

Matematički, model kauzalne podrške svakako nije najjednostavnija teorija kauzalnog učenja koju neko može da zamisli, ali je ideja na kojoj počiva u suštini jednostavna i elegantna. Model su u tri eksperimenta evaluirali Grifits i Tenenbaum i pokazali njegovu superiornost u odnosu na $\triangle P$ i kauzalnu moć (eksperimenti 1-3, Griffiths & Tenenbaum, 2005; njihov eksperiment 4 koristi donekle izmenjen model). Meta-analiza Peralesa i Šenksa iz 2007. godine, međutim, nalazi veoma malo empirijske podrške za ovaj model: kauzalna podrška ima ubedljivo najniži $R^2$ u uzorku od 114 analiziranih eksperimenata kauzalnog učenja[46] u ovoj studiji.

Ovako definisan bejzijanski model kauzalne podrške doveo je do određenih nesuglasica među pristalicama teorije kauzalnih modela. Lu i saradnici (Lu et al,

2007, 2008) primećuju da pitanje učenja kauzalne strukture nije isto što i pitanje ocene snage kauzalne veze. Ako se problem rešava bejzijanskim pristupom, prirodni način njegovog rešavanja u okviru teorije kauzalnih modela jeste izračunavanje *a posteriori* distribucije kauzalne moći uzroka $C$ pod pretpostavkom da kognitivni sistem ima neku relevantnu ocenu distribucije njene verovatnoće apriori. Konkretno, ako formulišemo ovu *a posteriori* verovatnoću (bez normalizacione konstante) kao

$$P(c|D, graf\,1) \propto \int_0^1 P(D|b, c, graf\,1)P(b, c|graf\,1)db \tag{33}$$

onda kognitivni sistem treba da izračuna očekivanu vrednost (tj. prosek) *a posteriori* distribucije kauzalne moći $c$ kao

$$\bar{c} = \int_0^1 c \cdot P(D|c, graf\,1)dc \tag{34}$$

gde se izraz sa desne strane jednačine (33) dobija integraljenjem po kauzalnoj moći pozadinskih uzroka $b$ u jednačini (32). Prema Luovoj i saradnicima, prosečna vrednost *a posteriori* distribucije kauzalne moći predstavlja ocenu intenziteta kauzalnog odnosa koju bi kognitivni sistem trebalo da donese. Međutim, Luova i saradnici prave dodatnu pretpostavku u odnosu na model kauzalne podrške koji su definisali Tenenbaum i Griftis. Podsetimo se da model kauzalne podrške pretpostavlja da su apriori verovatnoće kauzalnih moći $c$ i $b$ *uniformne*, tj. da kognitivni sistem nema nikakvu pretpostavku o distribuciji kauzalne moći potencijalnih uzroka u svojoj okolini. Luova i saradnici predlažu model prema kome pre izlaganja procesu kauzalnog učenja kognitivni sistem pretpostavlja da su uzroci *nužni* i *dovoljni*[47] da izazovu svoje posledice. Takva pretpostavka oličena je u specifičnoj distribuciji apriori verovatnoća koja snažno favorizuje uzroke visoke kauzalne moći. Reder i Milovanović su u studiji iz 2007. godine potvrdili sličnu pretpostavku (Rehder & Milovanović, 2007). U složenoj studiji revizije kauzalnih uverenja sa više potencijalnih uzroka oni su ustanovili da je dobijene eksperimentalne podatke moguće objasniti tek ukoliko se pretpostavi da kognitivni sistem u proces kauzalnog učenja ulazi sa pretpostavkom da su potencijalni uzroci nužni i dovoljni, odn. da imaju visoku kauzalnu moć apriori. U studiji Redera i Milovanovića takva prepostavka nije eksplicitna, kao što je to slučaj u studiji Luove i saradnika, već ona sledi kao jedini mogući zaključak posle ocene parametara apriori distribucija kauzalnih moći u odgovarajućem bejzijanskom kauzalnom modelu.

Inkorporirajući pretpostavku o apriori intenzivnim kauzalnim moćima u bejzijanski model kauzalnog učenja, Luova i saradnici su pokazali njegovu superiornost u predikciji eksperimentalnog nalaza većeg broja studija, uključujući tu rezultate tri originalna eksperimenta pored podataka prikupljenih u meta-analizi Peralesa i Šenksa. Bejzijanski model sa odgovarajućim apriori distribucijama kauzalne moći postiže $R^2$=.88 sa ljudskim procenama intenziteta kauzalne moći u većem broju eksperimentalnih studija iz meta-analize Peralesa i Šenksa. Ne treba izgubiti iz vida suštinsku razliku između ovog modela i modela kauzalne podrške Tenenbauma i Grifitsa: dok model Luove i saradnika predstavlja *a posteriori kauzalnu moć* dobijenu procesom bejzijanske inferencije, model kauzalne podrške pretpostavlja da je ocena intenziteta kauzalne moći istovetna sudu o evidenciji koju kovarijacija pruža za odluku o realnoj kauzalnoj strukturi.

*Komentari o debati o racionalnosti kauzalnog učenja.* Od početka 90-ih godina do danas, teorija kauzalnih modela, odn. primena kauzalnih grafova parametrizovanih odgovarajućim kauzalnim moćima, ostvarila je potpunu dominaciju kao normativni pristup problemu kauzalnog učenja u psihologiji. Uzimajući u obzir snažne empirijske dokaze da je ljudski kognitivni sistem preko kovarijacije osetljiv i na strukturu kauzalnih odnosa, kao i da pokazuje osetljivost na strukture složenije od onih u kojima učestvuju samo jedan uzrok i jedna posledica, nemoguće je potceniti doprinos teorije kauzalnih modela. U pitanju je prava promena paradigme, i prava naučna revolucija za kognitivne nauke, ako se uzme u obzir da primena kauzalnih modela počinje da nadilazi one probleme na koje je inicijalno primenjena (Tenenbaum, Griffiths & Kemp, 2006, Kemp & Tenenbaum, 2008, Tenenbaum et al, 2011). Grafički kauzalni modeli sa odgovarajućim parametrizacijama danas predstavljaju okosnicu nove paradigme u kompjutacionoj kognitivnoj psihologiji, paradigme koja omogućava modeliranje široke klase kognitivnih fenomena u zajedničkom formalnom jeziku kauzalnih mreža. U odnosu na pitanje kauzalnog učenja, možda sve što smo naveli navodi na zaključak da asocijacionističke modele jednostavno treba zameniti kauzalnim. Najvažnije ideje asocijacionističkih modela poput RW već dobijaju reformulaciju u odgovarajućim kauzalnim modelima; njihove prednosti, poput mogućnosti da objasne graduirane funkcije postepenog asocijativnog učenja sa procesijom potrekpljivanih i nepotkrepljivanih pojava uslovnih stimulusa, lako se uklapaju u dinamičke verzije kauzalnih modela (up. Danks, Griffiths & Tenenbaum, 2003).

Međutim, prethodni pregled empirijskih performansi asocijacionističkih i

kauzalnih modela, uz doprinos različitih ponuđenih heuristika (koje su sve, više ili manje, motivisane više asocijacionističkom nego kauzalnom logikom), sprečava nas da jednostavno popustimo pred elegancijom, racionalnim opravdanjem ili matematičkom koherencijom kauzalnih teorija. Normativni kauzalni modeli poput kauzalne moći ne omogućavaju predikciju eksperimentalnih nalaza u onoj meri u kojoj jednostavne heuristike poput diskutovanih H (Hattori & Oaksford, 2007) ili EI (Busemeyer, 1991, prema Perales & Shanks, 2007) to čine. Zašto bi kognitivni sistem birao da problem kauzalne indukcije, toliko važan adaptacioni problem u neizvesnoj okolini prožetoj probabilističkom kauzalnošću, rešava primenom veoma složenih arhitektura i zahtevnih reprezentacija koje odlikuju kauzalne modele, ako na raspolaganju može da ima jednostavne heurističke „prečice“ ka izračunavanju intenziteta kauzalnih odnosa? Ovako predstavljen, problem se svodi na debatu o racionalnosti u klasičnom ruhu: heuristici, ili normativna rešenja? Međutim, problem je ovde, verujemo, dublji i složeniji.

Slično kao u slučaju odlučivanja u uslovima rizika i neizvesnosti, pokazuje se da *više različitih formalnih, matematičkih modela pokazuju iste ili slične empirijske performanse*. Bejzijanske verzije kauzalnih modela, poput onog koji predlažu Lu i saradnici, još su složenije od standardnih modela iz te familije, ali po empirijskim performansama jesu uporedive sa performansama prethodno diskutovanih heuristika. Uzimajući u obzir da kognitivne funkcije koje operišu kauzalnim odnosima svakako ne mogu da se svedu na asocijacionističke ili heurističke funkcije u celini, postoji opravdanje za tvrdnju da bi kognitivni sistem bio u prednosti ako bi i funkcije kauzalnog učenja i druge kauzalne funkcije (rezonovanje, npr) obavljao u zajedničkom reprezentacionom i kompjutacionom okviru. Jedini koherentan takav okvir zaista opisuje samo teorija kauzalnih modela. S druge strane, podsetimo se sledećeg: $\triangle P$, mera probabilističkog kontrasta, jeste normativna mera intenziteta kauzalnih odnosa; to je isto i kauzalna moć - *ali pod drugim uslovima*. Heuristike poput heuristike dva faktora (H) koju predlažu Hatori i Oaksford takođe su normativne pod određenim uslovima: uslovima u kojima P($C$) i P($E$) imaju izuzetno niske vrednosti, što Hatori i Oaksford sugerišu kao načelo environmentalne racionalnosti, posle argumentacije koja bi trebalo da nas uveri da struktura okoline u kojoj se kauzalno učenje odvija zaista zadovoljava uslove o kojima oni govore (Hattori & Oaksford, 2007). Bejzijanski model kauzalne moći opet jeste normativno opravdan: pretpostavka da kognitivni sistem u kazalnom učenju ima *a priori* verovanje da su svi nepoznati uzroci čije kontingencije još ne poznaje nužni

i dovoljni (dakle, veoma intenzivni) odgovara tzv. *hipotezi o jednostavnosti*, za koju se veruje da može da predstavlja normativno validan okvir za procese učenja (Chater & Vitany, 2003). Pitanje se nameće: šta je, uopšte, *cilj izračunavanja* koji kognitivni sistem treba da izvede, suočen sa problemom kauzalne indukcije? Drugim rečima: *šta je problem* koji kognitivni sistem rešava u situaciji koju mi formalno opisujemo kao Hjumov problem kauzalne indukcije? Različiti normativni modeli odgovaraju različitim ciljevima izračunavanja: ponovimo, ako je normativni model $\triangle P$, onda je cilj izračunati doprinos nekog potencijalnog uzroka efektu koji je već po dejstvom konglomerata pozadinskih uzroka; ako je to, ipak, kauzalna moć $p_c$, onda je cilj izračunati doprinos probabilističkog uzroka u izolaciji. Za još jednostavnije postavljanje ovog problema: ne zaboravimo da je ljudski kognitivni sistem okružen i pravim kauzalnim pravilnostima i čistim korelacijama. I jedna i druga situacija pred kognitivni sistem iznose iste podatke, one o kovarijaciji. Kako kognitivni sistem donosi odluku o tome da li treba da izračuna stepen kovarijacije (normativno rešenje za varijable koje nisu zaista kauzalno povezane) ili kauzalnu moć (normativno rešenje za varijable koje su zaista kauzalno povezane)? U nekim situacijama, kriterijum plauzibilnosti, baziran na prethodnom znanju koje može da omogući generisanje hipoteza o tome šta jeste, a šta nije pravi kauzalni odnos, može da pomogne u rešavanju ovog problema. Ali, u problemu kauzalne indukcije, čak onako kako se on postavlja u laboratorijskim uslovima, govorimo o nepoznatim potencijalnim uzrocima i posledicama. Pitanje, dakle, ostaje otvoreno.

Drugo pitanje, koje takođe smatramo neizbežnim - pitanje koje se uopšte ne postavlja u debati o racionalnim modelima kognitivnih funkcija - je pitanje o tome da li kognitivni sistem uopšte sebi postavlja *jedan* cilj izračunavanja, ili *postavlja više ciljeva izračunavanja* u određenoj problemskoj situaciji koju mi pokušavamo da analiziramo kroz logiku racionalne metodologije? Ukoliko se pokaže da se odgovor na ovo drugo pitanje nalazi upravo u postojanju višestrukih ciljeva izračunavanja u pokušaju rešenja konceptualno jedinstvenog problema, uzimajući u obzir slične empirijske performanse različitih normativnih i nenormativnih modela kauzalnog učenja, sasvim prihvatljiv odgovor bi bio da su svi oni, u odnosu na te višestruke ciljeve izračunavanja, opravdani, i u nekoj meri tačni modeli kauzalnog učenja[48]. Ovo je zaključak koji bi svaka standardna racionalna analiza odbacila. Pored ostalih motiva da se predloži metodologija racionalne analize, Anderson (1991b) ističe mogućnost da se njenom primenom prevaziđe problem arbitrarnih ograničenja kognitivnih modela - problem koji je u prošlosti vodio proliferaciji velikog broja

mehanicističkih modela *istih* kognitivnih funkcija. Prethodna analiza strukture okoline i ograničenja kompjutacionih resursa kojima kognitivni sistem raspolaže bila bi sredstva da se obezbedi jasno prepoznavanje *jednog* kompjutacionog problema koji sistem pokušava da reši, i tako motiviše razvoj *jedne* teorije rešavanja tog problema. Ipak, ova analiza inherentno pretpostavlja (a) da je ciljeve kognitivnog izračunavanja moguće opisati strogo i disjunktno, i (b) da u pogledu svakog konceptualno jedinstvenog problema adaptacije, poput problema odlučivanja ili kauzalnog učenja, kognitivni sistem uvek rešava jedan jedinstven, identičan kompjutacioni problem. U poglavlju VI, kako bilo, pokazujemo da racionalna analiza u svojoj unutrašnjoj logici ne samo da omogućava zaključak o postojanju više ciljeva izračunavanja, dopuštajući tako paralelnu egzistenciju više različitih modela „*iste*“ kognitivne funkcije, već nužno vodi ka takvom zaključku - ako se konsekventno i u potpunosti izvedu sve posledice njene konceptualne strukture kao argumenta i metodologije.

## 7.3 Epizodička memorija

Pitanje o mogućnosti normativne, racionalne teorije epizodičke memorije izuzetno je interesantno za našu diskusiju. Racionalna analiza epizodičke memorije bila je jedna od prvih koju je Džon Anderson sa saradnicima ponudio, pošto je eksplicirao metodologiju racionalne analize kao istraživačku strategiju u kognitivnoj psihologiji (Anderson, 1991b). Andersonova racionalna analiza memorije je značajna zbog toga što su u njenom razvoju ponuđene paralelno (a) normativna analiza memorijskih funkcija u odnosu na strukturu okoline, koja je *specifikovana empirijski* (Schooler & Anderson 1991, 1993, Anderson & Schooler, 1991, 2000) i (b) normativna analiza memorijskih funkcija kroz *analogiju sa algoritmima za rad veštačkih sistema za upravljanje informacijama* - sistema za koje Anderson veruje da su projektovani da rešavaju istu klasu problema koje treba da reši epizodička memorija u procesima adaptacije čoveka svojoj sredini (Anderson, 1989, Anderson & Milson, 1989). Iako postoji još modela epizodičke memorije koji memorijske procese izvlačenja (engl. *retreival*), reprodukcije (engl. *recall*) i prepoznavanja (engl. *recognition*) objašnjavaju kao optimalne procese (npr. REM model, Shiffrin & Steyvers, 1997, 1998, Steyvers, Griffiths & Dennis, 2006), jedino je Andersonova teorija proizvod dosledno sprovedene racionalne analize.

Predstavićemo najvažnije eksperimentalne nalaze na koje cilja racionalna analiza pamćenja. Najvažniji eksperimentalni nalaz je svakako *zakon stepena* koji odlikuje

(a) funkciju retencije (engl. *retention function*), odn. zaboravljanja i (b) funkciju vežbe (engl. *practice function*). Još Ebinghausova istraživanja (1885/1913) su pokazala da sa povećanjem vremenskog intervala između učenja određenog materijala i trenutka u kome se testira poznavanje tog materijala uspešnost u reprodukciji (odn. količina zadržanog u pamćenju) opada nelinearno. Ebinghausovi rezultati replicirani su u mnogobrojnim studijama reprodukcije i predstavljaju „potpise“ ljudske epizodičke memorije. Ebinghausov bihejvioralni test zadržavanja naučenih besmislenih slogova bio je formulisan kao merenje procenta uštede u učenju neophodnog da bi se postigao kriterijumski nivo posle određenog vremena od prvog učenja materijala. Stepeni zakon odlikuje i funkciju vežbe. U Ebinghausovom klasičnom istraživanju, o njemu svedoči nelinearni pad u broju ponavljanja koji je potreban da bi se naučile serije besmislenih slogova sa danima vežbe. Slika 15. ilustruje ove karakteristike zadržavanja informacija i uštede u učenju koje karakterišu ljudsko pamćenje. Sa desne strane odgovarajućih grafikona prikazani su isti podaci na logaritamskim skalama. Kao što je poznato, zakoni stepena uzimaju linearnu formu u logaritamskom prostoru promenljivih. Na dva desna grafikona Slike 15. prikazane su i linearne regresione funkcije za odnose prethodnog logaritmovanih Ebinghausovih prediktora i kriterijuma: linearni odnos je očigledan za funkciju retencije i funkciju vežbanja podjednako. Stepeni zakon funkcije vežbanja je jedna od osnovnih osobina ljudskog učenja uopšte; mnogobrojne studije posvećene su njegovom utemeljenju u fundamentlanim prepostavkama kognitivne obrade informacija (npr. Newell & Rosenbloom, 1981).

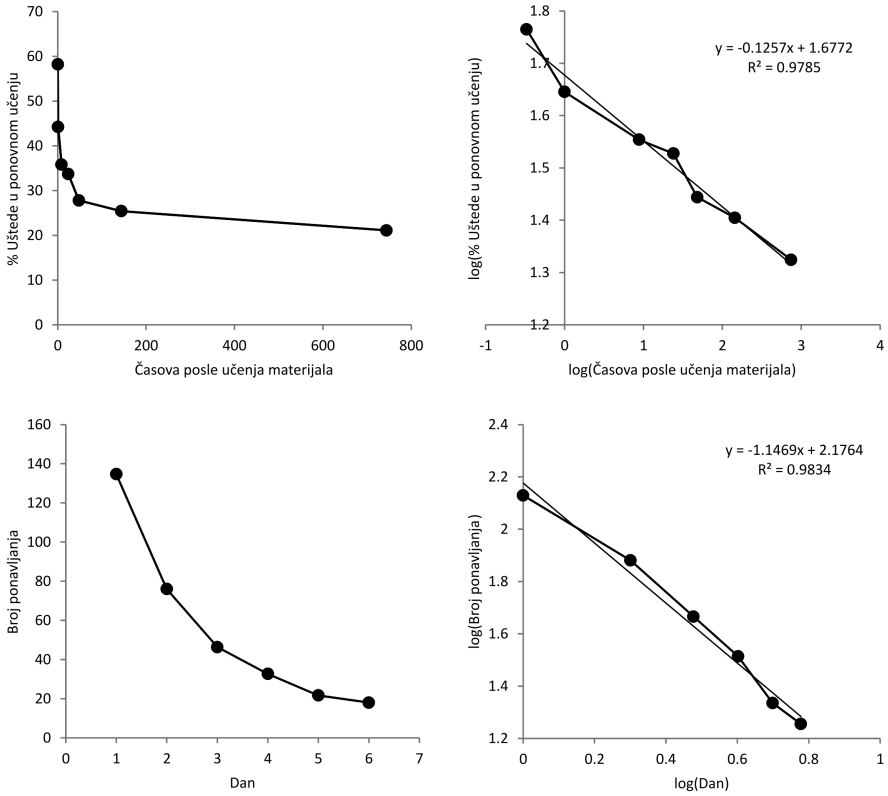
Stepeni zakon koji odlikuje funkciju retencije uzima oblik

$$P = a \cdot t^{-\alpha} \tag{35}$$

u linearnom prostoru, dok u logaritamskom ima oblik

$$log(P) = log(a) - \alpha \cdot log(t) \tag{36}$$

gde je $P$ odgovarajuća bihejvioralna mera zadržanog materijala (u savremenim istraživanjima pamćenja to će češće biti procenat uspešne reprodukcije nego Ebinghausova mera uštede u ponovonom učenju), $\alpha$ eksponent stepene funkcije i $a$ parametar koji skalira ovaj model; $t$ je, naravno, vreme.




Slika 15. *Ebinghausovi rezultati o uštedi u učenju sa protokom vremena (gore levo) i sa danima vežbe (dole levo).* Sa desne strane odgovarajućeg grafikona prikazan je isti nalaz na logaritamskim skalama zajedno sa odgovarajućom linearnom aproksimacijom. Zadržanost zapamćenog sa protokom vremena merena je procentom uštede u vremenu ponovnog učenja do kriterijuma (Ebbinghaus, 1885/1913, Poglavlje VII, sekcija 29). Rezultati uštede u učenju sa danima vežbe su uprosečeni brojevi ponavljanja devet serija od 12 slogova, tri serije od 24 sloga, i dve serije od 36 slogova kroz dane vežbe (Ebbinghaus, 1885/1913, Poglavlje VII, sekcija 31).

Stepena funkcija vežbe se najčešće iskazuje kao rastuća, ne opadajuća funkcija, ali je to samo pitanje načina na koji je izražena zavisna varijabla. U Ebinghausovim rezultatima, pa i na našoj Slici 15, ona je prikazana, neuobičajeno, kao opadajuća funkcija - ponovo zbog Ebinghausove odluke da kao zavisnu promenljivu odabere meru uštede (broj ponavljanja do kriterijuma učenja) nego neku meru progresa u učenju. Ponovo, sasvim je svejedno kako izražavamo ove funkcije: njihova funkcionalna forma, odn. zakon stepena, je ono što ostaje invarijantno pri različitim izborima operacionalizacije relevantnih varijabli. Prema tome, jednačina (35) podjednako odlikuje i funkciju vežbanja i funkciju retencije.

Treći eksperimentalan nalaz koji dominira istraživanjima reprodukcije u oblasti epizodičke memorije takođe je bio poznat i Ebinghausu, ali je njegova preciznija

forma ustanovljena tek u drugoj polovini XX veka. Reč je o efektima raspodeljenog vežbanja (eng. *spacing effects*), dokumentovanim u većem broju klasičnih studija (npr. Glenberg, 1976, prema Anderson & Schooler, 1991). Ovi efekti svedoče o tome da je reprodukcija naučenog bolja za materijal čije je vežbanje (učenje) raspodeljeno u određenom vremenskom intervalu, nego za materijal čije je vežbanje grupisano. Na primer, efekat raspodeljenog vežbanja svedoči o boljoj reprodukciji za materijal koji je vežban dva puta sa većim vremenskim intervalom između prvog i drugog vežbanja, nego za materijal koji je vežban dva puta sa manjim vremenskim intervalom između vežbanja. Vremenski interval u efektu raspodeljenog vežbanja može da bude zamenjen i merom broja ajtema prikazanih između dve ekspozicije istog ajtema koji se testira u reprodukciji. Interakcija sa još jednom promenljivom komplikuje ovaj nalaz, a reč je o vremenu koje prođe između poslednjeg prikazivanja test-ajtema i samog testiranja (ili, opet, broja ajtema prikazanih između poslednje ekspozicije i testa reprodukcije). U slučajevima kada je interval između poslednje ekspozicije ajtema i testiranja kratak, ispostavlja se da je reprodukcija bolja što je interval između prethodnih ekspozicija (učenja) ajtema bio manji; nasuprot ovome, što je interval između poslednje ekspozicije ajtema i testiranja duži, ispostavlja se da je reprodukcija bolja što je interval između prethodnih ekspozicija (učenja) bio veći[49]. Anderson i Skuler zaključuju da nalaz, generalno, svedoči o tome da je reprodukcija najbolja kada se interval između ekspozicija ajtema manje razlikuje - „poklapa“ - sa intervalom između poslednje ekspozicije i testa. Pre pokušaja racionalne analize epizodičke memorije Andersona i saradnika, nijedna ponuđena teorija nije mogla da obuhvati ove karakteristične rezultate i pri tom uspešno reprodukuje stepene zakone retencije i vežbanja. Racionalna analiza epizodičke memorije pokušava da objasni daleko više od ovde predstavljenih karakterističnih nalaza u istraživanju slobodne reprodukcije; Anderson (1989) navodi deset poznatih empirijskih fenomena u ovoj oblasti koje može da objasni jedinstven racionalni model pamćenja. Model koji razmatraju Anderson i saradnici nije eksplicitno fitovan na eksperimentalne podatke: svi zainteresovani za empirijsku performansu racionalne teorije epizodičkog pamćenja moraće da se zadovolje simulacijama tipičnih eksperimentalnih nalaza.

*Specifikacija problema pamćenja.* Osnovni adaptivni zadatak ljudskog pamćenja, prema Andersonu, jeste da organizuje u memoriji uskladištene podatke *na način koji optimizuje verovatnoću da će svaka uskladištena informacija biti neophodna u nekom narednom trenutku.* Očigledno, optimalan način organizacije počiva na onom uređenju memorije koje poštuje upravo verovatnoće da će svaka uskladištena

informacija biti potrebna u budućnosti: one informacije sa većom verovatnoćom da će biti potrebne trebalo bi da budu lakše i brže dostupne u odnosu na one informacije koje karakterište manja verovatnoća da će biti potrebne. Neka je $G$ neka vrednost koju kognitivni sistem stiče uspešno izvlačeći informaciju $A$ iz svoje memorije; neka je $C$ cena izračunavanja koju nosi pristup svakoj uskladištenoj informaciji. Ako je $p(A)$ verovatnoća da će $A$ biti potrebna u nekom budućem trenutku, za kognitivni sistem je optimalno da pretražuje memoriju uređenu po vrednostima $p$ za sve uskladištene informacije i stane kada je zadovoljen sledeći uslov:

$$C > p(A) \cdot G \tag{37}$$

tj. kognitivni sistem treba da zaustavi memorijsku pretragu u trenutku kada cena izvlačenja dodatnih informacija počne da prevazilazi očekivanu vrednost u slučaju uspešnog izvlačenja tražene informacije. Suština problema racionalne analize pamćenja se onda svodi na ocenu verovatnoća p($A$) - verovatnoća da će određena uskladištena informacija u memoriji biti potrebna u nekom trenutku posle njenog skladištenja. Anderson i saradnici problem ocene ovih verovatnoća specifikuju na sledeći način:

$$\frac{P(A|H,Q)}{P(\bar{A}|H,Q)} = \frac{P(A|H)}{P(\bar{A}|H)} \times \prod_{i \in Q} \frac{P(i|A)}{P(i|\bar{A})} \tag{38}$$

Izraz P($A|H,Q$) je zavisna verovatnoća da je neka memorija $A$ potrebna ako su poznati (a) $H$ - faktor njene *prethodne upotrebe* i (b) $Q$ - *skup znakova za prisećanje*, stimulusa u okolini koji svojom povezanošću sa memorijskim tragovima određuju *šta* je to čega se treba setiti u nekom trenutku. Objasnimo osnovnu jednačinu racionalnog modela u manje formalnom jeziku. Racionalna analiza memorije tvrdi da dve komponente utiču na to kolika je verovatnoća da je neka memorija $A$ potrebna u nekom kontekstu. Prva komponenta je istorija njene prethodne upotrebe: generalno, informacije koje su u prošlosti bile potrebne i korisne, u budućnosti će nastaviti to da budu. Videćemo uskoro da Anderson i saradnici zaista daju neke ubedljive empirijske podatke da to jeste tako. Druga komponenta se odnosi na skup stimulusa koji se nalaze u okruženju u onom kontekstu u kome kognitivni sistem treba da pronađe određenu informaciju u memoriji. Taj skup se označava sa $Q$: njegov uticaj na verovatnoću da je upravo $A$ memorija koja je kognitivnom sistemu potrebna se izražava preko stepena povezanosti svih njegovih članova

(svih relevantnih znakova u sredini u datom kontekstu) sa osobinama uskladištene memorije $A$. Ukoliko se neka osoba nalazi u parku, sedeći na klupi, pod kišobranom, i njen prijatelj ili prijateljica je pitaju da se seti imena neke ptice, njeno pamćenje će pre reprodukovati goluba (koji se često sreće u kontekstu gradskih parkova) nego orla ili pingvina. Sada možemo da protumačimo jasnije formalizam iskazan jednačinom (38). Ceo izraz sa njene leve strane predstavlja odnos verovatnoće zavisne od $H$ i $Q$ da je neka informacija $A$ iz memorije potrebna i da nije potrebna u datom trenutku. Pod pretpostavkom da veze elemenata skupa znakova $Q$ (označenih sa $i$) sa memorijom $A$ nisu zavisne od faktora prethodne upotrebe te memorije $A$, jednačina (38) se razvija u dva multiplikativna člana sa desne strane. Prvi je jednostavno odnos zavisnih verovatnoća da je $A$ potrebna i da nije potrebna ($\bar{A}$) u zavisnosti od njene prethodne upotrebe $H$. Drugi član je nešto komplikovaniji. Izraz $\mathrm{P}(i|A)$ je zavisna verovatnoća da je u kontekstu $Q$ prisutan neki znak $i$ ako je dat memorijski trag A. Ova verovatnoća se u racionalnom modelu formuliše kao funkcija asocijativne veze između određenih znakova i memorijskih tragova, i tu formulaciju ovde ne razvijamo dalje (up. Anderson & Milson, 1989, za detaljan razvoj ovog mehanizma). Izraz $\mathrm{P}(i|\bar{A})$ je zavisna verovatnoća da je u kontekstu $Q$ prisutan neki znak $i$ za date sve druge tragove osim $A$. Tako drugi član izraza sa desne strane jednačine (38) predstavlja proizvod odnosa ove dve verovatnoće i u suštini izračunava stepen korespondencije nekog memorijskog traga $A$ sa skupom znakova za prisećanje prisutnih u kontekstu $Q$. Osnovna pretpostavka primene racionalnog modela pamćenja je da postoji monoton odnos između $\mathrm{P}(A|H,Q)$ i $\mathrm{P}(\bar{A}|H,Q)$ koji izračunava osnovna jednačina modela i bihejvioralnih mera u eksperimentima dugotrajne memorije. Pretpostavljeni odnos je pozitivan u slučaju da je bihejvioralna mera procenat uspešne reprodukcije, i negativan u slučaju vremena latencije (vremena potrebnog za izvlačenje ajtema $A$ iz memorije).

Objašnjenjem jednačine (38) objašnjena je suština racionalne analize memorije Andersona i saradnika. Primenom racionalnog modela moguće je simulirati sve prethodno diskutovane robustne empirijske efekte u istraživanju epizodičke memorije. Više pažnje u narednim redovima posvećujemo kritici pokušaja Andersona i saradnika da statističkom analizom nekoliko izvora informacija pokažu da pod opisanim uslovima ljudsko pamćenje praktično ne čini ništa drugo do da precizno statistički mapira distribucije informacija u relevantnom okruženju. Od fundamentalnog značaja za razumevanje dalje analize je činjenica da je verovatnoća buduće upotrebe neke uskladištene informacije $A$ funkcija njene prethodne upotrebe

*H*.

*Ljudska epizodička memorija i struktura informacija u našoj okolini.* U seriji istraživanja Skuler i Anderson pokušavaju da sprovedu analizu strukture informacija u okolini u kojoj ljudski kognitivni sistem rešava problem optimalne organizacije zapamćenog u epizodičkoj memoriji (Anderson & Schooler, 1991, 2000, Schooler & Anderson 1991, 1993, 1997). Podsetimo se da je ovaj korak karakterističan za racionalnu analizu: da bi se razumeo problem koji kognitivni sistem pokušava da reši (Marov kompjutacioni nivo 3), i tako postavila ograničenja algoritamskoj teoriji rešavanja tog problema (Marov nivo 2), potrebno je razumeti strukturu sredinskih informacija koja konstituiše određeni problem adaptacije (Anderson, 1991b). Okolinu, za potrebe ovih istraživanja, Anderson i Skuler operacionalizuju kroz masovne izvore informacija koji nas okružuju u savremenom životu, koristeći sledeće izvore informacija: (a) frekvenciju pojave određenih reči u naslovima *Njujork tajmsa*, u rasponu od dve godine, (b) frekvenciju pojave određenih reči u korpusu dečjeg govora CHILDES, i (c) frekvenciju autora različitih poruka koje je putem elektronske pošte primio Džon Anderson u periodu od četiri godine. Osnovna hipoteza ispitivanja statističke informacione strukture okoline u pristupu Skulera i Andersona je da će izračunavanjem verovatnoća da se neka informacija $A$, koja se već prethodno pojavljivala u okolini, pojaviti ponovo, pokazati da informacionom strukturom ljudske okoline dominiraju iste empirijske pravilnosti koje su karakteristične za ljudsku epizodičku memoriju. Ovako specifikovana verovatnoća potpuno odslikava verovatnoću definisanu osnovnom jednačinom racionalnog modela memorije - jednačinom (38), dakle - ukoliko se pretpostavi da ona zavisi samo od faktora prethodne upotrebe neke informacije, $H$, ali ne i od $Q$, skupa znakova za prisećanje koje obuhvata drugi član sa desne strane jednačine[50].

Statistički podaci iz navedenih izvora, pokazuje se, zaista poštuju iste zakonitosti koje slede iz osnovne jednačine racionalnog modela memorije, i koje u empirijskim istraživanjima predstavljaju stabilne, više puta replicirane efekte: zakon stepena u opisu (a) funkcije retencije i (b) funkcije vežbe i (c) efekat raspodeljene vežbe. Anderson i Skuler ovo pokazuju na sledeći način: na osnovu statističkih podataka o frekvencijama reči u naslovima *Njujork tajmsa*, određuju faktor $H$ prethodne upotrebe neke reči kao verovatnoću (frekvenciju) njene upotrebe u prethodnih 100 dana, a na osnovu ove verovatnoće, pokušavaju da daju predikciju verovatnoće upotrebe takve reči u 101. danu. Ovako specifikovan faktor prethodne upotrebe $H$ daje izvanrednu predikciju verovatnoće javljanja neke reči u narednom trenutku

merenja, a isti nalaz se dobija i za druga dva izvora informacija koji su analizirani u ovim studijama. Dalje, Anderson i Skuler pokazuju da je verovatnoća da se neka reč *A* pojavi u naslovima *Njujork tajmsa* upravo stepeno opadajuća funkcija intervala koji je protekao od njene prethodne upotrebe u istom izvoru informacija. Funkcija koja opisuje određene osobine statističkog izvora informacija u okolini, prema tome, ima istu formu kao i funkcija retencije u ispitivanjima ljudske epizodičke memorije (v. Anderson & Schooler, 1991, Slika 7). Uz odgovarajuće definicije ključnih varijabli, i druga dva analizirana izvora informacija pokazuju istu zakonitost. Konačno, Anderson i Skuler uspevaju da pokažu da i nalazi poput interakcije efekta raspodeljenog vežbanja sa intervalom do testa slobodne reprodukcije imaju svoju „refleksiju“ u informacionoj strukturi ljudske okoline. Kako bi pronašli analogiju u odgovarajućem prirodnom eksperimentu raspodeljene vežbe, analiziraju samo one reči koje su se u prethodnom periodu posmatranja u tri korpusa pojavile dva puta, dalje birajući reči tako da interval od njihove druge do treće pojave varira. Anderson i Skuler dolaze do zaključka da verovatnoće javljanja takvih reči narednog dana - dakle, verovatnoća da će one biti potrebne u informacionom sistemu koji se analizira - kvalitativno odgovaraju profilima verovatnoće uspešne slobodne reprodukcije do kojih je došao Glenberg u studiji iz 1976 (Slika 16). Ako je prvi nalaz, koji pokazuje da strukturu izvora informacija u ljudskoj okolini karakteriše stepeni zakon, povezujući verovatnoću da će neka informacija biti potrebna sa vremenskim intervalom koji prolazi od njene prethodne upotrebe, na neki način i očekivan[51], drugi nalaz, koji pokazuje da informaciona struktura sredine pokazuje komplikovane osobine koje prate složene interakcije faktora u eksperimentalnim nacrtima istraživanja epizodičke memorije je u najmanju ruku fascinantan.

*Ljudska epizodička memorija i optimalna rešenja za skladištenje informacija.* Drugi pristup izgradnji racionalne teorije pamćenja polazi od pretpostavke da između ljudskog pamćenja i veštačkih sistema koji optimizuju upotrebu informacionih resursa postoji snažna analogija. Alan Bedli, jedan od najznačajnijih psihologa koji su ikad proučavali memoriju, u knjizi „*Ljudsko pamćenje*“ ovako započinje diskusiju pojma izvlačenja informacija iz memorije: „*Često se tvrdi kako pamćenje podseća na bogatu biblioteku, što je analogija koja ima svoja ograničenja, ali koja može biti vrlo korisna. Jedna izrazita sličnost pamćenja i biblioteke jeste stepen do koga će raditi efikasno ukoliko je informacija u njih pohranjena na strukturisan i sistematičan način, pri čemu će izvlačenje informacije zavisiti od tog početog „katalogizovanja“, odnosno kodovanja.* [...] *Efikasan sistem katalogizacije*

*i izvlačenja morao bi da omogući svim [pomenutim] čitaocima da dođu do [takve] knjige. Upravo isto tako, potrebno je da se informacija u ljudsko pamćenje smesti na način koji će omogućiti da joj se pristupi zbog različitih upotreba*" (citirano prema Baddeley, 1997/2004, prevod D. Lalović). Ne možemo da zamislimo bolji uvod u razvoj kompjutacionih mehanizama racionalne teorije pamćenja Džona Andersona od ovih reči Alana Bedlija. Proučavajući predložena rešenja u oblasti optimizacije organizacije knjiga u bibliotekama, Anderson predlaže kompjutacioni opis rada epizodičke memorije u rešavanju problema optimalne organizacije i procesa izvlačenja informacija iz memorije. Zbog relativno složenih formalnih osobina ovu teoriju nećemo predstavljati detaljno. Anderson i Milson projektuju kompjutacione mehanizame koji obezbeđuju objašnjenje dve komponente osnovne jednačine modela (38): izračunavanje faktora prethodne upotrebe neke informacije u memoriji $H$ i verovatnoće da je skup znakova za prisećanje $Q$ upravo takav kakav je određen trenutnim kontekstom ako je $A$ tražena informacija (Anderson, 1989, Anderson & Milson, 1989). Razvoj kompjutacionog mehanizma koji objašnjava efekat prethodne upotrebe $H$ diktiran je analogijom sa rešavanjem istog problema u slučaju organizacije bibliotečkih fondova. Anderson i Milson pretpostavljaju da je potreba za nekom informacijom suštinski određena Poasonovim procesom koji analiziraju bejzijanski, kroz upotrebu gama *a priori* distribucija sa odgovarajućim parametrima. Na osnovu ovog statističkog mehanizma razvija se objašnjenje efekata frekvencije i recencije ajtema u epizodičkoj memoriji. Daljim prilagođavanjem modela Poasonovog procesa objašnjavaju se efekti raspodeljenog vežbanja u oblasti ljudske epizodičke memorije. Objašnjenje efekta kontekstualnih znakova učenja $Q$ prati uobičajeni pristup u modeliranju pamćenja i učenja, pristup kojim se ovi efekti modeliraju na osnovu određenog stepena poklapanja osobina koje sadrži trenutni kontekst i osobina koje sadrže uskladišteni memorijski tragovi među kojima se nalazi i ciljana informacija u memoriji.

Rezultati pristupa koji polazi od analize strukture informacija u okolini (Schooler & Anderson 1991, 1993, 1997, Anderson & Schooler, 1991, 2000) i pristupa koji polazi od modeliranja ljudskog pamćenja kao optimalnog sistema za organizaciju i izvlačenje informacija (Anderson, 1989, Anderson & Milson, 1989) konvergiraju u korespondenciju sa jedinstvenom strukturom eksperimentalnih rezultata u istraživanjima epizodičkog pamćenja. Dok se u prvom pristupu pokazuje da ljudsko pamćenje pokazuje iste osobine kao i sredina koja generiše informacije koje treba pamtiti, u drugom pristupu se pokazuje da model ljudskog

pamćenja kao optimalnog sistema za skladištenje i izvlačenje informacija može da generiše iste te pravilnosti. Detaljnija studija kompjutacionog pristupa razvoju racionalne teorije memorije, ipak, otkriva da njena matematička konstrukcija počiva na velikom broju pretpostavki, od kojih za neke i sami autori primećuju da su arbitrarne. U modeliranju sistema složenosti kakva je ljudska epizodička memorija, možda su arbitrarne pretpostavke o kompjutacionim mehanizmima neizbežne. To ne otklanja opasnost od toga da je razvijen arbitraran model koji obuhvata komplikovane eksperimentalne rezultate tek zahvaljujući fleksibilnosti obezbeđenoj velikim brojem slobodnih parametara. Implemetacija racionalne teorije za potrebe simuliranja eksperimenata slobodne reprodukcije obuhvata 11 parametara, od kojih su devet slobodni, a dva potiču iz osobina eksperimentalnog nacrta. Za poređenje, standardni model u fizici elementarnih čestica sadrži između 19 i 25 paramatera, dok klasični kosmološki model sadrži samo tri[52] - oba precizno objašnjavaju skoro sve poznate fizičke interakcije u oblasti primene. Možda je neizbežno i to da ovakvi modeli memorije mogu da predvide samo kvalitativne trendove slične realnim eksperimentalnim rezultatima, a ne i da precizno modeliraju kvantitativne rezultate eksperimenata (mada su u slučaju nekih eksperimenata predikcije Andersonovog modela i više nego dobre na nivou kvantitativne analize).

*Komentari o racionalnosti epizodičke memorije.* Racionalna analiza pamćenja, sprovedena dosledno kroz analizu strukture okoline u kojoj ljudsko pamćenje (po pretpostavci) rešava problem optimalnog skladištenja informacija, povezana je sa veoma značajnim i interesantnim problemom. U jednom od radova u ovoj seriji istraživanja, Anderson i Skuler navode sledeće:

„*Jedna od kritika upućenih tom istraživanju je da sva tri domena koja su proučili Anderson i Skuler [...] obuhvataju ljudsku komunikaciju. Neko bi mogao da pomisli da je ljudska komunikacija određena ljudskim pamćenjem. Tako može delovati cirkularno to da osobine ljudskog pamćenja mogu da se predvide osobinama tih baza podataka. U pokušaju da se ovakve kritike razviju do u detalje javljaju se problemi. Na primer, jedan od termina u bazi podataka iz Njujork tajmsa je Čelendžer, i odnosi se na eksploziju Čelendžera*[53]*. Pomalo je bizarno misliti da je ljudsko pamćenje izazvalo eksploziju Čelendžera i samim tim njegovu pojavu u Njujork tajmsu. Svakako, bilo bi lepo imati baze podataka slobodne od uticaja ljudskog pamćenja*“ (citirano prema Anderson & Schooler, 2000, naš prevod).

Pomisliti da je ljudsko pamćenje ma kako moglo da izazove eksploziju Čelendžera

je, naravno, nonsens. Međutim, upotreba određenih informacija u ljudskoj komunikaciji izvesno jeste kauzalno povezana sa njihovim pamćenjem. Sledeći logiku same racionalne analize, ljudsko pamćenje se optimalno organizuje upravo prateći frekvenciju upotrebe relevantnih informacija. Optimalno organizovane verovatnoće koje opisuje jednačina racionalnog modela (38) određuju (a) da će u odgovarajućem kontekstu biti prizvana ona informacija iz memorije koja tom kontekstu najviše odgovara, kroz faktor $Q$, ali i (b) to da će informacije sa većim verovatnoćama da će biti potrebne biti i one sa većim verovatnoćama u upotrebi - faktor prethodne upotrebe $H$ doprinosi tome. Uticaj faktora konteksta $Q$ ne rešava ovaj problem: samim tim što su neke informacije češće od drugih, odn. što se neki referenti evociraju sa većom verovatnoćom, jer su relevantniji u diskursu ljudske komunikacije od drugih, i distribucija znakova za prisećanje koja konstituiše kontekst $Q$ će biti zavisna od verovatnoće prethodne upotrebe[54]. Mi češće koristimo reči više frekvencije zato što one referiraju na relevantnije - frekventnije - referente u našoj okolini, i *vice versa*, naše pamćenje je organizovano tako da olakša pristup rečima više frekvencije jer ćemo njih češće koristiti. Dok se ograničavamo na informacionu okolinu koju prozvodi diskurs ljudske komunikacije, objašnjavati strukturu ljudske informacione okoline strukturom ljudskog pamćenja, ili objašnjavati informacionu strukturu ljudskog pamćenja informacionom strukturom okoline, isto je i - žalosno za analize poput Andersonove i Skulerove - jeste cirkularno. Unutar ova dva koimplikativna diskursa, kognitivni sistem i njegova informaciona okolina potpuno su spregnuti. Teoretičari iz škole Gibsonove ekološke psihologije bili su dobro svesni ove činjenice o koimplikativnosti organizama i njihovih ekoloških niša, o čemu svedoči i formulisanje ključnih teorijskih pojmova gibsonijanske teorije, pojmova *afordanse* i *efektiviteta*, kao simetričnih i koimplikativnih (Shaw & McIntyre, 1974, Michaels & Carello, 1981). Autori enaktivističke orijentacije koju smo diskutovali u II delu ove rasprave ni ne pretpostavljaju ništa drugo do to da su kognitivni sistemi i odgovarajuće sredine prožeti skupom relacija koje potpuno relativizuju njihovu individualnost kao sistema u prirodi. Radovi autora ovakvih shvatanja se od strane mejnstrima kognitivne psihologije, u kome racionalna analiza igra jednu od vodećih uloga, posmatraju kao konceptualno toliko udaljeni od standardnih eksplanatornih paradigmi da se praktično ni ne razmatraju.

Način da se ova cirkularnost u racionalnom modelu pamćenja otkloni nalazi se u ispitivanju mogućnosti upravljanja efektom faktora $Q$ - skupa znakova za prisećanje. Verujemo da je jedini put da se otkloni cirkularnost iz ovakvog modela

pamćenja ta da se demonstrira efekat faktora - nezavisnog od $H$, $Q$ i strukture informacija u sredini - koji vodi ka *promenama* distribucija verovatnoće informacija u sredini, *pa samim time i u pamćenju* koje mapira te distribucije. U rešavanju složenih problema i kreativnim procesima, na primer, ljudski kognitivni sistem nužno mora da restruktuira poznate načine pretrage informacija da bi došao do prethodno nepoznatih rešenja za probleme adaptacije. Na taj način on zapravo oblikuje strukturu konteksta, definisanu kroz faktor $Q$ u racionalnom modelu.

## 7.4 Rezonovanje i suđenje

Diskusija racionalnosti ljudske moći rezonovanja i donošenja sudova o ishodima neizvesnih događaja je, pored diskusije o racionalnosti odlučivanja, sigurno dala najveći kritički doprinos savremenoj debati o racionalnosti uopšte. Od sredine 70-ih godina XX veka, kada su Kaneman i Tverski definisali istraživački program *heuristika i inklinacija*[55] (Tversky & Kahneman, 1974) do danas, pokušaji objašnjenja bihejvioralnih odstupanja od (očekivanih) normativnih rezultata procesa donošenja sudova proizveli su nepreglednu količinu hipoteza, teorija i naučnih radova. Nešto pre starta paradigme heuristika i inklinacija, Piter Voson je formulisao čuveni *zadatak selekcije* (poznat još i kao *Vosonov zadatak sa četiri karte*, Wason, 1968). Zadatak selekcije, koji je u prvim istraživanjima demonstrativno ukazivao na nesposobnost ljudskog rezonovanja da ispuni normativne uslove logičkog zaključivanja, predstavlja verovatno najviše diskutovanu pojedinačnu eksperimentalnu paradigmu u istoriji psihologije uopšte. Ove dve linije istraživanja, prva posvećena suđenju a druga rezonovanju, često su se susticale i presecale, stvarajući okosnicu debate o racionalnosti u drugoj polovini XX veka. Našu pažnju posvećujemo prvo donošenju sudova, programu heuristika i inklinacija Tverskog i Kanemana, i njegovim kritikama. Zatim diskutujemo racionalnost u rezonovanju kroz analizu pokušaja da se razumeju odgovori ispitanika u paradigmi Vosonovog zadatka selekcije. Konačno, pažnju za kratko posvećujemo najnovijim istraživanjima rezonovanja u oblasti kauzalnih odnosa.

*Suđenje: istraživački program heuristika i inklinacija.* Ogroman broj eksperimentalnih istraživanja demonstrirao je nesavršenost ljudskog kognitivnog sistema pred jednostavnim zadacima ocene numeričkih vrednosti, ili verbalnih, simboličkih analogija verovatnoća. U paradigmatičnoj proceduri, Kaneman i Tverski bi ispitanicima opisali Lindu - najslavniju imaginarnu ženu u istoriji kognitivne

psihologije - na sledeći način:

„*Linda ima 31 godinu, solo je, otvorena i veoma bistra. Diplomirala je filozofiju. Kao studentkinja bila je duboko zauzeta pitanjima diskriminacije i socijalne pravde i takođe je učestvovala u antinuklearnim demonstracijama*“ (citirano prema Tversky & Kahneman, 1982, naš prevod).

Posle ovakvog upoznavanja sa Lindom, Kaneman i Tverski su u velikom uzroku studenata postavili pitanje tome da li je Linda pre:

A. *Bankarska službenica.*

B. *Bankarska službenica koja je aktivna u feminističkom pokretu.*

i došli do rezultata koji danas nosi ime *greške konjukcije* (engl. *conjunction fallacy*): 86% njihovih ispitanika smatra da je verovatnije da je Linda danas bankarska službenica koja je aktivna u feminističkom pokretu, uprkos logički jasnoj čijenici da je skup (B) žena koje su bankarske službenice i aktivne u feminističkom pokretu *podskup* skupa (A) žena koje su bankarske službenice, te je nužno da $\mathrm{P}(B) < \mathrm{P}(A)$. Kada su umesto studenata osnovnih studija ovakav zadatak Kaneman i Tverski postavili postdiplomcima psihologije, oko 50% njih je počinilo istu grešku[56]. Tek kada bi ispitanicima pored ove dve alternative o Lindi ponudili veći broj drugih opisa (osam, u originalnom istraživanju), greška konjukcije bi nestala (Kahneman & Tversky, 1982). Efekat je demonstriran na najrazličitijim materijalima, uključujući i pažljive eksperimentalne manipulacije smišljene da se ispitanicima ukaže na činjenicu da je verovatnoća pripadanja većim skupovima *a priori* veća od verovatnoće pripadanja manjim skupovima. Uprkos takvim manipulacijama, efekti poput upravo predstavljenog opstaju. Kaneman i Tverski su pokazali da u ovakvim i sličnim situacijama ljudi donose sud o verovatnoći oslanjajući se na svoju *procenu sličnosti tog događaja sa klasom događaja* za koju veruju da im on pripada na osnovu eksperimentalnog konteksta (Tversky & Kahneman, 1983). U primeru sa Lindom, ovo objašnjenje sugeriše da ispitanici zaključuju kako je Linda veoma slična klasi osoba koje bi mogle biti i bankarske službenice i feministkinje, i prelazeći preko regularnosti koje slede teoriju verovatnoće zaključuju da je verovatnije da Linda jeste i bankarska službenica i feministkinja. Primer zaključivanja koji smo upravo diskutovali predstavlja primer upotrebe *heuristike reprezentativnosti* u donošenju sudova: sud o verovatnoći se donosi na osnovu stepena u kome je određeni događaj prepoznat kao tipičan u klasi slučajeva kojoj bi mogao da pripada (Tversky & Kahneman, 1983). Zahvaljući tome što ljudi donose sudove

oslanjajući se na heuristike, a ne na normativna pravila, njihovi sudovi pokazuju određene inklinacije koje se na bihejvioralnom nivou prepoznaju kao odstupanja od normativno racionalnih odgovora.

Pored heuristike reprezentativnosti - možda najviše diskutovane heuristike u debati o racionalnosti - Kaneman i Tverski su u ranoj fazi razvoja programa heuristika i inklinacija formulisali još dve očigledne heurističke strategije. Jednu predstavlja *heuristika dostupnosti*: kognitivni sistem će često sud o verovatnoći nekog događaja doneti na osnovu toga koliko je *brz i olakšan memorijski pristup takvom događaju* (Tversky & Kahneman, 1973). Nesreće u avionskom saobraćaju, nažalost, najčešće vode ka velikom broju žrtava, bivaju medijski propraćene i temeljno analizirane od strane eksperata u javnosti, te se dobro upamte od strane pripadnika opšte populacije koju psihologija proučava. S druge strane, statistički posmatrano daleko češće nesreće u automobilskom saobraćaju nikada ne plene toliko pažnje, i konsekventno bivaju ređe upamćene. Dobiti od ljudi sud da je avionska nesreća verovatnija od automobilske, što predstavlja uobičajen odgovor, posledica je primene heuristike dostupnosti: primeri avionskih nesreća se brže javljaju u našem pamćenju, razvijajući tako utisak da su češće, i samim tim - verovatnije. Primena *heuristike referentne tačke i podešavanja*[57] se takođe lako demonstrira. U paradigmatičnom istraživanju, ljudima postavljamo pitanje o tome da ocene npr. prosečni uzrast ljudi u nekoj nepoznatoj ili manje poznatoj populaciji. Odgovor se daje tako što se pred ispitanikom zavrti točak koji po obodu ima ispisane brojeve koji predstavljaju moguće ocene prosečnog uzrasta. Točak se zaustavlja na određenoj vrednosti, a ispitanik treba da ga podesi na vrednost koja predstavlja njegov odgovor. Paradoksalno ili ne, ispitanici koji podešavaju vrednost točka koja se zaustavila na npr. 25 daju manje ocene prosečnog uzrasta ljudi u nekoj populaciji nego ispitanici koji podešavaju vrednost točka koji je zaustavljen na vrednosti npr. 60. Kognitivni sistem sudove ne donosi u nekom apsolutnom reprezentacionom sistemu, već u referentnom reprezentacionom sistemu u kome je donošenje sudova zavisno od polazne, referentne tačke. Tako se, zahvaljujući primenama heuristika dostupnosti i referentnih tačaka ponovo konstatuju inklinacije u procesu donošenja sudova. Prepoznaje se veza heuristike referentnih tačaka i podešavanja sa fenomenima suđenja o fizičkim i simboličkim veličinama (npr. *efekat semantičke kongruencije*, up. Banks, Clarck & Lucy, 1975, Rot & Kostić, 1993a, 1993b, 1995, Milovanović, 1996). Program heuristika i inklinacija kao istraživačka paradigma i te kako je aktivan i posle više od četrdeset godina od kako

je započet radovima Tverskog i Kanemana. Ovaj istraživački program predstavlja paradigmatičnu poziciju pristalica ograničene racionalnosti u debati o racionalnosti; heuristike u suđenju koje su otkrili Kaneman i Tverski tako predstavljaju prototipske eksplanatorne strategije ovog istraživačkog programa. Sistematizaciju, pregled i obuhvatnu kritiku savremenih gledišta u okviru ovog programa daju Kaneman i Frederik (Kahneman & Frederick, 2005).

Kaneman i Frederik su ponudili koherentan eksplanatorni okvir za upotebu heuristika u vidu procesne teorije *zamene atributa*. Prema teoriji zamene atributa, kognitivni sistem u donošenju sudova može da se orijentiše prema (*i*) cilju određivanja *ciljanog, normativnog atributa* o kome treba doneti sud, što je najčešće teži zadatak, i prema (*ii*) cilju određivanja *heurističkog atributa*, koji je lak i dostupan, ali često može da predstavlja neodgovarajuće rešenje za neki problem adaptacije. Na primer, u slučaju heuristike reprezentativnosti, normativni, ciljani atribut - onaj koji sadrži sam zadatak odn. problem adaptacije - jeste verovatnoća. Ali, zaključivanje po pravilima teorije verovatnoće nije kompjutaciono lak zadatak, i sasvim u skladu sa Sajmonovim idejama Kaneman i Frederik zaključuju da će se kognitivni sistem orijentisati prema drugom, lakše ostvarivom cilju. U slučaju heuristike reprezentativnosti, tako dolazi do zamene (teško odredive) verovatnoće (lako odredivom) sličnošću između instance i klase kojoj ona pripada. Ovakva strategija, očigledno, vodi ka inklinacijama u procesu donošenja sudova (Kahneman & Frederick, 2002). Frederik i Nelson predstavljaju izuzetnu eksperimentalnu demonstraciju ovog procesa. Ispitanicima su postavili sledeći zadatak: ukoliko bacimo loptu određenog prečnika u kocku koja je taman tolika da lopta može da stane u nju, koliki procenat zapremine kocke će zauzeti ta lopta? Zadatak je u suštini veoma lako rešiv primenom osnovnih geometrijskih znanja koja se savladaju još u osnovnoj školi, ali se ispostavlja da kognitivni sistem u zadatku u kome treba brzo da donese ovakav sud čini nešto drugo. Drugoj grupi ispitanika postavljeno je pitanje: ako krug određenog prečnika ubacimo u kvadrat koji je taman toliki da taj krug može da stane u njega, koliku površinu kvadrata će okupirati krug? Normativno tačan odgovor za prvo pitanje (odnosa zapremina lopte i kocke) je 52%, a rezultati ovog istraživanja pokazuju da ispitanici u prvom zadatku daju odgovor 74% koji je skoro potpuno isti kao odgovor 77% koji daje druga grupa ispitanika za problem odnosa površina kruga i kvadrata. „Težak“ *ciljni atribut zapremine* zamenjen je u donošenju suda „lakim“ odn. *heurističkim atributom površine* (Frederick & Nelson, 2004, prema Kahneman & Frederick, 2005).

Diskusije o primeni heuristika u donošenju sudova vodile su ka formiranju danas standardnih *teorija dva sistema*. Kanoničku formu objašnjenja ovim tipom teorija dao je Daniel Kaneman u svojoj lekciji povodom prijema Nobelove nagrade iz ekonomije 2002. godine, „*Mape ograničene racionalnosti*“ (Kahneman, 2002). Teorije dva sistema ponudilo je više autora, u različitim formama, i primeni na različite probleme (za pregled up. Osman, 2004). Kanemanova sintetička verzija teorije dva sistema predstavlja suštinu svih sličnih teorija. Prema ovom shvatanju, ljudski intuitivni sudovi o numeričkim veličinama poput verovatnoća i frekvencija nalaze se negde između brzih, potpuno automatizovanih, visoko paralelnih procesa percepcije, i sporih, serijalnih procesa rezonovanja. Tverski i Kaneman su celu oblast istraživanja koju su konstituisali istraživačkim programom heuristika i inklinacija posmatrali kao *kognitivnu primenu perceptivnih principa na simboličke reprezentacije*. Procesi rezonovanja, koji su vođeni pravilima i omogućavaju normativno zaključivanje, čine *Sistem 2* donošenja sudova i rezonovanja. *Sistem 1* donošenja sudova je zasnovan na brzim asocijativnim procesima i automatizovan, i njegov efekat se ogleda u primeni heurističkih rešenja. Zadatak *Sistema 2* jeste i nadgledanje, kontrola rada *Sistema 1*: kada vremenska ograničenja i kontekst to omogućavaju, *Sistem 1* može da koriguje greške *Sistema 2*; u suprotnom, ljudi su prepušteni brzim, intuitivnim i ne uvek tačnim sudovima koje bez mnogo napora donosi *Sistem 2*.

*Pametne heuristike: Gigerencerova kritika Kanemana i Tverskog.* Jedna od paradigmatičnih tvrdnji pristalica ograničene racionalnosti jeste da se ljudski kognitivni sistem u mnogim situacijama pokazuje nesposobnim za bejzijansku inferenciju, donoseći sudove o verovatnoćama koji sistematski odstupaju od *a posteriori* verovatnoća izračunatih po ovom fundamentalnom pravilu teorije verovatnoće. Čuveni *taksi problem* će poslužiti kao ilustracija (Tversky & Kahneman, 1972, prema Bar-Hillel, 1980):

„*Dve taksi kompanije rade u datom gradu, Plava i Zelena (prema boji taksija koje voze). 85% taksija u gradu su Plavi, i preostalih 15% su Zeleni. Taksi je bio umešan u noćni incident u kome je udario osobu posle čega je vozač nastavio i pobegao. Svedok je kasnije identifikovao taksi kao Zeleni taksi.*

*Sud je testirao sposobnost svedoka da razlikuje Plave i Zelene taksije pod uslovima noćnog viđenja. Pronađeno je da je svedok bio u stanju da identifikuje svaku boju u oko 80% slučajeva, ali ju je pomešao sa drugom bojom u oko 20% slučajeva. Kolike mislite da su šanse da je odbegli taksi zaista bio Zeleni, kako je tvrdio svedok?*“

(citirano prema Bar-Hillel, 1980, koja je preuzela problem iz Kahneman & Tversky, 1972, naš prevod).

Primena Bejzove teoreme iz teorije verovatnoće nalaže da se problem reši na sledeći način. *A priori* verovatnoća da je taksi Zeleni je $\mathrm{P}(z) = .15$, dok je *a priori* verovatnoća da je taksi Plavi, $\mathrm{P}(p) = .85$. Verovatnoća da je svedok zaista tačno opazio da je taksi Zeleni na osnovu perceptivnih podataka $D$ koji su njemu bili dostupni, $\mathrm{P}(D|z)$, iznosi .80, odn. onoliko koliko je svedok uspešan u prepoznavanju boja noću. Verovatnoća, dakle, da je svedok opazio taksi pogrešno, svedoći da je u pitanju Zeleni taksi iako je on bio Plavi, $\mathrm{P}(D|p)$, iznosi .20 - kolika je verovatnoća da svedok greši u identifikaciji boja u noćnom viđenju. Zavisne verovatnoće $\mathrm{P}(D|z)$ i $\mathrm{P}(D|p)$ - odn. verodostojnosti u Bejzovoj teoremi - intepretiramo kao verovatnoće da su perceptivni podaci svedoka $D$ proizvedeni Zelenim taksijem - to je $\mathrm{P}(D|z)$ - i da su ti perceptivni podaci $D$ proizvedeni Plavim taksijem - $\mathrm{P}(D|p)$. Prema Bejzovoj teoremi, *a posteriori* verovatnoća da je taksi koji je svedok opazio zaista bio Zeleni je onda

$$P(H_1|D) = \frac{P(D|H_1) \times P(H_1)}{P(D|H_1) \times P(H_1) + P(D|H_2) \times P(H_2)} =$$

$$P(z|D) = \frac{P(D|z) \times P(z)}{P(D|z) \times P(z) + P(D|p) \times P(p)} =$$

$$P(z|D) = \frac{.80 \times .15}{.80 \times .15 + .20 \times .85} = .413$$

odn. verovatnoća da je opaženi taksi koji je uzrokovao nesreću iznosi 41%. Veliki broj eksperimentalnih istraživanja sa sličnim problemima pokazao je da ljudske procene verovatnoća *veoma* odstupaju od aposteriornih verovatnoća izračunatih prema normativnom pravilu Bejzove teoreme (Koehler, 1996). Osnovna greška koju ljudi čine u proceni ovih verovatnoća jeste *zanemarivanje a priori verovatnoća* (engl. *base-rate neglect*), i to predstavlja jednu od najviše diskutovanih inklinacija u donošenju sudova. Na primer, u *taksi problemu*, veliki broj ljudi će tvrditi da je verovatnoća da je svedok tačno prepoznao dati taksi oko 80% - koliko iznosti stepen njegove uspešne identifikacije boja tokom noći. Međutim, realna verovatnoća mora da uzme u obzir činjenicu da su Zeleni taksiji *a priori* daleko manje frekventni od Plavih taksija. Bejzova teorema predstavlja normativno pravilo računa verovatnoće koje pokazuje kako se *a priori* verovatnoće inkorporiraju u ovakve procese donošenja

sudova.

Markantan doprinos diskusiji o intuitivnoj primeni Bejzove teoreme u zaključivanju dali su Gigerencer i Hofrag ustanovljavajući empirijski nalaz koji se danas smatra jednim od najvažnijih u oblasti suđenja uopšte (Gigerenzer & Hoffrage, 1995). Gigerencer i Hofrag su pretpostavili da je u mnogim situacijama za ljudski kognitivni sistem lakše da radi sa *frekvencijama događaja* - apsolutnim merama broja događaja u nekoj klasi - nego sa verovatnoćama. Polazeći od činjenice da algoritmi za rešavanje istih kompjutacionih problema uzimaju drugačije forme u funkciji formata u kojima primaju ulazne podatke, Gigerencer i Hofrage su formulisali Bejzovu teoremu u formatu frekvencija, a pored toga ponudili i nekoliko njenih pojednostavljenja koja efektivno daju dobre aproksimacije *a posteriori* verovatnoća u velikom broju slučajeva. Pošto su ispitanicima predstavili tipične probleme koji zahtevaju bejzijansku inferenciju u formatu frekvencija, pokazali su da broj ispitanika koji primeni normativno rešenje (čije ocene su bliske realnim *a posteriori* verovatnoćama) naglo raste u odnosu na studije u kojima se koriste verovatnoće (oko 50% ispitanika u studijama Gigerencera i Hofragea daje normativno adekvatne, bejzijanske odgovore).

Uopšte, ovakva strategija konstrukcije „pametnih heuristika" - heuristika koje ne vode *isključivo* ka inklinacijama, već dobro aproksimiraju normativna rešenja, karakteristična je za Gigerencerov pristup ekološke racionalnosti koji smo diskutovali u Delu II ove rasprave. Tokom 90-ih godina, Gigerencer je temeljno kritikovao istraživački program Kanemana i Tverskog iz nekoliko perspektiva (Gigerenzer 1991b, 1993, 1994, 1996). Prvo, Gigerencer pokazuje da rezultati do kojih dolazi program heuristika i inklinacija nisu stabilni, što ilustruje upravo reformulacija problema bejzijanske inferencije iz formata verovatnoća u format frekvencija. Drugo, Gigerencer tvrdi da Kaneman i Tverski greše u izboru normativnih principa u odnosu na koje ocenjuju sudove koje ispitanici donose kao pogrešne. Gigerencer primećuje da ni u matematičkoj statistici, kao normativnoj disciplini, ne postoji opšte slaganje oko normativnih principa. Većina problema, primećuje Gigerencer, koje koriste Kaneman i Tverski se odnose na donošenje sudova o *jednom slučaju* (kao u primeru Linde); među matematičarima i filozofima uopšte ne postoji koncezus oko toga da li se pravila teorije verovatnoće (i kako) primenjuju u situacijama kada se diskutuju singularni slučajevi. Konačno, Gigerencer primećuje da su heurističke strategije poput reprezentativnosti, dostupnosti i određivanja referentne tačke nedovoljno specifikovane, te da zbog toga omogućavaju arbitrarne interpretacije

eksperimentalnih nalaza (ove kritike Gigerencer sumira u diskusiji sa Kanemanom i Tverskim u časopisu „*Psychological Review*“, Gigerenzer, 1996).

Prema idejama Gigerencera, ljudski kognitivni sistem ne rešava adaptivne probleme optimizacijama, koje uvek vode ka normativnim rešenjima, ali su kompjutaciono izuzetno skupe. U ovome se slažu istraživački programi Kanemana i Tverskog i njegov. Međutim, Gigerencer ne veruje da su heuristici kojima kognitivni sistem raspolaže dizajnirani tako da vode pravo u inklinacije i greške koje su dokumentovali Kaneman i Tverski. Heuristici su, prema njemu, sredstva koja su evolucijom dizajnirana da brzo, efikasno i aproksimativno rešavaju adaptacione probleme (Todd & Gigerenzer, 2000, 2007, Gigerenzer, 2008). Gigerencer nas poziva da razmislimo o dizajnu mogućih takvih heuristika i da u razvoju heurističkih teorija analiziramo koje podatke i u kakvoj formi kognitivni sistem ima šanse da izdvoji iz okoline. Tako sud o ograničenoj racionalnosti dobija različite konotacije kod Kanemana i Tverskog, s jedne, i Gigerencera, s druge strane: u paradigmi heuristika i inklinacija, kognitivni sistem čoveka praktično ide iz greške u grešku, dok u paradigmi ekološke racionalnosti on jeste ograničeno racionalan ali samo u smislu primene ekološki valdinih heuristika koje ipak aproksimiraju adaptivno relevantna rešenja.

*Racionalne osnove reprezentativnosti.* Racionalnu, bejzijansku analizu reprezentativnosti kao dominante heuristike u rešavanju mnogih problema pokušali su da sprovedu Tenenbaum i Grifits (Tenenbaum & Griffiths, 2001). Prema našem mišljenju, ova racionalna analiza reprezentativnosti suočava se sa ozbiljnim problemima. Tenenbaum i Grifits polaze od određenja problema za koje opravdano može da se postavi pitanje da li predstavlja *isti* problem kao onaj koji po Kanemanu i Tverskom rešava heuristika reprezentativnosti. Poznato je da će ljudi donositi različite sudove o tome koji od sledećih nizova bacanja novčića je zaista proizveden bacanjem fer (50/50) novčića: GPGPPPG, GGGGGGG, GPPPGGG (G - „glava“, P - „pismo“). Teorija verovatnoće nam kaže da je svaka od ovih sekvenci, naravno, podjednako verovatna, ali naša intuicija prkosi tom normativnom stavu. Niz GGGGGGP ne deluje kao da je moglo da ga proizvede bacanje fer novčića. Tenenbaum i Grifits uočavaju da je niz poput GPGPPPG *reprezentativniji* za proces slučajnog bacanja fer novčića od niza poput GGGGGGP, i da je to razlog što ljudi različito sude o verovatnoćama različitih, podjednako verovatnih sekvenci G i P. Ipak, postavljajući problem na ovaj način, Tenenbaum i Grifits u eksperimentalnom testu racionalne teorije reprezentativnosti koju ćemo sada predstaviti ne postavljaju

svojim ispitanicima pitanje o tome koja sekvenca je *verovatnija*, već koja sekvenca je *više reprezentativna*. Originalni problemi Kanemana i Tverskog ciljali su na procenu verovatnoće određenog događaja, pa je i heuristika reprezentativnosti razvijena je u odnosu na odstupanje od normativne ocene verovatnoće - ne same reprezentativnosti, koja se u nacrtima Kanemana i Tverskog prepoznaje kao *nezavisna*, *ne kao zavisna varijabla* (up. Kahneman & Frederick, 2005).

Tenenbaum i Grifits ne čine mnogo do toga da kao meru reprezentativnosti ponude (u bejzijanskoj teoriji i filozofiji nauke dobro poznatu) meru *logaritamskog odnosa verodostojnosti* (engl. *log-likelihood ratio*). Ova mera predstavlja normativnu ocenu doprinosa objašnjenju određenih opservacija od strane jedne u skupu alternativnih hipoteza. Uopšte, test logaritamskog odnosa verodostojnosti se često koristi (podjednako i od strane frekvencionista) da bi se donela odluka o tome koji od dva (ili više) modela bolje objašnjava podatke dobijene određenim eksperimentom. Taj test ima sledeću formu:

$$logL(D,h) = log\frac{P(D|h)}{P(D|\bar{h})} \tag{39}$$

i ima očiglednu interpretaciju: u pitanju je logaritam[58] verodostojnosti, odn. verovatnoća da su osmotreni podaci $D$ generisani modelom $h$ i da su osmotreni podaci $D$ generisani nekim drugim modelom (dakle, $\bar{h}$). Tenenbaum i Grifits testiraju ovaj model na podacima o reprezentativnosti sekvenci bacanja novčića i pokazuju da on visoko korelira sa bihejvioralnim procenama reprezentativnosti. Međutim, pored konfuzije oko izbora nezavisne i zavisne varijable u ovom radu, model ima i drugih problema. Na prvom mestu, da bi se primenio, test logaritamskog odnosa verovatnoća zahteva da sistem koji ga primenjuje (to je najčešće psiholog, ekonomista ili statističar zainteresovan za poređenje dva ili više modela nekih podataka) *generiše hipoteze* o tome šta je sve moglo da proizvede podatke. Generisanje hipoteza već za sam slučaj sekvenci G i P bacanja novčića je netrivijalan problem. Kako kognitivni sistem može da pretpostavi koje su svi mogući procesi koji su mogli da generišu sekvence koje opaža - „*novčić koji uvek pada na G*“, „*novčić koji 50/50 pada na G i P*“, „*novčić koji pada tri puta na G pa tri puta na P onda dva puta na G i pet puta na P i u krug*“ - i tako dalje? Tenenbaum i Grifits ovo pitanje rešavaju tako što pojednostavljuju prostor hipoteza, uključujući u analizu samo one koje su visoko karakteristične (izgledaju tipično) u odnosu na stimuluse (sekvence G i P) koje koriste u eksperimentu. Možda nije iznenađujuće što ovakva

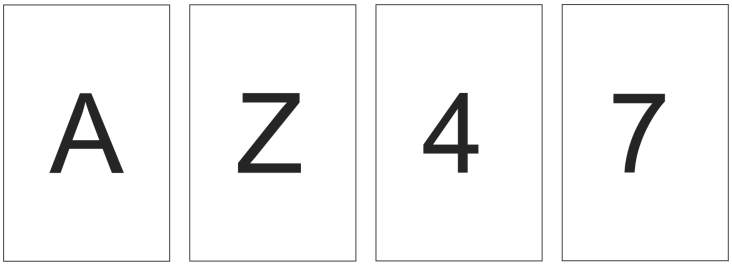
cirkularnost između metodoloških procedura i modela eksperimentalnih nalaza vodi ka visokoj korelaciji između predloženog modela i bihejvioralnih procena.

Ipak, empirijski rad o predikcijama svakodnevnih informacija koji su Grifits i Tenenbaum predstavili 2006. godine, a koji smo diskutovali prilikom uvođenja koncepta racionalnih bejzijanskih teorija u II delu naše rasprave, izvesno predstavlja najjaču do sada pruženu evidenciju protiv dominantog stava da su ljudski sudovi prožeti inklinacijama i da se oslanjaju na heuristike. Podsetimo, Grifits i Tenenbaum su iskoristili ekološki relevantnu situaciju u kojoj su jednostavno pitali ljude da donesu predikcije odgovarajućih događaja na osnovu ma kog znanja koje bi oni mogli o tim događajima da imaju. Jednostavan bejzijanski model, daleko plauzibilniji od njihovog racionalnog modela reprezentativnosti, uspeo je da pruži objašnjenje za ljudske sudove koji se očekuju u svakodnevnim situacijama (Griffiths & Tenenbaum, 2006). Eksperimentalna paradigma koja je korišćena u ovoj studiji jasno se razlikuje od tipičnih eksperimentalnih paradigmi koje su koristili Kaneman i Tverski, upućujući nas da se zapitamo u kojoj meri problem proceduralne invarijantnosti - odn. način na koji su adaptivni problemi u eksperimentalnim situacijama predstavljeni ispitanicima - utiče na naše teorijske zaključke o funkcijama prosuđivanja.

Najubedljiviji argumenti pristalica ograničene racionalnosti nalaze se eksperimentalnim rezultatima koje je donela oblast donošenja sudova. Po našem mišljenju, do sada nije ponuđena racionalna analiza koja bi objasnila pod kojim, i kakvim uslovima, prosuđivanje o verovatnoćama može da se shvati kao normativno adekvatan proces, bez obzira na to da li se kao norma posmatra račun verovatnoće ili neka alternativna, strukturom okoline definisana norma adaptacije. Nalaze o optimalnom donošenju sudova poput Grifitsovih i Tenenbaumovih još treba replicirati i ispitati njihovu stabilnost u odnosu na različite metodologije da se do njih dođe. Analize Gigerencera, koji je i sam zastupnik stavova paradigme ograničene racionalnosti, ukazuju tek na to da određeni heuristici mogu da budu motivisani varijablama koje se odnose na strukturu okoline (format u kome se predstavljaju podaci, npr) i aproksimativno tačni u odnosu na nisku cenu izračunavanja koju nose.

*Rezonovanje u Vosonovom zadatku selekcije.* Istraživanja ljudskog deduktivnog rezonovanja su takođe prepuna demonstracija iracionalnosti u odnosu prema normativnim standardima logičkog rasuđivanja. Krajem 60-ih godina XX veka, Piter Voson je formulisao eksperimentalnu paradigmu koja će odlučujuće odrediti buduća

istraživanja i teorijske predloge u oblasti rezonovanja (Wason, 1968). Vosonov *zadatak selekcije*, u svom kanoničkom obliku, izgleda ovako: pred ispitanikom se nalaze četiri karte, obeležene slovima i brojevima. Na primer, karte su predstavljene kao na Slici 16. Ispitaniku se objašnjava da postoji pravilo vezano za raspored slova i brojeva sa jedne i druge strane karata, na primer: „*Ako karta ima samoglasnik na jednoj strani, onda na drugoj strani ima paran broj*“. Pravilo je uvek tipa *ako p, onda q,* te po logičkoj formi prati *Modus Ponens*, jedino pravilo inferencije u formalnom sistemu iskaznog računa.




Slika 16. *Vosonov zadatak selekcije.* Za dato pravilo poput „*Ako karta ima samoglasnik na jednoj strani, onda na drugoj strani ima paran broj*“, ispitanik treba da odredi koje karte treba okrenuti da bi se proverilo da li pravilo važi.

Ispitanik u zadatku selekcije treba da pokaže koje karte treba okrenuti da bi se proverilo da li prethodno formulisano pravilo stoji, ili ne. Oslanjajući se na *Poperov kriterijum falsifikacije* (Popper, 1935/2002), prema kome hipotezu u strogom smislu reči nije moguće potvrditi, već samo *opovrći*, postoji samo jedno normativno adekvatno rešenje Vosonovog zadatka. Neka je pravilo zaista formulisano kao u primeru, „*Ako karta ima samoglasnik na jednoj strani, onda na drugoj strani ima paran broj*“. Označimo činjenicu da „*karta ima samoglasnik na jednoj strani*“ sa $p$, i činjenicu da „*karta ima paran broj na jednoj strani*“ sa $q$; odgovarajuće karte u zadatku selekcije zovemo $p$-kartom i $q$-kartom. Primer na Slici 16. tako redom sadrži $p$-kartu, *ne*-$p$-kartu, $q$-kartu i *ne*-$q$-kartu. Isključivo izbor $p$-karte i *ne*-$q$-karte predstavlja proceduru koja može da opovrgne pravilo. Okretanjem $p$-karte dobijamo informaciju koja potvrđuje pravilo, ako se sa druge strane nalazi paran broj; ako se sa druge strane ne nalazi paran broj, pravilo je automatski opovrgnuto. Okretanjem *ne*-$q$-karte možemo da falsifikujemo pravilo: ako ona sa druge strane sadrži samoglasnik, pravilo ne stoji; ako sadrži suglasnik, pravilo je ponovo potvrđeno. Okretanje *ne*-$p$ karte potpuno je neinformativno: pravilo ne govori ništa o tome šta se nalazi na poleđini karata koje imaju suglasnik sa jedne

strane. Okretanjem $q$-karte pravilo možemo da potvrdimo (ako je samoglasnik sa druge strane), ili da o njegovom važenju ne saznamo ništa (ako je suglasnik sa druge strane).

Prva eksperimentalna istraživanja zadatkom selekcije pokazala su da *samo 4% ispitanika* normativno rešava ovaj zadatak; Džonson-Lerd i Voson su ustanovili da su drugačiji izbori od normativnog daleko češći: oko 46% ispitanika bira kombinaciju $p$-karta i $q$-karta, samo $p$-kartu okreće oko 33% ispitanika, a $p$-kartu, $q$-kartu i *ne*-$q$-kartu njih 7% (Johnson-Laird & Wason, 1970). Ovakav nalaz motivisao je jednu od najdužih i najkompleksnijih diskusija u debati o racionalnosti. Rezultati eksperimenata koji primenjuju ovu osnovnu formu Vosonovog zadatka - koja se naziva *apstraktnom formom* zbog upotrebe samo slova i brojeva - dosledno repliciraju nalaze Džonson-Lerda i Vosona (up. Oaksford & Chater, 1994, za sumarni pregled rezultata većeg broja ovakvih studija).

Do interesantne promene u rezultatima, međutim, dolazi sa uvođenjem tzv. *tematskih formi* zadatka selekcije. U tematskim formama se koristi konkretan, verbalni materijal koji zamenjuje apstrakne simbole lišene značenja u osnovnoj formi zadatka. Istraživanja upotrebom tematskih formi odnose se na tzv. *deontičke zadatke selekcije*, u kojima se umesto apstraknih pravila vezanih za raspored slova i brojeva koriste pravila koja denotiraju određene *obaveze*, *dozvole* i sl. sheme u kojima je logički veznik materijalne implikacije zamenjen sa „*mora*“, „*treba*“, „*nužno je da*“ i sl. Na primer, pravilo „*Ako osoba pije pivo, mora da ima bar 19 godina*“ je jedna forma deontičkog pravila koja može da se koristi u zadatku selekcije. U ovako postavljenom zadatku selekcije, čak 70% ispitanika daje normativno adekvatne odgovore, odn. bira one karte koje zaista imaju potencijal da opovrgnu pravilo (Griggs & Cox, 1982, prema Kostić, 2006). Čengova i Holiuk su odnosu na ovakve verzije Vosonovog zadatka postavili teoriju *pragmatskih shema rezonovanja* prema kojima se vode odgovori ispitanika (Cheng & Holyoak, 1985). Pragmatske sheme rezonovanja ne predstavljaju opšta sintaksička pravila rezonovanja koja su nezavisna od sadržaja o kome se rezonuje - takva nezavisnost je osobina formalnih logičkih sistema. Ipak, one nisu ni puka primena prethodno zapamćenih pojedinačnih situacija u kojima je određena strategija rezonovanja bila korisna. Pragmatske sheme rezonovanja predstavljaju rezultate indukcije kroz veći broj slučajeva, ali indukcije koja se zaustavlja u granicama određenog sadržaja, i najbolje ih predstavljaju upravo primeri različitih obaveza, dozvola i kauzalnih pravilnosti kojima smo okruženi u svakodnevnici. U eksperimentu kojim demonstriraju efekat

pragmatskih shema, Čengova i Holiuk od ispitanika traže da zauzmu perspektivu osobe koja vrši pasošku kontrolu na granici, proveravajući da li ulazna dokumenta svih putnika zadovoljavaju sledeće pravilo: „*Ako sa jedne strane dokumenta stoji da putnik ulazi u zemlju, među bolestima navedenim na drugoj strani dokumenta mora da se nalazi kolera*[59]“. Ispitanicima su prikazane karte sa terminima „ULAZ“, „TRANIZIT“, „KOLERA, TIFUS, HEPATITIS“ i „TIFUS, HEPATITIS“. Dvema grupama ispitanika dat je identičan zadatak, osim što je jednoj grupi dodatno objašnjeno da se pravilo odnosi na to da ako putnik ulazi u zemlju, a nije tek u tranzitu, on mora da između ostalih vakcinacija dokaže i zaštitu protiv kolere. U toj grupi ispitanika, njih skoro 90% dalo je normativno adekvatan odgovor, za razliku od grupe kojoj nije pruženo dopunsko objašnjenje o značenju i značaju pravila, u kojoj je oko 60% ispitanika dalo normativan odgovor. Kao što vidimo, u tematskim zadacima selekcije je broj ispitanika koji daju normativno adekvatne odgovore u svakom slučaju veći nego u primeni apstraktne verzije zadatka.

Deontičkim verzijama Vosonovog zadatka vratićemo se nešto kasnije. Racionalnu analizu zadatka selekcije koju sada predstavljamo sproveli su Oaksford i Čater 1994. godine. Njihov doprinos predstavlja jednu od najplodnijih racionalnih analiza u tradiciji ove metodologije.

*Racionalna analiza Oaksforda Čatera.* Racionalna analiza zadatka selekcije počiva na bejzijanskoj teoriji *optimalne selekcije podataka* - teoriji koja je i normativno adekvatna klasi problema kakvu predstavlja Vosonov zadatak. U većini primena bejzijanskih modela, pretpostavlja se uniformna distribucija *a priori* verovatnoće da je neka od razmatranih hipoteza „tačna“, tj. da je upravo ona generisala podatke $D$ koji su rezultat naših opservacija. Da je distribucija verovatnoće hipoteza uniformna znači to da su *a priori* sve hipoteze podjednako verovatne. U racionalnoj analizi često se pretpostavlja suprotno: da *a priori* distribucija hipoteza prati neku empirijski ustanovljenu, ili tek teorijski plauzibilnu, distribuciju *environmentalno relevantnih informacija.* Distribucije prethodne upotrebe reči u relevantnim izvorima informacija u racionalnom modelu pamćenja Andersona i saradnika su upravo takve *a priori* distribucije. Bejzova teorema nam onda omogućava da izračunamo a posteriori verovatnoće da je podatke $D$ generisala neka od hipoteza koje razmatramo, polazeći od tih *a priori* verovatnoća. U optimalnoj selekciji podataka postavljamo „inverzni“ problem: formulisana je određena hipoteza, a problem se sastoji u *selekciji eksperimenta* koji bi najviše doprineo našem saznanju o tome u kojoj meri je ta hipoteza tačna. Vratimo se

Vosonovom zadatku. Svaku moguću kombinaciju okretanja karata koje se nalaze pred ispitanikom možemo da tretiramo kao jedan eksperiment kojim ispitanik pokušava da prikupi podatke koji su korisni za konfirmaciju i falsifikaciju hipoteze *ako p, onda q*. Oaksford i Čater primećuju da optimalna selekcija podataka možda predstavlja normativni okvir na koji se oslanja kognitivni sistem u Vosonovom zadatku - ako se o testu važenja *Modus Ponens* pravila ne razmišlja *logicistički*, već *probabilistički*. Pogledajmo detaljnije o čemu je reč.

U optimalnoj selekciji podataka, normativni cilj koji treba ispuniti jeste selekcija onih podataka (onog eksperimenta) koji će voditi ka *najvećoj očekivanoj redukciji neizvesnosti* u odnosu na hipotezu koja se testira. Neizvesnost se formalizuje jezikom Šenonove teorije informacija (Shannon, 1948):

$$I(H_i) = -\sum_{i=1}^{n} P(H_i) \cdot log_2 P(H_i) \tag{40}$$

i u tom smislu vidimo da je hipoteza shvaćena kao jedna distribucija verovatnoće. Kako je to moguće izraziti jednu hipotezu oblika *ako p, onda q* probabilistički, kao distribuciju verovatnoće? Na primer: za pravilo *ako je p, onda q*, jasno je da ako je $p$ već dato, $q$ mora da ima najmanje onu verovatnoću da se pojavi kao i $p$, pošto pravilo propisuje da $q$ sledi $p$ po materijalnoj implikaciji. Na ovaj način je moguće formulisati zavisne i zajedničke distribucije verovatnoća za varijable u svim logičkim pravilima. Tako, pristup Oaksforda i Čatera odlikuje prelaz za *logicističke* na *probabilističku teoriju* rezonovanja, prelaz koji su ovi autori u više teorijskih radova označili kao fundamentalan za teoriju rezonovanja (Oaksford & Chater, 1998, 2001). Drugim rečima pravilo *ako p, onda q* treba interpretirati kao distribuciju verovatnoće nad događajima (p, q, $\neg p$,$\neg q$) koju ono određuje[60]. Ovo je ujedno i jedini način da se rezonovanje po *Modus Ponensu* poveže sa teorijom optimalne selekcije podataka. Konkretno, probabilističke verzije pravila kao što je Modus Ponens mogu da se izraze preko kontingencijske tabele na sledećin način (Oaksford & Chater, 1994, 2001):

Tabela3. Probabilistička interpretacija logičkih argumenata.

| | $q$ | $\neg q$ |
|---|---|---|
| $p$ | a(1-$\varepsilon$) | a$\varepsilon$ |
| $\neg p$ | b-a(1-$\varepsilon$) | 1-b-a$\varepsilon$ |

gde su verovatnoće događaja $p$ i $q$ date sa $a = \mathrm{P}(p)$, $b = \mathrm{P}(q)$, a sa $\varepsilon$ verovatnoća *izuzetaka* od pravila *Modus Ponens*, $\mathrm{P}(\neg q|p)$. Čitalac može sam da ustanovi da ovako definisane verovatnoće formiraju pravu distribuciju verovatnoća i odgovaraju probabilističkom opisu kondicionala koji predstavlja *Modus Ponens*: $\mathrm{P}(q|p)$.

Posle prikupljanja određenih podataka $D$ - to je okretanje karata u Vosonovom zadatku - neizvesnost o statusu hipoteze se meri zavisnom informacionom entropijom koja je jednostavno data sa

$$I(H_i|D) = -\sum_{i=1}^{n} P(H_i|D) \cdot log_2 P(H_i|D) \tag{41}$$

gde je $\mathrm{P}(H_i|D)$, očigledno, mera verodostojnosti hipoteze $H_i$ pod podacima $D$. Dakle, dok $H_i$ predstavljaju probabilistički izrazi pravila koja se testiraju, podaci $D$ predstavljaju ishode okretanja karata u Vosonovom eksperimentu. Svaki izbor, svaka kombinacija od npr. dve karte koje mogu da se okrenu u skupu od četiri prikazane definiše jedan eksperiment za koji je potrebno odrediti, u odnosu na njegove moguće ishode, u kojoj meri može da umanji neizvesnost da je testirana hipoteza tačna (ili ne). Ona je, opet, data sledećim bejzijanskim izrazom

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_j P(D|H_j)P(H_j)} \tag{42}$$

Posle izračunavanja $\mathrm{I}(H_i|D)$, kognitivnom sistemu preostaje da ustanovi koliki je doprinos podataka D smanjenu redukcije neizvesnosti o statusu hipoteze $H_i$ kao

$$I_g = I(H_i|D) - I(H_i) \tag{43}$$

što je samo razlika između neizvesnoti, tj. količine informacija (entropije) vezane za hipotezu pod posmatranim podacima i neizvesnosti koja je bila vezana za istu tu hipotezu *pre* nego što su podaci osmotreni (tj. pre nego što su karte okrenute). Naravno, pošto pre nego što se eksperiment okretanja karata uopšte izvede, kognitivni sistem ne može da izračuna ni $I(H_i|D)$ ni $I(H_i)$, pa umesto njih mora da iskoristi očekivanu vrednost ovih izraza u odnosu na sve moguće ishode eksperimenta:

$$E(I_g) = E(I(H_i|D) - I(H_i)) \tag{44}$$

Kao što je već rečeno: što je veća $\mathrm{E}(I_g)$, očekivana redukcija u neizvesnosti hipoteze $H_\mathrm{i}$ *posle selekcije određene kombinacije karata*, to je jača i evidencija da je upravo pravilo izraženo tom hipotezom tačno.

U okviru ovog probabilističkog modela, Oaksford i Čater su uspeli da postave problem na način koji im omogućava da odrede u kojoj meri je, u testiranju pravila *Modus Ponens*, informativno okretanje svake pojedinačne karte u Vosonovom zadatku. Njihovi zaključci ukazuju na sledeće: (*i*) $p$-karta je informativna u onoj meri u kojoj je osnovna (*base-rate*) verovatnoća da se javi događaj $q$, $\mathrm{P}(q)$, *niska*, a istovremeno, informativnost $p$-karte je u velikoj meri nezavisna od verovatnoće javljanja $p$, $\mathrm{P}(p)$; (*ii*) $q$-karta je informativna u onoj meri u kojoj su i $\mathrm{P}(q)$ i $\mathrm{P}(p)$, osnovne verovatnoće javljaja događaja $p$ i $q$, *niske*; (*iii*) *ne*-$q$-karta je informativna kada je $\mathrm{P}(p)$, osnovna verovatnoća događaja $p$, visoka, i nezavisna od verovatnoće javljanja događaja $q$, $\mathrm{P}(q)$; (*iv*) *ne*-$p$-karta, očekivano, nije informativna ni pod kakvim uslovima. Podsetimo se da najveći broj ispitanika u apstraknim formama Vosonovog zadatka ne okreće $p$-kartu i *ne*-$q$-kartu, što je normativno adekvatno u odnosu na primenu pravila Modus Ponens, već $p$-kartu i $q$-kartu (46% ispitanika prema Džonson-Lerdu i Vosonu; preko ovog broja, još 33% ispitanika testira samo $p$-kartu, Johnson-Laird & Wason, 1970). Kombinacija okretanja ove dve karte je zaista najinformativnija za optimalni test hipoteze - izražene kao probabilističke verzije pravila *Modus Ponens* - u uslovima u kojima su događaji $p$ i $q$ *retki*. Upravo ovo zapažanje konstituiše racionalno objašnjenje velikog broja normativno neadekvatnog izbora $p$-karte i $q$-karte u Vosonovom zadatku: Oaksford i Čater tvrde da našu realnu okolinu karakteriše retkost, odn. mala *a priori* verovatnoća singularnih događaja - ma kojih $p$ i $q$. Ova pretpostavka, koju Oaksford i Čater nazivaju *hipotezom retkosti* (Oakfsord & Chater, 1994), od njihovog rada predstavlja jednu od najčešće korišćenih hipoteza o distribuciji environmentalno relevantnih informacija u metodologiji racionalne analize. U sekciji 7.2 posvećenoj kauzalnom učenju videli smo da se Oaksford i Čater oslanjaju na istu ovu hipotezu u formulisanju heuristike $H$ kojom objašnjavaju bihejvioralne procene intenziteta kauzalnih odnosa. Posvetićemo ovoj hipotezi sada nešto više pažnje.

Rezon koji vodi usvajanju hipoteze o retkosti, tvdnji da su pojedinačni događaji koji uzimaju učešća u našim verbalnim opisima sveta *a priori* veoma malo verovatni,

je sledeći. Uzmimo za primer argument u kom učestvuju dva koncepta, kategorija PTICA, i glagol *leteti*: ako je X PTICA, onda X *leti*. Oaksford i Čater primećuju sledeće: kada govorimo o singularnom događaju, kao što je događaj da vidimo neku pticu, ili da upotrebimo reč „*ptica*“ u diskursu, govorimo o događaju koji je izuzetno niske verovatnoće u odnosu na to šta smo sve mogli, i šta sve inače koristimo u diskursu komunikacije. Podjednako, taj rezon važi i za glagol *leteti* koji smo iskoristili u primeru. Jezički kodifikovane kategorije koje koristi ljudski kognitivni sistem svet dele na veoma fin, *precizan način*: u njemu, zapravo, ima veoma malo ptica, i veoma malo objekata koji lete, a mi svejedno imamo reči kojima označavamo te veoma male kategorije događaja koji nas okružuju. Prema hipotezi o retkosti, dakle, osnovno stanje probabilističke mašinerije kognitivnog sistema u zadacima rezonovanja, suđenja i sličnim jeste stanje u kome su *a priori* verovatnoće singularnih dođaja veoma niske. Mekenzi i Mikelsenova daju izvanrednu ilustraciju rezona koji opravdava hipotezu o retkosti (McKenzie & Mikkelsen, 2000) kroz diskusiju Hempelovog *paradoksa konfirmacije* (Hempel, 1945). Hempel je primetio sledeće: argument tipa „Svi GAVRANI su CRNI“ može da se prepiše u formu: „Ako je nešto GAVRAN, onda je to CRNO“. Kao kontrapozicija ove hipoteze, logički ekvivalentna forma „Ako nešto nije CRNO, to nešto nije GAVRAN“ dopušta da pojava svake NE-CRNE stvari koja je NE-GAVRAN predstavlja konfirmaciju početne hipoteze. Pošto su hipoteze logički ekvivalentne, opservacije plavih trkačkih automobila, crvenih mini-suknji ili belih zečeva su sve podjednako konfirmatorne informacije za hipotezu „Svi GAVRANI su CRNI“: neko može da sedi kod kuće, nikada u životu ne vidi gavrana, i logički validno prikuplja evidenciju da bi gavran morao da bude crn. Mekenzi i Mikelsenova navode da su filozofi pronašli način da Hempelov paradoks otklone upravo ako se on shvati kao stvar stepena - kako ga shvatamo u bejzijanskoj filozofiji nauke. U tom smislu, pošto su NE-CRNI NE-GAVRANI, odn. celokupan ostatak opservabilnog sveta koji isključuje CRNE GAVRANE, poptuno uobičajeni, stepen u kome oni doprinose konfirmaciji hipoteze „Svi GAVRANI su CRNI“ je izuzetno nizak, a stepen konfirmacije do kog dovodi opservacija CRNIH GAVRANA izuzetno visok - proprocionalno meri u kojoj su crni gavrani, ipak, retki u svakodnevnom okruženju.

U novijim radovima, probabilistički pristup rezonovanju proširen je u analizu složenog silogističkog rezonovanja koje ovde ne razmatramo (Oaksford, Roberts & Chater, 2002, Oaksford & Chater, 1999).

*Darvinijanski algortimi Lide Kosmajdes.* Lida Kosmajdes, rodonačelnica savremenog pravca *evolucione psihologije* (koji u prethodne dve decenije potiskuje i zamenjuje nekada popularnu paradigmu *sociobiologije*), ponudila je značajno alternativno objašnjenje efekata koje proizvode tematske verzije Vosonovog zadatka. Podsetimo se, Čeng i Holiuk su objašnjavali veći procenat normativnih odgovora u tematskim zadacima pojmom pragmatskih shema, semantičkih reprezentacija koje ispitanicima omogućavaju da znanja iz realnih situacija aktiviraju u rešavanju zadatka selekcije. Ideja koju je ponudila Kosmajdesova je donekle drugačija, i veoma značajna zbog uvođenja principa evolucione racionalnosti u raspravu o rezonovanju.

Rešavanje veoma važnih adaptivnih zadataka je tokom evolucije ljudskog kognitivnog sistema moralo da dovodi do *intezifikacije selekcionih pritisaka* da upravo takve klase zadataka budu uspešno rešavane. Ova činjenica procesa evolucije, koja lako otklanja lamarkijanske argumente i zabune vezane za „mogućnost genetskog nasleđivanja naučenih ponašanja“, opšte je mesto u savremenoj evolucionoj biologiji (Monod, 1971). U trenutku kada se neki selekcioni pritisak pojavi, boljom adaptacijom reaguju oni organizmi koji su slučajnim procesom genetskih mutacija pripremljeni na bolji adaptivni odgovor u kontekstu tih novih selekcionih pritisaka. Podižući tako svoju *inkluzivnu adaptivnu vrednost*, praktično postajući „konkurentiji“ na tržištu evolucionih proizvoda, bolje prilagođeni organizmi, osposobljeni za rešavanje novih klasa problema, doprinoseći svojim novim, više adaptivnim ponašanjima *podižu već postojeći nivo selekcionih pritisaka* na sve jedinke u odgovarajućoj ekološkoj niši. Proces evolucije algoritama za rešavanje određenih važnih klasa problema adaptacije tako je kroz princip povratne sprege podizao selekcione pritiske koji su favorizovali njega samog, vodeći ka značajnim prednostima za organizme čije su genetske mutacije i ukrštanja omogućavale adaptacije u tako već određenom pravcu. Zahvaljući ovakvoj prirodi procesa prirodne selekcije, lingvista Noam Čomski, u slučaju ljudskog jezika, a posle njega više evolucionih psihologa, mogli su da tvrde da je proces evolucije u slučaju ljudskog kognitivnog sistema vodio ka pojavi usko specifičnih domena kognitivnih funkcija - *modula* - specijalizovanih za rešavanje određenih klasa adaptivnih problema. Kanoničku verziju hipoteze o modularnosti dao je američki filozof Džeri Fodor u klasičnom eseju „*Modularnost uma*“ 1983. godine (Fodor, 1983). Posle originalne Fodorove verzije modularnosti iz 1983, diskusija o njenoj pravoj prirodi oživela je teorijske sukobe lokalizacionista i holista u raspravi o lokalizaciji kognitivnih funkcija, i tako dobila nov zamah koji je održava i danas.

Savremeno shvatanje o modularnosti, karakteristično za pravac evolucione psihologije, podrazumeva da su modularno organizovane kognitivne funkcije specijalizovane u odnosu na *sadržaj* nad kojim operišu (Cosmides & Tooby, 1987, Tooby & Cosmides, 2005). Drugim rečima, *ne postoji* jedan kognitivni modul, određena kompjutaciona, algoritamska realizacija kognitivnih funkcija u centralnom nervnom sistemu, koji je posvećen rezonovanju *uopšte*, *generalno* - *nezavisno od sadržaja toga o čemu se rezonuje*. Selekcioni pritisci koje je tokom svoje evolucije trpela ljudska vrsta nisu bili takve prirode da bi specijalizacija generalnih, apstraktnih kognitivnih sposobnosti predstavljala optimalan odgovor na te pritiske. Sistemi formalne logike, na primer, posledice su socio-kulturne evolucije čoveka koje su tek najnovijeg datuma. Za poređenje relevantnih evolucionih skala: oko 99% svoje istorije naši evolucioni preci proveli su u uslovima *Pleistocena* (aproksimativno, od pre oko dva i po miliona do pre oko 11.700 godina), u lovačko-sakupljačkim zajednicama; tek pre oko 10.000 godina nalaze se prvi tragovi poljoprivrede. Lida Kosmajdes iznosi argumente prema kojima je evolucija kognitivnih modula koji vrše funkcije rezonovanja morala da se odvija u *kontekstu socijalne kooperacije* između pojedinačnih pripadnika ljudske vrste, jer je upravo socijalna kooperacija predstavljala ono što je njih markantno izdvajalo u odnosu na druge vrste kojima su bili okruženi u rešavanju adaptivnih problema (Cosmides, 1985, 1989). Značaj *socijalnih ugovora*, i metoda da se proveri njihovo važenje - metoda koje je ljudski kognitivni sistem mogao da evoluira pred selekcionim pritiscima da uspešno učestvuje u socijalnoj kooperaciji - prema Kosmajdesovoj predstavlja osnovu na kojoj je procesom evolucije izgrađena ljudska sposobnost rezonovanja.

Za razliku od pragmatskih shema Čengove i Holiuka, kognitivni modul čija je funkcija provera validnosti socijalnih ugovora, teorijski konstrukt koji Kosmajdesova primenjuje u analizi zadatka selekcije, nije proizvod indukcije specifične u nekom domenu iskustva, već predstavlja evolucijom izgrađeni kompjutacioni mehanizam. Taj kompjutacioni mehanizam, koji na Marovom trećem nivou analize moramo da analiziramo u kontekstu socijalnih situacija, na algoritamskom i neurofiziološkom nivou jeste implementacija *darvinijanskih algoritama* (Cosmides, 1989): adaptivnih procedura za obradu relevantnih informacija. Kosmajdesova ističe da je u kontekstu provere validnosti socijalnih ugovora ključna *procedura za otkrivanje prevaranata* (engl. *cheater detection*). Sa stanovišta organizma koji mora da maksimizuje svoju adaptivnu vrednost pod datim ograničenjima, a čiji fokalni skup adaptivnih problema predstavljaju socijalne situacije, od ključnog je značaja da ne rasipa

resurse u kooperacijama u kojima druga strana preti da ne ispuni ono na šta nju obavezuje takva socijalna kooperacija. U teoriji igara, slična situacija se elegantno analizira u paradigmatičnoj igri sa strategijom „*besplatnog puta*“ (engl. *free rider*) u odnosu na pitanje o tome da li je za svakog pojedinca u društvenoj zajednici racionalno, ili nije, da plaća za javne usluge koje koristi (Binmore, 2007). Tako, ako se ispitanicima postavi zadatak selekcije koji je tematski vezan za socijalne ugovore, uključujući pravilo tipa *ako A uradi p, B će uraditi q*, a zatim se od njih traži da pravilo testiraju kroz karte ($p$) „A je uradila $p$“, ($q$) „B je uradila $q$“, ($ne$-$p$) „A nije uradila $p$“ i ($ne$-$q$) „B nije uradila $q$“, teorija socijalnog ugovora predviđa da će oni okrenuti $p$-kartu („A je uradila $p$“) i $ne$-$q$-kartu („B nije uradila $q$“), jer samo te dve karte uopšte mogu da otkriju osobe koje potencijalno varaju u kontekstu socijalne kooperacije. To je ujedno i normativan odgovor prema standardnoj teoriji falsifikacije hipoteza. Međutim, ako situacija nalaže analizu *zamenjenog socijalnog ugovora*, odn. pravila oblika „B će uraditi $q$, ako A uradi $p$“, odgovori ispitanika će se promeniti, jer sada samo $ne$-$p$-karta („A nije uradila $p$“) i $q$-karta („B je uradila $q$“) predstavljaju situacije u kojima uopšte mogu da se pojave osobe koje ne poštuju socijalni ugovor. Ovakav izbor, međutim, nije normativno adekvatan u testu *Modus Ponens* pravila. Socijalni ugovori ne moraju da prate ovakva pravila reciprociteta u socijalnim odnosima. Kosmajdesova ispitanicima daje socijalne zakone tipa „*Ako osoba jede korenje kasave, onda ima tetovažu na licu*“, i (zamenjeni ugovor) „*Ako osoba ima tetovažu na licu, onda jede kasava korenje*“. Eksperimentalne instrukcije ispitanicima objašnjavaju da u datoj kulturi samo oženjeni muškarci - koji imaju tetovažu na licu - smeju da jedu korenje kasave, poznato po afrodizijačkom dejstvu. U seriji eksperimenata sa upotrebom socijalnih ugovora, Kosmajdesova pokazuje da procenat ispitanika koji daju određeni tip odgovora konzistentno prati predikcije teorije socijalnog ugovora. Ono što je interesantno u radu iz 1989. godine je da Kosmajdesova, na osnovu pregleda velikog broja studija u kojima su korišćeni tematski materijali (sa, kako sama kaže, „kontradiktornim rezultatima“), tvrdi da nijedan tematski zadatak koji nije bio formulisan kao neki oblik socijalnog ugovora *nikad nije dao robustne i replikabilne rezultate* (Cosmides, 1989).

*Kauzalno rezonovanje.* Pre nego što damo komentar na debatu o racionalnosti u oblasti rezonovanja i suđenja predstavićemo tek ukratko najnovije rezultate iz oblasti kauzalnog rezonovanja. Proučavanje kauzalnog reznovanja je dobilo na značaju u prvim godinama XXI veka, pošto je tokom 90-ih kognitivna psihologija otkrila veliku eksplanatornu moć modela kauzalnih mreža - predstavili smo ih u sekciji 7.2

o kauzalnom učenju - u objašnjenju najrazličitijih perceptivnih i kognitivnih procesa. Normativni okvir za kauzalno rezonovanje je tako dat teorijom kauzalnih modela, a studije kauzalnog rezonovanja predstavljaju eksperimentalne pokušaje da se ispita u kojoj meri ljudski kognitivni sistem poštuje složena pravila kauzalnih modela u svom rezonovanju o kauzalnim odnosima.

U ispitivanju kauzalnog rezonovanja, ključna je distinkcija između *opservacija* i *intervencija* u kauzalnim mrežama (Pearl, 2000). Pogledajmo ponovo model (f) kauzalnog lanca u trećem redu na Slici 12. Podsetimo se i osnovnog normativnog uslova, odn. kauzalnog Markovljevog uslova, koji tvrdi da su verovatnoće javljanja određenih varijabli u kauzalnoj mreži nezavisne od njihovih ne-potomaka *uslovno u odnosu na verovatnoće svojih roditelja.* Ovaj normativni uslov diktira sve promene verovatnoća javljanja određenih varijabli u nekom kauzalnom modelu pod intervencijama: intervencija je, podsetimo se, situacija u kojoj spoljnim dejstvom postavljamo vrednost neke promenljive u mreži, čime ona biva „odsečena“ od dejstva svojih roditelja i predaka. U modelu kauzalnog lanca varijabla X utiče na Y i varijabla Y na varijablu Z: $X \rightarrow Y \rightarrow Z$. Pretpostavimo da se varijabla X (jedina egzogena u ovom modelu) javlja sa nekom verovatnoćom P(X); sledi da se Y javlja sa verovatnoćom P(Y|X), i da se Z javlja sa verovatnoćom P(Z|Y,X). Kompletna zajednička distribucija verovatnoće za celu mrežu kauzalnog lanca je onda data sa

$$P(X, Y, Z) = P(Z|Y, X) \cdot P(Y|X) \cdot P(X) \tag{45}$$

Izraz (45) bi trebalo da je i intuitivno jasan: verovatnoća Z zavisi od verovatnoća njenih predaka i roditelja, X i Y, verovatnoća Y od verovatnoće njenog roditelja X, a X ima samo svoju osnovnu (engl. *base-rate*) verovatnoću javljanja jer se ne nalazi pod kauzalnim dejstvom drugih promenljivih. Zamislimo sada da spoljnom intervencijom fiksiramo vrednost Y tako da je ona uvek pristuna; obeležimo to sa Y=1. Ova operacija se u jeziku teorije kauzalnih mreža naziva *do*-operatorom (od engl. „*to do*“ - činiti). Dakle, posle *do*(Y=1), u strukturi kauzalnog lanca dolazi do sledeće promene: varijabla Y više nije povezana sa svojim uzrokom X, jer je ona sada isključivo pod dejstvom nekog spoljnog faktora; tako verovatnoću da se varijabla Z pojavi, koja je bila određena verovatnoćama X i Y, sada moramo da svedemo na njenu uslovnu verovatnoću samo u odnosu na Y: P(X|Y), jer Z više nema načina da kauzalno deluje na nju (preko Y). Pod intervencijom *do*(Y=1), izraz (45) mora da se promeni u

$$P(X, do(Y = 1), Z) = P(Z|Y = 1) \cdot P(X) \tag{46}$$

Zamislimo sada drugačiju situaciju, u kojoj nema nikakve intervencije, već smo *opservacijom* ustanovili da je vrednost varijable Y=1. Šta nam struktura kauzalnog modela govori, pored te opservacije? Pošto je Y=1, odn. stalno prisutna, vrednost Z više ne može da zavisi od varijable X, baš kao u slučaju intervencije $do$(Y=1). Ali, pošto nismo intervenisali, već samo opazili da je Y=1, ne možemo da zaključimo da je kauzalni lanac između X i Y prekinut, već da ostavimo otvorenu mogućnost da je upravo kauzalno dejstvo X prouzrokovalo da Y=1. Tako, izraz (45) posle opservacije da je Y=1 uzima oblik

$$P(X, Y = 1, Z) = P(Z|Y = 1) \cdot P(Y = 1|X) \cdot P(X) \tag{47}$$

Očigledno, (46) i (47) nemaju istu formu. Eksplanatorna moć modela kauzalnih mreža upravo leži u mogućnosti opisa ovakvih interakcija između strukture kauzalnih modela, intervencija i opservacija. U eksperimentima kauzalnog rezonovanja, ispitanici prvo uče određene kauzalne modele, i pošto ih nauče, dobijaju informacije o određenim verovatnoćama koje opisuju prisustvo i odsustvo varijabli u modelu. Zadaci koji im se daju zahtevaju od njih da donesu sud o tome da li će neka varijabla biti prisutna ili ne posle određene intervencije ili opservacije, ili sa kojom verovatnoćom će se neka varijabla javiti posle određene intervencije ili opservacije u modelu. Pošto smo videli da se distribucije verovatnoća varijabli različito menjaju posle kauzalnih intervencija i čistih opservacija, distinkcija koju uvodi eksperimentalna manipulacija jednog ili drugog tipa je krucijalna. U većem broju eksperimenata, pokazano je da ljudi normativno adekvatno rezonuju o stanjima varijabli u kauzalnim modelima, i to selektivno u odnosu na intervencije i opservacije (Waldmann & Hagmayer, 2005, Sloman & Lagnado, 2005, Hagmayer, Sloman, Lagnado & Waldmann, 2007).

Naravno da postoje i eksperimentalni nalazi koji pokazuju da ljudsko kauzalnog rezonovanje ne poštuje normativne principe teorije kauzalnih modela. U eksperimentalnoj proveri rezonovanja prema kauzalnom Markovljevom uslovu Reder koristi više različitih materijala koji svi uzimaju zajedničke forme kauzalnih viljuški (odn. mreža zajedničkog uzroka) ili kauzalnih lanaca. Svaka varijabla u ovakvim modelima ponovo može da uzme binarne vrednosti, npr. da bude u stanju „visoka“

ili „niska“, „velika“ ili „mala“, „pristuna“ ili „odsutna“. Ispitanici prvo uče o kauzalnim zakonima koji važe između datih varijabli, a zatim im se prikazuje kauzalni model sa nekompletnim informacijama: npr, u modelu zajedničkog uzroka, dat je opis kauzalne mreže za koju znamo da je prisutan uzrok *C*, i jedna od posledica *E1*, ali ne znamo ništa o vrednosti druge posledice *E2*. U drugom modelu, dat je npr. opis iste takve mreže za koju znamo da je prisutan samo uzrok *C*. Pitanje za ispitanike je da odrede u kom od ta dva modela je verovatnije da će se pojaviti i druga posledica *E2*, čija vrednost nije definisana ni za jedan od dva modela. Pogledajmo ponovo mrežu zajedničkog uzroka (h) na Slici 12. Kauzalni Markovljev uslov diktira da posle fiksiranja vrednosti uzroka *C*, koja je fiksirana u oba modela pred ispitanicima, dve posledice zajedničkog uzroka *E1* i *E2* postaju statistički nezavisne: pojava jedne više ne govori ništa o pojavi druge, već su verovatnoće i jedne i druge zavisne samo od kauzalne moći uzroka. Ne poštujući normativno adekvatno kauzalno rezonovanje, Rederovi ispitanici u najvećem broju donose zaključak da će model u kome je prisutna jedna posledica, *E1*, biti i model u kome je verovatnije da se javi i druga posledica, *E2* (Rehder, 2006).

*Komentari o racionalnosti rezonovanja i suđenja.* Istraživački programi heuristika i inklinacija, s jedne, te alternativni programi ekološke racionalnosti, evolucione racionalnosti i „racionalne racionalnosti“ pristalica bejzijanskih racionalnih teorija, s druge strane, danas se svi čvrsto drže svojih pozicija. Osnovna linija demarkacije je stav prema tome da li je moguće da su ljudske kognitivne funkcije suđenja i rezonovanja uopšte posledica optimizacionih procesa, kao što pretpostavljaju pristalice bejzijanskog programa i (u nekoj meri) evolucione racionalnosti, ili su te funkcije prepuštene heuristikama koje svakako nisu optimalna rešenja. I Gigerencer, i pristalice tradicionalne paradigme heuristika i inklinacija, složiće se da kognitivne funkcije suđenja i rezonovanja nisu proizvod optimizacionih procesa. Gigerencerova rešenja će biti u „pozitivnom duhu“: heuristici ipak imaju ekološku validnost, i rešavaju adaptivne probleme bar aproksimativno. Rezultati programa heuristika i inklinacija nas ostavljaju sa daleko pesimističnijom slikom jednog „veoma ograničeno racionalnog“ kognitivnog sistema: Kaneman i Frederik predlažu da je u pitanju adaptivni organ koji slobodno zamenjuje teške (ali realne) ciljeve adaptacije lakšim (ali pogrešnim) ciljevima. Interesantno je da pristalice programa heuristika i inklinacija u njegovoj klasičnoj formi, određenoj prevashodno istraživanjima Tverskog i Kanemana, retko kada izlaze sa odgovorom na pitanje kako je moguće da sa grubim ograničenjima naših kognitivnih sistema koja opisuju, ljudi ipak žive

prilagođeno svojoj sredini? Naravno da postoje odgovori na ovakva i slična pitanja, ali često, proučavajući radove u ovoj tradiciji istraživanja, neko može da stekne utisak da kognitivni sistem u rezonovanju i suđenju praktično nije ništa drugo do skup loše projektovanih rešenja koji obiluje greškama i inklinacijama.

Oblast suđenja i rezonovanja pokazala se tvrdim orahom za racionalnu analizu. S jedne strane, noviji eksperimentalni rezultati koje predstavljaju Grifits i Tenenbaum ukazuju na to da problema, možda, uopšte i nema. Ipak, pored eksperimentalnih nalaza koji govore suprotno, nagomilanih decenijama rada, treba sačekati dopunske testove njihove metodološke paradigme suđenja. S druge strane, njihov pokušaj racionalne analize reprezentativnosti, po našem sudu, predstavlja markatan neuspeh ove metodologije. Izneli smo osnovni razlog za to - i to je, u suštini, razlog koji čini celu diskusiju toliko komplikovanom: kognitivni sistem mora da ima neku „supermoć“ kojom bira koja će hipoteze generisati kako bi ih testirao kao moguće generativne modele podataka pred njim, ako racionalno donosi sud o nekom neizvesnom događaju. Nemoguće je pretpostaviti da kognitivni sistem u svakom relevantnom kontekstu za ovaj tip analize pretražuje ceo skup hipoteza, jer je on beskonačan. Skup hipoteza koji se pretražuje, dakle, mora da bude suštinski ograničen *a priori*. Koja bi to ograničenja bila, i kakvo bi bilo njihovo poreklo - induktivno, ili nasleđeno iz evolucione prošlosti, ostaje otvoreno pitanje. Interesantno je da je Tenenbaumova i Grifitsova racionalna analiza reprezentativnosti zapravo jedna od formalno najjednostavnijih analiza ove vrste koja je ponuđena. Podsetimo se, oni od kognitivnog sistema zahtevaju samo da izračuna test logaritamske verodostojnosti, što verovatno i nije previše zahtevno za određenu arhitekturu na algoritamskom nivou opisa. Naravno - pod uslovom da su relevantne hipoteze unapred generisane. Do tada, one su *deus ex machina* ovakvog pristupa, duh koji problem rešava metodološki cirkularnom odlukom u kojoj statističari pretpostavljaju da crna kutija koju testiraju generiše iste hipoteze koje generišu oni koji je testiraju.

Konačno, pristup Oaksforda i Čatera pokazuje izuzetan uspeh u objašnjenju zadatka selekcije. Primer toga kako je moguće skup generativnih hipoteza ograničiti na relevantan način jeste njihovo oslanjanje na hipotezu o retkosti[61]. Ova hipoteza je, zanimljivo, opravdana i normativno (setimo se diskusije Hempelovog paradoksa), a deluje da može da dobije i opravdanje u ekološki relevantnom setingu za analizu ljudskog saznanja. Ipak, ni ovakva hipoteza nije bez problema. U bilo kom okruženju, perceptivni sistem čoveka neprekidno prima informacije o ogromnom

broju pozadinskih stimulusa koji su repetitivne prirode: u gradovima, mi smo neprestano okruženi ili zidovima ili ulicama, zgradama i kućama, automobilima, i mnogim drugim singularnim događajima na koje se hipoteza o retkosti očigledno ne odnosi. Da li bi ljudsko rezonovanje u Vosonovom zadatku, shvaćeno kao optimalna selekcija podataka, pokazalo iste bihejvioralne profile u slučaju stimulusa koji nisu retki, već veoma verovatni u svakodnevnom okruženju? Možda je naša kritika na ovom mestu preoštra: postoji mogućnost da se ovaj problem savlada pozivanjem na uobičajeni pristup modeliranju slične situacije u učenju, definisanjem neprestano pristunih, repetitivnih ili kontinuiranih stimulusa kao „pozadinskih" - oni su, npr, uvek uključeni u kompjutacione analize modela poput *Reskorla-Vagner*. Ni analiza Oaksforda i Čatera, dakle, ne može u potpunosti da odgovori na jedno od ključnih pitanja koje se postavlja pred svaki pokušaj racionalne analize: kojom metodom kognitivni sistem određuje šta je u datom okruženju *relevantno*, tj. šta predstavlja elementarne skupove događaja na koje se primenjuju optimalne probabilističke analize?

Činjenica preko koje je nemoguće preći jesu rezultati programa heuristika i inklinacija koji se odnose na interakcije između semantičkih faktora (sličnost i tipičnost) sa kompjutacionim ciljevima koji se odnose na normativne atribute poput verovatnoća. Dominacija asocijativnih sistema u ovoj oblasti jasno je i nedvosmisleno demonstrirana: *Sistem 2*, kako ga opisuje Daniel Kaneman, odnosi prevagu nad *Sistemom 1* u ogromnom broju slučajeva. Ne prvi put u analizi debate o racionalnost, naše je mišljenje da se sistematskoj analizi proceduralne invarijantnosti poklanja nedovoljna pažnja. Kako je moguće dobiti *toliko* diskrepatne nalaze u suđenju o verovatnoćama kakve dobijaju Tverski i Kaneman s jedne, i Tenenbaum i Grifits (optimalne procene verovatnoća), s druge strane? Na osnovu tih diskrepantnih, skoro simetričnih bihejvioralnih potpisa koje rad kognitivnog sistema ostavlja, nameće se zaključak da u dve tek donekle različite eksperimentalne paradigme on ne rešava isti problem.

Kako, onda, formulisati metodologiju koja bi nam pomogla da efektivno odlučujemo o tome *koji je kompjutacioni cilj* kognitivnog sistema u funkciji procedure kroz koju prikupljamo bihejvioralne podatke o njemu i relevantnih informacija kojima on može da raspolaže? To pitanje, nažalost, još uvek je bez odgovora. Mi ćemo mu posvetiti punu pažnju u trenutku kada budemo zatvarali ovaj kritički pregled debate o racionalnosti.

## 7.5 Koncepti I: funkcija kategorizacije

Debata o racionalnosti kognitivnih funkcija ljudskog konceptualnog sistema nije razvijena kao u prethodnim domenima naše diskusije. Postoji više razloga za to, od kojih se kao najznačajniji nameće ogromna složenost kognitivnih funkcija učenja koncepata, kategorizacije i primene koncepata u indukciji, analogiji, konceptualnim kombinacijama i metaforama. Drugi razlog, takođe povezan sa kompleksnošću ove oblasti, predstavlja činjenica da se kroz proučavanje konceptualnog sistema prožima proučavanje drugih viših kognitivnih funkcija: učenja i pamćenja, pre svega. Nemoguće je empirijski pristupiti ljudskom kognitivnom sistemu drugačije nego kroz „interfejs“ metodologija proučavanja učenja, pamćenja i rezonovanja. Ljudski koncepti se u svim ovim interakcijama sa drugim višim kognitivnim funkcijama ponašaju specifično. Učenje koncepata, iako poštuje neke regularnosti kao i svi ostali tipovi učenja, puno je specifičnosti. Tako je, na primer, demonstrirano da primenom identičnog eksperimentalnog nacrta onom koji u drugim oblicima učenja demonstrira fenomen blokiranja, u učenju koncepata do blokiranja ne dolazi (Bott, Hoffman & Murphy, 2007). Kategorijalna indukcija, proces zaključivanja o prisustvu određene karakteristike među pripadnicima jedne kategorije na osnovu informacije o tome da ona odlikuje jednog (ili manji broj) drugih pripadnika iste te kategorije, opet predstavlja sasvim specifičan vid induktivnog rezonovanja. Fenomene vezane za kategorijalnu indukciju nije moguće objasniti tek poznavanjem odgovarajuće statističke informacije o distribuciji neke karakteristike $f$ u određenom uzorku posmatranja članova neke kategorije. Pamćenje, opet suštinski povezano sa konceptima, posvećuje im poseban memorijski sistem - semantičku memoriju - koja se ni metodološki ne tretira na isti način kao epizodička memorija o kojoj smo prethodno diskutovali[62]. Više simboličke funkcije poput analoškog rezonovanja i konceptualnih kombinacija uopšte nije ni moguće diskutovati van konteksta kategorijalno organizovanog sistema. Ljudski koncepti, ako možda i nisu „*u stanju nereda*“ - kako ih karakteriše jedan od vodećih istraživača u ovoj oblasti (Murphy, 2002) - sasvim sigurno jesu svet za sebe. Činjenica da je za ljudski konceptualni sistem teško, ili nemoguće, pronaći odgovarajuću analogiju u drugim psihološkim funkcijama, ili prirodnim, biološkim procesima čija optimizacija inspiriše pokušaje racionalne analize, upućuje na ograničenja racionalne analize u ovoj oblasti.

Pokušaj dosledne racionalne analize funkcije kategorizacije - koja je samo jedna od funkcija koju igraju koncepti u ljudskoj semantičkoj memoriji - sproveo je Džon Anderson 1991 (Anderson, 1991c). Ovoj analizi se nećemo posvetiti detaljno kao u

slučaju Andersonove racionalne analize pamćenja. Ona danas svakako ne predstavlja okosnicu pokušaja da se funkcije konceptualnog sistema objasne adaptacionističkim principima u odnosu na ciljeve kognitivnog sistema i strukturu informacija u okolini u kojoj te ciljeve treba zadovoljiti. Ovu racionalnu analizu kategorizacije temeljno je kritikovao Gregori Marfi (Murphy, 1993), ukazujući na to u kojoj meri kompleksnost kognitivnog sistema i raznolike funkcije koje on obavlja prevazilaze jednostavnost pretpostavki od kojih je Anderson pošao. Više pažnje ćemo posvetiti pokušaju da identifikujemo sve principe konceptualne organizacije koji uopšte sadrže neki normativni predlog - neku logiku prema kojoj bi struktura i funkcije konceptualnog sistema, u svoj svojoj složenosti, *trebalo* da uzimaju određenu formu. Takva diskusija teorijskih principa predloženih u ovoj oblasti do sada nije sprovedena.

Skoro je nemoguće povesti teorijsku diskusiju o konceptualnom sistemu bez uvođenja nekih standardnih terminoloških distinkcija. Dakle, termine *koncept* i *kategorija* koristimo sinonimno, iako LISICA, koja je koncept, jeste i *entitet* odn. *član* ili *pripadnik* u kategoriji ŽIVOTINJA. Neka posebna, jedinstvena mačka - zovimo je *Brunhilda* za potrebe primera - nije koncept (i ako je deo naše semantičke, ne epizodičke memorije, ako se ispostavi da godinama živimo sa njom i poznajemo njeno ponašanje i narav do u detalje; Marfi ukazuje na to da ljudi, u tom smislu reči, imaju koncepte i za pojedine entitete koji ne formiraju nikakve šire kategorije, Murphy, 2002). U iskustvu smo imali ko zna koliko „epizodičkih mačaka" - od kojih naše epizodičko pamćenje pamti samo neke - ali nijedna od njih posebno, čak i ako je se sećamo do u detalje, ne čini koncept MAČKE. Tako, entitetima nazivamo ona bića koja su članovi (pripadnici) kategorija; pojedine *entitete* označavamo *italic* tipografijom, a ŠTAMPANIM SLOVIMA - KATEGORIJE, kao u: *Brunhilda* jeste MAČKA. Slično, *Lili* jeste LISICA, *Brunhilda* jeste ŽIVOTINJA, ali i LISICA jeste ŽIVOTINJA.

U narednim redovima jedna pretpostavka će implicitno pratiti sve naše diskusije. Sve teorije učenja koncepata, kategorizacije i primene konceptualnog znanja razvijene u kognitivnističkoj tradiciji od početka 70-ih godina do danas baziraju reprezentacije koncepata na reprezentacijama njihovih *karakteristika*. Kognitivni procesi kategorizacije - u najrazličitijim verzijama u kojima su ponuđeni na tržištu ideja - u suštini uvek određuju da li je neki entitet pripadnik neke kategorije tako što proveravaju u kojoj meri odlike koje učestvuju u deskripciji tog entiteta karakterišu i odlike koje učestvuju u deskripciji kategorije. Implicitna pretpostavka svih ovakvih teorija - a to će reći, *svih* kognitivnističkih teorija kategorizacije - jeste da su te

odlike po kojima se poređenja vrše *date*. Kognitivni sistem suočen sa entitetom iz kategorije poput DOMAĆE MAČKE nekako „zna“ da su odlike mačke te da ona ima uši, šape, rep, da je brza, vešta, da ima krzno, da voli da se mazi, da može da ogrebe, da glasno mjauče kada je gladna i tako dalje. Ove odlike su, po pretpostavci, stabilne, date, i uskladištene u ljudskoj semantičkoj memoriji. Slično, suočen sa sasvim novim, nepoznatim entitetima koje eksperimentatori u laboratorijama opisuju kroz skup određenih dimenzija, kognitivni sistem zna koje su to dimenzije, izdvaja ih iz opisa entiteta pred njima i barata njihovim vrednostima pokušavajući da jedne entitete u zadatku učenja kategorija svrsta u jednu, a druge u drugu kategorije, na osnovu njihovih sličnosti i razlika po relevantnim dimenzijama. Dakle, implicitna pretpostavka cele oblasti istraživanja koja je pred nama je ta da su karakteristike entiteta i kategorija koji se nalaze pod lupom analize fiksne, jasno definisane i date. U narednoj sekciji (7.6) videćemo kakve probleme može da donese preispitivanje ove pretpostavke.

*Normativni karakter principa konceptualne organizacije.* Već u ranoj fazi proučavanja koncepata u kognitivnoj psihologiji, Eleanor Roš je predložila *principe kognitivne ekonomije* za koji je verovala da odlikuju organizaciju koncepata (Rosch, 1978). Prvi princip koji predlaže Eleanor Roš (koji nazivamo principom kognitivne ekonomije u užem smislu) tvrdi da konceptualni sistem teži tome da formira maksimalno informativne kategorije uz minimalnan utrošak kognitivnih resursa. Veza sa strategijom racionalne analize više je nego očigledna u slučaju ovog principa koji predlaže Rošova. Kognitivni sistem koji koristi sve diskriminacije među stimulusima u svojoj okolini vrši perfektnu diskriminaciju, ali tamo gde takva diskriminacija nije neophodna, on troši resurse bez potrebe. Tako, moguće je zamisliti da je organizacija konceptualnog sistema posledica strategije zanemarivanja nebitnih razlika među pojedinim entitetima i izbegavanja mogućnosti da partikularije u potpunosti preplave raspoložive kognitivne resurse. Ovakav princip može da objasni ljudsku konceptualnu organizaciju kao vid optimalnog odgovora na adaptivne pritiske sredine, ali sam po sebi ne može da objasni sve bitne osobine ljudskog kognitivnog sistema. Kategorijalna organizacija nije samo tu da nas podseti da smo bića ograničenih kognitivnih kapaciteta. Bez formiranja pojmova i mogućnosti generalizacije, nikada ne bismo mogli da formiramo matematičke koncepte poput koncepta prave (koja je beskonačna) ili tačke (koja ima nula dimenzija). Drugi princip koji predlaže Eleanor Roš je *princip opažene strukture sveta*, koji tvrdi da karakteristike određenih entiteta koji nas okružuju

nisu uniformno i nezavisno distribuirane u sredini. Kako sama Roš to ilustruje, organizam koji raspolaže motoričkim planom za akciju sedanja, u svojoj okolini će primetiti da objekti koji odgovaraju toj funkciji *dele zajednički skup karakteristika* (određene su visine, stabilni, nisu previše meki da bi se deformisali pod težinom itd). Neke karakteristike javljaju se češće među pripadnicima određenih kategorija, a neke ređe, i pri tom u objektivnoj strukturi naše sredine mogu da se prepoznaju korelacije u pojavi različitih karakteristika: Rošova tvrdi da je suštinsko za razumevanje konceptualnog sistema uvideti da on u svojim internim reprezentacijama mapira ovakvu strukturu sveta. Oba principa koje predlaže Rošova su, bez ikakve sumnje, tesno povezana sa logikom racionalne analize (ovo primećuje još Korter, Corter, 1991, u komentaru na Andersonovo izlaganje programa racionalne analize, Anderson, 1991). Tu je pretpostavka o strukturi informacija u okolini (princip opažene strukture), kao i pretpostavka o optimizaciji u odnosu na ograničene kognitivne resurse (princip kognitivne ekonomije). Ovakvi principi se u savremenim diskusijama o konceptualnom sistemu nazivaju *principima konceptualne koherencije* (Murphy & Medin, 1985). Konceptualni sistem, tvrdi Rošova, koji funkcioniše u skladu sa ova dva principa težiće da formira koncepte tj. kategorije koji predstavljaju *prototipove* - sumarne reprezentacije informacija o članovima kategorije - koji maksimimizuju distinktivnost neke kategorije u odnosu na druge kategorije, sadržeći informacije o onim karakteristikama koje su visoko reprezentativne za pripadnike te kategorije, a nisko ili nimalo reprezentativne za njene ne-pripadnike.

*Teorije prototipova i primeraka.* *Teorija prototipova*, predložena u radovima Rošove i saradnika (Rosch 1975, 1978, Rosch & Mervis, 1975), ubrzo je od strane Džemsa Hemptona elaborirana u jednu od standardnih teorija koncepata (Hampton, 1979, 1997, 2001, 2006). Prototipovi predstavljaju sumarne reprezentacije pripadnika određene kategorije koje su ostvarene kroz reprezentaciju *karakterističnih* odlika te kategorije. Same odlike mogu biti ponderisane tako da prototip nosi i informaciju o značaju pojedinih odlika za prisustvo u nekoj kategoriji (Hampton, 1979). Dakle, naš koncept MAČKE je, prema ovom shvatanju, naš prototip mačaka: to je interna reprezentacija u semantičkoj memoriji koja se sastoji od skupa ponderisanih karakteristika koje odlikuju razne mačke. Teorijske analize po pravilu izostavljaju da činjenicu da teorija prototipova ima *normativno poreklo* u principima kategorizacije koje je predložila još Eleanor Roš. Teorija prototipova može da se ostvari u matematički različitim, fleksibilnim modelima (Nosofsky, 1992, Smith & Minda, 2000). Krajem sedamdesetih godina, uticajnim radom Medina i Šefera

uveden je model učenja koncepata i kategorizacije koji će označiti početak dugačke dominacije tzv. *teorija primeraka* (engl. *exemplar theories*, Medin & Schaffer, 1978). Teorije primeraka predstavljaju suštu suprotnost teoriji prototipova. Mi verujemo da je upravo fokus na teorije primeraka obeshrabrio pokušaje racionalne i normativne analize - čiji su koreni prisutni u logici razvoja teorije prototipova - u proučavanju koncepata i kategorizacije. Zato je važno da na ovom mestu ukažemo na poreklo teorije primeraka i motivaciju da se ona uvede. Medin i Šefer su proučavali učenje koncepata standardnom procedurom: definisali su dve kategorije kroz četiti binarne karakteristike i od ispitanika tražili da kroz ponovljene pokušaje učenja sa fidbekom nauče da klasifikuju devet stimulusa.

Tabela 4. Struktura 5-4 problema u učenju kategorija

kategorija A

| D1 | D2 | D3 | D4 |
|---|---|---|---|
| 1 | 1 | 1 | 0 |
| 1 | 0 | 1 | 0 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 |
| 0 | 1 | 1 | 1 |

kategorija B

| D1 | D2 | D3 | D4 |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 |

Raspodela karakteristika kroz devet stimulusa korišćenih u fazi treninga u eksperimentu 2. Medina i Šefera iz 1978. godine data je u Tabeli 4. Svaki red u Tabeli 3. predstavlja opis po jednog pripadnika kroz četiri binarne karakteristike (odn. „dimenzije“ - D1-4 u Tabeli 4); kategorija A ima pet pripadnika, kategorija B četiri. U učenju koncepata i kategorizaciji koriste se najraznovrsnije vrste stimulusa, od opisa pojmova preko skupa verbalnih karakteristika do vizuelnih stimulusa. Studija Medina i Šefera koristila je vizuelne stimuluse sa jasnim, distinktivnim odlikama čiju strukturu opisuje Tabela 4. Problem učenja i kategorizacije kategorija A i B u literaturi je dobio ime „5-4 struktura“ ili „5-4 problem“ zbog učestale upotrebe upravo ove strukture kategorija u eksperimentalnim studijama (Smith & Minda, 2000, daju pregled rezultata dobijenih primenom ove strukture u 30 studija objavljenih u osam naučnih radova).

Kategorijalna struktura 5-4 ima sledeće osobine: kategorije A i B su razvijene postepenim narušavanjem dva idealna prototipa, 1-1-1-1 za kategoriju A i 0-0-0-0 za kategoriju B. Ako posmatramo pripadnike A, videćemo da među njima dominiraju vrednosti 1 na četiri dimenzije opisa, dok među pripadnicima kategorije B

dominiraju vrednosti 0. To da pripadnici kategorija ne odgovaraju savršeno prototipovima je ekološki opravdana pretpostavka: nijedna realna mačka neće imati takvu distribuciju odlika koncepta MAČKE koji omogućava generalnu reprezentaciju svih članova kategorije. U kategoriji A, četiri pripadnika imaju tri vrednosti 1 na četiri dimenzije; jedan pripadnik, međutim (u drugom redu Tabele 4) ima podjednak broj vrednosti 1 karakterističnih za A i vrednosti 0 - karakterističnih za drugu kategoriju. Kategorijom B, s druge strane, dominiraju vrednosti 0, ali dva njena pripadnika (prvi i drugi red u Tabeli 4) ipak imaju dve vrednosti 1 i dve vrednosti 0, pa bi mogli biti i pripadnici kategorije A (kao što bi pripadnik A u drugom redu Tabele 4 po istom principu mogao biti pripadnik B). U trećem redu tabele je pripadnik kategorije B sa većinom karakterističnih vrednosti za kategoriju, dok je četvrti pripadnik prototip kategorije (0000). Ovakva kategorijalna struktura, dakle, ne definiše granice između kategorija jasno: tri od devet ovako definisanih koncepata su dvosmisleni u tom smislu reči što prema pravilu „Imati najmanje tri od četiri vrednosti 1 da bi se bio član A, i tri od četiri 0 vrednosti da bi se bio član B“ ne mogu jasno da se svrstaju u neku od kategorija. Struktura 5-4 problema ilustruje *nelinearno separabilne kategorije*: ako se njihovi pripadnici predstave u N-dimenzionalnom prostoru vrednosti svih N karakteristika, *ne postoji linearna diskriminantna funkcija* (prava u slučaju dve dimenzije, površ u tri ili hiperpovrši u višedimenzionalnim prostorima) koja bi bez izuzetaka podelila takav prostor na deo koji okupiraju isključivo pripadnici A i deo gde su isključivo pripadnici B. Podsetimo se da model teorije prototipova predstavlja ponderisanu reprezentaciju karakteristika koje odlikuju neki koncept. Dakle, taj model linearno kombinuje karakteristike neke kategorije kao $\alpha_1 \cdot f_1 + \alpha_2 \cdot f_2 + ..., \alpha_n \cdot f_n$ da bi izračunao u kojoj meri se koncept sa odlikama $f_1$, $f_2$, .., $f_n$ preklapa sa prototipom kategorije koji je određen ponderima $\alpha_1, \alpha_2$, .., $\alpha_n$. Očigledno, standardni model teorije prototipova ne može da izađe na kraj sa doslednim učenjem kategorijalne strukture poput 5-4 koja nije linearno separabilna. Ipak, kategorijalna struktura 5-4 pažljvo je dizajnirana da zainteresovanima za analizu ljudske kategorizacije zagorča život: ako neko jednostavno zanemari informacije koje nosi dimenzija D2, kategorije postaju linearno separabilne: sve što je potrebno jeste primeniti pravilo „Imati najmanje dve od tri vrednosti 1 da bi se bio član A, i dve od tri vrednosti 0 da bi se bio član B“ na D1, D3 i D4. Učenje ovakve kategorijalne strukture se sprovodi po blokovima: u svakom bloku se prikazuje svih devet koncepata i od ispitanika traži da odredi koji od njih se nalazi u kojoj kategoriji. Posle svakog odgovora ispitanika, daje mu

se fidbek o ispravnosti njegove kategorizacije (npr. „Da, to je član kategorije A“, ili „Pogrešili ste, to je član kategorije B“). Pošto se postigne određeni kriterijumski nivo učenja (na primer dva uzastopna bloka uspešne kategorizacije bez grešaka), sedam novih koncepata - definisanih na istim dimenzijama - se koristi u zadataku transfera da bi se ispitalo koliko su naučene strukture kategorija generalizovane i stabilne. U prvom eksperimentu sa 5-4 strukturom, samo 56% ispitanika uspelo je da nauči da razlikuje kategorije A i B posle 16 blokova učenja (Medin & Schaffer, 1978)!

Problemi učenja ovako neobičnih kategorija, i činjenica da one ni u principu ne mogu da budu naučene primenom teorije prototipova, motivisala je razvoj teorija primeraka. Prema teoriji primeraka, ne postoji jedna jedinstvena reprezentacija bilo kog koncepta. Koncepte u semantičkoj memoriji čine uzorci posebnih, jedinstvenih primeraka odgovarajućih kategorija koji su zapamćeni. Proces kojim kognitivni sistem određuje da li neki novi entitet pripada kategoriji A ili nekoj drugoj kategoriji sastoji se od tri koraka. U prvom, sistem izračunava sličnost između tog entiteta i svih zapamćenih pripadnika kategorije A. Sistem to čini poredeći stepene prisustva (izraženosti) relevantnih karakteristika između novog entiteta i uskladištenih primeraka, oslanjajući se na multiplikativnu funkciju sličnosti (uskoro ćemo pojasniti). U drugom koraku, kognitivni sistem vrši istovetna poređenja novog entiteta sa drugim kategorijama koje se nalaze u kontekstu i kojima bi on potencijalno mogao da pripada, dok u trećem koraku donosi odluku o tome u koju kategoriju svrstava novi entitet. Ova teorija eksplicitno je postavljena u poznatom radu Medina i Šefera iz 1978, pod imenom *kontekst-model kategorizacije*, ali je njena kanonička forma određena *generalizovanim kontekst-modelom* Roberta Nosofskog (GCM, Nosofsky, 1984, 1986, 1988, 1991, Nosofsky & Zaki, 2002). Ovaj model predstavlja centralnu matematičku konstrukciju u teorijama učenja koncepata i kategorizacije pa ćemo ga predstaviti formalno. Svi mogući entiteti predstavljeni su vrednostima karakteristika: one mogu biti binarne, uzimati više diskretnih vrednosti ili poticati sa kontinuiranih skala merenja. U višedimenzionalnoj reprezentaciji nekog novog entiteta $x_i$ i već uskladištenog primerka neke kategorije $x_j$, *distanca* između ta dva se izračunava u ponderisanoj euklidskoj geometriji na sledeći način:

$$d_{ij} = \sqrt{\sum_{m=1}^{N} w_m |x_{im} - x_{jm}|} \tag{48}$$

gde je $d_{\mathrm{ij}}$ distanca između $x_{\mathrm{i}}$ i $x_j$, suma prolazi $m = 1, 2, .., N$ dimenzija

(karakteristika) po kojima se poređenje vrši, a $w_m$ je ponder koji određuje značaj dimenzije $m$ za kategoriju u kojoj se nalazi uskladišteni primerak $x_j$. Jednačina (48) predstavlja običnu euklidsku distancu između dve tačke definisane koordinatama na $m$ dimenzija u N-dimenzionalnom prostoru u kojoj je svaka dimenzija ponderisana faktorom $w_m$ koji govori o njenom značaju tj. doprinosu u opisu određene kategorije. Poznato je da se distance između stimulusa opisanih na nekim objektivnim skalama psihološki preslikavaju u sličnosti *nelinearno*, prateći (aproksimativno) Šepardov univerzalni zakon generalizacije (Shepard, 1987). Sledeća jednačina uključuje to skaliranje objektivnih distanci izračunatih u (48) u psihološke distance:

$$\eta_{ij} = e^{-c \cdot d_{ij}} \tag{49}$$

Prema Šepardovom zakonu, psihološke distance - sličnosti, označene sa $\eta_{ij}$, jesu eksponencijalno opadajuća funkcija objektivnih, fizičkih distanci $d_{\text{ij}}$: parametar $c$ određuje precizan oblik ovog preslikavanja. U teoriji primeraka, primena Šepardovog zakona je fundamentalna. Za razliku od modela prototipova koji linearno kombinuju karakteristike kao $\alpha_1 \cdot f_n + \alpha_2 \cdot f_2 + ..., \alpha_n \cdot f_n$ , modeli primeraka karakteristike kombinuju multiplikativno, tako da odstupanje novog entiteta po samo jednoj dimenziji opisa uskladištenih primeraka neke kategorije u pamćenju može da ima dramatičan efekat na ukupnu sličnost između tog primerka i novog entiteta. Ako se sličnost izračuna kao linearna funkcija karakteristika, što predstavlja jednačina (48), onda jednačina (49) eksponencijalnim skaliranjem ima isti efekat kao da je sličnost izračunata multiplikativnom funkcijom. Podižući nivo analize sa algoritamskog na kompjutacioni, Nosofsky u razvoju generalizovanog kontekst-modela tako doprinosi unifikaciji teorije kategorizacije i učenja kategorija sa teorijom generalizacije koju je postavio Šepard. Konačno, kognitivni sistem, posle izračunavanja sličnosti novog entiteta sa prethodno zapamćenim primercima svih kategorija u koje on potencijalno može da spada, donosi odluku o tome u koju ga kategoriju svrstava primenom *Lusovog aksioma izbora* (Luce, 1977):

$$P(A|x_i) = \frac{\sum_{j \in A} \eta_{ij}}{\sum_C \sum_{j \in C} \eta_{ij}} \tag{50}$$

Jednačina (50) izračunava $P(A|x_i)$, verovantoću da je korektna kategorija $A$ ako je dat entitet $x_i$. Brojilac jednačine (50) predstavlja sumu sličnosti novog entiteta $x_{\text{i}}$ sa svim primercima kategorije $A$, a imenilac sumu takvih suma prema svim

pripadnicima svih kategorija u kontekstu nekog eksperimenta. Konačno, kognitivni sistem odlučuje u koju kategoriju spada novi entitet prema tome koju od njih odlikuje najviša verovatnoća $P(A|x_i)$ izračunata na ovaj način.

Primena modela primeraka u učenju koncepata ima dve osnovne posledice. Prvi je taj što ovi modeli objašnjavaju kako je moguće naučiti nelinearno separabilne kategorije uopšte. Konekcionističke neuronske mreže sa algortimom povratne propagacije signala, diskutovane u II Delu naše rasprave, postižu isti cilj: bez problema uče i linearno i nelinearno separabilne kategorije (Rogers & McClelland, 2004). Druga bitna posledica primene modela primeraka je u tome što oni postižu izuzetno dobre rezultate u fitovanju eksperimentalnih nalaza u standardnim paradigmama. Postoji više razloga zašto treba biti oprezan u vezi ovih zaključaka. Prvi je taj što modeli primeraka koriste *kompletnu statističku informaciju* u određenom eksperimentu učenja koncepata ili kategorizacije. Očigledno, koriste se informacije o svim dimenzijama i svim vrednostima na njima, jer reprezentacija koncepata putem uskladištenih primeraka ne „uprosečuje“ kompletnu informaciju kako ti čine modeli prototipova, već je u stanju da je očuva u potpunosti. Nije teško zaključiti da će eksperimentalne rezultate bolje objašnjavati teorija koja koristi kompletnu statističku informaciju od teorije koja pretpostavlja određeni stepen kompresije takve informacije. Upravo to je osnovna kritika teorije primeraka: u njoj, koncepti prestaju da postoje. Ako je neki koncept reprezentovan putem određenog uzroka memorisanih entiteta, on će zaista sačuvati svu prethodno dostupnu informaciju o distribuciji odlika u toj kategoriji. Ali, po (fundamentalnoj) teorijskoj pretpostavci, nešto kao što je koncept predstavlja *apstrakciju*, *generalizaciju* dostupnih distribucija informacija u sredini - ne njihovu kopiju. Teorija primeraka tako u potpunosti gubi odnos prema osnovnom teorijskom pitanju, a to je pitanje *šta* su ljudski koncepti i kategorije - pitanje na koje mora da se odgovori pre nego što se zapita o racionalnosti njihove organizacije i primene. Teorije primeraka kao da žrtvuju princip ekonomičnosti (u užem smislu) Rošove, pretpostavljajući da kognitivni sistem teži tome da uskladišti sve informacije[63] o entitetima koje sreće kako bi ih iskoristio u kategorizaciji, da bi maksimalno zadovoljili drugi princip Rošove: da kognitivni sistem mora da mapira korelacije između atributa koje odlikuju prirodnu raspodelu entiteta u svetu koji ga okružuje. Tako možemo da modele prototipova posmatramo kao rezultat interakcije dva principa kategorizacije koje je predložila Rošova, a modele primeraka kao deskriptivne teorije u oblasti kategorizacije koje bolje objašnjavaju eksperimentalne podatke, žrtvujući pri tom

određene normativne principe (prvi princip kognitvne ekonomije po Rošovoj). Vidimo da je ipak moguće voditi diskusiju u oblasti koncepata i kategorija pozivanjem na odnos prema određenim normativnim principima.

Ipak, status modela primeraka kao deskriptivnih teorija u oblasti koncepata i kategorizacije nije ista kao npr. status teorije izgleda u oblasti odlučivanja u uslovima rizika i neizvesnosti. Dok se teorija izgleda prema osnovnim teorijskim pojmovima odnosi isto kao i normativna teorija, teorija primeraka u potpunosti napušta ideju da uopšte postoji tako nešto kao što je koncept - ideju očuvanu u teoriji prototipova - da bi obezbedila veću eksplanatornu moć. U vezi teorija primeraka objektivno može da se postavi pitanje da li su u pitanju teorije učenja koncepata, ili teorije pamćenja i memorijskih efekata na učenje, ili oba. Teorije primeraka razvijene su da bi objasnile neobične i teške kategorijalne strukture kao što je struktura 5-4 problema. Marfi sa pravom primećuje u kojoj meri struktura nelinearno separabilnih kategorija ne odgovara strukturi prirodnih kategorija kakve su proučavali Rošova i saradnici (Murphy, 2002, str. 103-104). Nelinearno separabilne kategorije nisu realistične: distribucije karakteristika u prirodnim kategorijama, kategorijama koje su jezički kodifikovane i ekološki relevantne, mahom je linearno separabilna. Razvijeni da objasne ekološki manje validne situacije, prelazeći preko fundamentalnih teorijskih odrednica, modeli primeraka su disproprocionalno dominirali istraživanjima i literaturom u oblasti učenja koncepata i kategorizacije od ranih 80-ih do najnovijih dana. Ovde imamo jasan primer situacije u kojoj je mogućnost upotrebe elegantne matematike - spregnuta sa nereprezentativnim eksperimentalnim dizajnom i teorijski implauzibilnim pretpostavkama - skrenula oblast naučnog istraživanja od centralnog značaja za psihologiju u diskusije koje nisu karakteristične za savremenu debatu o racionalnosti. Naravno, niko ne može da opovrgne argument po kome ljudi ipak nekako uče i strukture poput 5-4 i slične. Odgovor ovakvoj kritici je jasan: prvo treba objasniti šta su koncepti uopšte i kako ljudski kognitivni sistem uči i reprezentuje ekološki relevantne kategorije, a onda razmišljati o takvim problemima. Zašto učenje kategorija u 5-4 problemu uopšte mora da predstavlja *jedinstven problem učenja kategorija*? Sasvim je moguće da u učenju struktura poput 5-4 učestvuje epizodičko pamćenje onih entiteta koji se u ranim fazama učenja opažaju kao izuzeci od pravila koja se testiraju. Marfi na više mesta u skorijoj monografiji posvećenoj konceptima sugeriše ovo objašnjenje (Murphy, 2002). Posledice po teorije primeraka, ako je ovo objašnjenje tačno, jeste da su one i statističke teorije učenja, i statističke teorije pamćenja, i pomalo teorije koncepata. Eksperimentalna studija iz

1998. pokazala je da u ranim fazama učenja koncepata model prototipova zapravo ima prednost nad modelom primeraka u objašnjenju proporcije kategorizacija u kategoriju A ili B (Smith & Minda, 1998), sugerišući da modeli primeraka svoju moć ostvaruju tek u kasnijim fazama učenja, kada je moguće da ispitanici bolje pamte prikazane pripadnike kategorija A i B i tako koriste veću proprociju ukupne statističke informacije u eksperimentu. Meta-analiza 30 eksperimenata koji su koristili 5-4 strukturu pokazuje da modeli prototipova zapravo veoma malo zaostaju u eksplanatornoj moći od daleko složenijih modela primeraka (Smith & Minda, 2000) kada se njihova izračunavanja sličnosti koriguju prema multiplikativnim funkcijama (što je prvi predložio Nosofsky, 1992).

Principi kategorizacije Eleanor Roš ni u kom slučaju nisu jedini teorijski principi koji mogu da se razmatraju kao normativna osnove teorije koncepata i kategorija. Pod istim imenom principa *kognitivne ekonomije*, Kolins i Kilijan su u ranim istraživanjima organizacije ljudskog semantičkog pamćenja (Collins, & Quillian, 1969) predložili da je *hijerarhijska*, *taksonomska organizacija* njen osnovni organizacioni princip. Taksonomska organizacija koncepata je intuitivno očigledna, i sa tog stanovišta ona zadovoljava uslov normativne adekvatnosti. Svi pojmovi su uključeni u kategorije veće širine, i takva organizacija se primenjuje rekurzivno na svim nivoima hijerarhijske organizacije koncepata: *Maca* je Papagaj, Papagaj je Životinja, Životinja je Živo biće, Živo biće je Univerzalni Koncept (koncept Bića, ma čega čemu se pripisuje egzistencija). Teorija hijerarhijskih mreža Kolinsa i Kilijana pretpostavljala je da su koncepti u ljudskoj semantičkoj memoriji organizovani upravo po hijerarhijskom principu. Međutim, ubrzo su eksperimentalna istraživanja pokazala da je teorija hijerarhijskih mreža neodrživ psihološki princip. Teorija hijerarhijskih mreža Kolinsa i Kilijana pogrešno je predviđala da će brzina prepoznavanje entiteta kao pripadnika određenih kategorija biti funkcija broja hijerarhijskih nivoa kojima su oni razdvojeni (Smith, Shoben & Rips, 1974). Ljudski konceptualni sistem, pokazala su istraživanja Eleonor Roš i saradnika, pokazuje preferentnost prema određenom srednjem sloju u hijerarhijskoj organizaciji kategorija, sloju koji se naziva *osnovnim nivoom kategorizacije* (engl. *base level*, Rosch, 1978, Rosch et al, 1978). Objašnjenje za postojanje osnovnog nivoa kategorizacije sastoji se u interakciji dva principa. Prvi princip se odnosi na *distinktivnost* koju određena kategorija postiže: na najvišim nivoima hijerarhijske organizacije koncepata, kategorije su maksimalno distinktivne. Reći da je nešto Životinja pravi maksimalni kontrast u odnosu na saznanje da je nešto drugo

Biljka. Drugi princip se odnosi na *informativnost* koju određena kategorija postiže, i u tom smislu je informativnost koncepata najviša na najnižim slojevima hijerarhijske organizacije. Reći da je nešto Čivava određuje osobine tog entiteta precizno: ukoliko znamo da je *Mrvica* pripadnik kategorije Čivava, mi možemo preciznije da odredimo njene osobine i pre nego što je upoznamo, mnogo preciznije nego ako znamo da je ona Pas, Sisar ili Životinja. Na osnovnom nivou kategorizacije, koji se nalazi „usred" hijerarhijskih struktura kategorija, postiže se *optimalan odnos između distinktivnosti i informativnosti koncepata*, što je fundamentalno objašnjenje zašto je to preferentni nivo kategorizacije. Marfi i Medin na sledeći način izražavaju ovaj interaktivni princip kategorizacije: „*... organizam programiran da iskoristi prednosti korelacija između atributa će težiti tome da formira kategorije koje imaju visoku sličnost unutar kategorija i nisku sličnost između kategorija kao posledicu detekcije korelacija*" (citirano prema Murphy & Medin, 1985, naš prevod). Tako, kategorije poput Voća ili Nameštaja nisu optimalne u tom smislu, ali kategorije poput Jabuka ili Stolica jesu. Ukoliko pretpostavimo da kognitivni sistem u organizaciji koncepata pokušava da maksimizuje njihovu distinktivnosti i informativnost istovremeno, uočavamo da se moguća ravnoteža između ta dva zahteva nužno nalazi negde između najopštijih i najspecifičnijih nivoa u hijerarhijskoj organizaciji. To objašnjava zašto je pristup semantičkoj memoriji olakšan za koncepte sa ovog nivoa kategorizacije - i mnoge druge empirijske nalaze koji ukazuju na prednost kategorija sa osnovnog nivoa (Murphy & Smith, 1982, Murphy & Brownell, 1985, Morris & Murphy, 1990). Sve ovo ne znači da konceptualni sistem nije organizovan hijerarhijski, već da to verovatno nije njegova organizacija u optimalnom stanju u kome se on nalazi kada reprezentuje prirodne kategorije koje ga okružuju. Činjenica je da ljudi lako i intuitivno prepoznaju da su koncepti organizovani hijerarhijski, ali kao i u slučaju teorije primeraka u učenju i kategorizaciji koncepata, moguće je da drugi psihološki procesi - procesi deduktivnog rezonovanja, u ovom slučaju - doprinose mogućnosti da se koncepti organizuju i hijerarhijski. Najnovije studije posvećene su pitanju da li kognitivni sistem koristi strategiju izvlačenja informacija o naučenim kategorijama po hijerarhijskom principu neposredno posle učenja. U seriji eksperimenata, od kojih su eksperimentalnim procedurama u nekim od njih eksplicitno forsirane strategije učenja hijerarhijske strukture koncepata, Marfi, Hempton i Milovanović nisu uspeli da pokažu da ljudski kognitivni sistem konzistentno koristi hijerarhijske informacije u verifikovanju iskaza o odnosima karakteristika i koncepata sa različitih

nivoa hijerarhijske organizacije. Zaključak njihove studije je da ljudi, verovatno, ne koriste hijerarhijsku organizaciju koncepata kao dominantu strategiju pristupu informacijama iz konceptualnog sistema (Murphy, Hampton & Milovanović, 2012).

Posvetimo na trenutak ponovo pažnju formulaciji osnovnog principa kategorizacije kako ga formulišu Marfi i Medin 1985. Prema drugom principu Rošove, principu opažene strukture sveta, entitete koji nas okružuju odlikuju neuniformne, korelirane distribucije karakteristika na osnovu kojih se oni svrstavaju u moguće kategorije. Ako (plauzibilno) pretpostavimo da su karakteristike korelirane upravo tako da postoji pozitivna korelacija između pripadnika iste kategorije, a negativna između pripadnika različitih kategorija, vidimo da je ovaj princip povezan sa karakteristikom distinktivnosti. Prema Marfiju i Medinu, konceptualni sistem koji se bazira na korelaciji karakteristika težiće da maksimizuje razlike između kategorija, a sličnosti unutar njih. Iako je ovaj princip moguće dovesti u pitanje (Milovanović, 2000), on jeste plauzibilan, u tom smislu reči što bi njegova primena, pod jednakim svim drugim uslovima, rezultirala u semantičkoj memoriji koja bi bila optimalno organizovana za zadatke kategorizacije i prepoznavanja. Ipak, konceptualni sistem mora da održi neki stepen distinktivnosti između koncepata koje reprezentuje - tendencija ka formiranju kategorija koje odlikuje visoka sličnost između njihovih pripadnika tako u nekom trenutku prestaje da bude funkcionalna. Milovanović je, u pokušaju da reši ovaj problem, predložio da se princip maksimalne sličnosti koriguje principom prema kome kognitivni sistem pokušava da razvije kategorije koje su *maksimalno informativne* - u smislu u kome informaciju koju nosi neka distribucija verovatnoće kao entropiju definiše Šenonova teorija informacija (Shannon, 1948). Prema ovom shvatanju, tendencija da se da se maksimizuju sličnosti između pripadnika neke kategorije sukobljava se sa tendencijom da se diversifikuje struktura sličnosti unutar kategorija. Ako bi kognitivni sistem dosledno minimizovao sličnosti između entiteta u nekoj kategoriji, na kraju bi postigao uniformnu distribuciju nad nekom minimalnom vrednošću sličnosti koja bi bila potpuno neinformativna (tj. takva kategorija bi se nalazila u stanju maksimalne entropije). Poštujući ovakav princip organizacije, kognitivni sistem bi formirao koncepte koje bi odlikovao optimalan balans između visoke sličnosti između njihovih pripadnika i difersifikovane strukture distribucije sličnosti između njih. Takva struktura bi obezbedila maksimalnu diskriminativnu moć u odnosu na potrebu za optimalnim odgovorom sistema u slučaju potrebe da se znanje o nekom konceptu reorganizuje u skladu sa novim informacijama. Ujedno, tako organizovani

koncepti bi bili maksimalno informativni u odnosu na ograničenja koja nameću odnosi sličnosti i distinktivnosti (različitosti) sa drugim pripadnicima iste kategorije (Milovanović, 2000). Ovako predložen princip može da se interpretira kao rekurzivna interakcija tendencije za distinktivnošću i tendencije za sličnošću na svim nivoima konceptualne organizacije. Jezikom teorije informacija, organizacija konceptualnog sistema bila bi rezultat pokušaja da se postigne *optimalna entropija* u odnosu na više ciljeva konceptualne reprezentacije realnog okruženja. U raznim varijantama, principi kognitivne ekonomije koje je predložila Eleanor Roš, verujemo, mogu da se interpretiraju kao normativne osnove za teoriju konceptualne organizacije.

*Pristup baziran na znanju ili „teorija teorija".* Dobro poznat teorijski rad Marfija i Medina iz 1985. otvorio je novo poglavlje u istraživanjima konceptualne organizacije u kognitivnoj psihologiji (Murphy & Medin, 1985). Marfi i Medin izvode opsežnu kritiku svih fundamentalnih teorijskih rešenja za problem organizacije i reprezentacije koncepata i zaključuju da oni pate od ozbiljnih nedostataka te da ne mogu biti adekvatni da u potpunosti objasne sve funkcije ljudskog konceptualnog sistema. Okosnicu njihove kritike čini prepoznavanje činjenice da su pristupi preko koreliranih karakteristika, pristupi bazirani na konceptu sličnosti i pristupi implicitni u teorijama kategorizacije ili cirkularni ili nedovoljno konceptualno ograničeni i tako prepušteni variranju u svojim eksplanatornim strategijama u zavisnosti od konteksta svake posebne analize. Principe organizacije konceptualnog sistema možemo da razumemo samo polazeći od činjenice da su koncepti u ljudskom semantičkom pamćenju deo *širokih eksplanatornih mreža*, koje uključuju *kauzalne odnose*, i koji obezbeđuju konceptualnu koherenciju kategorija pružajući razloge odn. *objašnjenja* za to zašto su entiteti organizovani na jedan određeni način a ne drugačije. Nešto je PTICA ako poseduje odgovarajuće karakteristike te kategorije živih bića. Međutim, kognitivni sistem čini mnogo više od prostog skladištenja distribucije takvih karakteristika (teorija prototipova) ili memorisanja konkretnih pripadnika kategorije (teorija primeraka). Ptice mogu da lete zahvaljujući tome što imaju lake kosti, perje, aeordinamički oblikovan anatomski sklop. Ta karakteristika, da one mogu da lete, kauzalno omogućava to da one mogu da se gnezde na visokom drveću i da neke od njih učestvuju u velikim sezonskim migracijama. Korelacije između karakteristika, čiji značaj naglašava drugi princip Rošove, princip opažene strukture sveta, Marfi i Medin proširuju eksplanatornim principima tih korelacija. Dok je model sveta koji opaža kognitivni sistem kod Rošove samo model kovarijacije bitnih varijabli, kod Medina i Marfija to postaje generativni kauzalni model koji objašnjava

poreklo korelacija između karakteristika. Ovaj pristup je u literaturi različito označavan kao „*teorija teorija*“ i „*pristup baziran na znanju*“ (engl. *Knowledge View*, Murphy, 2002). Od prvih empirijskih istraživanja u okviru ove paradigme do do danas više studija je eksperimentalno potvrdilo da ljudi brže i lakše uče kategorijalne strukture kada eksperimentalni uslovi obezbeđuju mogućnost da se distribucije karakteristika povežu sa prethodno poznatim eksplanatornim, teorijskim znanjima (Murphy & Allopenna, 1994; Rehder & Ross, 2001; Waldmann, Holyoak, & Fratianne, 1995; Wattenmaker, Dewey, Murphy, & Medin, 1986).

Najegzaktnija formulacija teorijske paradigme Marfija i Medina ostvarena je u Rederovoj *generativnoj teoriji kategorizacije* (Rehder, 2003a, 2003b, 2007). Reder koristi teoriju kauzalnih modela koju smo predstavili u sekciji 7.2 o kauzalnom učenju da bi formulisao teorijska znanja o odnosima između karakteristika koncepata kao probabilističke kauzalne mreže. Primena kauzalnih modela tako obezbeđuje (a) formalizaciju strukturalnih, eksplanatornih odnosa između karakteristika koje odlikuju određene koncepte i kategorije, i (b) objašnjenje probabilističke strukture kategorija, odn. činjenice da se različite karakteristike kod određenih pripadnika iste kategorije javljaju samo sa određenim verovatnoćama. Jezik teorije kauzalnih modela savršeno odgovara reformulaciji principa konceptualne koherencije koja je postala neophodna posle kritike Marfija i Medina. Više matematičkih modela koji odgovaraju ovom teorijskom pristupu je ponuđeno od 90-ih do danas, ali se generativna teorija kategorizacije izdvaja po svom oslanjanju na „čist“ teorijski model kauzalnih mreža za koji, kao što smo videli, postoji normativno opravdanje. Međutim, ni u diskusiji ove teorije, kao ni u diskusiji pristupa Marfija i Medina uopšte, istraživači se nisu fokusirali na racionalnu analizu i analizu adekvatnosti normativnih pretpostavki. Prethodna teorijska diskusija između pristalica i kritičara teorija primeraka kao da je celu raspravu o organizaciji ljudskog konceptualnog sistema „izmestila“ iz debate o racionalnosti. Mi smatramo da je Rederova generativna teorija idealan model za pokušaj racionalne analize koncepata i kategorizacije, upravo zbog oslanjanja na formalni model koji se u drugim oblastima kognitivne psihologije uzima kao normativno opravdan.

Generativna teorija kategorizacije, jednostavno, počiva na ideji da su karakteristike koje opisuju koncepte međusobno povezane generativnim kauzalnim odnosima. Pogledajmo ponovo kauzalne modele predstavljene na Slici 12. u sekciji 7.2. Predlog Marfija i Medina da se korelacije karakteristike u kategorijama objašnjavaju njihovim vezama u eksplanatornim shemama u generativnoj teoriji

se operacionalizuje tako što neke karakteristike igraju ulogu uzroka, a neke ulogu posledica. Činjenica da ptice mogu da lete je posledica više njihovih karakteristika. Te karakteristike se, predstavljene kao čvorovi u kauzalnim mrežama, onda tretiraju kao generativni uzroci posledične karakteristike letenja. Uzimanjem u obzir većeg broja karakteristika, gradi se kauzalni model celokupne kategorije, a njegova probabilistička struktura tj. distribucija verovatnoće da određeni pripadnici te kategorije (npr. *Pingvin*, *Albatros*, *Kos*, *Orao* u kategoriji PTICA) poseduju određene karakteristike određena je parametrizacijom odgovarajućeg kauzalnog modela. Parametrizacija koju koristi Reder u generativnoj teoriji izomorfna je parametrizaciji koju je koristila Čengova u razvoju teorije kauzalne moći. Dakle, verovatnoće javljanja generativnih uzroka i kauzalne moći kojima su ponderisane njihove veze prema karakteristikama koje generišu određuju kompletnu probabilističku strukturu kategorije. Polazeći od teorije kauzalnih mreža kao normativnog okvira za analizu kauzalnosti i pretpostavke Marfija i Medina da u formiranju kategorija kognitivni sistem mapira kauzalnu strukturu svoje okoline, govorimo o čisto normativnom modelu. Ipak, postoje eksperimentalni nalazi koji mogu da se interpretiraju kao odstupanja od normativnog okvira u metodološkim paradigmama karakterističnim za proučavanje koncepata i kategorija. U tom svetlu interesantan je *efekat primarnosti uzroka* (Redher & Kim, 2006). Rederova generativna teorija predviđa da će značaj pojedinih karakteristika u sudovima o pripadnosti određenoj kategoriji biti povezana sa brojem uzroka koji generišu tu karakteristiku u odgovarajućem kauzalnom modelu. Eksperimentišući sa veštačkim (ali plauzibilnim) kategorijalnim strukturama, Reder i Kim pokazuju da karakteristike koje u kauzalnim modelima zavise od više uzroka zaista više utiču na određivanje stepena pripadnosti kategorijama koje ispitanici u njihovim ogledima uče. Međutim, pokazuje se da u ispitivanju kategorija koje imaju strukturu kauzalnog lanca (Slika 12 (f)) ili mreže zajedničkog uzroka (Slika 12 (g)), ispitanici najveći značaj daju primarnim uzrocima - onim generativnim karakteristikama u kauzalnom modelu koji utiču na druge karakteristike ali se nalaze najdublje u kauzalnoj strukturi, odn. karakteristikama koje nemaju drugih uzroka u modelu (Rehder & Kim, 2006). Iako teorija kauzalnih modela omogućava dovoljno fleksibilnu parametrizaciju da statistički fituje ovakve nalaze, pretpostavka o tome da će veći uticaj imati one karakteristike koje imaju više uzroka - i koja je empirijski potvrđena za ostale karakteristike u ovakvim modelima - je nesaglasna sa efektom primarnosti uzroka. Reder i Kim spekulišu da je ovaj nalaz moguće objasniti ako se pretpostavi jedna verzija *psihološkog*

*esencijalizma* po kojoj ljudi projektuju egzistenciju još dubljeg, „skrivenog“ uzroka koji generiše prve opservabilne uzroke neke kategorijalne strukture i simulacijama pokazuju da takva hipoteza zaista može da objasni efekat primarnosti uzroka. Ideja da su različite prirodne kategorije koje predstavlja ljudski konceptualni sistem *esencijalizovane* u različitoj meri karakteristična je za Rederov razvoj generativne teorije (Rehder, 2007). Na primer, kategorije živih bića su kod odraslih, prosečno obrazovanih ljudi esencijalizovane, jer počivaju na pretpostavci da jedna jedina definišuća karakteristika ma koje biološke vrste - njen DNK kod, naime - kauzalno generiše sve njene opservabilne karakteristike tokom razvoja. Kategorije ne moraju da budu esencijalizovane, a Reder predlaže i hipotetički „teorijski scenario“ procesa esencijalizacije kategorija sa uklapanjem novih informacija u kategorijalnu strukturu (Rehder, 2007). Bitno je držati na umu da u ovakvim raspravama niko ne podržava teoriju metafizičkog esencijalizma po kome entiteti u Univerzumu zaista imaju „skrivene suštine“. Hipoteza psihološkog esencijalizma predstavlja tvrdnju da ljudski kognitivni sistem reprezentuje kategorije kroz proces postepene esencijalizacije koji ne predstavlja ništa drugo do proces u kome se celokupna kauzalna struktura kategorije vremenom dovodi u zavisnosti od jednog, definišućeg kauzalnog faktora. Ova pretpostavka, svakako, nije normativne prirode: ona se uvodi *a posteriori* da bi objasnila neke od eksperimentalnih nalaza i nije ni na koji način nužan deo normativnog okvira formalizma kauzalnih mreža.

Pored Rederove primene kauzalnih modela u razvoju generativne teorije kategorizacije, u seriji novijih radova grupa naučnika okupljenih oko Džošue Tenenbauma i Toma Grifitsa razvila je teoriju kauzalnih mreža u formu *kauzalnih gramatika* koje se koriste za opis arbitrarno kompleksnih domena znanja kojima upravljaju kauzalni zakoni (Tenenbaum, Griffiths, & Kemp, 2006, Tenenbaum, Griffiths, & Niyogi, 2007, Griffiths, & Tenenbaum, 2007, Griffiths & Tenenbaum, 2009). Koren razvoja ovog formalizma nalazi se u teoriji kauzalne podrške Grifitsa i Tenenbauma koju smo razmatrali u sekciji 7.2 posvećenoj kauzalnom učenju. Novi formalizam proširuje ideju bejzijanske analize kauzalnih modela na mnoge probleme indukcije tj. učenja uopšte, a ograničenja za njega predstavljaju praktično samo kompjutaciona ograničenja organizma koji ga implementira. Formalizam koji koriste ove bejzijanske, kauzalne teorije toliko je moćan da može da se shvati kao predlog opšte formalne kompjutacione mašinerije viših kognitivnih procesa uopšte. Ipak, zbog velike kompleksnosti teorije kauzalnih gramatika, složenih pretpostavki i mnogih nerešenih pitanja koja okružuju i manje kompleksne bejzijanske formalizme

od ovog (up. Jones & Love, 2011), nećemo ih uključiti u našu diskusiju organizacije konceptualnog sistema.

*Komentari o racionalnosti organizacije konceptualnog sistema u odnosu na funkciju kategorizacije.* Naša diskusija racionalnosti organizacije ljudskog konceptualnog sistema ograničena je iz više perspektiva. Pre svega, tu je problem inherentan ovoj oblasti naučnog rada, u kojoj debata o racionalnosti nije vođena elaborirano i dosledno kao u prethodnim oblastima koje smo diskutovali. Mi smo pokušali da identifikujemo neke razloge zbog kojih je to tako. Po našem mišljenju, nekritički razvoj teorija primeraka, koji je skoro dve pune decenije određivao fokus istraživanja u ovoj oblasti, osnovni je razlog zašto je mogućnost normativne, racionalne analize kategorizacije zanemarena. Drugo, mi nismo posvetili punu pažnju teorijama sličnosti[64]. Sličnost je fundamentalno povezana sa svim funkcijama konceptualnog sistema. Eksplanatornu moć sličnosti u teorijama učenja koncepata i kategorizacije oštro su kritikovali Marfi i Medin u klasičnom teorijskom radu iz 1985. Međutim, treba držati na umu da cela diskusija o normativnim pretpostavkama u ovoj oblasti može da se povede u potpunosti iz perspektive analize samo ovog pojma. Šta više, teorije sličnosti predstavljaju retku grupu kognitivnih teorija koje su aksiomatizovane: tzv. geometrijske teorije sličnosti poštuju aksiomatiku metričkih prostora (up. MDS modele, Borg & Groenen, 2005), dok set-teorijska teorija sličnosti Tverskog počiva na sopstvenom aksiomatskom sistemu (Tversky, 1977). Verovatno je, posle svega do sada iznetog u ovoj tezi, redundatno dodati da bihejvioralni podaci retko ili nikad ne zadovoljavaju predložene aksiomatske okvire teorija sličnosti. Konačno, najrazličitije formalne teorije učenja koncepata i kategorizacije prevazilaze jednostavne formulacije koje smo mi uveli. Ipak, mi se u našoj raspravi po pravilu usresređujemo samo na one formalizme koji sprovode racionalnu analizu na normativnim osnovama, i deskriptivne modele koje se razvijaju u diskursu kritike takvih normativnih teorija; komplikovani matematički modeli učenja kategorija i kategorizacija ni u kom slučaju nemaju te osobine. Skorija analiza koju je predstavio Feldman ukazuje na mogućnost da će normativni modeli ipak početi ozbiljnije da se razmatraju u ovoj oblasti. Njegovi rezultati pokazuju da kategorije definisane preko jednostavnih pravila logičkog računa prvog reda (konjukcije, disjunkcije i negacije) u učenju predstavljaju subjektivni napor za ispitanike (meren preko proporcije tačnih kategorizacija) koji je direktno proprocionalan jednostavnoj meri logičke kompleksnosti pravila karakterističnih za određenu kategorijalnu strukturu (tzv. *Boolean complexity*,

Feldman, 2000). Ovako jednostavan formalizam koji se pokazuje prediktivnim za učenje velike klase najrazličitijih kategorijalnih struktura ne može da se posmatra drugačije do kao dobrodošlo osveženje u oblasti kojom dominiraju teorijski manje plauzibilni i normativno manje motivisani, veoma komplikovani matematički modeli.

Anderson je, kao što je navedeno, sproveo jedan pokušaj racionalne analize kategorizacije (Anderson, 1991c). Prema ovoj analizi, osnovni cilj konceptualnog sistema je predikcija karakteristika novih entiteta koje srećemo u našoj okolini. Već ova specifikacija je u konfliktu sa realnošću u kojoj su ciljevi ljudskog konceptualnog sistema mnogostruki: podjednako se može tvrditi da je njegov osnovni cilj jednostavno određivanje kategorijalne pripadnosti nekog novog entiteta u celini. Anderson pitanje kategorizacije - koje smo upravo postavili - rešava tvrdnjom da je kategorijalna pripadnost nekog entiteta samo još jedna njegova karakteristika, naime, karakteristika da je njegova reprezentacija u semantičkoj memoriji povezana sa imenom kategorije kojoj pripada. Ponovo, teorijski nije jasno da li je ovakvo određenje plauzibilno ili ne, iz istih razloga iz kojih nije jasno šta okuplja entitete u nekoj kategoriji u modelima primeraka. Konceptualni sistem obezbeđuje funkcije kategorijalne indukcije, analoškog mišljenja, deduktivnog rezonovanja i kreativnih konceptualnih kombinacija koje ćemo uskoro razmatrati. Andersonova racionalna analiza je, prema tome, suviše specifična i trpi od nedostatka opštosti. Neke od Andersonovih pretpostavki skoro u potpunosti odgovaraju opisu karakteristika prirodnih vrsta koje daju još Rošova i saradnici, ali Anderson svoje pretpostavke ne dovodi eksplicitno u vezu sa principima kategorizacije Eleanor Roš. Marfi u prepostavkama racionalne analize kategorizacije prepoznaje još tvrdnji čija je plauzibilnost upitna, poput one o nezavisnosti distribucija karakteristika unutar kategorija, za koju primećuje da u slučaju hijerarhijske organizacije koncepata ni teorijski ne može da bude zadovoljena, Murphy, 1993); Gigerencer je kritikovao istu pretpostavku sa stanovišta ekološke validnosti (Gigerenzer, 1991a). Ostatak racionalne analize kategorizacije počiva na još hipoteza sumnjive plauzibilnosti, poput formulisanja *a priori* verovatnoća u tipičnom bejzijanskom modelu koje se odnose na tendenciju kognitivnog sistema da dva (ma koja) entiteta uopšte svrsta u istu, ili različite kategorije. Nije jasno zašto bi ljudski konceptualni sistem uopšte počivao na takvom parametru u funkcijama kategorizacije odn. zašto bi bilo korisno - osim za primenu u Andersonovom racionalnom modelu - unapred odrediti verovatnoću da neke dve stvari u svetu pripadaju istoj, a ne različitim kategorijama.

Naš zaključak je da u domenu organizacije ljudskog konceptualnog sistema

normativne principe jeste moguće prepoznati, i da njih treba eksplicitno graditi polazeći od pretpostavki koje je napravila još Eleanor Roš u danas klasičnoj liniji istraživanja. Ove principe Marfi i Medin proširuju u paradigmi „teorije teorija“, za koju smo videli da ima potencijalno elegantnu matematičku formalizaciju u jeziku kauzalnih modela - jeziku za koji je ponuđeno normativno opravdanje. Ipak, ovde govorimo o principima koji još uvek nisu formalizovani tako da predstavljaju „postavku problema“ racionalne analize: mnogo posla ostaje na polju unifikacije različitih ideja i podvođenja manje generalnih principa pod principe višeg reda. Ne možemo da zaključimo da normativni okvir za analizu konceptualne organizacije i kategorizacije postoji, već tek da se on nazire. Ako se jednom takav konzistentan, normativni okvir za organizaciju konceptualnog sistema zaokruži, biće moguće sistematski testirati njegovu racionalnost. Nažalost, upotreba ekološki manje validnih eksperimentalnih nacrta, poput 5-4 problema u učenju kategorija, nije način da se takva diskusija povede. Veliki deo eksperimentalnog rada koji nije usmeren na testiranje modela primeraka usmeren je na testiranje predikcija drugih matematičkih modela kategorizacije, bez prethodnog pokušaja da se sprovede njihova racionalna analiza. Ostatak eksperimentalnih napora usmeren je na otkrivanje novih efekata u poznatim paradigmama. Polazeći od kompleksnosti ljudskog konceptualnog sistema - a govorimo o najkompleksnijem teorijskom konstruktu savremenih kognitivnih nauka - možda nije čudno što je takve nalaze, sa malim modifikacijama postojećih eksperimentalnih nacrta i dovitljivim dizajnom novih stimulusa i procedura, moguće proizvesti svako malo. Može se reći da je oblast koncepata i kategorija jedna od najkreativnijih oblasti savremene kognitivne psihologije, u tom smislu; ipak, čini se da je te kreativnosti možda i previše. Oblast se nalazi u stanju rasparčanosti u kome tek poneliko istraživačkih grupa može da razvije zajednički jezik u okvirima iste ili slične eksperimentalne metodologije i pristupa modeliranju odgovarajućih procesa.

Prepreku poduhvatu racionalne analize organizacije konceptualnog sistema i njegovih funkcija svakako predstavlja neslaganje oko prirode odgovarajućeg teorijskog okvira za takvu analizu, ili nezainteresovanost da se u ovoj oblasti uopšte postavi takav okvir. Dok se diskusija bude vodila polazeći od internih kognitivnih modela - pretpostavljajući jedan, drugi ili treći vid komplikovane reprezentacije bez reference na ishode bihejvioralnih testova koji bi mogli da ih falsifikuju - jedinstven teorijski okvir normativne prirode neće biti razvijen. Druga prepreka je svakako činjenica da bi taj okvir predstavljao formalnu konstrukciju zavidne kompleksnosti. I dalje govorimo o problemima za koje smatramo da su makar i u principu rešivi i

koji mogu da ocrtaju plan nekog budućeg, konzistentnijeg istraživačkog programa u oblasti koncepata i kategorija. Sada prelazimo na analizu veoma teških problema vezanih za funkcije ljudskog konceptualnog sistema - problema za koje nismo sigurni da li u teorijskim paradigmama savremenih kognitivnih nauka uopšte mogu da pronađu egzaktna rešenja.

## 7.6 Koncepti II: kreativna funkcija

Na početku prethodne sekcije, naveli smo sledeće ograničenje pod kojim se vodi diskusija funkcija kategorizacije: svi standardni pristupi oblasti koncepata i kategorija - pristupi u kojima smo pokušali da pronađemo osnove za racionalnu analizu - pretpostavljaju da su *karakteristike koje definišu koncepte date i fiksne*. Eksperimentalna metodologija koja dominira oblašću u potpunosti je zasićena ovom pretpostavkom. U tipičnom eksperimentu učenja kategorija, eksperimentator će kroz instrukcije objasniti ispitaniku da se koncepti sadrže od određenih karakteristika na koje treba da obraća pažnju i na osnovu njih donosi odluke o tome koji stimulus pripada kojoj kategoriji. Kada eksperimentalne instrukcije nisu ovoliko eksplicitne, stimulusi - vizuelni ili verbalni - su generisani tako da su njihove karakteristike očigledne za ispitanika. Sve teorije kategorizacije i učenja koncepata predstavljaju formalne modele koji kao svoj input uzimaju skupove karakteristika. Modeli prototipova ih onda ponderišu i sabiraju, modeli primeraka računaju sličnosti između stimulusa kroz eksponencijalno skaliranje euklidskih distanci u prostoru karakteristika, kauzalni modeli generativne teorije ih tretiraju kao varijable u kauzalnim mrežama. Skupove karakteristika koji definišu pojmove odlikuje (*i*) konačan i nepromenljiv broj karakteristika i (*iia*) određeni raspon vrednosti koje mogu da uzimaju kontinuirane karakteristike (poput raspona zasićenosti boje, ili svetline vizuelnih stimulusa) odn. (*iib*) konačan i nepromenljiv broj vrednosti koje mogu da uzimaju diskretne karakteristike (npr. vrsta odeće = suknja/pantalone). Standardni pristupi konceptima i kategorijama ne postavljaju pitanje na koji način kognitivni sistem uopšte donosi odluku o tome koji skup karakteristika, sa kojim rasponom vrednosti, u kom kontekstu, ulazi u deskripciju nekog pojma - deskripciju koja postaje objekat, odn. input kompjutacionih procesa kategorizacije o kojima standardni pristupi imaju mnogo više da kažu.

U oblasti proučavanja koncepata i kategorija malo se pažnje poklanja činjenici da su karakteristike koje odlikuju pripadnike neke kategorije *i same kategorije*.

Uzmimo, na primer, tipičan opis stimulusa u eksperimentu statističkog učenja kategorija preko verbalnih karakteristika. Recimo da ispitanici treba da nauče da razlikuju dve kategorije njima nepoznatih životinja. Recimo da je karakteristika koja odlikuje četiri od pet pripadnika kategorije A da „imaju šape“, dok ista ta karakteristika odlikuje samo dva od pet pripadnika kategorije B. U eksperimentalnoj situaciji karakterističnoj za istraživanje teorija koncepata diskutovanih u prethodnoj sekciji, binarna karakteristika „ima šape/nema šape“, ili „ima šape/ima peraja“, tretira se isključivo kao jedna binarna varijabla koja uzima vrednost 0 ili 1, u zavisnosti od toga kako je kodirana za primenu nekog modela kategorizacije. U ekološki relevantnim uslovima problem reprezentacije je daleko složeniji. Karakteristika „ima šape“ implicira sledeće pitanje: *kakve šape* imaju pripadnici nove kategorije životinja koju treba naučiti? Da li su to šape psa, ili šape slične šapama medveda, ili su to šape veverice, ili su slične šapama mačaka, zatim, kojih mačaka, da li su to šape poput šapa domaćih mačaka, ili su to velike šape sa oštrim kandžama pripadnika klasa velikih mačaka? Koncepte prirodnih vrsta, dalje, odlikuje izvestan stepen konzistencije između forme koju uzimaju njihove razne karakteristike. Velika životinja ne može imati male šape. Boja, ili šara tih šapa će biti konzistentna sa bojom ili šarom ostatka tela kod ogromnog broja sisara koji imaju šape. Ta konzistencija važi i za druge „delove“ ovih životinja, koji se po analitičkoj pretpostavci računaju u karakteristike na osnovu kojih se o kategorijama životinja uči, preko kojih se one prepoznaju i kategorišu. Problem koji opisujemo „u principu“ mogu da inkorporiraju standardne teorije opisane u prethodnoj sekciji; kažemo „u principu“ jer ni iz daleka nije jasno kako bi se bilo koja od njih nosila sa problemom ogromne varijabilnosti karakteristika prirodnih vrsta, između kojih ponekad tek fini detalji određuju distinktivne razlike koje razdvajaju jednu kategoriju od drugih. U oblasti *ekstrakcije karakteristika*, koja se daleko više proučava u prepoznavanju vizuelnih objekata nego u konceptualnim kategorizacijama[65], poklanja se pažnja ovim pitanjima. Funkcija algoritama za ekstrakciju karakteristika je upravo da na osnovu nestruktuiranog inputa izvrše matematičku analizu koja bi izdvojila invarijante višeg reda, korisne u opisu više drugih, sličnih, donekle varijabilnih inputa. Nijedan algoritam predložen u studijama ekstrakcija karakteristika ni izbliza ne može da se nosi sa realnom varijabilnošću i kompleksnošću objekata u svetu koji nas okružuje.

Iza ovog zapažanja o fundamentalnoj nestabilnosti teorijskog koncepta karakteristike krije se još teži problem: na koji način, uopšte, kognitivni sistem

uspeva da otkrije novu karakteristiku, novu kategoriju, novi predikat koji može da iskoristi u opisu nekih događaja u svom okruženju? Prema standardnoj simbolicističkoj paradigmi, kognitivni sistem predstavlja implementaciju nekog formalnog sistema, a kognitivna psihologija je poduhvat „inženjerisanja unazad" tog formalnog sistema. Ipak, svaki formalni sistem se zasniva na zatvorenom, *konačnom skupu elementarnih simbola* koje može da reprezentuje i na njih primeni odgovarajuća pravila inferencije. Nove, prethodno nepoznate karakteristike o kojima sistem uči, moraju da budu svodive na neku vrstu kombinatorijalnog opisa izraženog u elementarnim simbolima - „karakteristikama" - koje sistem poznaje. Šins, Goldstoun i Tibo su učenje pod ovakvom pretpostavkom nazvali učenjem kroz *fiksne skupove* karakteristika, i suprostavili mu učenje koje su nazvali učenjem kroz *fleksibilne prostore* karakteristika (Schyns, Goldstone & Thibaut, 1998), ali kao što ćemo videti, u ovoj retkoj teorijskoj studiji koja dovodi u pitanje pretpostavku o fiksnim skupovima karakteristika nisu proučavali najteži oblik problema koji može da se postavi.

Ako se svako učenje o nekom novom konceptu svodi na otkrivanje načina na koji su njegove „nove" karakteristike kodiraju u kombinacije iz elementarnog, fiksnog skupa karakteristika odn. simbola, onda *nikakvog novog učenja uopšte nema*. Važna teorijska studija Šinsa, Goldstouna i Tiboa se obraća lakšem problemu. Oslanjajući se na praktično samo tri eksperimentalne studije (Schyns & Rodet, 1997, Schyns & Murphy, 1994, Rodet & Schyns, 1994), oni diskutuju slučajeve u kojima kognitivni sistem u zadatku kategorizacije selektivno izdvaja iz manje strukturiranih stimulusa karakteristike koje su *funkcionalne* za kategorizaciju koja se uči. Ovakvi eksperimenti koriste stimuluse koji nisu karakteristični za oblast koncepata i kategorija: manje su struktuirani, a njihove karakteristike, koje nisu unapred poznate ispitanicima, moraju da se izdvoje iz celine stimulusa. Te karakteristike nekada predstavlju konjukcije drugih karakteristika, tako da kognitivni sistem mora da donosi odluke o tome da li je neka celina u stimulusu relevantna karakteristika, ili je samo neki njen deo relevantna karakteristika, ili je ona tek deo veće celine koja je relevantna karakteristika. Manji broj ovakvih eksperimentalnih studija demonstrirao je da kognitivni sistem u učenju kategorija izdvaja iz nestruktuiranih celina one karakteristike koje su funkcionalne za određenju kategorizaciju, odn. da prepoznaje koji delovi stimulusa su informativni, a koji nisu. U zavisnosti od toga kakva je struktura kategorija koje treba da nauči, kognitivni sistem može iz istih stimulusa da ekstrahuje različite karakteristike - ako su te različite celine u stimulusima po

sebi prediktivne za kategorizaciju koja se uči (up. *princip funkcionalnosti* Šinsa i Marfija, Schyns & Murphy, 1994).

Međutim, i u ovakvim eksperimentalnim studijama postoji elementarni, fiksni skup karakteristika koji ispitanici, u krajnoj liniji, mogu da prepoznaju. Da takvog skupa karakteristika nema, teško da bi bilo moguće sprovesti eksperimente po nacrtima koje koriste Šins, Marfi i Rode u pomenutim studijama (Schyns & Rodet, 1997, Schyns & Murphy, 1994, Rodet & Schyns, 1994). Zato Šins, Goldstoun i Tibo greše kada ovakve procese diskutuju kao procese *kreiranja karakteristika*. Tu nema nikakvog pravog kreiranja novih predikata u opisu realnog okruženja: na početku svoje studije, Šins, Goldstoun i Tibo objašnjavaju da su u pitanju efekti viših kognitivnih procesa - procesa učenja kategorijalnih struktura pokušaj-po-pokušaj uz fidbek - na perceptivne procese. U pitanju je složen proces kroz koji viši kognitivni procesi učenja kategorija testiraju hipoteze o mogućim karakteristikama kroz perceptivne procese, odn. organizuju perceptivni input sa višeg nivoa funkcionisanja. Još jednom, pravog *otkrivanja* novih karakteristika, nikakvog kreiranja karakteristika, tu nema.

Briljantna studija Visnievskog i Medina o ulozi teorijskog znanja u učenju kategorija demonstrirala je efekte interpretacija karakteristika koji do danas nisu inkorporirani u formalne modele kategorizacije (Wisniewski & Medin, 1994). Visnievski i Medin su kao stimuluse koristili dečje crteže ljudskih figura iz poznatog Haris-Gudinaf testa crteža ljudske figure. U složenom eksperimentalnom nacrtu, izdvojili su grupe crteža koji su mogli da se opišu i preko konkretnih karakteristika („ruke stoje uz telo ili je na licu osmeh“ i sl.), i preko asptraktnih karakteristika (crtež je detaljan, ili je crtež neobičan, ili ljudi na crtežu izvode određene akcije, i sl.). Dvema grupama ispitanika su postavili zadatak učenja tako definisanih kategorija (bez davanja prethodnih informacija o karakteristikama koje određuju kategorije). Jednoj grupi ispitanika je rečeno da uče da razlikuju crteže kreativne dece od crteža dece koja nisu kreativna, a drugoj da uče da razlikuju crteže dece iz grupe 1 od crteža iz grupe 2. U prvoj grupi ispitanika, smisleni nazivi kategorija su doveli do snažnog efekta prethodnog znanja ili prepostavki o tome kako bi mogli da izgledaju crteži kreativne dece. Taj efekat je doveo do toga da su u objašnjenju svojih kategorizacija ispitanici iz te grupe koristili mnogo više apstraktnih nego konkretnih karakteristika - crteži kreativne dece su „detaljni“, „apstraktniji“, „kreativniji“. Iako su mogli da nauče kategorije preko konkretnih karakteristika kao u ma kom drugom eksperimentu učenja kategorija, ispitanici

iz ove grupe su, zahvaljujući povezivanju karakteristika crteža sa pretpostavkama o tipičnim crtežima kreativne dece, ekstrahovali apstraktne, i u mnogo manjoj meri konkretne, karakteristike. Situacija u grupi koja je učila iste kategorije bez smislenih naziva za njih bila je upravo obrnuta. U daljim eksperimentima, Visnievski i Medin su demonstrirali još upečatljivije efekte prethodnog znanja na učenje kategorija. Potpuno ista karakteristika - crtež odela na određenoj ljudskoj figuri - bila je interpretirana kao „uniforma za rad na farmi“ ili „gradska uniforma“ (tipična uniforma koju nose npr. predstavnici gradskih komunalnih službi) od strane grupe ispitanika koji su učili kategorizaciju crteža predstavljenu kao crteže gradske i seoske dece. Ako bi neki crtež svrstali u kategoriju crteža za koju im je rečeno da su ih nacrtala deca iz grada, dali bi interpretaciju da figura nosi „uniformu za rad na farmi“; u slučaju da dobiju negativni fidbek, odn. da nauče da taj crtež spada u drugu kategoriju, ispitanici bi kasnije menjali interpretaciju, tvrdeći da figura ima na sebi „gradsku uniformu“. Potpuno ista ljudska figura u pokretu je od strane ispitanika koji su učili da razlikuju crteže gradske od crteža seoske dece interpretirana kao „figura koja se penje u igralištu“, dok je od strane ispitanika koji su učili da razlikuju crteže kreativne dece od crteža dece koja nisu kreativna interpretirana kao „figura koja pleše“. Kada se učenje kategorija odvija u odnosu na određenu teorijsku pretpostavku o odnosima koji postoje u domenu učenja, ta teorijska pretpostavka ne dovodi samo do selekcije onih karakteristika koje su relevantne, već do *interpretacije tih karakteristika kako bi se one prilagodile semantici koju ta teorijska pretpostavka podrazumeva.* Ovde sve vreme govorimo o *istim crtežima* i *istim karakteristikama* čija se interpretacija menjala tokom učenja u zavisnosti od toga na koji način je mogla da bude asimilirana u teorijske pretpostavke o odgovarajućim kategorijama. Još 1994. godine, kada su Visnievski i Medin objavili svoju studiju, oni su morali da konstatuju da standardni modeli kategorizacije nemaju nikakvih mehanizama da inkorporiraju ovakve procese - danas, skoro dvadeset godina posle objavljivanja njihovih rezultata, situacija je nepromenjena.

*Komentari o racionalnosti kreativne funkcije konceptualnog sistema.* U visoko varijabilnim kontekstima, znanje o konceptima postaje proces ekstrakcije novih karakteristika vođen funkcionalnošću koja uslovljava prelazak objekata i pojmova iz jednih u druge kategorije, odn. primenu deskriptivnih predikata koji nisu dozvoljeni u jednom kontekstu na dekonstektualizovane objekte i koncepte. Kako se ta dekonstualizacija odvija? Ni standardna simbolicistička paradigma, ni konekcionistička, emergentistička paradigma ne vide ni trag koji bi vodio

ka putu prema odgovoru na ovako postavljen problem. U svim teorijskim usmerenjima kompjutacionističke paradigme, ovaj problem je moguće rešiti samo pretpostavljajući da na nekom nivou opisa postoji dovoljno apstraktan, generalan i konačan skup elementarnih predikata (ili složajeva aktivacije, u neuronskim mrežama, ili stabilnih stanja, u dinamičkim sistemima koji bi bili primenjeni na simboličke opise, i tako dalje) na čije se kombinacije svode svi drugi zamislivi predikati. Dodajte ovome skup pravila (karakteristike višeg reda odn. „karakteristike nad karakteristikama" u konekcionizmu, složene hijerarhijski organizovane atraktore u nelinearnim dinamičkim sistemima, i tako dalje) koji bi morao da postavlja ograničenja na skup dozvoljenih i nedozvoljenih primena određenih predikata na određene koncepte u određenim kontekstima. Ukoliko govorimo o potpuno novim karakteristikama, karakteristikama koje kognitivni sistem stvara tokom kategorizacija, analoškog mišljenja, rešavanja problema i drugih procesa, ni složena konstrukcija koju smo ovde skicirali ne radi posao. Svaki simbolički sistem, bez obzira na to da li je realizovan kao formalni sistem u simbolicističkoj paradigmi, ili nelinearni dinamički sistem u konekcionističkim tj. emergentističkim pristupima, potpuno je ograničen sopstvenom strukturom i informacijama koje nosi njegov input. Toliko je složen problem o kome govorimo.

U XIX veku, Lobačevski je morao da dekontekstualizuje jedan od najintuitivnijih pojmova koji je ljudska vrsta uopšte ikada razvila, pojam euklidskog prostora. Ova, u istoriji nauke čuvena dekontekstualizacija u kojoj je prostor prestao da bude „običan" prostor u kome se dve paralelne prave nikada ne preseku, ipak je zadržala koncept geometrije, apstrahujući ga na novu analitičku ravan sa koje se on danas koristi. Argument koji bi tvrdio da su samo konceptualni sistemi genijalnih pojedinaca sposobni za takve apstrakcije nije održiv: posle otkrića Lobačevskog, studenti matematike i na osnovnim studijama uspešno savlađuju koncepte neeuklidskih geometrija, a tek nekoliko decenija posle njegovog rada Minkovski i Ajnštajn pokazuju da fundamentalna struktura fizičkog sveta odgovarajuću deskripciju dobija upravo kroz koncept neeuklidskih geometrija, što je takođe činjenica koju savlađuju studenti fizike na osnovnim studijama. Suprotno dva i po milenijuma dugačkoj eksplicitnoj tradiciji da je prostor euklidski, suprotno nezamislivo dugačkoj tradiciji svih bioloških vrsta sa vizuelnim sistemom sličnim ljudskom, kroz koji smo i razvili intuiciju o opštoj formi prostora kao euklidskog. Konceptualni sistem koji nije sposoban za dekonstualizaciju koncepata i objekata u meri koja omogućava otkriće novih koncepata nije onaj konceptualni sistem koji

odlikuje čoveka. Zapravo, upravo ta sposobnost dekontekstualizacije i transpozicije koncepata i objekata iz jednog u drugi deskriptivni jezik kojom se uvodi novi predikat, ta sposobnost koju toliko ne razumemo, verovatno je deo fundamentalnog određenja ljudskog mišljenja i ključna među distinktivnim karakteristikama ljudskog kognitivnog sistema.

U prethodnoj sekciji smo pokušali da skiciramo obrise normativnog okvira za analizu racionalnosti kategorizacije. Naše mišljenje ostaje da je takav normativni okvir moguć, i da ga je moguće izgraditi kroz pažljivu sistematizaciju i unifikaciju do sada predloženih principa konceptualne koherencije. Međutim, diskusija u ovoj sekciji pokazala je da je problem kompleksniji od načina na koji je on postavljen u odnosu na funkcije kategorizacije u standardnim teorijama. Pretpostavimo da i postoji dogovor oko normativnog okvira za analizu kategorizacije koji diktira da ljudski kognitivni sistem stvara uvek maksimalno informativne kategorije, koje odlikuje visoka sličnost unutar kategorija i što veća distinktivnost (različitost) između kategorija, mapirajući objektivno korelacije atributa koje prepoznaje u okolini, kao što je predložila Eleanor Roš. Dodajmo tome principe konceptualne koherencije koji slede iz rada Marfija i Medina, formalizovane na odgovarajući način. Na primer, ideja bi mogla da bude da sve opisane principe kategorizacije kognitivni sistem pokušava da realizuje u minimalnom kauzalnom modelu koji omogućava inferencije o svim koreliranim atributima i realizuje maksimalno informativne kategorizacije. Ne tvrdimo da je razvoj ovakvog normativnog okvira jednostavan - on već u prvom koraku nailazi na tzv. *problem okvira* koji decenijama koči razvoj veštačke inteligencije (Shanahan, 2009) - ali se čini da je on bar u principu moguć[66]. Problem okvira, koliko god složen, već je formalno tretiran u okviru aksiomatskih sistema revizije verovanja (Gardenfors, 1988). U krajnjoj liniji, neka mogućnost normativne formalizacije se nazire. Problem koji preostaje je sledeći: kakva god bila formalna realizacija normativnog okvira kategorizacije, ako do nje dođemo, ona će realizovati ma koji semantički sistem koji mi budemo zahtevali da realizuje. Međutim, kakav god semantički sistem ona bude realizovala, a recimo da je to neka kodifikacija informacija koje dobro opisuju adaptivno relevantnu strukturu okoline čoveka, takva formalizacija nikada neće omogućiti dekonstualizaciju tog sistema u meri koja bi dovela do otkrića *suštinski novih deskriptivnih predikata* odn. karakteristika.

Džeri Fodor, verovatno poslednja linija odbrane simbolicističke paradigme i kompjutacionizma i pred ovakvim kritikama, u knjizi „*Jezik uma*“ formulisao je opšti

dokaz prema kome je, zbog problema upravo diskutovanih, svako *učenje koncepata* u suštini nemoguće (Fodor, 1975, Piatelli-Palmarini, 1975). Prema Fodoru, tako nešto kao što je učenje koncepata uopšte ne postoji: kognitivni sistem čoveka, zahvaljujući specifikaciji genetskog koda čoveka, zaista dobija urođen konačan, nepromenljiv skup hipoteza i sve što uopšte čini jeste testiranje tih hipoteza u odnosu na empirijsku stvarnost. Rezultati tih testova, koje mi kolokvijalno nazivamo učenjem, predstavljaju uklapanje novih informacija u već postojeće formalne strukture kroz selekciju odgovarajuće kombinatorijalne sheme koja njihov opis uklapa u opis putem fundamentalnih predikata. Fodor, niti bilo ko drugi, nikada nije specifikovao takav skup predikata koji bi objašnjavao sva učenja za koja se ljudski kognitivni sistem pokazuje sposobnim. U odnosu na posledice gledišta koje zastupa Fodor, Gi Selerje iz Međunarodnog centra za genetičku epistemologiju Univerziteta u Ženevi je tokom debate između Žana Pijažea i Noama Čomskog u opatiji de Roajamon 1975 primetio: „Primenimo li to na biologiju, Fodorova teorema dovodi do još neverovatnijih zaključaka, od kojih jedan sasvim očigledno razbija svako dalje razmišljanje o toj temi. Prvi je da darvinovski mehanizmi nisu sposobni da proizvedu niz formi koje vode do vrste kakva je *homo sapiens*. [...] Drugi zaključak je da vrsta kakva je *homo sapiens* u stvari ne postoji; što ću nazvati Fodorovim *cogito ergo non sum*“ (citirano prema Piatelli-Palmarini, 1975, prevod Vesna Polovina).

# 8 Refleksije o debati

Naš nepotpun kritički pregled debate o racionalnosti viših i simboličkih kognitivnih funkcija ipak se dotakao skoro svih značajnih perspektiva iz kojih se pruža pogled na ovu centralnu teorijsku diskusiju u kognitivnoj psihologiji. Naš cilj sada jeste ukazivanje na duboke probleme koji su u debati o racionalnosti najčešće implicitni, te razvoj što jasnije empirijske slike kognitivnog sistema koja proizlazi iz sinteze različitih teorijskih pozicija i eksperimentalnih nalaza koji ih podržavaju ili falsifikuju. Pokazaćemo da debata o racionalnosti *summa summarum* otkriva jednu empirijsku sliku kognitivnog sistema koja nije koherentna ako se posmatra iz pojedinačne perspektive ma koje od pozicija iz spektra teorija racionalnosti i ograničene racionalnosti. Ono što nužno sledi jeste zahtev za promenom paradigme.

## 8.1 Problem adekvatnosti normativnog okvira

Pitanje *adekvatnosti normativnog okvira* u odnosu na koji se donose sudovi o racionalnosti ljudskog kognitivnog sistema je jedno od klasičnih pitanja u debati o racionalnosti. Ovu problematiku smo već predstavili prilikom uvođenja metateorijskog okvira za analizu, na kraju II dela ove rasprave. Da li se problematika menja posle našeg kritičkog pregleda rezultata i argumenata iz više oblasti istraživanja?

*Problem selekcije cilja kognitivnog izračunavanja.* Prema Andersonu, videli smo, specifikacija kompjutacionih ciljeva koje kognitivni sistem treba da ispuni je od najvećeg značaja za racionalnu analizu. Tek kada znamo *šta* pokušava kognitivni sistem da uradi, tj. koji kompjutacioni problem rešava, znamo i kako da pristupimo problemu specifikacije jednačina koje povezuju parametre tog problema sa optimalnom bihejvioralnom funkcijom.

U celini, rešenje zahteva još i specifikaciju *resursa* koji su na raspolaganju kognitivnom sistemu, što je korak koji izostaje u skoro svakoj racionalnoj analizi diskutovanoj u prethodnom pregledu. Sajmonovo inicijalno shvatanje ograničene racionalnosti, videli smo, suštinski je počivalo na ovom uvidu. Specifikacija kompjutacionih resursa kognitivnog sistema je notorno težak zadatak. Naše poznavanje kompjutacionih moći centralnog nervnog sistema, koji po pretpostavci implementira algoritme za rešavanje problema adaptacije, je na veoma niskom nivou. Oceniti koliko je za takav sistem kompleksna primena određenog rešenja adaptivnih problema je isključivo u domenu aproksimacije. Međutim, ni takve aproksimacije se ne sprovode u naučnoj praksi racionalne analize. Više puta tokom našeg pregleda debate o racionalnosti mogli smo da se uverimo da pristalice racionalnih teorija ne pridaju mnogo pažnje ovom problemu, implicitno pretpostavljajući da je kognitivni sistem u stanju da iznese ma koji kompjutacioni pritisak koji bi mogli da izvrše nekada više, nekada manje složeni modeli racionalne bejzijanske inferencije. Retki autori (kao Maloney, 2002) primećuju da složeni bejzijanski algoritmi često mogu da predstavljaju netrivijalne kompjutacione probleme.

Vratimo se pitanju selekcije ciljeva izračunavanja. Kroz najmanje dve diskusije u pregledu debate o racionalnosti uverili smo se u sledeće: *kompjutacioni ciljevi kognitivnog sistema realno ne moraju da budu, i verovatno nisu, jedinstveno određeni.* Podsetimo se naše diskusije problema kauzalnog učenja. Pred kognitivnim sistemom se uvek nalazi jedna ista informacija: informacija o kovariranju

varijabli. To je definicija Hjumovog problema kauzalne indukcije. Kako na osnovu puke korelacije između promenljivih kognitivni sistem odlučuje o tome koja od njih je indikator pravog kauzalnog odnosa, a koja tek kovarijacija? Racionalna analiza zahteva od nas da sprovedemo podrobnu analizu strukture okoline u kojoj se ponašanje odvija pre nego što specifikujemo teoriju o tome kako kognitivni sistem rešava neku klasu problema, i ta analiza strukture okoline nam u slučaju kauzalnog učenja nedvosmisleno govori sledeće: neke kovariacije u okolini predstavljaju puke kovarijacije, dok neke od njih predstavljaju trag stvarnih kauzalnih odnosa. Kognitivni sistem, dakle, mora da ima neku hipotezu o potencijalnim uzročno-posledičnim odnosima *unapred definisanu* da bi u odnosu prema detekciji određene kovarijacije odlučio da li treba da izračunava kauzalne moći, kako sugeriše normativna teorija kauzalnih modela, ili treba da izračunava kovarijacije, kako to sugerišu (opet normativne!) teorije izračunavanja kovarijacije. Kognitivni sistem neće uvek imati formiranu takvu hipotezu: šta više, Hjumov problem kauzalne indukcije potpuno gubi na značaju iz perspektive savremenih teorija kauzalnosti *ako je takva hipoteza unapred data*. To onda više nije problem kauzalne indukcije. Ono što se nameće posle ovakve analize je sledeće: kognitivni sistem, sasvim racionalno, u pogledu izvora kovarijacije u svojoj okolini, može da postavi *uporedo dva kompjutaciona cilja*, izračunavanja snage kauzalnog odnosa i izračunavanja kovarijacije. Sistem koji ne zna unapred koji od dva problema treba da rešava racionalno postupa ako pokuša da reši oba, i na osnovu ma kojih drugih informacija do kojih može da dođe - npr. kauzalnih intervencija, odn. eksperimentisanja sa fenomenima, u slučaju kauzalnog učenja - *a posteriori* odluči da je li izvor kovarijacije indikator pravog kauzalnog odnosa, ili puke kovarijacije. Podsetimo se da odluka o tome koji je normativni okvir adekvatan varira i sa manje apstraknim problemima od ovog. Dok je mera probabilističkog kontrasta, kao normativna mera kovarijacije, normativna i kao mera snage kauzalnog odnosa u slučaju da se ona ocenjuje u kontekstu drugih izvora kauzalnosti, mera kauzalne moći je normativna samo u kontekstu ocene snage kauzalnog uticaja izolovanog uzroka. Da li kognitivni sistem uvek zna *unapred* koja od ove dve situacije je relevantna za adaptaciju? Ako ne, racionalna strategija koju on može da implementira jeste izračunavanje obe mere uporedo. Uzimajući u obzir da teorija kauzalnih modela supsumira i teoriju kovarijacije (svaki sistem koji je izračunao kauzalnu moć izračunao je i probabilistički kontrast), ne deluje kao da bi rešavanje ova dva problema adaptacije istovremeno za kognitivni sistem bilo mnogo više zahtevno od

njihovih nezavisnih tretmana.

Vratimo se diskusiji odlučivanja u uslovima rizika i neizvesnosti. Normativna teorija očekivane korisnosti predstavlja optimalan odgovor kognitivnog sistema koji je suočen sa objektivnim distribucijama verovatnoća potencijalnih dobitaka i gubitaka. Videli smo da modeli ograničene racionalnosti, poput RDU modela odn. kumulativne teorije izgleda kao njihove najmoćnije implementacije, objašnjavaju bihejvioralne nalaze bolje od teorije očekivane korisnosti (mada ni iz daleka onoliko bolje može da se stekne utisak ako se proučava samo naučna periodika psihologije, ne uzimajući u obzir rezultate do kojih dolaze eksperimentalni ekonomisti). Međutim, teorija ograničene racionalnosti poput teorije izgleda nema dobru korespondenciju sa konceptima racionalne analize. Na primer, ona uopšte ne daje odgovor na pitanje šta je cilj izračunavanja, tj. koji to problem rešava kognitivni sistem koji menja model očekivane korisnosti modelom teorije izgleda? U čemu je njegova dobit? Čime je motivisano ponašanje koje odstupa od normativnih standarda? Vratićemo se, u V i VI delu naše teze, ovim problemima vezanim za teoriju izgleda. Sada treba obratiti pažnju na sledeće: i u slučaju odlučivanja u uslovima rizika kognitivni sistem može legitimno i racionalno pred sebe da postavi više ciljeva izračunavanja uporedo. Normativna teorija očekivane korisnosti daje normativno očekivane rezultate u situaciji za koju su je fon Nojman i Morgenštern i projektovali, a to je situacija u kojoj su distribucije verovatnoća dobitaka (i gubitaka, da trivijalno proširimo klasičnu EU) *objektivne*. Sevidžov subjektivistički okvir dopušta da donosioci odluka imaju svoja subjektivna uverenja o tome kakve su te distribucije verovatnoća. Pretpostavimo sledeće: donosilac odluka živi u nekoj okolini koja se empirijski bar kvalitativno poklapa sa našim realnim ekonomskim okruženjem, u kome je *a priori* manja verovatnoća zarade 500 dinara nego 400 dinara, 400 dinara nego 300 dinara, i tako dalje. Ovakva ekonomska okruženja određuju distribucije bogatstva koje su ekonomistima dobro poznate, poput Paretove distribucije koja predstavlja univerzalni nalaz o funkciji stepenog opadanja verovatnoće da neka individua u nekom ekonomskom sistemu raspolaže određenim resursima[67]. Veoma mali broj individua u realnim ekonomskim sistemima raspolaže veoma velikim udelom kapitala, i obrnuto: veoma veliki broj individua deli ono što preostaje. U takvom okruženju, racionalno za kognitivni sistem koji odlučuje o izboru između neizvesnih monetarnih dobitaka i gubitaka jeste da kroz svoja subjektivna verovanja inkorporira sliku te okoline u svoje odluke. Međutim, situacija se menja ako neko ima razloga da veruje da mu je *sa objektivnim verovatnoćama* ponuđen loz koji sa 5% donosi

gubitak 5 dinara, a sa 95% milion dolara. Ponuda ovakvog loza je, imajući na umu ekološku relevantnost, sumanuta: ko bi ponudio takvu igru (neko mora da isplati milion dolara u 95% slučajeva)? Donosilac odluka, suočen sa ovakvim problemom, sa pravom bi postavio pitanje da li je u pitanju neka prevara ili šala: svoje čuđenje bi mogao da izrazi inkorporacijom prethodnih verovanja o pravilnostima ekonomskog okruženja u odluke koje donosi. Pretpostavimo da kognitivni sistem *ne zna* da li se distribucije verovatnoća koje se nalaze pred njim povinuju regularnostima okoline koju je imao prilike da nauči, ili predstavlju *prima facie* objektivne verovatnoće koje nije potrebno korigovati. Koja strategija je za onda njega racionalna? Naravno: ona koja uporedo ispunjava dva kompjutaciona cilja, jedan pod pretpostavkom da su verovatnoće objektivno date, i drugi pod pretpostavkom da te verovatnoće treba subjektivno korigovati u odnosu na prethodno iskustvo u relevantnoj sredini. Koji normativni okvir je onda relevantan? Nijedan, i oba: i onaj koji propisuje odlučivanje u skladu sa objektivnim verovatnoćama, i onaj koji dopušta subjektivne verovatnoće odn. inkorporaciju prethodnog iskustva u odluke. Ponovo kao u slučaju kauzalnog učenja, između Sevidžove teorije i fon Nojman-Morgnešternove teorije postoji jasan kontinuitet: Sevidžova subjektivistička teorija supsumira teoriju očekivane korisnosti. Plauzibilna pretpostavka da slični kompjutacioni problemi zahtevaju slične algoritme čini da opet možemo da zaključimo kako pritisak na kompjutacione resurse sistema ne bi mnogo porastao ako bi sistem pokušao da reši dva („objektivistički“ i „subjektivistički“) problema odlučivanja uporedo. Kakav je položaj deskriptivne teorije izgleda Tverskog i Kanemana u odnosu na mogućnost rešavanja više problema odlučivanja paralelno manje je jasno. U V delu naše rasprave u potpunosti ćemo se posvetiti ovom problemu.

Dva važna zaključka se nameću posle ove analize. Prvi, sasvim očigledan i implicitno prisutan u debati o racionalnosti, jeste taj da izbor normativnog okvira jedinstveno vrši i izbor cilja izračunavanja, koji onda usmerava celokupnu racionalnu analizu neke kognitivne funkcije. Drugi, neočekivan zaključak u odnosu na prethodni istorijat debate o racionalnosti, jeste *relativizacija pitanja adekvatnosti* normativnih okvira. Ako kognitivni sistem racionalno postupa pokušavajući da reši više kompjutacionih ciljeva paralelno, koji od normativnih okvira je zaista normativan? Odgovor upućuje na to da sam pojam *normativne adekvatnosti u ovom kontekstu gubi smisao*: normativni su svi okviri koji specifikuju rešavanje pojedinačnog cilja, a u celini nije normativan nijedan po sebi. Šta onda, ako išta, predstavlja normativni okvir za funkcije kognitivnog sistema? Da li je to nekakva

optimalna kombinacija više postojećih normativnih okvira[68]? Može li ona da ima jasno teorijsko utemeljenje? Ova pitanja ostavljamo za VI deo naše rasprave, gde ćemo pokazati da je odgovor na njih jedno uslovno „da“, koje vodi ka temeljnom promišljanju odnosa ma koje norme i funkcija kognitivnog sistema u odnosu na adaptivne probleme koje one rešavaju.

*Značenje normativnog: racionalnosti, adaptacija i optimizacija.* Debata o racionalnosti motivisala je više pokušaja da se odstupanja od normativnih kriterijuma kojima vode kognitivne funkcije čoveka opravdaju promenom samih normativnih kriterijuma. Prethodne stranice uverile su nas da racionalne bejzijanske teorije po pravilu posežu za ovakvim argumentom. Oaksfordova i Čaterova analiza racionalnosti u zadataku selekcije, na primer, uopšte nije moguća ako se problem ljudskog rezonovanja ne formuliše u probabilističkoj formi. Njihova racionalna analiza u odnosu na pitanje odstupanja ljudskog mišljenja od normativnih pravila logike zapravo nema drugi odgovor do odgovora koji glasi: ljudski kognitivni sistem uopšte ne rešava predstavljeni problem kao problem koji odgovara strukturi normativne logičke forme. Međutim problem je *de facto* postavljen kao logički problem. Sve što racionalna analiza poput ove može da nam saopšti jeste da kognitivni sistem ne izgleda toliko iracionalno koliko izgleda *ako se pretpostavi* da on pokušava da reši neki drugi problem. Očigledno, ovakvoj tvrdnji je potrebno jako opravdanje.

Osnovna strategija opravdavanja objašnjenja po kome kognitivni sistem, suočen sa zadacima određene forme, u stvari rešava neke druge zadatke, *zbog čega* se osmotreno ponašanje interpretira kao normativno neadekvatno, jeste pozivanje na argument prema kome kognitivni sistem u svakoj formalnoj strukturi problema prepoznaje *strukturu adaptivno relevantnih problema* za koje je „podešen“ da ih normativno rešava - ali u odnosu na normativne kriterijume za rešavanje tih „interpretiranih“, „latentnih“ problema. Samo ovaj argument opravdava objašnjenja koja daju Oaksford i Čater u oblasti rezonovanja (koja postaje oblast probabilističkog rezonovanja posle racionalne analize) ili ona koja daju Tenenbaum i Grifits u diskusiji reprezentativnosti (koja postaje diskusija testiranja hipoteza posle racionalne analize).

Tri koncepta su povezana u složenu mrežu odnosa u opravdanju racionalne analize. Koncept normativnog, posle racionalne analize, prelazi sa suprapersonalnog nivoa - nivoa na kome nalazimo formalne sisteme logike i teorije verovatnoće - na personalni, ili subpersonalni nivo analize. Na personalnom ili subpersonalnom

nivou analize, normativni kriterijum koji se odnosi na formalni opis problema koji kognitivni sistem *treba da reši*, onda zamenjuje neki drugi normativni kriterijum za problem koji kognitivni sistem *aktualno rešava*. Na taj način racionalna analiza zadržava svoju „racionalnost" - to je njena RACIONALNOST$_1$. Racionalnost kognitivnog sistema ili subjekta (zavisi da li želimo da govorimo na subpersonalnom, ili na personalnom nivou analize), odn. RACIONALNOST$_2$, onda prolazi kroz *proces restituisanja*, pokazivanjem da model odgovarajuće kognitivne funkcije rešava alternativni problem na normativno adekvatan način tako da se osmotreno ponašanje, što je moguće potpunije, poklapa sa predikcijama tog modela. Drugi koncept, koncept adaptacije, je koncept koji celom postupku opravdanja racionalne analize daje semantiku: on ga suštinski opravdava, što je neophodno - jer jedan normativni kriterijum, videli smo, mora nekako da „zameni" drugi. Očigledno, „zamena zadataka", teorijski potez koji vodi ka analizi formalno drugačije strukture problema od one koju definišu eksperimentalni nacrt i motivacija originalnog problema, mora da bude opravdana ne samo činjenicom da objašnjava osmotrene podatke (ako ih objašnjava) - jer onda bismo mogli ma koje osmotreno ponašanje da objašnjavamo ma kojim modelom koji ga statistički fituje - već i nekim faktorom koji *teorijski* motiviše samu zamenu. Korak racionalne analize u kome se zahteva analiza strukture okoline u kojoj se problem rešava je korak koji je inherentno povezan sa ulogom koju koncept adaptacije igra u opravdavanju eksplanatorne strategije celog postupka.

Kognitivni sistem, kao deo nekog organizma koji proučavamo, po pretpostavci mora da odredi ponašanje tog organizma da bi ga maksimalno prilagodio sredini, poštujući određena (energetska i informaciona) ograničenja. Koja tačno adaptacija je relevantna, određuje struktura okoline. Kod Oaksforda i Čatera, u slučaju analize zadatka selekcije, strukturu sredine opisuje hipoteza o retkosti. Već smo diskutovali ovu hipotezu i shvatili da je ona povezana sa savremenim, normativnim kriterijumima u probabilističkom rezonovanju. Hipoteza o retkosti, videli smo, može da bude zadovoljena u odnosu na opravdanje koje je ponuđeno za nju, ali se relativno lako nalaze i primeri koji je opovrgavaju. Sledeće zapažanje je veoma važno za našu diskusiju: za ma koji racionalni model koji inkorporira informacije o strukturi okoline i opravdava svoju normativnu adekvatnost preko koncepta adaptacije toj okolini, *uvek je moguće konstruisati okolinu u kojoj je on na taj način racionalan.* Jednostavno, svaka zamisliva struktura okoline implicira određenu klasu sistema koji su u njoj adaptirani, i obrnuto: za ma koju klasu sistema,

uvek je moguće „dizajnirati“ sredinu u kojoj su ti sistemi adaptivni. Tako smo u diskusiji epizodičke memorije pokazali da racionalna analiza Andersona i saradnika postaje istovremeno i teorija pamćenja i teorija informacija koje se pamte. Drugi pristup epizodičkoj memoriji Andersona i saradnika, pristup koji razvija teoriju o mehanizmima memorije preko analogije sa optimalnim veštačkim sistemima za menadžement podacima, ipak je samo još jedna mehanicistička kognitivna teorija koja obiluje slobodnim parametrima i *ad hoc* hipotezama o tome kako „funkcionišu neopservabilne stvari u glavi“. Može se reći da sam Anderson u ovom slučaju nije izbegao zamku za koju smatra da je racionalna analiza izbegava. Posle svega, ostaje sledeći zaključak: uz dovoljnu fleksibilnost hipotetskih konstrukata koje koristi teorija, i pogodne hipoteze o strukturi okoline u kojoj se ponašanje analizira, *uvek* može da se ostvari perfektna simetrija između te okoline i tog ponašanja. Adaptacija kognitivnog sistema tako postaje *koncept koji je potrebno opravdati*, dok je njena pretpostavljena uloga da bude *koncept koji opravdava* postupak racionalne analize.

Treći ključan koncept u mreži opravdanja racionalne analize je koncept *optimizacije*. Optimizacija je, jednostavno, proces kojim se rešava neki problem, u odnosu na neki kriterijum, *pod određenim ograničenjima resursa za rešavanje tog problema*[69]. Za data ograničenja u kompjutacionoj moći, određeni kompjutacioni cilj i određenu strukturu okoline u kojoj se on ispunjava, reći da kognitivni sistem optimizuje ponašanje znači reći da on rešava zadati problem uzimajući u obzir svoja kompjutaciona ograničenja i ograničenja koja nameće struktura okoline. Matematički, formalno posmatrano, problemi optimizacije su *izuzetno teški* problemi. Veći broj autora u debati o racionalnosti, na čelu sa Gigerencerom kao verovatno najglasnijim zastupnikom teze, tvrdi da je nemoguće da kognitivni sistem čoveka rešava adaptivne probleme dosledno primenjujući optimizacione algoritme. Čateru i Oaksfordu ovo ne smeta; oni objašnjavaju da se u racionalnoj analizi ni ne tvrdi da kognitivni sistem zaista izračunava optimalna ponašanja, već samo da na kompjutacionom nivou analize njegov proizvod - ponašanje - izgleda *kao da* makar aproksimativno optimizuje relevantne probleme. Ovi autori su spremni da teške poslove optimizacije prepuste heurističkim procedurama kojima je proces evolucije obdario kognitivni sistem (up. Chater & Oaksford, 2000; na stranu što su matematički problemi optimizacije toliko složeni da je većina praktično primenjljivih metoda u matematici - takođe heurističke prirode, tj. dolazi bez garancije da će za svaki zadati problem uspeti da izračuna najoptimalnije rešenje). U diskusiji odnosa racionalnosti i optimizacije, međutim, Čater i Oaksford se brzo zapliću u

cirkularna određenja: „*Optimalnost nije isto što i racionalnost.* [...] *Odluka o tome da li je određeno ponašanje racionalno ili ne zato zahteva više nego samo mogućnost da se za njega ponudi objašnjenje u terminima optimizacije. Tako racionalnost zahteva ne samo optimizaciju već optimizaciju nečega što je razumno.* [...] *ovo je očigledno cirkularno. Ali ako posmatramo racionalnost u terminima optimizacije, opšte koncepcije o tome šta su razumni kognitivni ciljevi mogu da se pretoče u specifične i detaljne kognitivne modele. Tako, program racionalne analize, ne dajući odgovor na konačno pitanje toga šta racionalnost jeste, u svakom slučaju pruža osnove za konkretnu i potencijalno plodnu liniju empirijskog istraživanja*“ (citirano prema Chater & Oaksford, 2000, naš prevod). Po našem mišljenju, program racionalne analize je nesumnjivo razvio empirijski plodnu paradigmu istraživanja, ali se suočava se većim problemom od toga što ne daje odgovor na pitanje šta je racionalno.

Prethodno diskutovan odnos koncepata racionalnosti i adaptacije ukazuje na to da adaptacija, jednom kada se definiše preko veze sredine sa ciljevima izračunavanja, „vaskrsava“ koncept racionalnost na personalnom ili subpersonalnom nivou analize. Već smo prethodno prepoznali kako se ta restitucija racionalnosti odigrava: jednom kada je dogovor o tome koji je alternativni, latentni, „pravi cilj“ kognitivnog sistema, dovoljno je pokazati da racionalni model razvijen u odnosu na taj cilj statistički dobro fituje bihejvioralne podatke. Uzimajući u obzir sada činjenicu da je, uz dovoljnu dovitljvost i kreativnost u razvoju hipoteza o sredini na koju se treba adaptirati, uvek moguće „dizajnirati“ sredinu kojoj odgovaraju određene adaptacije, jasno je da koncept optimizacije - proces koji fundamentalno povezuje kompjutacione ciljeve, adaptaciju i sredinu - *igra centralnu ulogu u određenju svega što je „racionalno“ u racionalnoj analizi.* Strategija racionalne analize više liči na potragu za odgovarajućom sredinom i odgovarajućim ciljevima da bi se zadovoljio *kriterijum optimalnosti* - koji se onda, kroz koncept adaptacije, izgubljen u korespondenciji sredine i ciljeva, projektuje kao koncept restituisane, adaptivne racionalnosti. Čater i Oaksford u izuzetnom teorijskom radu u filozofskom časopisu „*Synthese*“ slute da nešto sa celom teorijskom konstrukcijom nije na mestu, i ta slutnja se prepoznaje u redovima iz tog rada koje smo prethodno citirali (Chater & Oaksford, 2000). To nešto što ozbiljno nije u redu sa metodologijom racionalne analize, videli smo, nije lako izolovati u komplikovanoj teorijskoj strukturi ove metodološke strategije, ali se konačno svodi na činjenjicu da *ceo postupak nije dovoljno determinisan*: previše *ad hoc* odluka se nalazi u svim racionalnim

analizama, previše hipoteza koje obezbeđuju da optimizacioni procesi - procesi kojima je svejedno šta optimizuju - uspešno povežu sredinu i ciljeve u objašnjenju bihejvioralnih podataka. Činjenica da je neki racionalni model moguće optimizovati u odnosu na neki zadatak je sav sadržaj koji koncept racionalnosti saznanja dobija u racionalnoj analizi. Činjenica da će *uvek* postojati model koji će optimizovati vezu između neke sredine i nekih kompjutacionih ciljeva zahteva još samo da se pronađu adekvatna opravdanja za izbor te sredine i tih ciljeva kako bi model postao racionalan. Da racionalna analiza, vođena koracima koje preporučuje Anderson, omogućava nedvosmislenu selekciju ciljeva i analizu strukture relevantne okoline, nikada se ne bismo suočili sa problemom postojanja nekoliko racionalnih analiza *istih* kognitivnih funkcija, od kojih sve pretenduju na normativnu adekvatnost. Oblast kauzalnog učenja, u kojoj se koriste skoro *isključivo racionalni modeli bez slobodnih parametara* - što nas oslobađa potencijalne konfundacije u argumentu jer tu očigledno višestrukost modela *nije* posledica proliferacije hipotetskih konstrukata - pruža najdirektniji dokaz za to da racionalna analiza nije u stanju da nedvosmisleno reši problem koji postavlja. U oblasti rezonovanja, videli smo, odstupanja od norme takođe mogu da se racionalizuju na više načina: aktivacijom pragmatskih shema, darvinijanskih algoritama ili izborom optimalnog eksperimenta pod hipotezom o retkosti koju predlažu Oaksford i Čater. U oblasti teorije odlučivanja, višestrukost objašnjenja ključnih eksperimentalnih podataka je još šira nego u oblastima proučavanja kauzalnosti i rezonovanja: pođimo samo od teorija inspirisanih radom Dejvida Bela o razočarenju i žaljenju kao faktorima odlučivanja (Bell, 1988) ili Birnbaumovih modela transfera pažnje (Birnbaum, 2011), koji objašnjavaju sva robustna odstupanja od normativne racionalnosti - pozivanjem na potpuno različite kognitivne mehanizme.

Na ovom mestu je potrebno povući jasnu distinkciju između prethodnih kritika panglosijanskih „racionalizacija“ grešaka i inklinacija koje otkriva empirijsko istraživanje viših i simboličkih kognitivnih procesa od argumenata iznetih ovde. Stain (Stein, 1996, prema Stanovich, West & Toplak, 2011) zaključuje da po pretpostavkama panglosijanaca - u koje se praktičari savremenih racionalnih analiza svakako ubrajaju - samo empirijsko proučavanje nekog kognitivnog zadatka kao da otkriva normu prema kojoj on treba da se obavi. Ovaj rezon je u potpunosti primenljiv na sve racionalne analize koje smo mi diskutovali: njega otkriva proces zamene *prima facie* strukture kognitivnog zadatka strukturom *latentnog zadatka* koji se onda opravdava kao problem adaptacije koji kognitivni sistem *treba* da reši.

Naš argument, očigledno, govori nešto drugo: da je opravdavanje takvog poteza pozivanjem na adaptivnu vrednost latentnog zadatka *uvek cirkularno*. O tome svedoči višestrukost mogućih kognitivnih ciljeva - te samim tim, i mogućih adaptacija - koje se otkrivaju pri pokušaju racionalne analize *istih* kognitivnih zadataka[70]. Da će neka adaptacija odgovarati nekoj sredini u odnosu na neki cilj je tautologija, a u ponovljenim potragama za takvim tripletom sredina-adaptacija-cilj metodologija racionalne analize demonstrira da se iz tautologije može izvesti bilo šta.

Jedna važna opomena će se naći na kraju ovog razmatranja o odnosu optimizacije, adaptacije i racionalnosti. Naime, savremena evoluciona biologija, nauka koja svim društvenim naukama pa i psihologiji otvara pristup adaptacionističkoj eksplanatornoj paradigmi, nije bez ozbiljnih kritika te paradigme (up. Gould & Lewontin, 1979). Ideja da evolucioni procesi predstavljaju optimizacione procese je pretpostavka za koju se relativno lako pokazuje da je ne odlikuje nužnost. Ako uporedimo sa ovakvim kritikama ono što o ideji objašnjenja putem dinamičkih sistema, matematičkom formalizmu koji se savršeno uklapa sa adaptacionističkim objašnjenjem, pišu uticajni savremenici poput Volframa (Wolfram, 2002), videćemo da cela teorijska konstrukcija koja se proteže kroz evolucionu biologiju ka svim naukama koje „pozajmljuju“ njene eksplanatorne strukture, može da se uzdrma. Volfram daje ubedljive ilustracije sistema čije ponašanje nije uopšte intuitivno moguće razumeti kao vid „*zadovoljavanja ograničenja*“ (engl. *constraint satisfaction*, što je samo drugo ime za optimizaciju), na stranu nivo kompleksnosti umešan u dinamički opis mnogih čak intuitivno veoma jednostavnih prirodnih sistema. Njegova sugestija je da ćemo ponašanje kompleksnih sistema bolje razumeti sa stanovišta koje podrazumeva da su oni sami, inherentno, generatori slučajnih procesa, procesa koji kroz interakcije sa okolinom rezultiraju u obostranim adaptacijama, ali koji nisu nužno opisivi kao optimizacije ma kakve vrste.

## 8.2 Formalni modeli i bihejvioralni podaci

Problem odnosa realne kompleksnosti ponašanja i hipotetskih konstrukata na koje se oslanja neka kognitivna teorija, i sa njima povezan problem selekcije modela, već smo razmatrali u sekciji 6.1. Preostaje da donesemo naš sud o ovom problemu posle pregleda debate o racionalnosti viših i simboličkih kognitivnih funkcija koju smo predstavili.

U oblasti odlučivanja u uslovima rizika i neizvesnosti, koja proučava formalno

najjednostavniji problem u debati o racionalnosti, sve teorije koje smo diskutovali su čvrsto povezane sa bihejvioralnim podacima. Ta veza je ostvarena kroz same formalne konstrukcije ovih teorija. Sve teorije koje mogu da se realizuju u zajedničkom aksiomatskom okviru (kroz hijerarhiju uslova konzistencije otkupa), u rasponu od očekivane korisnosti do kumulativne teorije izgleda, omogućavaju da svi njihovi subjektivni parametri budu ocenjeni direktno iz opservabilnih izbora subjekata. To znači da su te teorije *direktno podložne neparamterijskoj oceni*: ako ne želimo da se obavežemo na rad sa nekom određenom familijom funkcija korisnosti, na primer, možemo neparametrijskim postupkom direktno da ocenimo gde leže tačke subjektivne korisnosti nekog donosioca odluka, koristeći samo posmatranja njegovih izbora (u odgovarajućem eksperimentalnom nacrtu, naravno). Ovo važi i za funkciju ponderisanja verovatnoća u teoriji izgleda, kao i za averziju prema gubicima (u nacrtima koji uključuju mešovite lozove). Videli smo da ni ova, veoma povoljna osobina teorija odlučivanja, ne pomaže mnogo u selekciji „prave teorije odlučivanja".

Problemi kauzalnog učenja i kauzalnog rezonovanja predstavljaju onoliko formalno komplikovanu problematiku koliko je komplikovana mreža kauzalnih odnosa koja se analizira. Pod normativnim uslovima koje je opisala Čengova za ocenu kauzalne moći jednog uzroka, a proširile Novikova i Čengova na slučaj jednostavnih kauzalnih interakcija, teorija kauzalnih modela jeste teorija koja obezbeđuje direktnu inferenciju subjektivnih verovanja o kauzalnoj moći. Ovde govorimo o teoriji *bez slobodnih parametara*, tako teorija kauzalne moći ni u principu ne može da inkorporira individualne razlike u oceni snage kauzalnih odnosa. Za razliku od teorija odlučivanja kod kojih su formalne konstrukcije precizno i bez ostatka povezane sa objektima merenja - jer izbori ispitanika predstavljaju direktnu implementaciju relacije preferencije - u oblasti kauzalnog učenja se koriste bihejvioralne mere koje nisu temeljno povezane sa objektima merenja, poput procena na skalama Likertovog tipa, ili direktnih numeričkih procena. Verujemo da bi teorije kauzalne indukcije mogle značajno da uznapreduju u nekom alternativnom aksiomatskom okviru koji bi ih *direktno* povezao sa opservabilnim podacima. Jasno je da bi rezultirajući modeli bila neka vrsta inkorporacije teorija kauzalne indukcije u teoriju odlučivanja, pošto ova druga već jasno povezuje svoje koncepte sa metodološkim procedurama. Priroda bihejvioralne metodologije je takva da je odlučivanje uvek najpogodniji proces za njenu operacionalizaciju: u krajnjoj liniji, sve metodologije karakteristične za kognitivnu psihologiju svode se na posmatranje odluka koje donose ispitanici (odluka o podeoku koji treba zaokružiti na skali

Likertovog tipa; odluka o trenutku u kome treba pritisnuti taster u procedurama merenja reakcionog vremena; odluke o redosledu u postupku rangiranja; odluke o kategorijalnoj pripadnosti u postupku sortiranja, itd).

Konstrukte teorija suđenja i rezonovanja, dok se one nalaze na kompjutacionom nivou analize, nije teško povezati sa opservabilnim podacima, zahvaljujući relativnoj jednostavnosti struktura podataka koje ispitanici daju u karakterističnim metodološkim procedurama u ovim oblastima. Heurističke procedure je teže povezati direktno sa opservabilnim podacima, pošto heuristici po pravilu podrazumevaju elaboraciju određenog internog algortima koji kognitivni sistem koristi - njih nalazimo na Marovom drugom nivou analize, koji zahteva mnogo dublji pogled u crnu kutiju od onog koji obezbeđuju bihejvioralne procedure. Teorija podrške (engl. *Support Theory*), koju su razvili Tverski i Keler (Tversky & Koehler, 1994) rešavajući problem reprezentacije subjektivnih verovatnoća da bi objasnili fenomene poput greške konjukcije, obezbeđuje direktne inferencije o subjektivnim sudovima na osnovu opservabilnih izbora. Ova teorija ima donekle jednostavniju formalnu strukturu od teorije izgleda, ali u naučnoj periodici nažalost ne nalazimo više interesovanja za nju, iako predstavlja direktno formalno otelotvorenje ideja o psihološkom procesu suđenja razvijenih u programu heuristika i inklinacija Tverskog i Kanemana.

Teorije epizodičkog pamćenja i konceptualne organizacije uopšte nema smisla diskutovati u odnosu na problem njihovog direktnog povezivanja sa bihejvioralnim podacima. Sve takve hipoteze koriste se kao manje formalna ograničenja u motivaciji izgradnje egzaktnijih, preciznijih modela sa većim brojem neopservabilnih parametara. Pod pretpostavkom da je u oblasti kategorizacije moguće doći do jedinstvenog normativnog okvira, pitanje je tačno kako bi on bio povezan sa bihejvioralnim podacima zbog problema sličnih onome koji smo diskutovali u sekciji 6.2 (o falsifikabilnosti određenih delova kompjutacionističkog programa). Svaki semantički reprezentacioni sistem, lako je to videti, nasleđuje probleme koje postavlja pitanje odnosa smisla i reference, problem čija predložena rešenja obuhvataju hipoteze koje je nemoguće testirati bihejvioralnim testovima. Ovo je posledica toga što je jedina opservabilna relacija ona između simbola (denotatora - reči, stimulusa) i denotata, dok je veza između denotatora i koncepta - ključnog hipotetskog konstrukta u ovim teorijama - neopservabilna (i arbitrarna). Uzimajući u obzir (a) konvencionalni karakter *svih* ovih veza i (b) činjenicu da isti odnos znak-označeno može da bude realizovan preko arbitranog broja neopservabilnih

konceptualnih, semantičkih reprezentacija, zaključujemo da bihejvioralni testovi nikad neće moći u potpunosti da omoguće testiranje hipoteza o pretpostavljenoj trijadnoj relaciji znak-koncept-označeno. Detalje analize upravo iznete tvrdnje ostavljamo za VI deo rasprave.

*O značaju neopservabilnosti kognitivnih funkcija uopšte.* Prethodno zapažanje o fundamentalnoj ograničenosti moći bihejvioralnih testova u odnosu na otkrivanje struktura semantičkih reprezentacija uvodi veoma ozbiljan metodološki problem u analizu racionalnosti saznanja. Ako pođemo od ograničenja da jedni kognitivni sistemi o stanjima drugih kognitivnih sistema saznaju samo na osnovu bihejvioralnih podataka, zaključujemo da je prirodna selekcija favorizovala one kognitivne sisteme koji onemogućavaju lak pristup sopstvenim kognitivnim funkcijama. Kognitivna psihologija ima dosta toga da kaže o ljudskom rezonovanju i suđenju u laboratorijskim uslovima, gde ispitanici pristaju da kroz serije bihejvioralnih odluka izraze neka svoja neopservabilna stanja. Međutim, kognitivna psihologija nema skoro ništa da kaže o čoveku koji naboranog čela sedi na klupi u parku i pokušava da reši neki težak matematički problem, ili ženi zagledanoj u more koja na obali razmišlja o tome kako da reši neka za nju važna, lična pitanja. Naravno, umovi ne saznaju jedni o drugima samo na osnovu bihejvioralnih podataka, već i na osnovu elaboriranih teorija o drugim umovima po modelu folk-psihologije, i informacijama koje kooperativno razmenjuju. S vremena na vreme, ljudski kognitivni sistemi namerno odašilju pogrešne informacije o sopstvenim internim, neopservabilnim stanjima; kolokvijalno rečeno, lažu. Tek u odnosu na ovaj, potpuno simbolički kontekst analize kognitivnih funkcija, vidimo u kojoj meri je kognitivni sistem čoveka evoluciono skrojen tako da može da *umanji predvidljivost svog ponašanja*, ili čak namerno navede druge, slične sisteme u svojoj okolini na pogrešne predikcije. Kako se to postavlja kao problem simboličke kooperacije između ljudi u svakodnevnici, kao *problem poverenja*, tako to predstavlja i bitno metodološko ograničenje kognitivne psihologije, pa i psihologije kao nauke uopšte. Želimo da budem potpuno precizni u određenju ovog, čini nam se, veoma važnog problema. Ovde nije u pitanju problem socijalne prirode koji bi mogao da podrazumeva nekooperativnost subjekata kognitivne psihologije u eksperimentalnim situacijama - npr. svesno, namerno davanje slučajnih, pogrešnih ili nevalidnih odgovora u bihejvioralnim eksperimentima. Ove govorimo o suštinskom pitanju u kojoj meri je prirodni dizajn ljudskog kognitivnog sistema razvijen tako da ga zaštiti od uvida u njegovo funkcionisanje, odn. uvida koji bi omogućio predikciju njegovog ponašanja.

U VI delu ove teze, problem će postati precizniji, kada u analizu uvedemo neke bitne koncepte teorije igara. Teorija igara savršeno predviđa koncept *racionalnog subjekta koji minimizuje predvidiljivost sopstvenog ponašanja.* Mi smo za sada ovaj problem identifikovali na simboličkoj ravni uvidom u fundamentalnu ograničenost bihejvioralnih testova da omoguće selekciju između alternativnih semantičkih reprezentacija nekog kognitivnog sistema. Problem koji diskutujemo, naglasimo, nije ni problem ograničenja introspektivne metode koji je diskutovan u istorijski ranim fazama eksperimentalne psihologije. Upravo diskutovan problem nastaje tek kada se jasno ocrtaju granice bihejvioralne metodologije i moći bihejvioralnih testova da obezbede selekciju naših hipoteza o neopservabilnim, hipotetskim konstruktima.

Tokom ljudske evolucije, malo pažnje je, sasvim očigledno, posvećeno pitanju toga da li će njegova stabilna evolutivna forma biti zgodna za naučno proučavanje. Nasuprot tome, postoji jasna evoluciona motivacija da se razvije rešenje u formi kognitivnog sistema koji minimizuje odavanje informacija o sopstvenim funkcijama - radeći tako, nažalost naučnika, protiv mogućnosti sopstvene naučne analize. Nije u pitanju zavera boga i majke prirode protiv nauke, dakle, već činjenica prirodne evolucije da će takvi sistemi biti bolje adaptirani na kompetitivne uslove sredine. Kognitivna psihologija je malo pažnje posvetila pitanju ovog fundamentalnog ograničenja - pitanju u kome se prožimaju njene metodološke osnove sa prirodom sistema koji proučava.

*Racionalna analiza kao strategija redukcije broja mehanicističkih modela.* Predlažući metodologiju racionalne analize, Anderson je kao jednu od njenih prednosti prepoznao to što, jasnim određenjem kompjutacionih ciljeva kognitivnog sistema i analizom okoline u kojoj te ciljeve treba postići, ona omogućava redukciju broja mehanicističkih, algoritamskih modela kognitivnih funkcija (Anderson, 1991). Podsetimo se, u sekciji 6.1 smo razmatrali veoma težak problem čija se suština sastoji u tome što potencijalno ogroman broj mogućih formalnih modela može da realizuje isto ponašanje koje je predmet naših opservacija. *Problem identifikabilnosti*, kako ga naziva Anderson, racionalna analiza rešava tako što umesto da nagađa o neopservabilnim procesima u crnoj kutiji kognitivnog sistema, temeljno proučava okolinu u kojoj on donosi odluke - koja je mnogo lakša za opservaciju i formalni opis. Posle kritičkog pregleda debate o racionalnosti, jasno je da metodologija racionalne analize nije doprinela redukciji broja mogućih modela. Međutim, zanimljiv je način na koji ona to nije uspela da učini. Proliferacija algoritamskih modela kognitivnih funkcija, posebno izražena u oblastima pamćenja i kategorizacije, direktna je

posledica stepena kreativnog napora istraživača u odgovarajućim oblastima, te broja slobodnih parametara i hipotetskih konstrukata u odgovarajućim matematičkim modelima koje je naučna zajednica odlučila da može da istrpi (pokazujući se veoma velikodušnom u tom pogledu). Takvoj igri, poptuno je jasno, kraja nema. S druge strane, mi smo u pregledu debate o racionalnosti već videli da različite teorije o tome kako kognitivni sistem rešava konceptualno iste, jedinstvene probleme, polažu pravo na adaptacionističko opravdanje svojih tvrdnji. Ponovo se srećemo sa problemom selekcije kompjutacionih ciljeva: tačno *šta* treba da izračuna kognitivna mašinerija naših umova se pokazuje kao pitanje sa mnogo više odgovora od broja koji je očekivan. U prethodnih dvadeset godina istraživanja, zaključak se nameće, racionalna analiza nije uspela da pronađe rešenje za problem identifikabilnosti.

## 8.3 Problemi proceduralne i deskriptivne invarijantnosti

Više puta tokom diskusije smo ustanovili da se problemima kršenja proceduralne invarijantnosti ne poklanja dovoljna pažnja. Problemi deskriptivne invarijanosti su u oblasti odlučivanja u uslovima rizika i neizvesnosti dobili donekle adekvatan tretman u teoriji izgleda Kanemana i Tverskog. Funkcije korisnosti sa različitim osobinama (konkavnost i konveksnost) za dobitke i gubitke, uz averziju prema gubicima, objašnjavaju bar osnovne efekte kršenja deskriptivne invarijantnosti. U ovoj oblasti, međutim, efekte zamene preferencija ne može da inkorporira nijedna od standardnih teorija odlučivanja. Ovi efekti, podsetimo se, nastaju kao posledica promene *procedure* ispitivanja izbora: za neke lozove, ponašanje u oceni monetarnih ekvivalenata i izboru između dva loza nije konzistentno.

U oblasti odlučivanja u uslovima rizika postoje teorijski značajni primera kršenja proceduralne invarijantnosti koji nisu diskutovani u naučnoj periodici. U merenju monetarnih ekvivalenata rizičnih lozova, postupak u kome ispitanik za loz tipa ($x$,$p$;$y$,1-$p$) treba da odredi minimalnu cenu po koji bi ga prodao, ili minimalni iznos koji bi platio da se osigura od odigravanja loza koji potencijalno donosi gubitak, dominiraju dve metode. Sistematske studije ovom metodologijom su retke; veći broj lozova ispituju samo tri takve studije, studija Tverskog iz 1967 (Tversky, 1967), Tverskog i Kanemana iz 1992 (Tversky & Kahneman, 1992) i Gonzalesa i Vua iz 1999 (Gonzales & Wu, 1999). Studija Tverskog iz 1967. godine koristi *metodu direktne numeričke ocene* monetarnog ekvivalenta. Upotrebom Beker-Degrut-Maršakove procedure (skr. BDM procedura, Becker, DeGroot &

Marschak, 1964, up. Tversky, 1967 za primenu u eksperimentalnoj psihologiji) i realnog odigravanja i isplate podskupa lozova koje analizira, Tverski obezbeđuje iskrenost odgovora svojih ispitanika. U studiji iz 1992. godine, međutim, Tverski i Kaneman menjaju proceduru merenja monetarnih ekvivalenata. Ispitanicima se prikazuje određeni loz oblika ($x$,$p$;$y$,1-$p$) pored koga je prikazana skala sigurnih iznosa sa logaritamskim intervalima od najnižeg mogućeg do najvišeg mogućeg ishoda koji taj loz može da donese. Procedura zahteva od ispitanika da označe sve sigurne monetarne iznose koje ne bi prihvatili u zamenu za loz i sve sigurne ishode koje bi prihvatili. Softver za izvođenje eksperimenta je forsirao internu konzistentnost odgovora, u tom smislu što nije bilo omogućeno da se npr. za određeni loz odbije sigurni iznos od 5$, zatim prihvati sigurni iznos od 10$, a onda odbije sigurni iznos od 15$. Pošto bi ispitanici tako doneli svoje odluke o prihvatljivim i neprihvatljivim cenama loza, Tverski i Kaneman pokušavaju da uzmu finiju meru monetarnog ekvivalenta. Posle prvog skupa izbora ispitanika, kompjuter bi skalu sigurnih ishoda zamenio finijom skalom raspona od 25% niže vrednosti do 25% više vrednosti od one vrednosti na kojoj je ispitanik počeo da prihvata odgovarajuće sigurne iznose u zamenu za rizični loz. Konačno, mera monetarnog ekvivalenta je srednja vrednost intervala između poslednje odbijene i prve prihvaćene vrednosti u nizu sigurnih ishoda u drugom, finijem merenju. Tverski i Kaneman navode da je prednost ove procedure u tome što se umesto direktne numeričke ocene monetarni ekvivalent izvodi na osnovu opservabilnih izbora ispitanika. Studija Gonzalesa i Vua iz 1999. godine (Gonzales & Wu, 1999) koristi istu metodologiju kao studija Tverskog i Kanemana iz 1992, ali ponavlja proceduru preciznijeg merenja sve do rezolucije od 1$. Ovi detalji u metodološkim procedurama svakako nisu nebitni, posebno ako se uporede rezultati do kojih dolazi Tverski 1967. sa rezultatima do kojih dolaze kasnije studije. Pošto se izvedu funkcije ponderisanja verovatnoća za studije iz 1992. i 1999. godine, sa novom metodom merenja monetarnih ekvivalenata, dobijaju se prilično očigledne inverzne-S funkcije kakve predviđa teorija izgleda Tverskog i Kanemana (Slika 4b). U studiji iz 1967. godine, Tverski je izveo subjektivne verovatnoće svojih ispitanika u odnosu na njihove ocene monetarnih ekvivalenata (podsetimo se, to je oko deset godina pre nego što je razvijena teorija izgleda). Rezultati njegove analize (slika 2, str. 32 u Tversky, 1967) otkrivaju subjektivne verovatnoće koje veoma blago precenjuju niske i potcenjuju visoke objektivne verovatnoće, što je standarni nalaz u oblasti, ali koje *ne otkrivaju nelinearnosti* karakteristične za inverznu-S funkciju ponderisanja verovatnoća teorije izgleda.

Problematiku proceduralne i deskriptivne invarijantnosti u oblasti kauzalnog učenja smo već detaljno diskutovali: pitanje procedure kojom se meri procena intenziteta kauzalne moći, tj. precizne eksperimentalne instrukcije kakva se daje ispitanicima, i pitanje deskripcije tj. forme prezentacija informacija, u ovoj oblasti bitno utiču na selekciju modela. Eksperimentalni rezultati su daleko od toga da pružaju ubedljivo svedočanstvo o tome da određeni formati prezentacije i eksperimentalne instrukcije indukuju kontekst procene koji odgovara formi određene teorije o intenzitetu kauzalnih odnosa, iako smo videli da postoje neke indicije (u tek tri eksperimentalne studije) da sumarni formati i kontrafaktualna pitanja odgovaraju modelu kauzalne moći Čengove. U oblasti donošenja sudova, verbalno kodirani problemi koje zadaju Tverski i Kaneman svojim ispitanicima dovode do grubih grešaka u odnosu na normativno adekvatne sudove; Tenenbaum i Grifits, postavljajući ispitanicima pitanje kroz metodu direktne numeričke procene određenih veličina zaključuju da su ljudski sudovi, nasuprot dve decenije tradicije ograničene racionalnosti, aproksimativno objektivni i da ih je moguće modelirati jednostavnim bejzijanskim modelom. U proučavanju epizodičke memorije i konceptualnog sistema, može se reći, vlada pravi metodološki haos u odnosu na prethodno diskutovane oblasti. Empirijske studije koje testiraju uže specifikovane hipoteze svoje metode konstitušu *ad hoc*, po potrebi; većina eksperimenata učenja kategorija, kategorizacije, ili neke verzije metode reprodukcije u pamćenju, jesu slični, ali variraju u raznim eksperimentalnim parametrima, izboru materijala, instrukcijama i načinu prezentacije stimulusa do te mere da to praktično onemogućava precizna poređenja kakvo smo mogli da izvedemo za metode merenja monetarnih ekvivalenata.

Veliki praktični problem je, činjenica da u naučnoj zajednici ne postoji koncenzus oko jedinstvenog metodološkog okvira koji bi obezbedio da eksperimentalni nalazi nesmetano komuniciraju kroz različite formalne modele. U nekim oblastima se takav okvir nazire; najprecizniji je svakako u odlučivanju i kauzalnoj indukciji. Za razliku od navedene dve oblasti, kada istraživači u oblasti kategorizacije i učenja kategorija ponude podatke većeg broja studija pripremljenih za meta-analizu (up. Smith & Minda, 2000, za izuzetan napor da se izvede meta-analiza učenja 5-4 strukture), možemo biti sigurni da se među rezultatima nalaze podjednako razni geometrijski oblici, shematizovana ljudska lica, crteži i fotografije. Da li toliko kompleksni kognitivni procesi, kao što su učenje kategorija i kategorizacija, mogu da budu invarijantni kroz različite sadržaje, fine varijacije u eksperimentalnim instrukcijama

ili procedure koje variraju odnos značaja ograničenja vremena reakcije i davanja tačnih odgovora - verovatno je pitanje koje ne vredi ni diskutovati.

Ono što predstavlja najveći problem u odnosu na fenomene kršenja proceduralne i deskriptivne invarijantnosti jeste to što posle registrovanja takvih fenomena, ako odnos modela koji se testiraju, deskripcije stimulusa i eksperimentalne procedure nije jasno i potpuno specifikovan, ne može da se donese zaključak o tome *šta se tačno promenilo* sa promenom deskripcije ili procedure. Pri tom, pitanje o tome koji je tačno cilj kognitivnog izračunavanja sa svakom promenom deskripcije i/ili procedure se ponovo otvara. Ispitanici, pokazale su ubedljivo mnoge studije odlučivanja u uslovima rizika, nisu konzistentni ni u davanju odgovora unutar jedne iste deskriptivne i proceduralne paradigme. Kako onda očekujemo da razumemo inkonzistentnost između različitih procedura i deskripcija? To može da bude omogućeno samo ako se detaljno formalno opišu i deskriptivni, i proceduralni aspekti eksperimentalne prakse u kognitivnoj psihologiji. Patrik Sapis, najznačajniji predstavnik semantičkog pristupa filozofiji nauke - doduše, na fusnoti svoje knjige iz 2002. - primećuje kako pitanje formalizacije eksperimentalnih procedura nije nikada dobilo adekvatan tretman u analizi naučne metodologije: „*Bilo bi takođe poželjno razviti modele eksperimentalnih procedura, ne samo rezultata. Zaista detaljan potez u ovom pravcu bi se nužno oslanjao na psihofizičke i njima srodne psihološke koncepte u opisu toga šta eksperimentalni naučnici zaista rade u svojim laboratorijama. Ovo značajno, fundamentalno pitanje* [...] *je tek malo sistematski razvijeno u literaturi filozofije nauke*“ (Suppes, 2002, str. 7, fusnota 3, naš prevod).

Čitalac koji ima iskustva u eksperimentalnoj praksi kognitivne psihologije se sada već možda pita da li naša rasprava postavlja pitanja čija kompleksnost prevazilazi i samu eksperimentalnu praksu i kompleksnost formalnih metoda koji se koriste u modeliranju eksperimentalnih nalaza. Njemu odgovaramo: kognitivna psihologija predstavlja suštinski projekat naturalizacije psihologije kao nauke, i kao takav, taj projekat mora da trpi najoštrije kritike kojima projekat zasnivanja jedne prirodne nauke može da se podvrgne. U odnosu na pitanja deskriptivne i proceduralne invarijantnosti, ta kritička diskusija traži odgovor na pitanje da li će kognitivna psihologija imati, ili neće imati, čvrsto utemeljene merne procedure za koje mogu da se odrede invarijante u odnosu na transformacije indukovane promenama opisa stimulacije ili eksperimentalne procedure. U suprotnom, kognitivna psihologija može i zauvek da ostane disciplina u kojoj „merenje“ podrazumeva prosto zapisivanje neke informacije u formi broja, što je svetlosnim godinama udaljeno od suštine

ideje merenja u prirodnim naukama. Sama „naučnost“ projekta - koji se toliko ističe svojom „naučnošću“ u odnosu na druge psihološke discipline i društvene nauke uopšte - tako može da se dovede u pitanje.

## 8.4 O egzistenciji reprezentativnih subjekata i individualnim razlikama

Posvetićemo sada pažnju ulozi koju u debati o racionalnosti igraju dva povezana problema. Prvi od njih se odnosi na pitanje interpretacije individualnih razlika u višim i simboličkim kognitivnim funkcijama. Drugi problem je novijeg datuma, do sada u literaturi uopšte nije dobio adekvatan teorijski tretman, a odnosi se na pitanje o postojanju reprezentativnog subjekta ma kog formalnog modela kognitivnih funkcija uopšte. Pod reprezentativnim subjektom nekog formalnog modela smatramo kognitivni sistem koji dosledno implementira neku kognitivnu funkciju onako kako je ona formalno definisana u okviru *jednog određenog modela*. Ispitanik čiji su svi odgovori reprezentativni za teoriju očekivane korisnosti je, tako, reprezentativni subjekt te teorije.

*Individualne razlike u višim i simboličkim kognitivnim funkcijama.* Proučavanje individualnih razlika u oblasti viših i simboličkih kognitivnih procesa, potpuno paradoksalno, nije ustaljena praksa. Situacija je paradoksalna zbog toga što se istraživanja ljudske inteligencije - paradigmatične oblasti psihologije individualnih razlika - skoro u potpunosti baziraju na upotrebi kognitivnih zadataka iz domena viših i simboličkih funkcija. Istraživači u oblasti inteligencije dobro poznaju tzv. Galtonovou paradigmu, prema čijoj savremenoj verziji se inteligencija shvata kao supstancijalno povezana sa brzinom obavljanja elementarnih kognitivnih zadataka[71]. U okviru debate o racionalnosti, tek su studije Stanoviča i Vesta (Stanovitch & West, 1998, 2000) otkrile na koji način proučavanje individualnih razlika može da da doprinos diskusiji odnosa između normativnih i deskriptivnih teorija. Paradoks u kome nedostaju istraživanja individualnih razlika koja se, u većini slučajeva, u kognitivnoj psihologiji tretiraju kao „šum“ koji tek ometa fino fitovanje modela na bihejvioralne podatke, još je dublji ako se posmatra iz perspektive evolucione racionalnosti koja značajno motiviše savremene racionalne analize. Proces evolucije, sasvim u skladu sa svojim inherentnim osobinama, morao je da *forsira prisustvo individualnih razlika*, retko kada vodeći ka postojanju jedne morfološke ili bihejvioralne crte koja bi odlikovala sve pripadnike iste populacije

(Tooby & DeVore, 1987).

Izuzetak svakako predstavlja oblast odlučivanja u uslovima rizika i neizvesnosti, gde se tradicionalno analiziraju podaci pojedinačnih subjekata. Naglasak nije toliko na analizi individualnih razlika karakterističnoj za diferencijalnu psihologiju, koliko na razvoju modela odlučivanja koji će objašnjavati odluke svakog pojedinačnog subjekta. To se postiže, naravno, inkorporacijom subjektivnih parametara. Posle svega što smo do sada izneli o ovoj oblasti, jasno je da nema govora o tome da se ponašanje dva subjekta modelira funkcijama korisnosti sa istim parametrima - jednostavno, jedan ispitanik može biti neutralan prema riziku, a drugi da pokazuje averziju (ili, ređe, sklonost) prema riziku. Neki ispitanici će pokazati veću, a neki manju averziju prema gubicima. Ovo su sasvim tipični eksperimentalni nalazi u oblasti odlučivanja.

Problem za analizu individualnih razlika svakako predstavljaju modeli bez slobodnih parametara. Svi normativni modeli ocene snage kauzalnog odnosa spadaju u ovu klasu modela. I dok je sa stanovišta statističke selekcije modela ova osobina veoma povoljna, jer odsustvo slobodnih parametara pokazuje da je reč o modelima zavidne jednostavnosti - što se „nagrađuje“ odsustvom statističke penalizacije u standardnim procedurama selekcije modela - modeli bez slobodnih parametara ne mogu ni u principu da objasne pojavu individualnih razlika. Jednostavno, oni računaju određenu funkciju samo na osnovu opisa sredine, odn. nekog kvantitativnog opisa stimulacije, pa prema tome ne uključuju izvore neizvesnosti koji bi poticali od karakterističnih razlika između subjekata. Već smo zaključili da bi proširenje ovakvih modela u forme koje mogu da obuhvate individualne razlike bilo sasvim dobrodošlo debati o racionalnosti.

Opsežna eksperimentalna studija Stanoviča i Vesta (Stanovich & West, 1998) ispitala je povezanost (a) *standardnih mera kognitivnih sposobnosti* i (b) *mera tendencija u mišljenju* sa merama koje se odnose na normativnu adekvatnost odgovora u tipičnim eksperimentima u debati o racionalnosti. Njihova studija obuhvata probleme tipične za oblast odlučivanja u uslovima rizika i neizvesnosti, suđenja i rezonovanja. Pod tendencijama u mišljenju podrazumeva se skup mera koje se primarno odnose na sklonost ka dekonstualizaciji problema kako je predstavljen, i za koje Stanovič i Vest veruju da pokrivaju četiri dimenzije tzv. *epistemičke racionalnosti*: epistemički apsolutizam, spremnost na promenu perspektive, spremnost na dekontekstualizaciju i tendenciju da se drugačija mišljenja prihvate kao evidencija. Mere kognitivnih sposobnosti koje koriste su klasični test

školskog uspeha (tzv. *SAT* skor), skor na Ravenovim matricama i test razumevanja pročitanog (up. Stanovich & West, 1998, za detalje). Skorovi kognitivne sposobnosti i tendecija u mišljenju su izračunati kao kompozitne mere na osnovu skorova pojedinačnih testova koje su ispitanici rešavali, a zatim su izračunate korelacije ovih skorova sa stepenom u kome su ispitanici davali normativno adekvatne odgovore u tipičnim eksperimentima suđenja i rezonovanja. Rezultati Stanoviča i Vesta pokazuju da postoje niske, ali *dosledno značajne pozitivne korelacije* između mera kognitivne sposobnosti i stepena u kome ispitanici daju normativno adekvatne odgovore u eksperimentima. Slično, takve korelacije postoje i sa merama tendencija u mišljenju, ali su one niže od onih između mera kognitivne sposobnosti i stepena u kome se daju normativno adekvatni odgovori. Na osnovu ove studije, Stanovič i Vest su obezbedili snažne argumente protiv nekoliko interpretacija razlika između normativnih kriterijuma i realnih odgovora koje ispitanici daju. Ove razlike nisu posledica nesistematskih, sporadičnih grešaka u izvođenju zadataka, jer da su u pitanju nesistematske greške izvođenja u odnosu na racionalnu kompetenciju koju ispitanici možda imaju, ne bi smela da se pojavi pozitivna korelacija između testova kognitivnih sposobnosti i stepena u kome su dati normativno adekvatni odgovori. Dalje, zaključuju Stanovič i Vest, da ispitanici daju zadacima drugačije interpretacije od onih za koje eksperimentatori veruju da bi trebalo da budu upotrebljene, ovakve sistematske korelacije ponovo ne bi smele da se pojave. Pošto ova dva izvora variranja sigurno ne objašnjavaju postojanje pozitivne korelacije između mera kognitivnih sposobnosti i stepena u kome se daju normativno adekvatni odgovori, Stanovič i Vest zaključuju da može biti reč samo o tome da ispitanici sa većim algoritamskim, odn. procesnim ograničenjima u manjoj meri uspevaju da daju normativno adekvatne odgovore.

Poptuno očigledna kritika koja se nameće već na prvi pogled na rezultate Stanoviča i Vesta je sledeća: ako pogledamo ma koji standardan test inteligencije, ili analiziramo meru školskog uspeha, uviđamo da su mere kognitivnih sposobnosti već unapred definisane kao mere stepena u kome je određena individua u stanju da pruži normativno adekvatne odgovore. Tako, rezultati Stanoviča i Vesta lako mogu da se interpretiraju kao rezultati koji govore o tome da su različiti načini merenja stepena u kome se daju normativno adekvatni odgovori međusobno korelirani, što je fundamentalni rezultat u oblasti merenja ljudske inteligencije koji je otkrio još Spirman (tzv. „*positive manifold*“, empirijska činjenica da ma koja dva testa inteligencije imaju pozitivnu korelaciju i osnova za teorijski

konstrukt *generalne inteligencije*). Prema našem mišljenju, ne potcenjujući nalaze Stanoviča i Vesta (a ponajmanje potcenjujući ogroman eksperimentalni napor koji su ovi istraživači uložili u svoje studije), verujemo da bi za debatu o racionalnosti korisnija bila diskusija individualnih razlika u smislu studije *različitih strategija* koje ispitanici koriste u tipičnim eksperimentalnim paradigmama. Svaku takvu strategiju možemo da shvatimo kao potencijalno operacionalizovanu odgovarajućim formalnim, matematičkim modelom - bilo normativne, bilo deskriptivne prirode. U tom smislu, verujemo, podaci o individualnim razlikama bi značajno mogli da doprinose debati o racionalnosti. Proučavanjem statističkih distribucija upotreba različitih strategija odn. distribucija upotrebe različitih kognitivnih modela u određenim eksperimentalnim paradigmama, saznali bismo više o različitim potencijalnim adaptacijama koje kognitivni sistemi pokušavaju u odnosu na svoju informacionu, epistemičku sredinu. Ovakvo istraživanje bi moglo da usmeri debatu o racionalnosti ka pitanju koje ona nikada nije postavila: da li postoji neka funkcionalnost u činjenici da kognitivni sistemi *sa nekom verovatnoćom biraju da iskoriste neku od raspoloživih kognitivnih strategija* u rešavanju adaptivno relevantnih problema? Da li postoji neka *distribucija verovatnoće koja karakteriše izbore ciljeva* koje kognitivni sistemi sebi zadaju u odnosu na konceptualno, formalno identične problemske situacije?

*O egzistenciji reprezentativnog subjekta.* Sa formalnim modelima koji ne počivaju na pretpostavci o postojanju reprezentativnog subjekta određene teorije kognitivnih funkcija smo se sreli u dva navrata. Studija Harisona i Rutstrumove, koju smo predstavili u sekciji o racionalnosti odlučivanja u uslovima rizika i neizvesnosti, eksplicitno kritikuje pretpostavku o postojanju reprezentativnih subjekata (Harrison & Rutström, 2008). Autori koriste tzv. *mešoviti statistički model* (engl. *mixture model*) kojim kombinuju teoriju izgleda i teoriju očekivane korisnosti da bi objasnili binarne odluke svojih ispitanika u eksperimentu izbora. Hibridni pseudo-normativni model koji su predložili Perales i Šenks takođe je vrsta mešovitog statističkog modela i omogućava da se oceni proprocija sa kojom jedan ili drugi model koji su uključeni u statističku mešavinu objašnjavaju odgovore ispitanika(Perales & Shanks, 2007). Te proporcije u kojima različiti modeli doprinose objašnjenju ponašanja onda su takođe parametri koji odlikuju različite subjekte ili grupe subjekata. Direktno mogu da se interpretiraju kao verovatnoće da će određeni kognitivni sistem, suočen sa određenim problemom, upotrebiti neku od raspoloživih strategija adaptacije u odnosu na taj problem. Standardni pristup podrazumeva da postoji *jedna* takva strategija adaptacije. Pristup koji napušta koncept reprezentativnog subjekta implicira da

isti kognitivni sistem može paraleleno da koristi *više* adaptivnih strategija u odnosu na jedan problem.

Sa stanovišta uobičajene psihološke analize matematičkih modela kognitivnih funkcija, upotreba mešovitih modela sigurno deluje egzotično. Međutim, skoro sve naše prethodne kritike rezultata do kojih se dolazi primenom racionalne analize ukazuju na to da funkcije kognitivnog sistema *vrlo verovatno mogu uspešno da opišu samo ovakve vrste modela.* Koherencija koja (ne) odlikuje (uvek) utemeljenje i razvoj pojedinačnih modela kognitivnih funkcija verovatno će morati da bude zamenjena na prvim pogled ne tako koherentnim i elegantnim mešovitim modelima. Takvi modeli imaju značajnu osobinu da mogu da objasne ponašanje kognitivnih sistema koji sa određenim verovatnoćama formulišu *različite* kompjutacione ciljeve, tako predstavljajući sisteme koji se *adaptiraju na svoju sredinu kroz paralelenu primenu različitih, više ili manje uspešnih adaptacija.*

Naš zaključak o tome da racionalna analiza kognitivnih funkcija ne vodi identifikaciji jedinstvenih kompjutacionih ciljeva, pa tako ni identifikaciji jedinstvenih formi kognitivnih adaptacija, implicira da pretpostavka o reprezentativnom subjektu nije tačna. To, dalje, implicira da su tek mešoviti statistički modeli u stanju da modeliraju realne kognitivne adaptacije. Ono što deluje kao složena statistička mašinerija koja ne nalazi pravo, „čisto“ teorijsko opravdanje, i kakvu na prvi pogled predstavljaju mešoviti modeli, postaće osnovni eksplanatorni koncept u našem predlogu pristupa problemu racionalnosti saznanja: U VI delu, videćemo da pozajmica teorijskih koncepata iz teorije igara daje mešovitim modelima onu koherenciju i elegantnost koja odlikuje tip modela za kojim tragaju racionalne analize.

## 8.5 *Summa summarum*

Debata o racionalnosti viših i simboličkih kognitivnih funkcija u svom dosadašnjem razvoju bila je motivisana pitanjem zašto se, u velikom broju slučajeva, ponašanje kojim upravlja ljudski kognitivni sistem pokazuje različitim od ponašanja koje predviđaju normativno adekvatni modeli simboličke logike i teorije verovatnoće. U skladu sa tim, debata je daleko više razvijena u oblastima u kojima postoji koncenzus o izboru normativnog okvira za diskusiju problema, i manje razvijena u onim oblastima, poput pamćenja i organizacije konceptualnog sistema, za koje ne postoje tradicionalni normativni okviri. Kognitivna psihologija se, svakako,

suočava i sa nekim veoma teškim problemima čiji se normativni analitički okvir ni ne nazire. Činjenica da normativni okvir nedostaje upravo u domenu konceptualne organizacije, ključnom za razumevanje svih funkcija koje se odnose na simboličke reprezentacije i značenje, veoma je važna i biće inkorporirana u naše dalje diskusije.

Pitanje adekvatnosti normativnog okvira prati celu debatu. Istraživači koji rade u okviru programa savremene racionalne analize usvajaju strategiju kojom se tradicionalni zamenjuju „latentnim“ normativnim okvirima koji odgovaraju rešenjima problema adaptacije sredini. Pokazali smo da ta eksplanatorna strategija pati od mnogih nedostataka. Na prvom mestu, ta strategija ne ispunjava nadu koju je sa sobom donela - nadu da će opasnost koja preti od beskonačne proliferacije broja mogućih kognitivnih modela biti ublažena. S druge strane, pristalice ograničene racionalnosti, posebno istraživačkog programa heuristika i inklinacija, ne mogu da ponude odgovarajući odgovor na pitanje kako su kognitivni sistemi, uopšte, prilagođeni odgovarajućoj epistemičkoj sredini. Ne zvuči previše prihvatljivim stanovište po kome evolucija ljudske vrste - koja zahvaljujući svojim kognitivnim funkcijama dominira nad svim drugim vrstama na Zemlji - rezultira u razvoju kognitivnih funkcija čiji su ishodi puni sistematskih odstupanja i grešaka.

Diskusija rezultata racionalne analize više kognitivnih funkcija pokazuje da u odnosu na probleme definisane istim strukturama environmentalnih informacija - poput rizičnih lozova ili kontingencijih tabela koje treba oceniti - kognitivni sistemi mogu da postave različite, podjednako opravdane, kompjutacione ciljeve. Ako sada za trenutak prestanemo da posmatramo kognitivni sistem kao stabilni rezultat evolucionog procesa, i promenimo perspektivu tako da ga posmatramo kao deo evolucionog procesa koji još uvek traje, i u kome razvoji svih kognitivnih funkcija nisu još stabilizovani, razumećemo da je rezultat poput prethodnog upravo očekivan. Određivanje tačnog značenja te *stabilnosti* u kontekstu analize racionalnosti kognitivnih funkcija, te usvajanje kriterijuma za njihovo razvrstavanje u odnosu na stepen stabilnosti koje postižu, biće centralni deo argumentacije koju razvijamo u VI, ključnom delu ove rasprave.

Činjenica da odluke pojedinaca u okviru iste eksperimentalne paradigme nisu perfektno konzistentne, kao i ubedljiva činjenica o postojanju očekivanih individualnih razlika u višim kognitivnim funkcijama, u potpunosti se uklapa u sliku kognitivnog sistema kao sistema koji jedinstvenim strukturama informacija u svojoj sredini pristupa iz više perspektiva, koristeći više strategija i usmeravajući sebe na više različitih ciljeva. Malobrojne, najnovije studije u kojima se napušta

pretpostavka o postojanju reprezentativnog subjekta određene formalne teorije neke kognitivne funkcije predstavljaju korak ka specifikaciji ovakvog shvatanja. Neophodna je teorijska paradigma u kojoj je takva slika kognitivnog sistema u potpunosti naučno jasna, a koja ne narušava koherentnost i kontinuitet kognitivne psihologije sa drugim prirodnim naukama.

Takva teorijska paradigma postoji, i mi ćemo pokušati da debatu o racionalnosti saznanja u VI delu reinterpetiramo konceptima koje je ta teorijska paradigma razvila. Do tada, debati o racionalnosti i pokušaju njene teorijske reinterpetacije ćemo prvo dati jasan istorijski kontekst u IV delu, a zatim posvetiti još neko vreme fundamentalnoj raspravi o odnosu normativnih i deskriptivnih modela odlučivanja u V delu naše rasprave.

# Deo IV

# Poreklo savremenog shvatanja racionalnosti saznanja

Celokupna konceptualna mreža koja organizuje naše diskusije racionalnosti saznanja, videli smo, u savremenoj paradigmi racionalne analize počiva na tri suštinska koncepta: konceptu *normativnog okvira*, konceptu *adaptacije* i konceptu *optimizacije*. Jedinstven skup pretpostavki savremene KKP dalje situira ovu mrežu koncepata u kompjutacionizam kao najširu, dominantu perspektivu razumevanja saznajnih procesa uopšte. Tako, procesi optimizacije su *kompjutacioni procesi* koje *implemetira* kognitivni sistem da bi zadovoljio *kompjutaciono određene* ciljeve adaptacije relevantnoj sredini. Normativni okvir, kada se jednom pređe sa njegove formalne analize na pitanje njegove implementacije, postaje skup ograničenja na dizajn mogućih reprezentacija i formu kognitivnih procesa koji njima upravljaju da bi se odredili optimalni bihejvioralni odgovori na adaptacione pritiske. Savremena racionalna analiza je u potpunosti otvorena za implementaciju teorijskih koncepata evolucione biologije. Dok evoluciona biologija tako daje semantiku KKP, opravdavajući adaptivni karakter kognitivnog sistema neprestanim sredinskim pritiscima koji zajedno sa njegovim ciljevima determinišu njegove

odgovore, matematička i logička revolucija prve polovine XX veka su joj pružile sintaksu: *mehanizam* kojim ona objašnjava *implementaciju funkcija* kognitivnog sistema u tom evolucionom, analitičkom okviru.

Savremena KKP je tako posledica nekoliko isprepletanih istorija koje se odnose na najmanje tri, nikako nezavisne, linije razvoja naučne misli od XVII veka do danas. Prva se odnosi na izuzetan uticaj teorije evolucije Čarlsa Darvina na razvoj psihološke nauke uopšte. Ovaj razvoj je isprepletan sa istorijom ideja o optimizacionim procesima u univerzumu uopšte, neraskidivo povezan sa diskusijama o značenju drugog zakona termodinamike i interpretacijom entropije u kontekstu analize živih sistema (Schrödinger, 1944, Monod, 1971/1983, Prigogine & Stengers, 1984). Savremenu KKP, sasvim u skladu sa dubokom vezom sa principima Darvinove teorije, tako danas odlikuje *funkcionalističko* shvatanje mentalnog sistema, pretpostavka toliko duboko usađena u njenu teorijsku konstrukciju da se od strane praktičara ona više i ne diskutuje.

Druga istorijska linija se odnosi na nešto širu i apstraktniju istoriju *mehanicističkih objašnjenja* živih sistema i psiholoških fenomena uopšte, liniju koja se razvija najmanje od Dekartovog mehanicističkog pogleda na rezonovanje i funkcije duše, te ranih spekulativnih modela poput Vukansonovog mehaničkog flautiste i mehaničke patke (oba predstavljena 1737, Boden, 2006, up. Glimcher, 2004 za diskusiju istorijske uloge ovih modela) do savremenih kompjuterskih simulacija kognitivnih funkcija. Ova istorijska linija kulminira u rezultatu o mogućnosti dizajna univerzalnih kompjutacionih mehanizama u prvoj polovini XX veka - najvažnijem naučnom rezultatu za zasnivanje KKP uopšte. Čomskom je, posle rezultata Tjuringa, Čerča, Klinija i drugih matematičara, preostalo da krajem 50-ih godina XX veka tek prodiskutuje posledice koje slede iz njegovih rezultata o mogućnostima kompjutacione implementacije formalnih sintaksičkih struktura da bi (skoro nenamerno, moglo bi se reći) zasnovao danas dominantu paradigmu KKP. Početak, koji deluje tako naglo ako se posmatra samo kao posledica debate Čomskog i Skinera (Chomsky, 1959/1967), zapravo je posledica dugotrajnih i komplikovanih istorijskih interakcija u rešavanju problema mehanizacije mišljenja, problema koji je opsedao još Lajbnica (Couturat, 1901/2002) u formi koja se malo razlikuje od savremene. Duboka kritička diskusija fundamentalnih problema matematičke logike i filozofije matematike od strane pristalica tri velike „škole“ sa početka XX veka: *logicizma*, *formalizma* i *intuicionizma* (Šikić, 1987), na presudan način je odredila probleme čije je rešavanje dovelo jednu briljantnu generaciju matematičara

do otkrića jedinstvene ideje o mehanizaciji rezonovanja uopšte.

Konačno, treća linija istorije nauke, čiji uticaj danas decidno određuje oblik debate o racionalnosti, jeste linija razvoja naših shvatanja pojmova verovatnoće, neizvesnosti i informacije. Debata o racionalnosti svoj izvor duguje raspravi o racionalnom izboru, koja je u potpunosti deo ove istorijske tradicije. Od prvih Paskalovih uvida u mogućnost formalizacije našeg suda o verovatnim događajima, Bernulijeve hipoteze o očekivanoj korisnosti, preko Bejzove teoreme, Laplasove sinteze do Kolmogorova i kulminacije u konceptualnoj revoluciji teorije igara Džona fon Nojmana, ova linija istorijskog razvoja bila je neraskidivo povezana sa analizom ljudskog suda i ponašanja. Ako savremena debata o racionalnosti pogleda unazad u svoju istoriju dalje od druge polovine XX veka, ona lako može sebe da interpretira kao specifičnu granu diskusije koja je započeta u istoriji teorije verovatnoće zapažanjima Paskala i Ferma polovinom XVII veka.

# 9 Kratka kritička istorija prirodnog uma

Redovi koji su pred nama nikako ne predstavljaju pregled relevantnih poglavlja istorije nauke. Uostalom, ogromna naučna periodika posvećena je boljem razumevanju i dokumentovanju istorije nauke upravo u oblastima relevantnim za našu diskusiju. Naš cilj je da kroz kraći osvrt na najznačajnije, presudne trenutke koji su kroz istoriju nauke oblikovali teorijski okvir racionalnosti saznanja, ustanovimo kako je taj okvir postao eksplanatorno koherentan u odnosu na naučne probleme koje je održava u fokusu rasprave. Naš cilj je da kroz diskusiju o istorijskom poreklu problema ocrtamo još jasnije teorijski okvir savremene debate o racionalnosti. Diskusija zato ne prati linearno istorijske razvoje, već polazi od zatečenog stanja u debati o racionalnosti i selektivno se obraća istoriji nauke u interpretaciji ključnih teorijskih koncepata.

Karakteristično za savremenu KKP jeste verovanje naučnog establišmenta - odn. njegovog dela koji je prihvata - da ona predstavlja uspešan projekat naturalizacije uma. Prema ovoj interpretaciji establišmenta, KKP preostaje razvoj *normalne nauke* (Kun) - progresivni eksperimentalni i matematički napori u sve složenijim sintezama teorijskih sudova o kompjutacionim mehanizmima ljudskog uma. Naredni redovi, osim što rasvetljavaju motive ove savremene sinteze, ukazuju na određene pukotine u njenoj teorijskoj konstrukciji, otkrivaju zaključke nekih nedorečenih analiza i postojanje implicitnih pretpostavki koje se retko, ili nikada, ne kritikuju.

Zato ih ne treba posmatrati kao istorijski prilog, fusnotu ovoj metateorijskoj i metodološkoj kritici debate o racionalnosti, već kao integralni deo te kritike i tekuće debate uopšte.

## 9.1 Naturalizacija uma I: mehanicistička tendencija

Sledeći odlomci iz Dekartovih „*Principa filozofije*“ otkrivaju prirodu najranije elaborirane tvrdnje o mogućnosti shvatanja ljudskog duha i ponašanja životinja i ljudi kao *mehaničkih sistema*:

„*Započeo sam posmatranjem svih snažnih i čistih pojmova koje naše razumevanje može da ima u pogledu materijalnih stvari, i sve što sam pronašao bili su naši pojmovi oblika, veličina i kretanja, i pravila prema kojima se ova tri međusobno modifikuju - pravila koja su principi geometrije i mehanike. Ovo me je vodilo ka sudu da svo ljudsko znanje prirodnog sveta mora biti izvedeno iz ova tri, zato što su svi preostali pojmovi koje mi imamo o čulno-opažajnim stvarima konfuzni i nejasni, i mogu samo da nam otežaju - ne pomognu - u našoj potrazi za saznanjem stvari van nas samih*“, i dalje:

„*Kada sam kasnije opazio u čulno-opažanjim objektima tačno iste one efekte koji su bili predviđeni mojim teorijskim pristupom, prosudio sam da su oni zaista bili efekti upravo takve interakcije tela koja nisu čulno-perceptivna; i u tome sam ostao uveren putem očigledne nemogućnosti da se nađe bilo koje drugo objašnjenje za njih. U razmišljanju o ovome značajno mi je pomoglo razmatranje artefakata. Ne raspoznajem nikakvu razliku između artefata i prirodnih tela osim što artefakti mahom rade kroz mehanizme koji su dovoljno veliki da bi ih naša čula lako opažala (moraju biti, ako su ljudi u stanju da ih proizvedu!)*“ (Descartes, 1664, Deo IV, §203, naš prevod).

Pristalice simbolicističke verzije programa veštačke inteligencije u kognitivnim naukama verovatno bi bez oklevanja potpisalale svaku od Dekartovih rečenica iz prethodna dva odlomka. Za razliku od Dekarta, kognitivni psiholozi druge polovine XX veka raspolagali su vrstom mehanizama - *mehanizama za obradu informacija* - čija je korespondencija sa Dekartovim analitičkim pristupom ljudskom umu još potpunija od analogija sa mehanikom koje je on mogao da povuče. Poput Dekarta, koji nije video nikakvu razliku između artefata i prirodnih tela, simbolicisti sa polovine XX veka nisu videli nikakvu razliku između prirodne i veštačke inteligencije - osim što je prva realizovana u mehanizmima za obradu informacija koji su proizvod

procesa evolucije, realizovani u neurofiziološkom supstratu ljudskih kognitivnih procesa - a druga u medijumu silikonskih, poluprovodničkih elemenata. Za razliku od Čomskog i Fodora, „neokartezijanaca“ XX veka, Dekart nije verovao da su razumevanje i produkcija novih rečenica mehanički procesi. Više funkcije ljudskog uma pripadaju eksplanatornim principima *res cogitas*, ne *res extensa*, koja predstavlja mehanicistički univerzum ispunjen etrom, mašinu koju je Dekart mogao da uporedi sa mehaničkim časovnicima koji su počinjali ubrzano da se razvijaju u Evropi XVII veka (Wright & Bechtel, 2007), ispisujući prethodne redove par decenija pre Njutnove velike sinteze („*Philosophiæ Naturalis Principia Mathematica*“ iz 1687) u prvom totalnom, konzistentnom mehanicističkom objašnjenju fizičkog sveta.

Još u Dekartovom mehanicizmu naziru se osobine simbolicističke paradigme razvijene u XX veku posle otkrića mehanizama za obradu informacija. Prvo, isto kao što Dekart nazire mogućnost unifikacije objašnjenja „*čulno-opažajnih objekata*“ istim principima koji objašnjavaju interakcije između fizičkih tela, principa geometrije i mehanike, simbolicistička paradigma je u teoriji folk-psihologije, po pretpostavci realizovanoj u nekoj vrsti računa verovanja, videla *naturalističku teoriju uma*. Kompjutacionizam, čijim ćemo se istorijskim korenima polako približavati, podsetimo, predstavlja teoriju o tome kako se formalni sistemi poput onih koje predlaže simbolicistička paradigma KKP *fizički realizuju*. Slično kao što Dekart nije osećao potrebu da povuče razliku između artefakata i prirodnih tela, potrebu koja dominirala zapadnim mišljenjem pre revolucije naučne racionalnosti XVII veka, simbolicisti ne osećaju potrebu za povlačenjem razlike između implementacije formalnih sistema u biološkom supstratu i u fizičkoj realizaciji koju predstavljaju digitalni kompjuteri, najsloženiji artefakti koje je ljudski um ikada konstruisao. Upravo to prelaženje (unifikacijom fenomena) preko granice prirodnog i veštačkog karakteriše strategiju naturalizacije uma koja predstavlja suštinsku odredbu savremene KKP. Ovakva strategija naturalizacije uma, međutim, *nije karakteristična i za pristalice emergentističkih paradigmi*: konekcionizma i teorije dinamičkih sistema. Videćemo da je razumevanje njihovog pokušaja naturalizacije uma, koji sa simbolicizmom deli zajednički kompjutacionistički okvir, moguće tek posle istorije prirodnih nauka XIX veka, istorije koja je otkrila jednu novu, čudniju, nepredvidljiviju vrstu mašina i mehanizama koje će emergentističko usmerenje odabrati za svoju analogiju između prirode i uma.

Jedan bitan deo odnosa savremene simbolicističke paradigme KKP, koja predstavlja najčistiju prepoznatljivu posledicu razvoja kartezijanskog mehanicizma,

ukazuje na problematiku koja je Dekartu samom bila strana, a čiju netrivijalnost u savremenim racionalnim analizama smo mi već diskutovali. Diskutujući pitanje koja supstanca - *res cogitas* ili *res extensa* - predstavlja izvor naših opažaja, Dekart - vođen idejom da senzacije nije moguće svesno stvarati - tvrdi da njihov izvor mora biti *res cogitas*, pošto svih efekata *res cogitas mi moramo biti svesni* (Jovanović, 1997). Njegovo *cogito ergo sum*, dakle, specifikuje *ego* kao misleći, epistemički subjekat sa odlikom svesti, *dakle na personalnom nivou analize* u odnosu na hijearhiju subpersonalnog, personalnog i suprapersonalnog nivoa psihološke teorije koju koristimo kao analitičkog sredstvo u našoj raspravi. U debati o racionalnosti, ogromna većina diskutovanih mehanicističkih objašnjenja kognitivnih funkcija počiva na subpersonalnom nivou: nivou na kome operišu brzi, nesvesni kompjutacioni kognitivni procesi koji tek svoje autpute, rezultate svojih izračunavanja predaju funkcijama koje ponašanje *izvode*. U nekom trenutku tog procesa „prenosa rezultata“, subjekat savremene kognitivne psihologije jeste svestan („*znam* koju sam odluku doneo“, „*znam* u koju kategoriju spada stimulus koji si mi prikazao“), ali je taj momenat u savremenoj KKP *daleko van fokusa analize*. KKP, videli smo, vođena mogućnošću bihejvioralne rekonstrukcije subjektivnih verovanja koju je započeo Remzi, a usavršili fon Nojman, Morgenštern, Sevidž i konačno Tverski i Kaneman, zadržava sve neophodne interne, neopservabilne procese na subpersonalnom nivou analize - nivou koji nije nivo Dekartovog mehanicističkog objašnjenja, i koji nije svestan. Problemi koji nastaju ovakvim „cepanjem“ epistemičkog subjekta[72] odnose se na probleme intuitivnosti normativnih okvira. Tačno *ko* u teoriji racionalnog izbora intuitivno prihvata aksiome racionalnog izbora: „neko“ na subpersonalnom nivou analize, ili svesni subjekat kome možemo bezbedno da se obratimo na personalnom nivou analize? Tamo gde složenost odgovarajućih formalnih normativnih okvira, ili formalnih deskriptivnih modela, prevazilazi intuitivno prihvatljiv nivo kompleksnosti za koji bismo očekivali da ljudski um može da toleriše u svakodnevnim aktivnostima, KKP otvara imaginarijum hipotetskih konstrukata na subpersonalnom nivou analize. To, nažalost, stvara nepremostive probleme u analizi adekvatnosti normativnih okvira. U slučaju konstrukata sa subpersonalnog nivoa koji mogu da se rekonstruišu kroz bihejvioralne funkcije (kao u većini teorija odlučivanja), reprezentaciona teorema bar predstavlja dokaz da bi takvi konstrukti mogli da postoje i „unutar ljudskog uma“. U slučaju teorija čije konstrukte nije moguće rekonstruisati kroz bihejvioralne funkcije direktno, a tu govorimo o većini teorija u okviru KKP uopšte, subpersonalni nivo analize ne

predstavlja ništa drugo nego rezervoar za bezbedno uklanjanje „semantičkog viška“ odgovarajuće teorije, konstrukt po pravilu prepušten budućem razvoju neuronauka koje će razumeti implementacioni nivo analize jednom kada kompjutacioni i algoritamski nivoi postanu jasni. Uzimajući u obzir istorijsku rekonstrukciju motivacije Dekartovog mišljenja u XVII veku, rekonstrukciju koja ukazuje na potrebu za jednom teorijom epistemičkog subjekta i razumnog, moralnog postupanja u epohi koja je bila fokusirana na probleme upravljanja društvom i državom (Jovanović, 1997), postaje jasnije zašto se Dekartov mehanicizam otkriva - u meri u kojoj se otkriva i koju određuje granica *res extensa* i *res cogitas* - na nivou analize *svesnog* saznajnog subjekta.

Pitamo se zašto, onda, u doba neposredno pre i posle Drugog svetskog rata, koje na odlučujuć način oblikuje naučnu praksu i kontekst u kome nastaju kibernetika, teorija igara i savremene kognitivne nauke, KKP bez obzira na sve očigledne probleme prihvata analizu na subpersonalnom nivou bez dublje teorijske ograde? Odgovor, verovatno, leži u jednoj pragmatičnosti tih odsudnih godina XX veka koja nije odlikovala pragmatičnost Dekartove epohe: u slučaju savremene KKP, koja nastaje u periodu u kome preku potrebu da se dobije Drugi svetski smenjuje preka potreba da se preživi Hladni rat, govorimo o epohi motivisanoj daleko više stvaranjem matematičkih modela koji probleme *rešavaju* nego potrebom da se *razume način na koji ih rešavaju* u odnosu na dublju teorijsku i filozofsku motivaciju. U prilog ovoj tezi, koju u različitim oblastima istorije nauke i na različite načine iznose i drugi autori (up. Mirowski, 2002, za primer matematičke ekonomije i teorije igara), govori oštra disproporcija između *fundamentalnog teorijskog interesa* 20-ih, 30-ih i 40-ih godina, kada pre II Svetskog rata nastaju rezultati poput Gedelove teoreme nepotpunosti ili Tjuringovih i Čerčovih univerzalnih izračunavanja, i visokog pritiska na *primenu* teorije igara, kibernetike i kompjuterskih nauka posle Drugog svetskog rata. Naučni i društveni kontekst u kome nastaje KKP, kontekst koji čine kibernetika, kompjutaciona lingvistika, program razvoja veštačke inteligencije, operaciona istraživanja i teorija igara, je kontekst neposredno usmeren na *primenu*. Većina ovih disciplina bile su izdašno finansirane od strane državnih i vojnih agencija u periodu neposredno po završetku Drugog svetskog rata, kada su pružale nadu o mogućnostima optimalnog diplomatskog, ekonomskog, informacionog, konvencionalnog i konačno nuklearnog odgovora stratezima dveju supersila - SAD i SSSR - koje su izronile kao jedini pravi pobednici najvećeg oružanog sukoba u istoriji čovečanstva (Mirowski, 2002). U uslovima hladnoratovske paranoje, praktični zadaci

su usmeravali nauku da traga za rešenjima koja će garantovati *efekte*, ne teorijsko opravdanje, ne koherenciju, ne duboko razumevanje fenomena koji se proučava. Nove, kibernetičke nauke, svojom evolucijom ka ideji o opštoj teoriji sistema, davale su perspektivu iz koje je ova motivacija bila praktično opravdana: ako su ljudska društva, uključujući ekonomske, političke i sve druge interakcije koje obuhvataju, opisiva u zajedničkom teorijskom i matematičkom okviru koji pruža opis i fizičkih i bioloških sistema, onda je njima moguće *upravljati* baš kao što je moguće upravljati i ma kojom drugom klasom sistema. Paradoksalno, kritika Skinerovog bihejviorizma od strane Čomskog uklonila je sa istorijske scene naučnu paradigmu koja je bila bazirana na ideji predikcije i upravljanja ponašanjem, da bi motivisala razvoj naučne paradigme koja *bolje predviđa ljudsko ponašanje* - zahvaljujući modelima koji su dopustili interne hipotetičke konstrukte i slobodne parametre. Probleme koji su bihejvioristi toliko pokušavali da izbegnu KKP je prihvatila bez ozbiljnije diskusije. Vreme u kome ona nastaje nije marilo za teorijske probleme sa kojima se mi suočavamo u ovoj diskusiji. Međutim to, ni u kom slučaju, nije jedini izraz njene nekompletnosti kao projekta naturalizacije epistemičkog subjekta.

Dekart sigurno ne predstavlja jedini izvor mehanicističkog pristupa kognitivnim sposobnostima u istoriji filozorije. Ideje iz tradicije britanskog empirizma i asocijacionizma sve redom su mehanicističke. Tomas Hobs je u „*Levijatanu*“ pošao od jedne nerazrađene, ali po svojoj skici sasvim savremene kompozicionalne kognitivne teorije u kojoj neki vide rani predlog koji odgovara savremenoj simbolicističkoj verziji KKP (Haugeland, 1981). Filozof i matematičar koji je najčešće zapostavljen u savremenim diskusijama KKP - kada se one uopšte obraćaju svojim istorijskim prethodnicama - Godfrid Lajbnic, maštao je o razvoju *characteristica universalis* - univerzalnog jezika za naučni i matematički opis stvarnosti koji bi prevazišao sve probleme u komunikaciji između istraživača (Couturat, 1901/2002). Lajbnic nije svoju *characteristica universalis* predstavio kao predlog teorije saznanja (iako je tvrdio da mu je cilj razvoj „*alfabeta ljudske misli*“) niti kognitivne (psihološke) teorije; ipak, od njegove ideje univerzalnog jezika ne može da se zamisli bolja istorijska prethodnica ideji formalnih sistema koja predstavlja okosnicu simbolicističkih kognitivnih teorija. Mehanicistička tendencija psihologije saznanja, prisutna u novovekovnom shvatanju subjektivnosti uopšte, vremenom je morala da zameni fizikalne eksplanatorne mehanizme simboličkim: u trenutku kada je to ostvareno, praktično je rođena savremena kognitivna psihologija.

## 9.2 Naturalizacija uma II: psihofizički zakoni i merenje

Objašnjenje mehanizmima i objašnjenje naučnim zakonima filozofija nauke izdvaja kao dve osnovne, komplemenetarne eksplanatorne strategije u naučnom diskursu (Psillos, 2007). Mehanicistička objašnjenja, karakteristična za simbolicističku KKP, povezuju savremenu standardnu paradigmu sa tradicijom novovekovnog mehanicističkog viđenja ljudskog duha uopšte. Međutim, tu tradiciju odlikuje jedan ozbiljan nedostatak. *Ona nije tradicija eksperimentalne nauke*, i zbog toga *nema razvijenu inherentnu vezu sa konceptom merenja.* Studenti psihologije će Dekarta, Loka, Lajbnica, Hobsa, Hjuma i druge upoznati najčešće u poglavljima istorije filozofije koja pripremaju teren za nastup psihologije u XIX veku, jer tek tada, saznaće, počinje govor o *naučnoj psihologiji*, zahvaljujući na prvom mestu eksperimentima Vebera (1846), Fehnerovoj psihofizičkoj sintezi (1860, sa objavljivanjem „*Elemenata psihofizike*“) i Vuntovom osnivanju prve psihološke laboratorije u Lajpcigu (1879). Ono što je karakteristično za zasnivanje psihofizike, kome ćemo nakratko posvetiti našu pažnju, jeste eksplanatorna strategija objašnjenja putem naučnog zakona. Ono što istorija psihologije propušta da učini - a neki autori smatraju da je to posledica nekritičkog tumačenja istorije psihofizike u Boringovoj uticajnoj „*Istoriji eksperimentalne psihologije*“ iz 1929 (i još uticajnijem drugom izdanju iz 1950, Masin, Zudini, & Antonelli, 2009) - jeste da *sagleda istoriju razvoja psihofizike kroz kontinuitet objašnjenja putem naučnih zakona*, koji Fehnerov rad povezuje direktno sa ovde detaljno diskutovanim radom Danijela Bernulija iz 1738.

Na ovom mestu potrebno je da izvršimo jednu prihvatljivu konceptualnu generalizaciju. Psihofizika je oblast percepcije, ne oblast koja se odnosi na više i simboličke kognitivne funkcije koje su ovde od interesa. Međutim, psihofizički zakoni, uzeti zajedno sa drugim formulacijama koje pretenduju na status naučnog zakona u psihologiji - poput Šepardovog univerzalnog zakona generalizacije (Shepard, 1987, 2001), Ebinghausovih funkcija zaboravljanja (up. sekciju 7.3), Hik-Hajmanovog zakona (Hick, 1952), funkcije korisnosti ili funkcije ponderisanja verovatnoća u teoriji odlučivanja i drugih - predstavljaju skup psiholoških objašnjenja koje povezuje jedinstvena forma. Sve ove funkcije jesu neka preslikavanja domena *objektivnog* u domen *subjektivnog.* Sve ove funkcije jesu instance objašnjenja putem naučnog zakona u psihologiji: njih definiše odnos bihejvioralnih odgovora, osobina stimulacije i subjektivnih parametara (ako ih određena funkcija sadrži).

Funkcije korisnosti i ponderisanja verovatnoća se u savremenom žargonu teorije odlučivanja ponekad nazivaju *psihoekonomskim funkcijama*, po direktnoj analogiji sa psihofizičkim funkcijama; čak i ova upotreba sličnih terminoloških rešenja upućuje na zajednički teorijski okvir ovakvih objašnjenja. Karakteristika svih ovih funkcija i zakona je ta da su one *potpuno nezavisne* od forme mehanicističkog objašnjenja koja bi mogla ili trebala da razjasni kako „interni mehanizmi uma“ proizvode ponašanje koje je opisano nekom od njih.

Kratka digresija u vezi pitanja kontinuiteta objašnjenja putem naučnog zakona u psihologiji će se pokazati korisnom. Interpretacije prisutne u istoriji psihologije navode na pogrešan zaključak da je rana istorija psihofizike oslonjena na prihvatanje teorijski klimavih koncepata poput *jedva primetnih razlika* (pojam uvodi Veber u formulaciji svojih empirijskih zapažanja). To nije tačno. Gustav Teodor Fehner nudi u svojim „*Elementima psihofizike*“ dve derivacije zakona koji danas nosi njegovo ime, od kojih se nijedna ne oslanja na pojam jedva primetnih razlika, niti na ma koji drugi interni, neopservabilni koncept, bilo da je reč o kvantu senzacije ili nekom mehanizmu koji bi objasnio zašto psihofizička funkcija uzima logaritamski oblik za koji je Fehner verovao da ga uzima (direktan dokaz ovoga na osnovu analize Fehnerovih „*Elemenata*“ daju Dzhafarov & Colonius, 2011, up. Masin, Zudini, & Antonelli, 2009). Dupla greška se zatim sastoji u prenebregavanju činjenice da je Fehner dobro poznavao Bernulijev rad iz 1738, da je u „*Elementima*“ diskutovao Bernulijeve principe *in extenso*, i da je u derivaciji psihofizičkog zakona iskoristio istu pretpostavku koju je Bernuli iskoristio u derivaciji logaritamske funkcije korisnosti. Tvrdnjom da je derivacija u „*Elementima*“ generalnija od Bernulijeve, jer se ne odnosi na moralnu vrednost (termin koji je koristio Bernuli za korisnost), već na ma koji modalitet oseta, Fehnerov tekst je izgleda uspeo da autore poput Boringa navede na to da datiraju početke naučnog objašnjenja putem zakona na 1860. godinu (mada, kako primećuju Masin, Zudini, & Antonelli, 2009, ne daje nikakva objašnjenja zašto bi Fehnerovu derivaciju logaritamskog zakona uzeli kao generalniju od Bernulijeve). Istorijska faktografija u ovom slučaju direktno svedoči da je objašnjenje putem naučnog zakona u psihologiji pogrešno datirano: ono počinje više od sto godina pre Fehnerovog rada, 1738. godine u radu Danijela Bernulija, radu kome treba priznati puno pravo prvenstva u istoriji matematičke, bihejvioralne analize ljudskih mentalnih funkcija. Ono što suštinski razdvaja Vebera i Fehnera od Bernulija jeste upotreba eksperimentalne metode. Bernulijeva analiza, bar u originalnom radu, nije podržana eksperimentalnim opservacijama.

Sa druge strane, ono što suštinski razdvaja Bernulija i Fehnera od savremenog tretmana problema jeste nepotpuna svest o značaju *utemeljenja struktura merenja* odn. odsustvo pokušaja da se *egzistencija osobina* subjektivnih konstrukata (senzacija) regularno poveže sa osobinama stimulacije preko odgovarajuće reprezentacione teoreme. Ovde pozivamo na veoma preciznu interpretaciju prethodno iznete tvrdnje. Fehner je i te kako svestan neophodnosti da njegovo merenje bude dobro utemeljeno; i te kako dobro poznaje gausijansku teoriju statističke greške merenja, uspostavljajući tzv. *klasičan* ili *fehnerijanski model merenja* koji mi danas koristimo u oceni psihofizičkih modela i modela odlučivanja (Link, 1994). Međutim, ono što još uvek ne odlikuje Fehnerovu analizu psihofizičkog problema jeste dublji, stroži poduhvat koji su poduzeli tek Remzi i drugi u XX veku: poduhvat dokaza *egzistencije osobina senzacija* (korisnosti, u kontekstu istorijskog razvoja ove ideje) na osnovu aksiomatskih pretpostavki koje se odnose *isključivo na opservabilne, bihejvioralne varijable.* Ako sada posmatramo razvoj psihofizičkih i psihoekonomskih funkcija od Bernulija u XVIII veku, preko Fehnera u XIX do Remzija, fon Nojmana i Morgnešterna u prvoj polovini XX veka, ugledaćemo jedan jasan kontinuitet ideje o merenju neopservabilnih subjektivnih stanja, kontinuitet koji povezuje ranu Bernulijevu ideju baziranu na nekoliko plauzaibilnih principa o odlučivanju, preko strože Fehnerove analize koja uključuje neizvesnost inherentno povezanu sa merenjem, do Remzijeve stroge analize koja prvo dokazuje egzistenciju funkcije koja se meri, a tek zatim upućuje na njen status kao naučnog objašnjenja. Ako se sad podsetimo diskusija o odlučivanju u sekciji 7.1, jasno vidimo da se savremeni problem kreće negde između Fehnera i Remzija: teorijama odlučivanja baziranim na reprezentacionim teoremama izmiče objašnjenje realnog ponašanja jer ne inkorporiraju greške merenja u svoju konstrukciju, što vodi najnovijim pokušajima unifikacije dva modela: fehnerijanskog modela i modela reprezentacione teoreme (zapravo, modela aditivne teorije združenog merenja, čije razmatranje izbegavamo u ovoj raspravi zbog velike složenosti koje bi njena ekspozicija zahtevala, up. sekciju 7.1).

Upravo ova nejasnoća po pitanju toga na koji način inkorporirati inherentnu neizvesnost procesa merenja u teoriju odlučivanja, i posledično, teoriju psihofizike - jer i ona mora da počiva na rekonstrukciji kroz bihejvioralne odgovore subjekta - predstavlja najjaču poziciju za ma koga ko hoće da kritikuje postojanje naučnih zakona u psihologiji uopšte. Status ovog problema i dan danas, koliko god to moglo da zvuči neprijatno, upućuje na to da je pošten i direktan odgovor na

pitanje o tome da li psihologiju odlikuju naučni zakoni kakvi odlikuju druge prirodne nauke jednostavan i glasi: ne. Negativni zaključci koje je u izveštajima iz 1938. i 1940. objavio Fergusonov komitet, osnovan 1932. od strane Britanske asocijacije za unapređenje nauke, sa zadatkom da „*razmotri i izvesti o mogućnosti kvantitativnih ocena senzornih događaja*“, bili su validni tada kao što su validni i danas (Mitchell, 1999). Fizičar Kembel, autor jednog od najznačajnijih doprinosa teoriji merenja uopšte, lako je ubedio ostatak komiteta sastavljenog od fizičara i psihologa u nemogućnost fundamentalnog merenja subjektivnih konstrukata koje bi zadovoljavalo sve uslove koje merenje u prirodnim naukama zadovoljava. Mičel, jedan od vodećih psihometričara današnjice, primećuje kako se cela diskusija odvija pre *„prave revolucije*“ razvoja aditivne teorije združenog merenja koju će izvesti statističari i psiholozi tek par decenija posle zaključaka Fergusonovog komiteta (Luce & Tukey, 1964). Mičel ne propušta da primeti kako i dan danas ogromna većina psihologa živi u ubeđenju da je fundamente merenja u psihologiji uspešno postavio Stenli Smit Stivens, čije se razumevanje merenja kao „dodeljivanja brojeva objektima posmatranja prema određenom pravilu“ smatra za paradigmatično pogrešan pristup razumevanju koncepta merenja uopšte (up. Mitchell, 1999).

Nažalost, ni primena aditivne teorije združenog merenja, na kojoj počivaju suštinski dokazi teorije izgleda Kanemana i Tverskog[73] nije u stanju da u potpunosti reši problem merenja psihoekonomskih funkcija. Podsetimo se, problem počiva na činjenici da bihejvioralni podaci nalaze način da naruše i teoriju tako elaborirane strukture poput teorije izgleda. Pitanje kako inkorporirati stohastički element u formalnu teoriju izbora - odn. teoriju reprezentativnog merenja - tako ostaje otvoreno, i proganja sve ostale oblasti merenja psihofizičkih i drugih sličnih funkcija podjednako. Objašnjenje koje bi se pozivalo na to da su zakoni psihologije (kao i zakoni savremene fizike, uostalom) inherentno statističke prirode gubi iz vida najtežu posledicu neusaglašenosti klasičnog i reprezentacionog modela merenja: činjenicu da svaki put kada je oboren neki aksiom teorije očekivane korisnosti, ili neki aksiom kumulativne teorije izgleda, nije falsifikovana samo jedna teorija odlučivanja, već je zajedno sa njom falsifikovana čitava teorija merenja neopservabilnih konstrukata na kojoj ona počiva. To je ono što čini savremeni problem toliko teškim.

Sada bi trebalo da je jasno da metodologija racionalne analize, u krajnjoj instanci, pokušava *unifikaciju* mehanicističkog objašnjenja i objašnjenja putem naučnih zakona u psihologiji. Na Marovom nivou 3 kompjutacione analize sprovodi se matematička analiza koja treba da poveže organizam (senzacije, verovanja i druga

neopservabilna stanja) sa sredinom (stimulusima) kroz optimalnu bihejvioralnu funkciju. Psihofizičke i psihoekonomske funkcije su tako viđene kao da predstavljaju optimalne bihejvioralne funkcije u odnosu na neki normativni, adaptivni standard. Na Marovom nivou 2 algoritamske analize suočavamo se sa problemom algoritamske rekonstrukcije tih optimalnih bihejvioralnih funkcija. Na osnovu niza matematičkih pretpostavki, projektuje se interni kompjutacioni sistem čiji rad *mehanicistički objašnjava* kako i zašto optimalne bihejvioralne funkcije u rešenju problema adaptacije uzimaju oblik koji uzimaju. Racionalna analiza pamćenja Andersona i saradnika koju smo diskutovali u sekciji 7.3 predstavlja čist primer objašnjenja u kome ove dve eksplanatorne strategije konvergiraju.

Vidimo da racionalna analiza predstavlja metodologiju sa izuzetno snažnom eksplanatornom strukturom, strukturom koja poštuje sve aspekte naučnog objašnjenja u kognitivnoj psihologiji i strukturom koja ih egzaktno pozicionira na različite, konceptualno ograđene nivoe analize. Međutim, sprovođenje racionalnih analiza u celini je retkost; kao što smo videli, one se najčešće zaustavljaju na kompjutacionim nivou 3. Povezivanje sva tri Marova nivoa u jedinstvenu racionalnu analizu neke kognitivne funkcije nikada nije ni pokušano. Ovo ne čudi ako se ne gube iz vida razmere problema koji nastaje usled izostanka jasne ideje o selekciji parametara modela. Posle kratke istorijske diskusije problematike merenja optimalnih bihejvioralnih funkcija u ovoj sekciji, postaje jasno da se još jedan sloj *veoma teških* problema umeće između tih funkcija i kompjutacionih eksplanatornih mehanizama. Upravo u odnosu na ove veoma teške probleme u fundamentima merenja relevantnih funkcija nalazi se i najveća pukotina u naizgled koherentnoj celini KKP kao projekta naturalizacije epistemičkog subjekta.

## 9.3 Formalizacija I: sud o verovatnoći

Naglasili smo već da je naše interesovanje vezano za analizu racionalnosti saznanja u okviru kognitivne psihologije *kao prirodne nauke*. Dok smo u prethodnim redovima rasvetlili poreklo njenih eksplanatornih mehanizama i diskutovali razvoj složenog problema merenja koji je odlikuje i danas, sada se okrećemo pitanju njene *formalizacije*. Način mišljenja koji ni u principu ne vodi ka formalizaciji tj. razvoju adekvatne matematičke deskripcije ne odlikuje prirodne nauke. Formalizacija u kognitivnoj psihologiji ima dva izvora. Prvi koji diskutujemo odnosi se na danas veoma moderan trend analize kognitivnih problema u terminima teorije verovatnoće

(Oaksford & Chater, 2001, 2009, Chatter, Tenenbaum & Yuille, 2006). Diskusije koje su istorijski vodile razvoju savremene teorije verovatnoće, kako primećuju neki savremeni autori u kognitivnoj psihologiji (Oaksford & Chater, 2009), zapravo su uvek bile filozofske diskusije o tome na koji način probabilistički modeli mogu da predstavljaju modele uma tj. ljudskog mišljenja. Drugi izvor formalizacije kognitivne psihologije, zanimljivo, istorijski je mlađi od teorije verovatnoće, ali je dominirao njenim razvojem sve do 90-ih godina XX veka kada probabilističke teorije počinju da postaju deo establišmenta; njegove duboke korene nalazimo u filozofiji matematike prve polovine XX veka, kada debate između inuticionista, logicista i formalista kulminiraju idejom o mogućnosti mehanizacije procesa matematičkog rezonovanja uopšte - idejom o automatskim, mehaničkim formalnim sistemima simbola, koju smo već predstavili u II delu naše rasprave, i koja je omogućila podjednako razvoj oblasti veštačke inteligencije i savremene kognitivne psihologije.

Savremena debata o racionalnosti, videli smo u III delu naše rasprave, u potpunosti je formulisana u terminima teorije verovatnoće. Teorija verovatnoće, posebno njena subjektivistička, bejzijanska interpretacija, danas predstavlja skoro univerzalno prihvaćen formalni jezik naučne deskripcije u kognitivnoj psihologiji. *Probabilistički obrt* - kako neki autori (Oaksford & Chater, 2009) nazivaju savremeni trend analize kognitivnih problema u probabilističkim terminima u kontrastu prema prethodno popularnim logičkim i semantičkim analizama - u stvari je događaj skorijeg datuma. Kako Oaksford i Čater to formulišu, pokazuje se da su zaključci (na prvom mestu u oblasti ljudskog rezonovanja, up. sekciju 7.4) o iracionalnosti ljudskih kognitivnih funkcija zapravo predstavljali posledice poređenja ljudskog rezonovanja sa *pogrešnim normativnim standardima* logike (Oaksford & Chater, 2001); kada su problemi jednom postavljeni u jeziku teorije verovatnoće, a kognitivni sistem shvaćen kao sredstvo za (makar kvalitativne) probabilističke inferencije, pokazalo se da je moguća reinterpretacija ključnih empirijskih nalaza kao racionalnih. Ogroman deo debate o racionalnosti, videli smo u III delu naše rasprave, upravo počiva na potezu koji podrazumeva probabilističku deskripciju problema adaptacije koji kognitivni sistem, po pretpostavci, rešava.

Istorijski, postavlja se pitanje zašto je toliko uspešan program probabilističke deskripcije kognitivnog sistema morao da čeka 90-te godine XX veka da bi postao deo establišmenta kognitivne psihologije? Odgovor se nalazi u ogromnoj konceptualnoj inerciji u oblastima veštačke inteligencije i kognitivne psihologije koja je nastupila posle inicijalnog optimizma usled otkrića mogućnosti univerzalnog

izračunavanja. Jedan značajan događaj u istoriji psihologije presudnije od drugih je odredio višedecenijsko, skoro unisono verovanje naučnika u mejnstrimu kognitivnih nauka o tome da formalno-logički opis kognitivnih problema nosi prednosti koje probabilistički opis ne može da stekne. Pre svega, otkriće mogućnosti mehanizacije procesa logičkog i matematičkog rezonovanja kroz automatske formalne sisteme odredio je proučavanje jezika (i simboličkih sistema uopšte) kao ključno polje istraživanja u kognitivnim naukama. Jasan program formalizacije sintakse prirodnih jezika koji je predstavio Čomski sredinom 50-ih godina XX veka (Chomsky, 1957) pokazao je da novi matematički aparat - aparat kompjuterskih nauka, upravo - može da ima potencijalno plodnu primenu u analizi toliko značajne problematike društvenih nauka. Čuvena kritika Skinerovog istraživačkog programa koji je ciljao na bihejvioristički opis verbalnog ponašanja od strane Čomskog (čijim ćemo se nekim aspektima posvetiti na kratko u VI delu ove rasprave) kao jedan od suštinskih argumenata sadržala je tvrdnju o tome da realizacija sintaksičkih struktura prirodnog jezika jeste kognitivni proces fundamentalno logičke, a ne probabilističke prirode (Chomsky, 1959/1967). Dok su se 80-ih godina čak i konekcionistički modeli - kao i dalje suštinski *deterministički modeli* - teško nosili sa kritikama simbolicističkog establišmenta (Fodor & Pylishin, 1988), teorija verovatnoće ostajala je tako tema interesantna praktično samo onim psiholozima koji su učestvovali u debati o racionalnosti *pre njene ekspanzije na oblasti van suđenja, odlučivanja i rezonovanja* - ekspanziji koja je karakteristika tek istorijski najnovijih razvoja 90-ih godina XX veka. Tako je, zahvaljujući uspehu koji je jedna paradigma obećavala držeći u rukama „makar i u principu“ rešenje problema kognitivnih nauka, pažnja skrenuta sa pojmova verovatnoće, neizvesnosti i informacije koji dominiraju savremenom probabilističkom verzijom KKP. Teorija igara, rana operaciona istraživanja, teorija informacija i čitav spektar ideja povezanih sa idejom da kognitivne probleme u indeterminističkom svetu može da reši samo kognitivni sistem koji je osetljiv na strukture verovatnoća u takvom svetu, decenijama su ostale van fokusa rasprave. Interesantno je da su neke od ovih ideja (npr. teorija informacija Kloda Šenona) bile veoma značajne u ranoj posleratnoj kognitivnoj psihologiji (up. Kostić, 2006). Dok je kognitivna psihologija razvijala simbolicističke modele kognitivnih procesa, jedna druga društvena nauka se razvijala zahvaljujući matematičkim naporima upravo u oblastima neposredno vezanim za teoriju verovatnoće poput teorije igara i teorije odlučivanja; reč je, naravno, o matematičkoj ekonomiji. Zato se savremena probabilistička KKP, u kojoj debata o racionalnosti predstavlja glavnu (ako ne i

jedinu) teorijsku tematiku, značajno oslanja na koncepte ekonomskih nauka, što je jasno već pri prvom pregledu debate o problemu racionalnog izbora. Matematičku ekonomiju i probabilističku KKP ne vezuje samo skup srodnih problema, već i izbor teorije verovatnoće kao suštinskog sredstva matematičke deskripcije; sve više i više se govori o istraživačkim programima koji jezik matematičke ekonomije koriste kao deskriptivno sredstvo problema kognitivnih nauka, bilo da je reč o neurofiziološkom nivou opisa, gde nalazimo istraživački program *neuroekonomije* (Glimcher, 2004, Glimcher, Camerer, Fehr & Poldrack, 2009), ili o klasičnom kognitivnom nivou opisa, gde se sve češće koristi izraz *psihoekonomija*[74]. Koliko je razvoj kognitivne psihologije do 90-ih godina XX veka karakterisao njen suštinski odnos prema lingvistici, vođen sličnošću simbolicističkih, formalno-logičkih deskriptivnih sredstava u dve nauke, toliko njen savremeni razvoj karakteriše njen odnos prema ekonomiji, sa kojom sada deli probabilističku deskripciju.

Ako se sada podsetimo problema koje analiziraju Bernuli još u XVIII veku, ili Remzi, fon Nojman, Morgenštern i Sevidž (između ostalih) u XX veku, jasno je da je njih danas moguće sagledati kao sastavni deo kognitivnih nauka. Metod merenja intenziteta verovanja koji je prvi predstavio Remzi, videli smo još u II delu ove rasprave, predstavlja fundamentalnu tvrdnju o vezi između verovanja - centralnog teorijskog pojma KKP - i opservabilnih varijabli uopšte, i kao takvog, nemoguće ga je ne sagledati kao deo celine jedne nauke o kognitivnim funkcijama. Prethodni redovi otkrivaju samo zašto se teorija verovatnoće u određenoj epohi razvoja kognitivne psihologije nalazila više u senci nego na glavnoj sceni teorijskih diskusija i eksperimentalnih analiza, ne način na koji je postala suštinski deo prirodne nauke o kognitivnim sistemima - jer ona je to oduvek bila.

## 9.4 Optimizacija: od časovnika ka termodinamičkoj mašini

U sekciji 8. „*Refleksije o debati*“ III dela naše rasprave detaljno smo diskutovali odnos pojmova adaptacije, optimizacije i racionalnosti. Ta diskusija je osvetlila centralni značaj koji odnos ovih teorijskih pojmova ima u metodološkoj paradigmi racionalne analize, ukazujući na to da je značenje „*racionalnosti saznanja*“ uvek moguće relativizovati upotrebem druga dva pojma. Nameće se zaključak da racionalne analize praktično poistovećuju značenje *racionalnosti* neke kognitivne funkcije sa značenjem *uspešne adaptacije*, a da u krajnoj liniji teško ili uopšte ne razlikuju koncept *optimizacije* od prethodna dva.

Istorijsko poreklo koncepata adaptacije i optimizacije u savremenom teorijskom

diskursu kognitivne psihologije ima dva izvora. Koncept adaptacije, naravno, poreklom je iz teorije evolucione biologije, a uticaj na mejnstrim psihološke misli ima još od radova Viljema Džemsa (James, 1890/1950); već smo napomenuli da funkcionalizam u savremenoj kognitivnoj psihologiji odavno ima status implicitnog, bezupitnog teorijskog stava. Metodologija racionalne analize neposredno je motivisana funkcionalističkim stavom. Koncept optimizacije, s druge strane, složenije je strukture, i izgleda se kao jasno određena konceptualna struktura otkriva tek XX veku. Tzv. panglosijanski adaptacionizam u evolucionoj biologiji predstavljao je stav prema koje je svaki adaptivni odgovor nekog živog sistema pritiscima koje na njega vrši sredina („selekcionim pritiscima“) uvek - ili bi trebalo da bude uvek - *optimalan odgovor* tog sistema tim pritiscima (Gould & Lewontin, 1979). Videli smo da savremena panglosijanska pozicija u debati o racionalnosti deli ista shvatanja. Da je neka akcija, neki kognitivni čin, neko ponašanje optimalno, znači da ono *predstavlja najbolji odgovor* motivisanog živog sistema na neko stanje sredine koji taj sistem može da pruži *uzimajući u obzir resurse koji su njemu na raspolaganju.* Resursi koji su nekom sistemu na raspolaganju u procesu adaptacije određuju *ograničenja* pod kojima se determiniše taj njegov optimalni odgovor. Matematički, proces optimizacije ne podrazumeva ništa drugo do nalaženje minimuma ili maksimuma određene funkcije. Sadržaj koji ovaj koncept nosi u empirijskim analizama određen je tek *funkcijom* čiji se se minimumi ili maksimumi traže. Na primer, donosilac odluka u nekoj sredini treba da donese niz odluka tako da maksimizuje ukupnu korisnost za sebe. On ne može da donese niz odluka koji bi mu doneo maksimalnu moguću korisnost jer se u sredini nalaze određena ograničenja koja ga sprečavaju u tome, ili su ta ograničenja vezana za njegove sposobnosti, vreme koje ima na raspolaganju i sl. Uzimajući u obzir sva ta ograničenja, on pokušava da odredi koje odluke treba da donese da bi došao do maksimalne moguće korisnosti *do koje može da dođe pod takvim ograničenjima.* Kada donosilac odluka pronađe koji skup odluka donosi maksimalnu očekivanu korisnost pod datim ograničenjima, on je rešio problem optimizacije; u savremeni teorijskim terminima biologije i kognitivne psihologije, on je tada odredio najbolju moguću adaptaciju kojom može da reaguje na određeni selekcioni pritisak koji sredina vrši.

U upravo opisanom značenju, koncept optimizacije pronalazimo kao predmet zasebne discipline u matematici XX veka - iako su, naravno, rešenja za probleme pronalaženja minimuma i maksimuma funkcija poznata još odavno. Ekonomske nauke su te koje su direktno motivisale razvoj proučavanja optimizacionih problema

u zasebnu matematičku disciplinu u XX veku: problem pronalaženja optimalnog odgovora na skup informacija koje definišu stanje određenog tržišta, pronalaženje optimalnog odgovora u strateškim interakcijama sa konkurentskim ili koalicionim okruženjem (teorija igara), pitanje optimalne alokacije resursa kao suštinsko pitanje ekonomije uopšte - svi ovakvi i slični problemi zahtevaju skup detaljno razrađenih, egzaktnih rešenja za probleme pronalaženja minimuma i maksimuma odgovarajućih funkcija na osnovu kojih ekonomski akteri treba da odrede svoje ponašanje tako da ono maksimizuje njihov profit (Mirowski, 2002). Motivacija nije mogla biti snažnija. Čini se da tek u drugoj polovini XX veka pojam optimizacije - shvaćen matematički - počinje da vrši upliv u biološke nauke i psihologiju. On prvo nalazi direktnu primenu u evolucionoj biologiji i bihejvioralnoj ekologiji, pošto se, posle teorijskih elaboracija koje pripadaju upravo drugoj polovini XX veka (Maynard Smith, 1982, up. poglavlje 11 u Glimcher, 2004), uviđa da on predstavlja perfektnu operacionalizaciju pojma adaptacije. U kognitivnoj psihologiji, koncept optimizacije, videli smo, predstavlja srž tvrdnje o racionalnosti kako se ona shvata u Andersonovoj paradigmi racionalne analize, a to znači da je u diskurs savremenih kognitivnih nauka ovaj pojam ušao *posredno* - upravo preko evolucione biologije i bihejvioralne ekologije, koje obe motivišu razvoj savremene racionalne analize.

Međutim, pojam optimizacije, kako ga sada razumemo, suštinski je vezan za još jednu, nešto stariju istorijsku liniju razvoja naučne misli. Druga polovina XIX veka dovela je do određenih suštinskih promena u studiji fizičkih sistema u okviru *termodinamike*, discipline koja tada počinje da beleži fascinantan razvoj, prenoseći ubrzano uticaj novih ideja koje tada niču u njenim okvirima na fiziku u celini. Formulacija statističke termodinamike - ili *statističke fizike*, uopšte - predstavljaće kulminaciju u ovom značajnom naučnom razvoju. Za razliku od klasične mehanike čiji je razvoj još Njutn postavio na čvrste osnove, proučavanje termodinamičkih sistema se u osvit industrijske ere suočavalo sa problemima koji nisu odgovarali rešenjima preko idealizacija tipa „materijalnih tačaka“ klasične mehanike i sl. Otkriće *nemogućnosti opisa ireverzibilnosti termodinamičkih procesa u terminima teorije klasične mehanike* - odn. nalaz da su idealne mašine koje ne rasipaju energiju *strogo nemoguće* - grubo se kosilo sa reverzibilnom slikom sveta klasične mehanike koja je dominirala fizikom od Njutna od XIX veka. Potreba da se razumeju važne makroskopske varijable poput temperature i takvi fenomeni poput propagacije toplote kroz termodinamičke sisteme diktirale su potragu za deskripcijom koja bi adekvatno povezala osobine mikro-strukture sa

osobinama makro-strukture sistema. Pojmovi poput ansambla, odn. mogućih stanja (definisanih, npr, pozicijama i brzinama/impulsima) čestica u sistemu, počinju da se razvijaju u potrazi za takvim opisom. Posle Klauzijusovog uvođenja pojma *entropije* (1865) i Bolcmanove statističke formulacije ovog koncepta, statistička termodinamika počinje da dobija konačnu matematičku formu u kojoj je i dan danas proučavamo na uvodnim kursevima fizike (Prigogine & Stengers, 1984). Pratiti trajektorije svih čestica u nekom gasu koji koristimo u proučavanju procesa propagacije toplote efektivno nije moguće; koncept entropije, statistički definisan upravo kroz broj različitih mikroskopskih stanja određenog sistema koji odgovaraju *istom* makroskopskom stanju tog sistema, obezbedio je „obuhvatanje“ svih efektivno neizvodivih opservacija u definiciju koja je i eksperimentalno i konceptualno omogućila analizu termodinamičkih problema. Prvi sistemi čije će makroskopske osobine biti nazvane *emergentnim* u filozofiji nauke bili su termodinamički sistemi u fokusu interesovanja fizičara poput Klauzijusa, Bolcmana, Gibsa, Maksvela i drugih u drugoj polovini XIX veka.

Termodinamički sistemi se nalaze u stanju *termodinamičke ravnoteže* kada nema promene u makroskopskim varijablama koje opisuju njihova stanje poput pritiska ili temperature; dva termodinamička sistema će, ako se dozvoli razmena toplote između njih, težiti stanju termodinamičke ravnoteže, a tzv. „nulti“ zakon termodinamike predstavlja zapažanje da je ova relacija tranzitivna: ako su dva tela pojedinačno u termodinamičkoj ravnoteži sa trećim telom, onda su ona i međusobno u termodinamičkoj ravnoteži. Interesantno je otkriće mogućnosti da se formuliše varijabla na makroskopskom nivou opisa termodinamičkih sistema čija *minimizacija* može da reprezentuje težnju tih sistema ka termodinamičkoj ravnoteži. Koncept *Gibsove slobodne energije*[75] je upravo takav koncept: sistem u okruženju kontrolisane temperature i kontrolisanog pritiska se nalazi u termodinamičkoj ravnoteži kada ima minimalnu Gibsovu slobodnu energiju. Termodinamička ravnoteža je stabilno stanje kome određeni sistem teži i koje konačno postiže tokom postepenih interakcija sa svojim okruženjem. To znači sledeće: termodinamički sistem je moguće opisati kao sistem koji *optimizuje* jednu makroskopsku varijablu - Gibsovu slobodnu energiju - i ta optimizacija reprezentuje njegovu težnju ka stabilnom stanju[76].

Razvoj termodinamike u XIX veku doveo je do jedne od revolucionarnih smena paradigmi u istoriji nauke; nisu promenjeni tek fokus rasprave, tek specifične metode, tek eksperimentalna praksa, već je jedna slika sveta *an generale* smenila drugu. Njutnov svet, u kome su sve transformacije dinamičkih sistema reverzibilne, i u

kome je informacija uvek u potpunosti očuvana sve do nivoa inicijalnih uslova, svet za koji je mehanički *časovnik* - struktura izdvojenih, krutih, dobro definisanih delova koji funkcionišu u perfektnoj harmoniji, simetrično u vremenu - predstavljala *sliku*, morao je da ustukne pred novom inspiracijom i intelektualnim interesovanjem koje ju je karakterisalo. Prelazilo se na svet industrijske ere, u kome će sve postati složenije i manje predvidljivo, svet čiji je suštinski simbol apstraktna *termodinamička mašina*, mašina koja gubi informacije tokom rada, teži stabilnosti u termodinamičkom ekvilibrijumu kroz složene interakcije sa svojom okolinom, ultimativno, zajedno sa svim drugim sistemima, sa Univerzumom u celini, proizvodeći ireverzibilni porast entropije, neprestano opominjujući da baš sve, uvek, ima svoju cenu.

Ako sada razumemo na koji način je fizika XIX veka postavila jasan diskontinuitet u način problematizovanja predmeta proučavanja u odnosu na problematizaciju karakterističnu za Njutna i period koji je sledio njegovoj velikoj sintezi, spremni smo da razumemo jednu sličnu promenu paradigme do koje je došlo u kognitivnoj psihologiji 80-ih godina XX veka. Kompleksno isprepletane teorijske paradigme konekcionizma, emergentizma, dinamičkih sistema i konstruktivističkih pristupa (poput enaktivističkog) u svojim najdubljim teorijskim osnovama sadrže upravo ideju koja je potpuna analogija prelasku sa *modela časovnika* kao slike sveta na model *termodinamičke mašine* koji odlikuje fiziku XIX veka. Konekcionističke neuronske mreže, objasnili smo, ne nude nikakvo objašnjenje kognitivnih fenomena na mikro-nivou analize: objašnjenja su u konekcionizmu emergentistička, odn. ona podrazumevaju da su kognitivna stanja na makro-nivou analize supervinijentna nad stanjima na mikro-nivou. Podsetimo se, mikro-stanja u neuronskim mrežama su subsimboličke prirode: ovi modeli pretpostavljaju da je moguće objasniti formalno-simboličko funkcionisanje kognitivnog sistema (koje simbolicistički pristup modelira direktno formalnim sistema) kao posledicu kompleksne evolucije masovnih, paralelnih interakcija između elementarnih, *neinterpretabilnih entiteta*. Postavka problema onda podrazumeva specifikaciju algebarskih struktura koje omogućavaju identifikaciju stabilnih stanja odgovarajućih dinamičkih sistema koji su na nekom nivou deskripcije izomorfni relevantnim sintaksičko/kompozicionalnim formalnim teorijama. Problem, ovako postavljen, je *notorno težak* (up. Petitot, 1995), i pitanje je da li za objašnjenje supervinijencije formalnih sistema kakvi npr. odlikuju strukturu ljudske kompozicione semantike ili sintakse prirodnih jezika uopšte postoje adekvatna matematička sredstva na sadašnjem nivou razvoje nauke. Procesi učenja neuronskih mreža - posmatranih kao dinamičkih sistema za koje je

proces učenja ništa drugo do temporalna evolucija određena njihovom arhitekturom, funkcijama aktivacije, i inicijalnim uslovima - predstavljaju jasnu analogiju su sa temporalnim evolucijama kompleksnih sistema koje nalazimo u statističkoj fizici. Neuronska mreža sa algoritmom povratne propagacije signala greške (v. sliku 7, II deo) koja je završila proces učenja daje optimalne odgovore na stanja sredine zato što je za njenu *funkciju greške* - funkciju koja meri odstupanja njenih odgovora od očekivanih odgovora - pronađen skup parametara takav da ona ima minimalnu vrednost u odnosu na sve moguće kombinacije stanja sredine koja mogu da se pojave na input čvorovima. Parametri u ovom problemu optimizacije su, naravno, težine veza između čvorova u mreži. Sada bi trebalo da jasno razumemo sledeće: konekcionistički modeli, kao i dinamički modeli u kognitivnoj psihologiji uopšte, *predstavljaju doslednu primenu prirodnih optimizacionih procesa na kognitivne probleme*. Sa stanovišta istorije nauke, onda, očigledna postaje analogija koja povezuje istoriju kognitivnih nauka sa istorijom fizike: u prelasku sa načina mišljenja koje odlikuje neokartezijanski pristup Čomskog, Fodora, Pilišina i drugih simbolicista u kognitivnim naukama ka načinu mišljenja koji ispituje simboličke strukture kao posledice kompleksnosti vremenske evolucije kognitivnih sistema, jasno se ocrtava *model naučne revolucije* koji je odlikovao smenu paradigmi u fizici XIX veka. Na neki način, dakle, vidimo da kognitivne nauke druge polovine XX veka kao da kroz svoju istoriju izvode rekapitulaciju istorije prirodnih nauka sa kojima pokušavaju da sebe postave u kontinuitet.

Posle ove diskusije, čitalac sa pravom može da se zapita: iz kojih razloga onda, u teorijskoj diskusiji o racionalnosti, nije posvećeno više prostora konekcionističkim modelima u raspravi koja se nalazi pred nama? Ako konekcionistički modeli predstavljaju otelotvorenja sistema koji kroz prilagođavanje rešavaju optimizacione probleme, a savremena debata o racionalnosti suštinski počiva na teorijskom modelu koji praktično izjednačava racionalnost sa adaptacijom i optimizacijom, nisu li konekcionističke, dinamičke, emergentističke teorije mogle da budu glavni akteri rasprave? Dva faktora su, verujemo, presudno uticala na to da konekcionisti i pristalice srodnih teorijskih pravaca praktično ni ne uzmu ozbiljnog učešća u debati o racionalnosti. Prvi je taj što su konekcionistički modeli postavljeni na Marovom algoritamskom nivou 2, i na tom nivou predstavljaju veoma složene klase modela. Povezivanje nivoa 2 u ovim modelima sa kompjutacionim nivoom 3 - što je korak u racionalnoj analizi koji se retko sprovodi i za mnogo jednostavnije modele od konekcionističkih - podrazumeva detaljne studije odnosa između interakcija

na mikro-nivou (nivo 2) i strukture temporalne evolucije neuronskih mreža kao dinamičkih sistema na makro-nivou. Ovakve analize su tek ekstremno složene i teorijski se diskutuju u kognitivnoj psihologiji praktično samo na nivou didaktičkih modela (mada neki konekcionisti uspevaju da ekspliciraju i teorijski iskoriste ove ovakve veze, up. Rogers & McClelland, 2004). Čitaocu ostavljamo samo da zamisli kompleksnost analize u kojoj bismo na osnovu ma kakvog aksiomatskog okvira pokušali da rekonstruišemo interne funkcije ovakvih kognitivnih modela, uspostavimo neophodnu reprezentacionu teoremu i otvorimo debatu u duhu one koja je odredila istoriju proučavanja problema poput racionalnog izbora. Dakle, jedna granica preko koje još nismo prešli - i koja nas, za sada, fundamentalno ograničava - jeste granica kompleksnosti. Drugi faktor za koji verujemo da je presudno uticao odsustvo konekcionista i pristalica dinamičkog pristupa uopšte iz debate o racionalnosti jeste, verujemo, izbor problema kojima su oni odlučili da posvete pažnju. Centralne teorijske rasprave u konekcionizmu su se skoro po pravilu fokusirale oko pitanja psiholingvistike (up. Pinker & Ulman, 2002, McClelland & Patterson, 2002) u kojoj debata o racionalnosti nije razvijena kao u drugim oblastima viših kognitivnih funkcija koje diskutujemo u ovoj tezi.

Konačno, konekcionistički i dinamički pristup su danas suočeni sa određenim problemima koje samo radikalna novina u njihovim teorijskim naporima može da otkloni; reč je problemu uspostavljanja *konstituentne strukture*, sa kojima ih je suočila još rana kritika Fodora i Pilišna (Fodor & Pylyshin, 1988, Fodor, 1997), koji pokazuje da konekcionistički modeli koje danas poznajemo nisu u stanju da dinamički „simuliraju“ neke suštinske odlike relacija koje se koriste u simbolicističkim teorijama, zbog čega efektivno ne mogu da objasne simboličke kognitivne procese koje pokušavaju da objasne. Savremena teorijska debata povtrdila je da, u formi koju sada uzimaju, konekcionistički modeli ne mogu da pruže odgovor na ovaj izazov (Marcus, 2001). Problem je prepoznat i prihvaćen od strane tzv. *implementacionih konekcionista* - koji napuštaju strogo određen program tzv. *eliminativnog konektivizma* (Smolensky, 1987, Prince & Smolensky, 1997) - priklanjajući se gledištu prema kome rezultat istraživanja treba da pokaže način na koji konekcionistički modeli u svojim faznim prostorima mogu da simuliraju formalno-simboličke strukture; ipak, ni implementacioni konekcionizam još nije ponudio zadovoljavajuće rešenje (Fodor & McLaughlin, 1990). Kao što smo već primetili, ovaj problem se - kada mu se pristupi matematički - otkriva kao izuzetno težak, i verovatno će čekati razvoj sasvim novih analitičkih sredstava kako bi dobio

zadovoljavajuć tretman (Petitot, 1995). Do tada, ostaće tačno da simbolicistička paradigma makar i samo u principu može da ponudi rešenja problema za koje dinamička, konekcionistička teorija još uvek nema razvijen analitički tretman.

## 9.5 Formalizacija II: elementarne intuicije uma

*Sekcija koja se nalazi pred nama zahteva apstraktniji odnos prema nekim problemima koje smo do sada diskutovali.* U II delu ove rasprave upoznali smo se sa fundamentalnim tvrdnjama KKP. Njen centralni teorijski program, onaj koji u najveći meri danas predstavlja establišment ove nauke, videli smo, shvata realizaciju kognitivnih procesa i funkcija kao eksplicitnih izračunavanja u automatskim, fizičkim sistemima simbola: zbog toga se često u teorijskim diskusijama naziva simbolicističkim pristupom[77]. Tradicija mišljenja kojoj pripada simbolicistički pristup, podsetimo se, potiče iz klasičnih rasprava u filozofiji matematike prve polovine XX veka. Stavovi tri „velike škole" filozofije matematike, logicizma, formalizma i intuicionizma tokom dvadesetih i tridesetih godina XX veka izoštrili su se u sukob mišljenja formalista i intuicionista; pristalice jednog i drugog pravca okupljale su genijalne matematičare, filozofe i logičare. Početni stav logicista - najkompletnije oličen Raselovom i Vajthedovom obimnom studijom „*Principia Mathematica*" (Whitehead & Russell, 1910, 1912, 1913) - da je sve matematičke tvrdnje moguće svesti na tvrdnje odgovarajućeg logičkog sistema, napušten je pred strožim i konceptualno još čistijim programima zasnivanja matematike formalista i intuicionista. Logicistički program, uz izmene koje bi bile praktično čisto semantičke prirode, mogao bi danas da se posmatra kao specijalan slučaj formalizma, dok s druge strane logicisti nisu ni razmatrali neka od pitanja i neke od pristupa koje će formalisti izneti pred intuicionističku kritiku i *vice versa.* Uprkos činjenici da je establišment savremene matematike po svojoj prirodi daleko više orijentisan formalistički, stavovi intuicionista i dan danas se ozbiljno studiraju; na stranu što su upravo napori onih koji su pokušavali da dosledno izgrade intuicionistički argument, poput Tjuringa, doveli do rezultata koje danas shvatamo kao kamene temeljce kompjutacionističke revolucije u nauci uopšte. Da bismo razumeli neke od problema sa kojima mi želimo da suočimo savremenu KKP, moramo da se vratimo na trenutak u ta slavna vremena filozofije i matematike i dotaknemo se nekih od najsloženijih pitanja koje je ljudski um uopšte diskutovao kroz istoriju mišljenja. Opšta karakteristika rasprave između škola filozofije matematike pre Drugog svetskog rata, u celini posmatrano, jeste zajednički napor na svim stranama da se pokaže

(ili opovrgne) mogućnost jednog minimalističkog, *konstruktivističkog* i *finitističkog* zasnivanja matematike[78]. Mi smatramo da se u argumentima ponuđenim tada, skoro sto godina pre nego što smo mi započeli ovu diskusiju, nalazi dovoljno razloga da se pred KKP postave *pitanja zasnivanja* veoma slična onima koja su tada postavljena pred matematiku i matematičara kao njenog delatnika.

Pitanje prirode matematičke inferencije uopšte široko je problematizovano posle Kantorove demonstracije postojanja transfinitinih kardinalnih brojeva većih od kardinalnosti skupa prirodnih brojeva (između 1874 i 1884). Da postoji jedna, *potencijalna beskonačnost*, sledi iz razvoja ordinalnih (rednih) brojeva po sebi, koje prirodno vodi uvođenju koncepta transfinitnih rednih brojeva u diskurs matematike, kako Hilbert zapaža: "*Do njih dolazimo prostim nastavljanjem brojanja preko redno prebrojivog beskonačnog, odnosno prirodno i jedinstveno determinisanim nastavljanjem uobičajenog konačnog nabrajanja*" (Hilbert, 1925):

$1, 2, 3, ...$
$\omega, \omega + 1, \omega + 2, ...$
$\omega \cdot 2, (\omega \cdot 2) + 1, (\omega \cdot 2) + 2, ...$
$\omega \cdot 3, (\omega \cdot 3) + 1, (\omega \cdot 3) + 2, ...$
.
.
.
$\omega^2, \omega^2 + 1, ...$
.
.
.
$\omega^2 + \omega, \omega^2 + \omega \cdot 2, \omega^2 + \omega \cdot 3, ...$
$\omega^2 \cdot 2, (\omega^2 \cdot 2) + 1, ...$
$(\omega^2 \cdot 2) + \omega, (\omega^2 \cdot 2) + (\omega \cdot 2), ...$
$\omega^3, ...$
$\omega^4, ...$
$\omega^\omega, ...$

Uvođenjem operacije tetracije (koristi se najčešće u ilustrativne svrhe u kursevima teorije transfinitnih rednih brojeva, up. Rucker, 1982/2005) dobijamo najbrži način da redno nabrajamo; pomoću nje najbrže "dostižemo" prve transfinitne redne brojeve. Oni još uvek nisu transfinitni kardinali, koji odgovaraju na pitanje

o količini, ali je već u ovom jednostavnom razvoju prvih transfinitnih brojeva prisutan osećaj obevaze prema finitističkom pristupu (Hilbert: "*... odnosno prirodno i jedinstveno determinisanim nastavljanjem uobičajenog konačnog nabrajanja*", Hilbert, 1925).

Drugim rečima, ako matematičar redno navodi prirodne brojeve, 1, 2, 3, .., oslanjajući se npr. na elementarno rekurzivno shvatanje nizanja prirodnih brojeva prema kome je sledbenik prirodnog broja prirodni broj, npr. $n \in \mathbb{N} \Leftrightarrow succ(n) \in \mathbb{N}$, u jednom trenutku on može da u univerzum diskursa uvede *transfinitni redni broj*, neko $\varepsilon$, oslanjajući se na samorazumljivost njegovog dosezanja u tom univerzumu diskursa koja počiva na očiglednoj, rekurzivnoj osobini nizanja prirodnih brojeva. *Kardinalni broj* nekog skupa predstavlja broj njegovih elemenata; na osnovu prethodnog, za skup prirodnih brojeva, $\mathbb{N}$, kažemo da ima prebrojivo beskonačno mnogo elemenata, u oznaci $\aleph_0$. Kardinalni broj skupa prirodnih brojeva $\aleph_0$, dakle, stoji za onu „intuitivnu beskonačnost" čiju predstavu dosežemo umom kada apstrahujemo osobinuređanja niza prirodnih brojeva kroz elementarnu rekurzivnu reprezentaciju. U carstvu matematičkog mišljenja nikakvih problema, verovatno, ne bi bilo, da nemački matematičar Georg Kantor, upotrebom danas dobro poznate tehnike *dijagonalizacije* (up. Rucker, 1982/2005) nije pokazao da se sa lakoćom konstruišu skupovi čiji su kardinalni brojevi *veći* od $\aleph_0$, te da je $\aleph_0$ tako tek prvi transfinitni kardinal u čitavoj procesiji transfinitnih kardinalnih brojeva koji danas predstavljaju univerzum proučavanja moderne teorije skupova.

Makar i površno upoznavanje sa prirodom Kantorovog postupka dijagonalizacije koja vodi ka ovim potpuno kontraintuitivnim zaključcima pomoćiće nam da motivišemo raspravu koja sledi. Organizujmo naše mišljenje na sledeći način: u prvoj koloni navodimo sve prirodne brojeve, pretpostavljajući da možemo sve skupove prirodnih brojeva, koje navodimo u drugoj koloni, da dovedemo u korespondenciju 1:1 sa njima:

| $\mathbb{N}$ | podskupovi od $\mathbb{N}$ | 0 | 1 | 2 | ... |
|---|---|---|---|---|---|
| 0 | neparni brojevi | **0** | 1 | 0 | ... |
| 1 | parni brojevi | 1 | **0** | 1 | ... |
| 2 | umnošci broja 3 | 0 | 0 | **0** | ... |
| ... | ... | ... | ... | ... | ... |

U prvom redu navodimo ponovo sve prirodne brojeve, a ćelije tabele popunjavamo sa 1 ukoliko broj iz prve kolone pripada skupu u odgovarajućem redu, i sa 0 u suprotnom. Biramo po jednu „osobinu“ prirodnih brojeva u drugoj koloni kao osnovu za svrstavanje nekih elemenata skupa prirodnih brojeva u novi skup, i potencijalno tako listamo sve osobine prirodnih brojeva. Skup svih takvih podskupova čini partitivni skup skupa prirodnih brojeva. Na ovaj način smo kodirali sve moguće podskupove od $\mathbb{N}$ nizom simbola 0 i 1; uzmimo sada sekvencu *dobijenu na dijagonali*, i zamenimo sve simbole 0 simbolom 1 i obrnuto, dobijajući tako sekvencu koja *izvesno reprezentuje neki podskup od* $\mathbb{N}$: *ona se svakako neće naći u tabeli koja pretenduje da izlista* sve *takve podskupove.* Jednostavno, razlikovaće se od podskupa koji reprezentuje prvi red na prvoj poziciji, od podskupa koji reprezentuje drugi red na drugoj poziciji, i tako *ad infinitum.* Zaključak? - Partitivni skup prirodnih brojeva, odn. skup svih podskupova $\mathbb{N}$*, ne može se dovesti u 1:1 korespondenciju sa prirodnim brojevima u* $\mathbb{N}$. Pošto su prirodni brojevi prebrojivi, njihov partitivni skup onda *mora biti neprebrojiv*, te njegova kardinalnost mora biti veća od $\aleph_0$; partitivni skup prirodnih brojeva dakle ima kardinalnost $\aleph_1$, što direktno vodi u sledeći zaključak: postoji *beskonačnost veća od beskonačnosti* koja nam govori koliko ima elemenata u skupu prirodnih brojeva.

Zaključak do kojeg nas je doveo izbor prethodnog misaonog puta, slobodno možemo to reći, na sudbonosan način je odredio stavove prema nekim od najdubljih pitanja kojima se filozofija ikad, uopšte obratila. Teren koji je odabran kao podloga za sučeljavanje mišljenja nije mogao biti bolji: najmanje od Kantove prve Kritike (1781) matematičko saznanje se uzima kao jedini raspoloživi egzemplar *izvesnog* saznanja uopšte; svaka teorija saznanja, dakle, maksimalno profitira ukoliko dokaže da je sposobna da fundira upravo najegzaktniju moguću formu inferencije: onu koja

odlikuje simboličko-logičku odn. matematičku misao. Naš argument, koji tek sada možemo da počnemo da izgrađujemo, jeste da ni pokušaj da se izgradi pozitivna, prirodna nauka o saznajnim procesima - *što KKP tvrdi da jeste* - ne može biti oslobođen obaveze da se odredi prema ovim pitanjima. Kantorov razvoj egzistencije transfinitinih kardinala izazvao je dve skoro oprečne reakcije u filozofiji matematike prve polovine XX veka, reakcije koje se ogledaju u stavu formalističke škole, s jedne strane, i intuicionističke škole, sa druge strane. Suština sukoba da se izraziti, posle prethodne demonstracije dijagonalnog postupka, relativno jednostavno. Za intuicioniste u filozofiji matematike, poput Bruera, argument nije validan i ne postoji nikakav transfinitni kardinal poput $\aleph_1$; argument se bazira na činjenici da ne postoji jednačina čije bi $\aleph_1$ bilo rešenje, ili, uopšte, da ne postoji nikakav *konstruktivan, finitistički postupak koji bi pokazao egzistenciju nekog* $\aleph_1$ *kao svoj rezultat* (Brouwer, 1912; v. belešku br. 79 uz ovu tezu). Intuicionisti će reći: Kantor ni na koji način nije demonstrirao egzistenciju nekog $\aleph_1$ *time što je pokazao da ne može biti drugačije u diskursu rasprave do da takav objekat postoji*; za intuicioniste, *objekat matematike će postojati ako postoji neposredan, formalizovan, konačan skup koraka u procesu izračunavanja koji kao svoj rezultat iznosi taj objekat.* Za formaliste u filozofiji matematike, argument jeste validan, pošto će oni, sledeći Hilberta, shvatiti matematiku uopšte kao igru simboličkog karaktera (Hilbert, 1925, von Neumann, 1931/1987) - i sada upućujemo čitaoca da se podseti definicija formalnih sistema i njihove uloge u osnovama savremene KKP diskutovane u II delu ove rasprave - u kojoj Kantorov argument ne čini ništa drugo do što pokazuje da uvođenje transfinitnog kardinalnog broja poput $\aleph_1$ *nije u suprotnosti ni sa čim što je već prethodno uvedeno u diskurs rasprave sredstvima koja su već bila raspoloživa u tom diskursu.* Ove komplikovane stavove, od suštinskog značaja za raspravu o osnovama ma kakve nauke o saznanju, sada možemo da sažmemo u oblik prepoznatljiv tradiciji koja je prisutna i danas: dok je intuicionističko zasnivanje matematike suštinski *psihološko*, zahtevajući da egzistencija svakog novog objekta koji se uvodi u diskurs rasprave bude demonstrirana tako *da subjekt koji razvija taj diskurs rasprave ima neposredno iskustvo sa tim objektom* (tj. da taj objekat predstavlja rezultat sleda operacija izračunavanja, što ga čini *neposrednim*), formalističko zasnivanje matematike počiva na suštinski semiotičkim osnovama, zahtevajući samo da sintaksička ograničenja ne budu narušena, odn. da se održi slaganje novouvedenog objekta u diskurs rasprave sa prethodno uvedenim objektima, *bez obzira na to u kojoj meri bi značenje novouvedenog objekta moglo da se problematizuje.*

Shvatanje matematike kao proučavanja formalnih sistema karakterisaće Hilbertovu formalističku školu mišljenja, za koju smo već konstatovali da je dominantno uticala na stavove savremenih matematičara, koji nemaju problema sa prihvatanjem teorije skupova koja barata transfinitnim kardinalnim brojevima. U II delu rasprave pokazali smo na koji način su formalni sistemi fundamentalni u razvoju KKP kao projekta naturalizacije nauke o saznanju. S druge strane, napori (makar u nekom trenutku) intuicionistički orijentisanih matematičara dovešće do razvoja koncepta univerzalnog izračunavanja (Turing, 1936): ono nikada neće demonstrirati egzistenciju transfinitinih kardinala (što intuicionista ne bi ni zahtevao), ali će zato demonstrirati da je svaka intuitivno izračunljiva funkcija izračunljiva unutar određenog formalnog sistema koji je ekvivalentan sistemu univerzalne Tjuringove mašine, otvarajući tako put ka konceptu mehaničkih, automatskih formalni sistema simboličkog izračunavanja: teoriji KKP, drugim rečima. Sve diskutovano do sada u ovoj sekciji predstavljalo bi samo izlet u poglavlje matematike i filozofije matematike *da jedna očigledna istorijska činjenica ne remeti tu sliku*, a ona upravo svedoči o sledećem: u istoriji razvoja standardne paradigme KKP negde dolazi do prelaženja preko problema zasnivanja, koji se odnosi na *izbor* između *semiotičkog odnosa* prema formalnim sistemima (karakteristika formalizma) i *psihološko-finitističkog odnosa* prema formalnim sistemima (karakteristika intuicionizma). Tu odluku je *neko* morao (ili *neko* mora) da donese, a ako to nije ljudski um - uz sve apstrakcije koje nisu strane ni empirijskoj metodologiji - ostaće nejasno *ko*; taj ljudski um je, podsetimo se, *određen kao predmet proučavanja KKP*.

Intuicionistički pristup zasnivanju matematike otvoreno se karakteriše kao psihološki, pri čemu se tu ne tvrdi egzistencija nekog konkretnog psihološkog subjekta, već neke vrste *idealizovanog uma* (Iemhoff, 2012) koji vrši operacije izračunavanja i u čijem iskustvu se javljaju njihovi rezultati. Međutim, koliko god ova apstrakcija odgovarala matematici koja se bori za svoju nezavisnost od psihologije kao takve - što je opravdano u razvoju ma koje formalne discipline - ostaje potpuno nejasno kakav stav bi ma kakva pozitivna psihologija trebalo da zauzme po tom pitanju. Ako ona podrazumeva proučavanje ma kakvog realizovanog, neidealizovanog uma, za koji pretpostavlja da je u stanju da izvede ono što intuicionista u matematici zahteva od njega - a podsetimo se da standardna paradigma KKP tvrdi *upavo to* - ona očigledno nasleđuje problem izbora između intuicionističkog i formalističkog stava. Međutim, u jednom fundiranoj pozitivnoj nauci ovakva vrsta problema *ne sme da postoji*: ako se ovakav problem u njoj

prepoznaje, ostaje samo da se konstatuje da je njena *separacija od filozofije nedovršena*. Simetričan argument može da se izvede ako se postavi pitanje odnosa zasnivanja neke pozitivne psihologije i Hilbertovog formalističkog stava. To što strategija razvoja formalizma izbegava da za semiotički sistem koji gradi uopšte postavi pitanje interpretacije (van čisto logičkog shvatanja interpretacije kao otelotvorenja u modelu neke teorije), odn. da postavi pitanje *egzistencije interpretatora*, ne može da ograniči potencijalnu pozitivnu nauku o saznanju da takvo pitanje ne postavi; ona je, naime, *u obavezi da o njemu govori*. Nastavimo ovim putem i shvatićemo da Hilbertovi napori, ako se analiziraju na najelementarnijoj ravni, analogno strogosti zahteva koji postavlja intuicionizam kad postulira idealizovani um matematičara, zavise od mogućnosti ma kakve (makar minimalne) semiotike, a to direktno znači survavanje rasprave u odnose unutar Ogden-Ričardsovog trougla i postavljanje pitanja prirode posredovanja između znaka i označenog, pitanja koje semiotika unisono prepoznaje kao psihološko pitanje (de Saussure, 1916/1977). Ako je tako: koja nauka onda fundira koju - filozofija matematike psihologiju (što je, čini se, red inferencije koji vodi nastanku savremene KKP u istorijskom sledu događaja), ili psihologija (semiotika) filozofiju matematike (pa i matematiku samu, što intuicionistima ne bi bilo strano)? Unutar savremene, simbolicističke KKP, sasvim očigledno, ovo pitanje uopšte ne može ni da se razmatra - iako ona *post hoc* uzima sebi za pravo da domen simboličkog i kompjutacionog odredi kao inherentan diskursu rasprave koju vodi.

## 9.6 *Zeitgeist* kompjutacionizma

Od revolucije naučnog racionalizma XVII veka, kako se kanonski još uvek određuje „početak“ epohe savremenog racionalnog promišljanja Univerzuma, do danas, grubo možemo da prepoznamo smenu tri globalne slike sveta. One obuhvataju sve metodološke odluke i reprezentuju implicitna ontička ubeđenja odgovarajućih epoha; treća od njih je aktualna epoha, epoha koju slute filozofske, logičke i matematičke rasprave prve polovine XX veka, ali koja se otkriva i osvešćuje tek po završetku Drugog svetskog rata. To je epoha kompjutacionog, kibernetičkog i kognitivnog: *epoha simboličke mašine*. Dok u svojoj genezi ona referira na duboke probleme zasnivanja matematike, logike, pa time i racionalne misli uopšte (up. 9.5), u drugom koraku ona se otkriva kao slutnja o mogućnosti pozitivne teorije saznanja u istraživačkim programima kognitivne psihologije i veštačke inteligencije (up. II deo), dok u trećem, najnovijem koraku, ona ispituje svoje granice kao slike

sveta, otkrivajući pretenziju ka opštem opisu Univerzuma u kompjutacionističkim terminima (up, 5.1.2, 5.2.2, Wolfram, 2002). Aktualna slika sveta u istorijskom diskursu naučnog racionalizma reprezentuje *epohu simboličke mašine*, kojoj prethode slike *epohe termodinamičke* i *epohe klasične mašine*. Najraniju, epohu klasične mašine, koju grubo datiramo do sredine XIX veka, do možda 1860-ih, odlikuje shvatanje Univerzuma kao potpuno reverzibilnog, determinističkog dinamičkog sistema. To je epoha koju najbolje predstavljaju Njutn i Laplas. Drugu, epohu termodinamičke mašine, odlikuje shvatanje Univerzuma kao ireverzibilnog i stohastičkog dinamičkog sistema. Treću, epohu simboličke mašine, teže je okarakterisati, jer je njena bitna odlika *početak matematizacije društvenih nauka* u kibernetici, teoriji igara i kompjutacionizmu: ona se ontički razlikuje od prethodne dve. Prostor-vreme fizike u njoj zamenjuje fazni prostor *ma kog dinamičkog sistema*; njen Univerzum nije fizički Univerzum, već simbolički Univerzum, Univerzum zamislivog; njega ne nastanjuju fizička tela, već simbolički objekti koji zauzimaju „*logička mesta*“ (izraz dugujemo Vitgenštajnu, Wittgenstein, 1921/1987). Koje fazne prostore, odn. koje klase sistema možemo da razmatramo u opštem jeziku kompjutacionih procesa? *Bilo koje*. U domenu važenja Čerč-Tjuringove teze, prema kojoj je klasa intuitivno izračunljivih funkcija ekvivalentna klasi funkcija koje može da izračuna Tjuringova mašina ili njoj ekvivalentan formalizam, bilo koji skup varijabli, bilo koja dinamička struktura, bilo koja trajektorija bilo kog dinamičkog procesa postaje predmet diskursa. Uopšte, za epohu kompjutacionizma, za svet koji se odslikava u simboličkoj mašini, bilo šta što ne može da se postavi kao algoritamski rešiv problem - uopšte nije problem, ako je tako nešto kao što bi bio algoritamski nerešiv problem *uopšte zamislivo*. U epohi kompjutacionizma sve je predmet kodiranja i izračunavanja; ona supstanciju i energiju zamenjuje informacijom, a naučni zakon - sintaksom, u procesu kroz koji koncept informacije postaje centralni koncept naučne analize uopšte: *kompjutaciona fizika*, *dekodiranje genetskog koda*, *formalni sistemi ljudskih simboličkih funkcija*, *ekvilibrijumi evolucione teorije igara u bihejvioralnoj ekologiji*, *optimizacija pod pritiskom selekcionih restrikcija u evoluciji*, sve klasični naučni problemi prošlih vremena - svi reformulisani u jeziku kompjutacionizma. Nije neophodno spekulisati zajedno sa Volframom o Univerzumu kao kompjuteru *de facto*. Kompjutacionizam je *zeitgeist* - ako nije paradigma, jer je preširok za paradigmu i u svakoj naučnoj disciplini se ostvaruje na svoj način - i ako nije jedinstvena teorija, već pre interpretativna shema, *kod* savremenosti od kraja Drugog svetskog rata do današnji dana. Kompjutacionistički *zetigeist*:simbolička

mašina kao oličenje vladajućih navika deskripcije i mišljenja - predstavlja najširi kontekst za istorijsku interpretaciju kognitivne psihologije kao projekta razvoja *naturalističke teorije uma*.

Prethodni redovi ukazali su na konceptualne izvore same KKP. Videli smo, međutim, da isti koncepti, već raspoloživi u istoriji nauke, čija je rekompozicija odredila establišment u posleratnoj kognitivnoj psihologiji, igraju ključnu ulogu *u određenju racionalnosti u KKP*. Ako ispitamo te koncepte na najdubljem nivou, gde nalazimo raspravu intuicionista i formalista u filozofiji matematike, i ingeniozan doprinos intuicionistički orijentisanih naučnika u razvoju koncepcije univerzalnog izračunavanja, suočavamo se sa dve grupe problema. S jedne strane, videli smo da te koncepte odlikuje neka vrsta *post hoc* karaktera u raspravi o kognitivnoj psihologiji, jer se u samom njihovom razvoju prelazi preko određenih pitanja čija relevantnost za psihološku teoriju saznanja ne bi mogla biti veća. Rešavanje pitanja zasnivanja - bez obzira da li se ono rešava na način bliži formalistima ili intuicionistima - onemogućava rešavanje pitanja zasnivanja kognitivne psihologije, jer se rešenje odnosi na *odluku* o prirodi interpretacije matematičkih i logičkih objekata. Ostaje nespecifikovano *ko*, i *kako*, donosi tu odluku; možda se subjekat saznajnog procesa koji bi kognitivna psihologija trebalo da proučava na samom početku logičkog razvoja rasprave ostavlja bespovratno izgubljenim za buduće analitičke napore koji će se odvijati u lavirintu matematičkih modela i teorija pošto je bar jedna suštinska odluka o njegovoj prirodi već doneta.

Posle proučavanja ovih istorijskih izvora, postaje jasno da KKP zauzima centralno mesto u konceptualnoj mreži savremenih nauka, a ne tek teorijski projekat naturalizacije psihološkog i simboličkog koji na neki način pati od nedostatka egzaktnosti kakva odlikuje fizikalne nauke. Generacije psihologija školovane su da posmatraju fiziku kao nauku koja pruža model prirodne nauke kakva bi i psihologija trebala da postane, dok smo danas svedoci vremena u kome jedna disciplina opisuje kognitivne funkcije terminima koji su za nju fundamentalni, u odnosu na njeno istorijsko poreklo, a prirodne nauke - koje su joj prethodno nametale naturalistički model - preuzimaju te termine i prilagođavaju svoje strukture opisu koji nastaje na osnovu njih. Psihologija se, u osvit kompjutacionističkog *zeitgeist-a*, kroz razvoj KKP transformisala u nauku informacione epohe istom brzinom, ako ne i brže od fizike. Međutim, nešto u teorijskoj konstrukciji KKP nije na svom mestu; to zna i oseća svaki naučnik u ovoj oblasti kada je suočen sa činjenicom da se (a) fundamentalna teorija nije promenila već više od pola veka, (b) da je fundamentalna

teorija postavljena tako da vodi u „očigledne zaključke“ o mogućnosti naturalističke realizacije i objašnjenja simboličkog, te da (c) rad pod fundamentalnom teorijom nije doveo do očekivanih rezultata kroz decenije - o čemu svedoči značaj kritika poput Drajfusove, koja već decenijama ne gubi na snazi (a da se ne oseća potreba da se sadržaj te kritike uopšte promeni). Mi smo u ovom pregledu istorijskih izvora savremene KKP ukazali na ono što smatramo najvećom preprekom u projektu naturalizacije psihologije saznanja: naime, *koncept merenja u ovoj nauci još uvek nije zasnovan sa sigurnošću koja odlikuje fizikalne nauke.* KKP je *samo teorija*, i to je njen najveći problem: to je teorija koja pati od ogromnog problema loše povezanosti svojih teorijskih koncepata sa naučnim opservacijama, problema koji se otkriva u utemeljenju suštinskih koncepata poput psihofizičke funkcije i funkcije korisnosti. Taj rascep između onoga što može biti mehanička teorija uma, i onoga što je neophodno da bi se takva teorija uma naturalizovala, videli smo, datira još od prvih rasprava racionalista i empirista o mogućnosti takve teorije. Takva vrsta rasprave, najmanje do rada Gustava Teodora Fehnera, nije osećala obavezu da uzme u obzir ograničenja koja slede iz postupaka merenja. Savremena KKP kao da je još jednom prešla preko tog problema, ostavljajući tako svet teorijskog i svet empirijskog na - *za prirodnu nauku* - nedopustivoj distanci jedan od drugog.

Posle ovih redova, teorijske sinteze KKP u II delu, i kritičkog osvrta na debatu o racionalnosti u III delu, trebalo bi da smo konceptualno potpuno pripremljeni za zaključke o debati o racionalnosti u savremenoj kognitivnoj psihologiji. Ipak, mi ćemo pre donošenja tih zaključaka napraviti još jedan korak. Taj korak će nas na trenutak postaviti u ulogu učesnika u samoj debati o racionalnosti. Nadamo se da ćemo u narednim redovima uspeti da razvijemo neke nove perspektive iz kojih može da se sagleda problem racionalnog izbora - tradicionalno centralni problem debate o racionalnosti. Za razvoj tih novih perspektiva neophodni su koraci eksperimentalne prakse.

# Deo V

# Teorija racionalnog izbora

Vraćamo se još jednom problemu koji nas prati kroz celu diskusiju: problemu odlučivanja u uslovima rizika. Naš cilj sada je da pokažemo kako je moguća jedna racionalna bihejvioralna teorija odlučivanja u uslovima rizika. Ovu „racionalizaciju“ bihejvioralne teorije odlučivanja sprovodimo da bismo je konceptualno poravnali sa teorijskim i metodološkim okvirom adaptacionistički motivisane racionalne analize. Naš predlog obuhvata i temeljnu kritiku teorije izgleda Tverskog i Kanemana - teorije za koju se veruje da je paradigmatičan i uspešan predstavnik teorija ograničene racionalnosti. Pošto demonstriramo da je moguće da teorija odlučivanja u uslovima rizika bude ujedno racionalna i u stanju da inkorporira nalaze koji su do sada svedočili o ograničenoj racionalnosti, pokazaćemo da takav rezultat vodi u zaključke koji su kontraintuitivni ako se posmatraju u okviru paradigme racionalnosti koju smo do sada diskutovali. To će nas prirodno voditi ka promeni paradigme i reinterpretaciji problema racionalnosti saznanja na jednoj novoj konceptualnoj ravni.

# 10 Formiranje verovanja u odlučivanju

Suočena sa lozom oblika (*x,p;y,1-p*) koji sa verovatnoćom *p* donosi iznos *x*, a sa verovatnoćom *1-p* iznos *y*, osoba može da se nađe u situaciji u kojoj treba da oceni monetarnu vrednost celog loza. U takvoj situaciji, kada se ona pita o minimalnoj ceni po kojoj bi takav loz prodala, mi kažemo da ta osoba određuje *monetarni ekvivalent* (ili *ekvivalent u izvesnosti*, engl. *certainty equivalent*, skr. CE) loza (*x,p;y,1-p*). Suočena sa dva loza oblika (*x,p;y,1-p*) i (*z,q;w,1-q*), osoba može da se nađe u situaciji u kojoj treba da donese odluku o tome koji od dva predstavljena loza bi pre odigrala. Ove dve situacije opisuju paradigmatične eksperimentalne metode koje se koriste u istraživanju odlučivanja u uslovima rizika. Prva situacija odgovara metodologiji merenja monetarnih ekvivalenta, u kojoj se od ispitanika traži da da direktnu numeričku procenu vrednosti loza (Tversky, 1967), ili da do te ocene dođe upotrebom procedure serijskih izbora (Tversky & Kahneman, 1992, Gonzales & Wu, 1999). Druga situacija odgovara metodologiji *eksperimenta izbora*, u kojoj ispitanici jednostavno donose binarne odluke između dva predstavljena loza. Teorije odlučivanja se po svojoj eksplanatornoj moći porede u odnosu na to u kojoj meri mogu da objasne odgovore ispitanika u ovim (i drugim) eksperimentalnim paradigmama.

Pretpostavimo da neka osoba pokušava da donese svoj sud o monetarnom ekvivalentu loza (50 $EUR,\frac{1}{2}$;51 $EUR,\frac{1}{2}$), dakle loza koji sa 50% verovatnoće donosi 50 ili 51 evra. Očekivana vrednost loza je 50.5 evra, tako da uz pretpostavku o određenom stepenu neizvesnosti u procesu suđenja, i uračunavajući određeni stepen averzije prema riziku, možemo da očekujemo ocenu monetarnog ekvivalenta od oko nešto manje od 50 evra za ovaj loz. Pretpostavimo sada da od ispitanika zahtevamo da donese sud o monetarnom ekvivalentu loza (50000 $EUR,\frac{1}{2}$;5 $EUR,\frac{1}{2}$). Situacija u kojoj se on nalazi je očigledno složenija od prethodne. Očekivana vrednost loza - dakle, vrednost koja ne uzima u obzir stepen averzije prema riziku ispitanika o kome je reč - iznosi 25002.5 evra. Uračunavajući određeni stepen averzije prema riziku (prema konkavnoj funkciji korisnosti), monetarni ekvivalent bi očekivano bio niži od ove vrednosti. Ni analiza odlučivanja u normativnoj teoriji očekivane korisnosti (EU) ni analiza kumulativne teorije izgleda (CPT) ne uzimaju u obzir sledeću mogućnost. Osoba koja treba da donese sud o monetarnom ekvivalentu loza (50000 $EUR,\frac{1}{2}$;5 $EUR,\frac{1}{2}$) može da rezonuje na sledeći način: uopšte, u ma kom relevantnom ekonomskom okruženju, verovatnoća da se dođe do dobitka od 50000

evra je sigurno niža od verovatnoće da se dođe do dobitka od 5 evra. Jednostavno, više plate su ređe od nižih plata, teže je dobiti bolje nego lošije plaćen posao, pod lakšim uslovima se dobija manji nego veći kredit od banke, verovatnoća da na ulici pored nas prođe osoba sa niskim ili prosečnim primanjima je daleko veća od verovatnoće da pored nas prođe prebogati biznismen. Intuicija o tome da je do manjih vrednosti lakše doći nego do većih jedna je od, verujemo, elementarnih ljudskih intuicija o ustrojstvu socijalnih odnosa i proteže se kroz primere od traženja adekvatno plaćenog posla, preko ocene zadovoljstva odnosom kvaliteta usluge i njene cene, do igara na sreću. U odnosu na distribuciju novca koji neko poseduje, kao osnovnog resursa čiju alokaciju proučava ekonomska nauka, prethodna intuicija postaje fundamentalna, dobro poznata činjenica koju opisuje *distribucija bogatstva* u ma kom čoveku poznatom ekonomskom okruženju. Ekonomske nauke dobro poznaju ovakve distribucije čije se proučavanje vezuje za ime slavnog italijanskog matematičara, ekonomiste i sociologa Vilfreda Pareta, koji je (verovatno) oko 1897. došao do klasičnog zapažanja o (aproksimativnom) odnosu prema kome je oko 20% građana kontrolisalo oko 80% vlasništva u Italiji, zapažanja koje kasnije generalizovano kao poznato pravilo 80-20 (Johnson, Kotz & Balakrishnan, 1994; up. Juran, J. M, 1950, prema Wood & Wood, 2005, za sud o tome da je princip 80-20 formulisan tek 1906). Pareto se posvetio detaljnoj matematičkoj studiji distribucije bogatstva, razvijajući klasu statističkih distribucija koje opisuju ovaj odnos i danas nose ime *Paretove distribucije*. Pretpostavimo, sasvim plauzibilno, da osobe koje pitamo o monetarnim ekvivalentima rizičnih lozova poput onih iz prethodnih primera uzimaju u obzir svoje prethodno iskustvo u relevantnom ekonomskom okruženju, i u donošenju suda o vrednosti loza oblika $(X, P), X = x_1, x_2, .., x_n,\ P = p_1, p_2, .., p_n,$ u taj sud inkorporiraju svoju reprezentaciju tog ekonomskog okruženja kroz distribuciju verovatnoće $P' = p'_1, p'_2, .., p'_n$. Distribucija $P'$ opisuje reprezentaciju verovatnoće da se nečije ukupno bogatstvo uveća za određenu monetarnu vrednost $x$ i, po pretpostavci, ta (memorijska) reprezentacija predstavlja rezultat prethodnog iskustva neke individue u ekonomskim interakcijama u odgovarajućem ekonomskom sistemu. Inkorporacijom verovatnoća $P'$ u svoj sud o monetarnom ekvivalentu datog loza $(X, P)$, donosilac odluka bi praktično donosio odluku o monetarnom ekvivalentu loza $(X, P'')$ - loza u kome je prethodno korigovao date verovatnoće $P$ niza vrednosti u $X$ inkorporacijom prethodnih, reprezentovanih verovatnoća $P'$, dolazeći tako do *a posteriori* verovatnoća $P''$ u *aposteriornom lozu* $(X, P'')$. Mehanizam bejzijanske inferencije se direktno nameće kao normativno

adekvatan metod za ovu korekciju verovatnoća koje sadrži loz o čijem se monetarnom ekvivalentu sudi, ili za korekciju verovatnoća na dva ili više lozova između kojih se vrši izbor.

U narednim redovima, izgradićemo jednu formu teorije odlučivanja u uslovima rizika koja kroz mehanizam bejzijanske inferencije inkorporira prethodna verovanja donosioca odluka u aktualne sudove koje on donosi. Ključni problem čije ćemo rešenje predstaviti jeste *problem formiranja verovanja u odlučivanju*. To je sledeći problem: donosilac odluka, suočen sa lozom $(X, P)$, koji poznaje relevantnu distribuciju vrednosti $X$ u svojoj okolini, u oznaci $P'$, može da (a) odluči da veruje da je distribucija verovatnoća data na lozu, $P$, objektivna, i donese odluku uopšte ne korigujući verovatnoće $P$ date na lozu, ili može da (b) odluči da veruje da distribuciju verovatnoća datu na lozu, $P$, treba da koriguje u skladu sa svojim verovanjima o verovatnoćama odgovarajućih vrednosti, $P'$, i odluku donese u odnosu na aposteriorni loz $(X, P'')$. Treća mogućnost je da donosilac odluka odluči da koriguje svojim *a priori* verovatnoćama $P'$ date verovatnoće $P$ na lozu *u određenom stepenu*, i tako donese odluku „kompromisno", na osnovu aposteriornog loza $(X, P'')$ gde su *a posteriori* verovatnoće $P''$ posledice bejzijanske inferencije *čiji je efekat posredovan tim stepenom u kome donosilac odluka veruje da treba da koriguje date verovatnoće P*. Preciznu formulaciju $(i)$ kognitivne reprezentacije *a priori* verovatnoća $P'$ i (b) *stepena* u kome donosilac odluka veruje da treba da koriguje date verovatnoće na lozu $P$ u odnosu na reprezentovane *a priori* verovatnoće $P'$ daje *teorija poverenja* u odlučivanju u uslovima rizika - teorija koju ovde predstavljamo.

U ovom opisu problema za koji teorija poverenja nudi rešenje jasno je da je u pitanju jedna tipična kognitivna teorija o *formiranju verovanja*. Gde je onda tu teorija odlučivanja? Bejzijansku teoriju odlučivanja na čijim osnovama gradi teorija poverenja formulisao je američki ekonomista Kip Viskuzi, pod imenom *teorije perspektivne reference* (Viscusi, 1989). Teorija poverenja koju predstavljamo jeste teorija formiranja verovanja u Viskuzijevoj bejzijanskoj teoriji perspektivne reference. Pokazaćemo da je rezultirajuća teorija odlučivanja u uslovima rizika u stanju da predvidi sve robustne fenomene koji svedoče o ograničenoj racionalnosti, uključujući i neke nove empirijske fenomene koje teorija izgleda ne može da objasni. Rezultirajuća teorija, u kojoj komponenta teorije poverenja objašnjava formiranje verovanja, a komponenta Viskuzijeve teorije procese inkorporacije tih verovanja u odlučivanje, prati normativno opravdanu, racionalnu teorijsku konstrukciju[79].

## 10.1 Viskuzijeva teorija

Viskuzijevu teoriju perspektivne reference ovde formulišemo u donekle drugačijoj notaciji od one koju koristi autor i oslanjajući se na minimalan broj plauzibilnih pretpostavki koje će poslužiti u razvoju teorije poverenja (i koje se ni na koji način ne kose sa logikom teorije perspektivne reference). Posmatrajmo jednostavan loz sa dva moguća ishoda ($x$,$p$;$y$,*1-p*), loz koji donosi ishod $x$ sa verovatnoćom $p$ i ishod $y$ sa verovatnoćom *1-p*. Odigravanje ovakvog loza možemo da interpretiramo na sledeći način. Kutija sadrži 100 kuglica od kojih je $100p$ belih i $100$(*1-p*) crnih. Loz se odigrava slučajnim izvlačenjem jedne kuglice; ako je izvučena kuglica bela (kojih ima $100p$), osvaja se $x$, a ako je izvučena kuglica crna (kojih ima $100$(*1-p*)), osvaja se $y$. Očigledno, distribucija verovatnoće koja karakteriše loz ($x$,$p$;$y$,*1-p*), jeste binomialna distribucija sa parametrom $p_{\mathrm{x}}$ i indeksom $N$:

$$p(x|p_x) = \binom{n}{x} p_x^x (1-p_x)^{n-x} \tag{51}$$

Verovatnoća $p_{\mathrm{x}}$ je verovatnoća osvajanja ishoda $x$ odigravanjem loza ($x$,$p$;$y$,*1-p*). Pretpostavimo da donosilac odluka ima razloga da veruje da osvajanje vrednosti $x$ i $y$ ima odgovarajuće *a priori* verovatnoće $p'_x$ i $p'_y$, tako da $p'_x + p'_y = 1$. Neka je njegovo verovanje u mogućnost osvajanja iznosa $x$ odigravanjem loza određeno njegovim verovanjem o verovatnoći $p'_x$, i neka to njegovo verovanje *a priori* o verovatnoći $p'_x$ sledi Beta distribuciju datu sa

$$p(p'_x) = \frac{1}{B(\alpha,\beta)} (p'_x)^{\alpha-1} (1-p'_x)^{\beta-1} \tag{52}$$

gde je sledeći član izraza u (52):

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{53}$$

beta funkcija (zadata preko gama funkcija); ovaj član jednačine (52) tek normalizuje funkciju gustine verovatnoće Beta distribucije (tako da je njen integral jednak jedan, te čini pravu funkciju gustine), pa Beta distribuciju *a priori* verovatnoće $p'_x$ možemo da pojednostavljeno izrazimo kao

$$p(p'_x) \propto (p'_x)^{\alpha-1} (1-p'_x)^{\beta-1} \tag{54}$$

Vidimo da, ako je *a priori* distribucija verovatnoće $p_{\mathrm{x}}$ izražena Beta distribucijom, a podaci o realizaciji $x$ uspešnih u $n$ pokušaja binomijalnom distribucijom $p(x|p_{\mathrm{x}})$ (jednačina (51)), *a posteriori* distribucija verovatnoće $p_{\mathrm{x}}$ do koje stižemo bejzijanskom inferencijom ponovo uzima oblik Beta distribucije:

$$p(p_x|x) = p(x|p_x)p(p_x) \propto p_x^{\alpha+x-1}(1-p_x)^{\beta+n-x-1} \tag{55}$$

Ovu *a posteriori* verovatnoću da će biti osvojen iznos $x$ na lozu (*x,p;y,1-p*) označavaćemo sa $p''_x$. Očigledno, $p''_y = 1 - p''_x$. U bejzijanskoj inferenciji, familije *a priori* distribucija koje posle primene na podatke (tj. verodostojnosti) izražene određenom distribucijom (u našem slučaju binomijalnom) vode ka *a posteriori* distribucijama iz iste familije, nazivaju se *konjugovanim a priori distribucijama* (Lee, 1989/2004). U našem slučaju, očigledno je da je Beta distribucija konjugovana *a priori* distribucija binomijalnoj distribuciji.

Pošto je očekivana vrednost Beta distribucije data sa $E(p'_x) = \frac{\alpha}{\alpha+\beta}$, te vrednost $\alpha + \beta$ očigledno možemo da izjednačimo sa ocenom ukupnog broja posmatranja na osnovu kog je formirano to *a priori* očekivanje $E(p'_x)$, Viskuzi (1989) koristi alternativnu parametrizaciju Beta distribucije sa parametrima $\theta = \frac{\alpha}{\alpha+\beta}$, $N = \alpha + \beta$. Dva nova parametra, $\theta$ i $N$, sada sadrže istu informaciju koju i originalni parametri $\alpha, \beta$: $\alpha = \theta N$, $\beta = N - \alpha$. Pojasnimo: parametar $\theta$ opisuje naše *a priori* očekivanje o tome da se ishod $x$ osvaja sa verovatnoćom $p'_x$, dok parametar $N$ predstavlja meru pouzdanosti tog očekivanja jer korespondira broju posmatranja (odn. veličini uzroka) na kome je to očekivanje formirano. Dakle, kažemo da je *a priori očekivanje* da će odigravanjem loza (*x,p;y,1-p*) biti osvojen ishod $x$ reprezentovano očekivanjem $\theta$ formiranom na $N$ posmatranja. Loz (*x,p;y,1-p*) sada reprezentuje nove informacije kojima treba korigovati *a priori* očekivanje kroz bejzijansku inferenciju: informacije da se $x$ osvaja u $100p$ slučajeva a $y$ u $100(1\text{-}p)$ slučajeva; ovo implicira da loz tretiramo kao da nosi informaciju o sto novih posmatranja, *što je psihološki plauzibilna pretpostavka pošto se lozovi skoro bez izuzetka opisuju na skali procenata.* Pošto $N = \alpha + \beta$ reprezentuje broj posmatranja na osnovu kojh je formirano $\theta$, a $\theta$, $N$ određuju *a priori* beta distribuciju $p'_x \sim \mathrm{Beta}(\alpha, \beta)$, možemo da odredimo *a posteriori* Beta distribuciju sa novim parametrima: $p''_x \sim \mathrm{Beta}(\alpha'', \beta'')$. Na osnovu jednačine (55) uviđamo da je $\alpha'' = \alpha + 100p - 1$, i $\beta'' = \beta + 100 - 100p - 1$ (odn.$\beta'' = (N+100-2) - \alpha''$): tako podaci kojima loz doprinosi *a priori* očekivanju da se osvoji ishod $x$ koriguju očekivanje $\theta = \frac{\alpha}{\alpha+\beta}$ proprocionalno broju slučajeva u

kojima loz donosi $x$, a ukupni broj posmatranja aproksimira zbir onog na kome je formirano *a priori* očekivanje, $N$, i onog kojim doprinosi loz - taj broj je sto, po pretpostavci.

Sumirajmo prethodno izneto. Dva parametra odlikuju donosioca odluka: $\theta$ i $N$, tj. njegovo očekivanje $\theta$ verovatnoće $p'_x$ osvajanja iznosa $x$ i, stepen $N$ u kome je siguran u to svoje očekivanje. Ova dva parametra nose istu informaciju kao i standardni parametri $\alpha, \beta$ Beta distribucije *a priori* verovatnoće $p'_x$. Donosilac odluka kome je predočen loz (*x,p;y,1-p*) koristi bejzijansku inferenciju da bi izveo parametre $\alpha'', \beta''$ *a posteriori* distribucije da se iznos $x$ osvaja sa verovatnoćom $p''_x$. Konačno, on uzima očekivanje sa *a posteriori* Beta distribucije, $\theta'' = \frac{\alpha''}{\alpha''+\beta''}$, kao svoju bejzijansku ocenu[80] *a posteriori* verovatnoće $p''_x$. Ta očekivana vrednost *a posteriori* verovatnoća $p''_x$osvajanja ishoda $x$ uzima oblik:

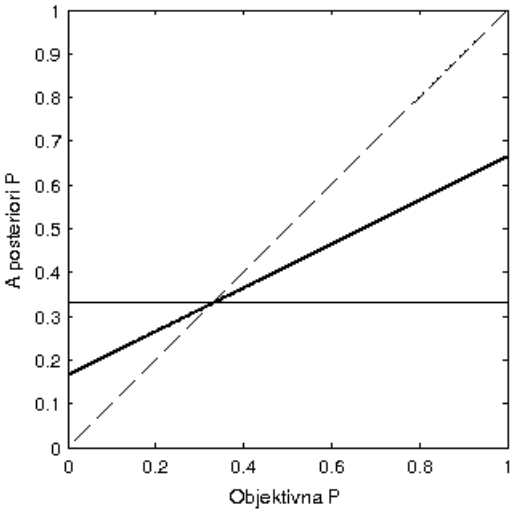
$$p''_x = \frac{Np'_x + 100p}{N + 100} \tag{56}$$

sa $N = \alpha + \beta$ kao i do sada. Ako pažljivije pregledamo (56), vidimo da je *a posteriori* verovatnoća osvajanja ishoda $x$ sa loza (*x,p;y,1-p*) samo suma ($i$) njenog očekivanja na osnovu *a priori* verovatnoće ($Np'_x$) i ($ii$) njenog očekivanja na osnovu verovatnoće koju specifikuje loz (*100p*), podeljena sa novom veličinom uzorka[81] *N+100*. Iz jednačine (56) sledi najbitnija osobina Viskuzijeve teorije perspektivne reference: *a posteriori* verovatnoće osvajanja ishoda na nekom lozu su *linearne funkcije verovatnoća koje su date na tom lozu*. Ako ispišemo (56) kao

$$p''_x = \frac{N}{N + 100}p'_x + \frac{100}{N + 100}p \tag{57}$$

vidimo da član $\frac{100}{N+100}$ predstavlja *nagib*, dok član $\frac{N}{N+100}p$ predstavlja *intercept* linearne funkcije koja povezuje *a posteriori* verovatnoću $p''_x$ sa verovatnoćom $p$ datom na lozu. Ako sada, za potrebe primera, pretpostavimo da je *a priori* verovanje nekog donosioca odluka da se ishod $x$ osvaja sa verovatnoćom $p'_x = .33$, sa indeksom $N$=100 koji predstavlja veličinu uzorka posmatranja na kome je to verovanje *a priori* formirano, odnos verovatnoće date na lozu („objektivne verovatnoće“) i *a posteriori* verovatnoće je kao što prikazuje Slika 17. Odnos je linearan, pri čemu su objektivne verovatnoće ispod nivoa *a priori* verovatnoće $p'_x = .33$ *precenjene*, a verovatnoće iznad tog nivoa *potcenjene*, što opisuje osnovni empirijski nalaz ponderisanja verovatnoća u odlučivanju koji opisuje i funkcija ponderisanja

verovatnoća u teoriji izgleda[82]. Zbog ove osobine, Viskuzi *a priori* verovatnoću $p'_x$ naziva *referentnim nivoom rizika*, što predstavlja pojam po kome cela teorija nosi ime, *teorija perspektivne reference*: procena rizika skopčana sa odlukom o vrednosti nekog rizičnog loza donosi se kroz proces bejzijanske inferencije iz *perspektive određenog referentnog nivoa rizika* koji je predstavljen *a priori* verovatnoćama.




Slika 17. *Linearno ponderisanje verovatnoća u Viskuzijevoj teoriji.* Isprekidana linija predstavlja liniju identiteta objektivnih i *a posteriori* („ponderisanih“) verovatnoća, dok podebljana linija predstavlja linearnu funkciju odnosa objektivnih i *a posteriori* verovatnoća za $p'_x = .33$ i $N$=100 prema jednačinama (56) i (57). Funkcija identiteta i funkcija ponderisanja verovatnoća seku se tačno u visini *a priori* verovatnoće $p'_x = .33$.

Vratimo se sada odlučivanju. Pošto je posle bejzijanske inferencije, koja polazi od toga da donosilac odluka poznaje svoje ocene *a priori* verovatnoća $p'_x$ i $p'_y$ za ishode $x$ i $y$ na lozu ($x$,$p$;$y$,*1-p*), on formirao aposteriorni loz $(x, p''; y, 1 - p'')$, Viskuzijeva teorija perspektivne reference predviđa da on očekivanu korisnost tog loza evaluira isto kao što se evaluira dati loz pod teorijom očekivane korisnosti:

$$EU(x, p''; y, p''_y) = p''_x u(x) + p''_y u(y) \tag{58}$$

gde je naravno $p''_y = 1 - p''_x$. Sledeće zapažanje je od ključnog značaja za razumevanje teorije perspektivne reference i kasnije diskusije teorije poverenja. Ako iskoristimo razvoj *a posteriori* verovatnoća prema jednačini (56) u izrazu za očekivanu korisnost *a posteriori* datu jednačinom (58), dobijamo

$$EU(x, p_x; y, p_y) = (\frac{Np'_x + 100p_x}{N + 100}) \cdot u(x) + (\frac{Np'_y + 100p_y}{N + 100}) \cdot u(y) \tag{59}$$

koja posle nešto algebarskog sređivanja može da se izrazi kao

$$EU(x, p_x; y, p_y) = \frac{N}{N + 100}(p'_x u(x) + p'_y u(y)) + \frac{100}{N + 100}(p_x u(x) + p_y u(y)) \tag{60}$$

odn.

$$EU(x, p_x; y, p_y) = \frac{N}{N + 100} EU'(x, p'_x; y, p'_y) + \frac{100}{N + 100} EU(x, p_x; y, p_y) \tag{61}$$

Jednačina (61) nam otkriva sledeće: očekivana korisnost *aposteriornog loza* - loza koji se evaluira u skladu sa *a posteriori* verovatnoćama $p''_x$, $p''_y$, jeste linearna kombinacija *a priori očekivane korisnosti* - izračunate iz *a priori* verovatnoća $p'_x$, $p'_y$ - i *date očekivane korisnosti*, izračunate iz verovatnoća $p_{\mathrm{x}}$, $p_{\mathrm{y}}$ koje specifikuje loz. Ova osobina teorije perspektivne reference *pokazaće se ključnom u našim diskusijama statusa racionalnosti odlučivanja* u uslovima rizika uopšte. Očekivana korisnost loza po ovoj teoriji, vidimo, predstavlja ponderisanu sumu (a) očekivane korisnosti loza izračunate na osnovu prethodnih verovanja o relevantnim verovatnoćama koje ima donosilac odluka i (b) očekivane korisnosti kakvu bi izračunao donosilac odluka prema fon Nojman-Morgnešternovoj teoriji, dakle donosilac odluka koji ne inkorporira nikakva prethodna verovanja u svoj sud o očekivanoj korisnosti.

Generalizujemo sada analizu Viskuzijeve teorije na lozove sa proizvoljnim brojem ishoda: $(X, P)$, $X = x_1, x_2, .., x_n$, $P = p_1, p_2, .., p_n$, gde donosilac odluka inkorporira svoju reprezentaciju relevantnog ekonomskog okruženja kroz distribuciju verovatnoće $P' = p'_1, p'_2, .., p'_n$ relizacije ishoda $X$ na lozu. Ukoliko je loz sa dva ishoda modeliran binomijalnom distribucijom, pa je prema tome analogan statističkom eksperimentu bacanja novčića koji ima verovatnoću $p_{\mathrm{x}}$ da padne na stranu (ishod) $x$ i $p_{\mathrm{y}}$ da padne na stranu $y$, loz sa proizvoljnim brojem ishoda $n$ se modelira bacanjem $n$-dimenzionalne kockice, što je statistički eksperiment koji opisuje *multinomijalna distribucija*, data sa

$$p(x|P) = \frac{N!}{x_1!x_2!...x_n!} p_1^{x_1} p_2^{x_2} ... p_n^{x_n} \tag{62}$$

gde vektor $x = x_1, x_2, .., x_n$ sadrži broj uspešnih realizacija redom ishoda $X = x_1, x_2, .., x_n$ sa loza $(X, P)$ na uzorku posmatranja $N = \sum_{i=1}^{n} x_i$, a vektor $P = p_1, p_2, .., p_n$ redom verovatnoće ostvarenja tih ishoda. Kao što je Beta distribucija konjugovana binomijalnoj distribuciji u bejzijanskoj inferenciji, tako je *Dirišleova distribucija* konjugovana multinomijalnoj distribuciji (Smith, 2010):

$$p(P|\alpha) = \frac{\Gamma(\alpha_1 + \alpha_2 + ...\alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_n)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} ... p_n^{\alpha_n - 1} \tag{63}$$

gde vektor $\alpha$, sa $\alpha_i > 1$ za svako $1<i<n$, predstavlja vrednosti intenziteta subjektivnih verovanja o verovatnoćama u vektoru $P = p_1, p_2, .., p_n$, na veličini uzorka $\alpha_0 = \sum_{i=1}^{n} \alpha_i$. Član jednačine (63) u kome učestvuju gama funkcije predstavlja tek izraz za normalizaciju distribucije. Očekivana vrednost ovde distribucije nalazi se na

$$E(p_i) = \frac{\alpha_i}{\alpha_0} = \frac{\alpha_i}{\sum_{i=1}^{n} \alpha_i} \tag{64}$$

*A posteriori* distribucija uzima ponovo oblik Dirišleove distribucije

$$p(\alpha|P) \propto p_1^{\alpha_1 + x_1 - 1} p_2^{\alpha_2 + x_2 - 1} ... p_n^{\alpha_n + x_n - 1} \tag{65}$$

a kao najbolju ocenu *a posteriori* verovatnoća $p_i''$ iz vektora $P$ uzimamo njihova očekivanja sa *a posteriori* distribucije $p(\alpha|P)$ prema jednačini (64). Ovime je Viskuzijeva teorija predstavljena u svojoj generalnoj formi u kojoj se odnosi na lozove oblika $(X, P)$ *sa proizvoljnim brojem ishoda.*

Sumirajmo ukratko generalni slučaj. Donosioca odluka odlikuju određena *a priori* verovanja o mogućnostima osvajanja ishoda $X$ sa loza $(X, P)$. Ta verovanja izražavamo pomoću Dirišleove distribucije: ona modelira prethoda verovanja donosioca odluka o verovatnoćama $P$ ishoda $X$ sa nekog loza koji sadrži $n$ ishoda. Loz $(X, P)$ onda predstavlja novu, dodatnu informaciju o verovatnoćama ishoda u $X$ - informaciju modeliranu multimodijalnom distribucijom. Verovatnoće date na lozu donosilac odluka koriguje u skladu sa svojim prethodnim verovanjima u procesu Dirišle-Multinomijalne bejzijanske inferencije da bi razvio svoja *a posteriori* verovanja o verovatnoćama ishoda u lozu $(X, P)$. U evaluaciji loza, on koristi generalizovanu formu jednačine (58):

$$EU(X,P) = \sum_{i=1}^{n} p''_{x_i} u(x_i) \tag{66}$$

Sada bi trebalo da je potpuno jasno da je suštinski problem Viskuzijeve teorije - problem kome Viskuzi, začuđujuće, ne posvećuje dovoljnu pažnju - kako odrediti subjektivna, *a priori* verovanja donosioca odluka o vrednostima u vektoru $\alpha$? Ove vrednosti opisuju njegova prethodna verovanja o verovatnoćama ishoda $X$ sa kojima može da se suoči na lozu $(X, P)$. Po pretpostavci koje se sve vreme držimo, ta prethodna verovanja su posledica iskustva u prethodnim ekonomskim interakcijama koje je određeni donosilac odluka imao u relevantnom ekonomskom okruženju. Viskuzi (1989) diskutuje dva slučaja. Prvi, koji naziva *slučajem simetrične referentne tačke*, podrazumeva da donosilac odluka polazi od *uniformne a priori* distribucije: ako njegova prethodna verovanja vrede nekih $n$ posmatranja, on u odlučivanju polazi od toga da su svi ishodi $X$ na lozu *a priori* podjednako verovatni sa verovatnoćama $\frac{1}{n}$. Ovakva analiza vodi u situaciju u kojoj su funkcije ponderisanja verovatnoća - za sve moguće lozove koji sadrže ishode iz $X$ - iste. Viskuzi diskutuje i drugi slučaj, koji naziva *slučajem višestrukih referentnih tačaka* (v. dodatak A.1 njegovom radu iz Viscusi, 1989) - naše dosadašnje izlaganje implicitno pretpostavlja taj slučaj. Pošto je plauzibilno pretpostaviti da donosioci odluka zaista imaju neka sistematska prethodna očekivanja o verovatnoćama različitih monetarnih ishoda, tj. da njihova očekivanja nisu uniformna, mi se u svim daljim analizama krećemo samo u okviru slučaja višestrukih referentnih tačaka. Interesantno je da empirijski testovi Viskuzijeve teorije perspektivne reference uopšte ne razmatraju ovaj slučaj, uprkos njegovoj *prima facie* plauzibilnosti; svi testovi koji su obuhvatali ovu teoriju (Harless, 1993, Blavatskyy, 2011) baziraju se na jednostavnom slučaju simetrične referentne tačke, pretpostavljajući uniformne *a priori* distribucije. U razvoju teorije poverenja mi ćemo pokazati da je prihvatljivim, racionalnim proširenjem ideje o višestrukim referentnim tačkama moguće osvojiti dopunsku eksplanatornu moć u teoriji perspektivne reference.

Fundamentalan problem je, dakle, dati tačnu specifikaciju parametara na osnovu kojih kognitivni sistem vrši proces bejzijanskog formiranja verovanja o verovatnoćama u evaluaciji loza $(X, P)$. U prvom koraku problem je specifikacija prethodnih verovanja o ishodima $X$ , odn. vektora $\alpha$ u jednačini (63) odgovarajuće Dirišleove distribucije koja modelira ta prethodna verovanja. Kao što ćemo uskoro videti, iz određenih komplikacija (koje je Viskuzi tačno predvideo i kratko diskutovao

u dodatku A1. njegovom radu, Viscusi, 1989) sledi dodatni, ne manje značajan problem koji ćemo morati da rešavamo. Polazeći od osnovne pretpostavke racionalne analize da kognitivni sistemi u rešavanju adaptivnih problema koriste informacije o strukturi relevantne sredine, sledeći analizu distribucije bogatstva koju je predložio još Vilfredo Pareto, i oslanjajući sa na neke dobro poznate pretpostavke matematičke psihologije, sada predstavljamo rešenje problema formiranja verovanja u odlučivanju u uslovima rizika.

## 10.2 Teorija poverenja

Teorija poverenja počiva na analizi strukture verovatnoća monetarnih dobitaka i gubitaka u realnom ekonomskom okruženju. Pretpostavljajući da su donosioci odluka kroz svoje prethodne ekonomske interakcije informisani o osobinama relevantnog okruženja, u teoriji poverenja se dalje pretpostavlja da oni kroz te interakcije razvijaju kognitivne (memorijske) reprezentacije odgovarajućih struktura verovatnoća. Te strukture verovatnoća su u teoriji poverenja reprezentovane odgovarajućom familijom distribucija verovatnoće. Konačno, teorija poverenja opisuje proces formiranja subjektivnih verovanja o verovatnoćama dobitaka i gubitaka na lozovima tipa $(X, P)$ primenom prethodno reprezentovane strukture verovatnoća iz relevantnog ekonomskog okruženja. Konačni rezultat procesa su *a priori* verovatnoće $P'_x$ koje odlikuju donosioca odluka u procesu bejzijanske inferencije ka *a posteriori* verovatnoćama $P''_x$ kako ga opisuje Viskuzijeva teorija perspektivne reference.

*Struktura informacija u relevantnom ekonomskom okruženju.* Istorija Paretove čuvene opservacije o distribuciji prihoda u određenom ekonomskom okruženju je složena, a sudovi o tome kako je Pareto došao do ove opservacije, i kada je tačno formulisao zakon koji danas poznajemo kao *Paretov princip*, ograničeni su usled izostanka prevoda odgovarajuće arhivske građe (ukoliko ona postoji; čak i najznačajnija Paretova dela su tek mestimično prevođena na engleski jezik). Neki izvori navode da je sama opservacija formulisana 1897. godine u Paretovom delu „*Cours d'Economie Politique*“ (Johnson, Kotz & Balakrishnan, 1994). Po svemu sudeći, prvo objavljivanje Paretovog principa, odn. formulacije statističke distribucije koja se danas naziva Paretovom distribucijom, pada 1906. godine sa objavljivanjem Paretove knjige „*Manuale di economia politica*“ (Juran, J. M, 1950, prema Wood & Wood, 2005). Kroz decenije diskusija ovaj princip se često pogrešno

pripisuje američkom ekonomisti Maksu Lorencu, koji je analizi distribucije bogatstva doprineo 1905. grafičkom metode prezentacije nejednakosti raspodele bogatstva u nekoj populaciji - razvojem dijagrama koji se po njemu danas zove *Lorencova kriva* (Juran, 1975). Tokom decenija, nauka je počela da prepoznaje Paretov princip u drugim društvenim i prirodnim pojavama (up. Newman, 2005), generalizujući ga tako van domena ekonomije gde se odnosi na distribuciju prihoda (ili bogatstva) u nekoj populaciji. Takve generalizacije se u savremenoj literaturi najčešće nazivaju *zakonima stepena* (engl. *power laws*). Važnu opservaciju da Paretov princip važi van domena ekonomije dugujemo američkom ekspertu u menadžmentu kvaliteta Džozefu M. Djuranu, koji je i skovao izraz „*Paretov princip*“ (Juran, 1975).

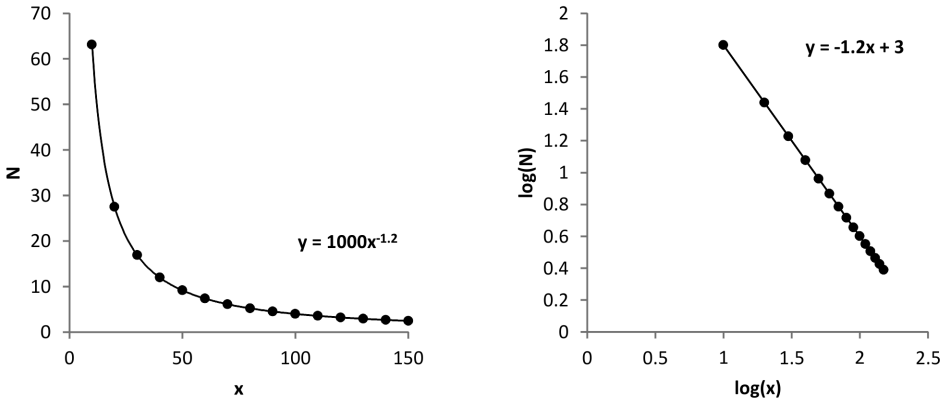
Proučavajući distribuciju bogatstva koje pripada individuama u nekom ekonomskom sistemu, Pareto je došao do zaključka da nju opisuje zakonitost oblika

$$N = Ax^{-a} \tag{67}$$

gde je $N$ broj individua čije je bogatstvo *veće ili jednako* vrednosti $x$, a dok se $A$ i $a$ se redom nazivaju Paretovom konstantom i Paretovim parametrom (Johnson, Kotz & Balakrishnan, 1994). Logaritamska forma ovog zakona omogućava da ga diskutujemo u linearnom obliku:

$$log(N) = log(A) - a \cdot log(x) \tag{68}$$

gde vidimo da $log(A)$ kontroliše intercept, a vrednost parametra $a$ nagib linearne funkcije koja opisuje odnos $N$ i $x$. Zahvaljujući ovoj osobini, najčešći oblik testiranja Paretovog principa sastoji se u logaritmovanju vrednosti $N$ i $x$ i diskusiji kvaliteta linearnog fita koji se dobija između njih. Konsekventno, ovo se smatra i metodom ocene vrednosti parametra $a$, odn. nagiba linearne funkcije funkcije između $N$ i $x$; noviji radovi upozoravaju na to da su ovi popularni testovi nedovoljno konzervativni i da učestalo otkrivaju važenje Paretovog principa u podacima koji ne pružaju dovoljno evidencije za njega (u odnosu na ocene parametara metodom maksimalne verodostojnosti, Clauset, Shalizi, & Newman, 2009). Slika 18. prikazuje hipotetički, idealizovan odnos između $N$ i $x$ za vrednosti $a = 1.2$ i $A = 1000$, i linearni odnos koji se otkriva logaritmovanjem ovih varijabli.




Slika 18. *Hipotetički, idealizovan zakon stepena.* Levi panel: na apscisi se nalazi vrednost nekih hipotetičkih primanja $x$, na ordinati broj ljudi koji u nekoj populaciji imaju primanja veća ili jednaka $x$. Zakon stepena odlikuje odnos na levom panelu. Desni panel: linearni odnos između logaritmovanih varijabli. Grafikoni su napravljeni na osnovu vrednosti $a = 1.2$ i $A = 1000$, jednačine (67) i (68).

Stepeni zakon (slika 18) opisuje odnos broja individua $N$ u nekoj populaciji čiji su prihodi veći ili jednaki od određene monetarne vrednost $x$ i same te monetarne vrednosti. Forma stepenog zakona jasno ukazuje na osnovnu karakteristiku ove distribucije: na veoma mali broj individua se raspodeljuje ogromna proporcija novca u nekom ekonomskom sistemu; veoma velikom broju individua pripada raspodela veoma malog ostatka raspoloživih prihoda. Empirijska posmatranja koja su nastavljena od Paretovog uvida do danas ubedljivo potvrđuju kvalitativnu formu ovog odnosa (Kleiber & Kotz, 2003). Podsetimo se, u sekciji 7.3 smo se već susreli sa stepenim zakonom u diskusiji racionalne analize pamćenja Andersona i saradnika.

Forma Paretovog principa, ili stepenog zakona, koju daju jednačine (67) i (68) nije normalizovana (nema integral pod krivom sa vrednošću 1), pa ne predstavlja pravu distribucije verovatnoće. Posle normalizacije (up. Kleiber & Kotz, 2003 za izvođenje), uobičajena forma u kojoj se zadaje *Paretova distribucija* je sledeća (Johnson, Kotz & Balakrishnan, 1994):

$$S(x) = P(X \geq x) = (\frac{x_{min}}{x})^q \tag{69}$$

Ovde je potrebno obratiti pažnju da je jednačina (69) data u dekumulativnoj formi, sledeći formu koju je Pareto odabrao u formulaciji svog principa: $S(x)$ označava *dekumulativnu funkciju verovatnoće* ($S$ je skr. od engl. *survivor function*), koja opisuje verovatnoću da je prihod $X$ *veći ili jednak* od neke vrednosti $x$. Vrednost

$q$ je Paretov parametar (u prethodnoj diskusiji označavan sa $a$) koji fundamentalno opisuje ponašanje ove funkcije verovatnoće, a $x_{\min}$ neki minimalan prihod koji se pojavljuje u datoj populaciji. Kumulativna funkcija verovatnoće je, očigledno,

$$F(x) = P(X < x) = 1 - (\frac{x_{min}}{x})^q \tag{70}$$

dok je gustina data preko izraza

$$p(x) = q x_{min}^q x^{-(q+1)} \tag{71}$$

Ovime je dat potpun opis strukture verovatnoće prihoda odn. distribucije bogatstva u nekom ekonomskom sistemu, sledeći klasičnu Paretovu analizu. Ekonomisti su predložili više različitih distribucija koje bi opisivale ovaj odnos, ali je analiza koja prati Paretovu distribuciju ubedljivo najviše zastupljena (v. Kleiber & Kotz, 2003 za pregled). Paretova distribucija ima karakteristike koje su sa stanovišta matematičke i konceptualne analize u najmanju ruku nezgodne. Ona spada u tzv. distribucije verovatnoće *teških repova*. Bez ulaženja u tehničke detalje, to znači da je verovatnoća ekstremnih događaja, npr. pojava individua koje zadržavaju ogromnu proporciju prihoda (relativno u odnosu na ostatak populacije) veća nego što bi uobičajeno bila na distribucijama koje imaju tzv. eksponencijalne repove, poput eksponencijalne ili normalne (Gausove) distribucije. Sama priroda stepenog zakona, ond. njegov stepeni - a ne eksponencijalni - pad, određuje ovu suštinsku karakteristiku Paretove distribucije. Ako posmatramo normalnu, Gausovu distribuciju, znamo da njeni repovi opadaju eksponencijalno brzo kako se udaljavamo od centralne tendencije na jednu ili drugu stranu. To čini pojavu ekstremnih događaja na Gausovoj distribuciji *veoma retkom*. Međutim, na Paretovoj distribuciji, kako vrednost prihoda koji posmatramo raste, verovatnoća da će se pojaviti individua kojoj će pripasti određeni prihod ne pada *tako brzo* kao što bi padala na nekoj distribuciji sa eksponencijalnim repovima (otud izraz „teški repovi“ - pod njima se nalazi „više verovatnoće“ nego što je uobičajeno kod drugih distribucija). Zbog ove osobine, Paretova distribucija se opire nekim uobičajenim analitičkim sredstvima matematičke statistike. Konkretno, za $q < 1$, ne postoji varijansa ove distribucije, za vrednost $q$ između 1 i 2 varijansa je beskonačna; ona postaje konačna tek za $q > 2$. Očekivana vrednost Paretove distribucije je beskonačna za $q < 1$, i konačna tek za $q > 1$. Uopšteno, $n$-ti moment Paretove

distribucije postoji samo ako je $n<q$.

Pored upravo diskutovane forme, koja se naziva *Paretovom distribucijom I tipa*, Vilfredo Pareto je predložio još dve povezane statističke distribucije koje danas nose njegovo ime. Distribucija koja je zbog jedne svoje osobine veoma pogodna za primenu u teoriji poverenja jeste *Paretova distribucija II tipa*[83]. Paretova distribucija, data jednačinama (69-71) za njenu dekumulativnu i kumulativnu funkciju te funkciju gustine verovatnoće, uključuje parametar $x_{\min}$ koji predstavlja neku najnižu vrednost prihoda koja se beleži u određenoj populaciji. U analizi odlučivanja monetarne vrednosti $X$ na lozovima $(X, P)$ uobičajeno mogu da uzmu raspon od 0 do neke konačne vrednosti. U Paretovoj distribuciji I tipa, međutim, $x_{\min}$ mora da uzima pozitivnu vrednost. Paretova distribucija II tipa je razvijena tako da se minimalan prihod nalazi u vrednosti nula: ona predstavlja translaciju Paretove distribucije I tipa po abscisi, do oordinate. Njena dekumulativna funkcija je zadata preko izraza

$$S(x) = \frac{b^q}{(x+b)^q} \tag{72}$$

gde je parametar $q$ i dalje Paretov parametar, a $b$ novi parametar koji je neophodno uvesti prilikom ove modifikacije Paretove distribucije; vidimo da parametar $x_{\min}$ više ne igra ulogu. Tzv. *standardna* Paretova distribucija II tipa uvek ima vrednost parametra $b$=1; mi ćemo koristiti upravo ovu, standardnu Paretovu distribuciju II tipa u teoriji poverenja tako da izbacujemo ovaj parametar iz svih formula i nastavljamo da radimo sa jednoparametarskom verzijom distribucije. Njena dekumulativna funkcija onda ima oblik

$$S(x) = \frac{1}{(x+1)^q} \tag{73}$$

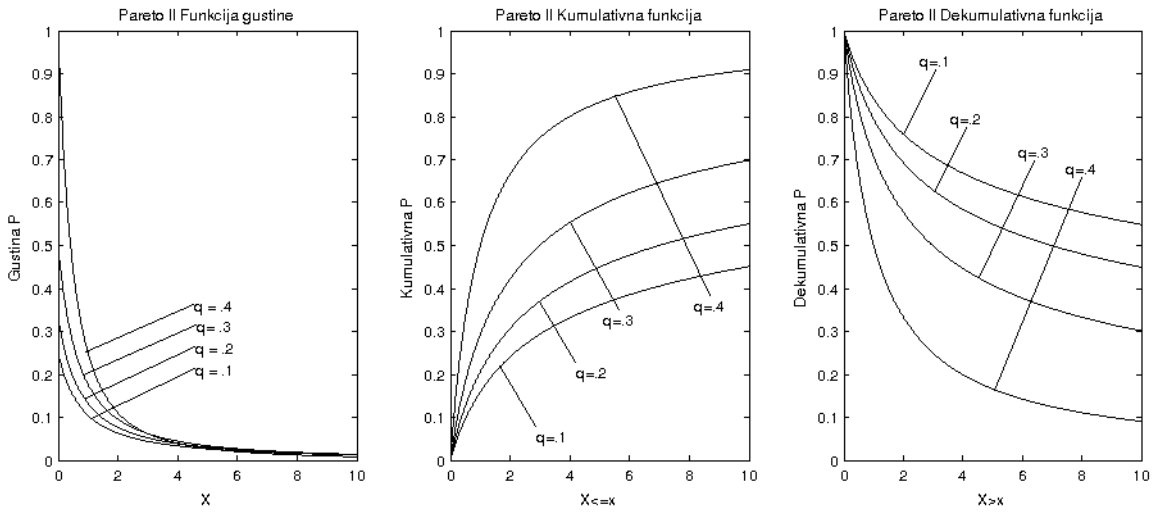
Kumulativna funkcija verovatnoće je, očigledno:

$$F(x) = 1 - \frac{1}{(x+1)^q} \tag{74}$$

dok je funkcija gustine data izrazom:

$$p(x) = q(1+x)^{-q-1} \tag{75}$$

za $x>0$, $q>0$. Slika 19. prikazuje ponašanje familije standardnih Paretovih distribucija II tipa za vrednosti parametra $q = .1$, $q = .2$, $q = .3$ i $q = .4$.




Slika 19. *Familija standardnih Paretovih distribucija II tipa.* Paneli s leva na desno: funkcija gustine verovatnoće, kumulativna i dekumulativna funkcija.

*Kognitivna reprezentacija monetarnih dobitaka i gubitaka.* Osnovna pretpostavka je da donosioci odluka memorijski reprezentuju svoja prethodna iskustva u ekonomskim interakcijama, što praktično znači da oni sempluju odgovarajuću distribuciju verovatnoća monetarnih dobitaka i gubitaka iz relevantne populacije događaja. Njihovu memorijsku reprezentaciju u teoriji poverenja karakteriše upravo diskutovana standardna Paretova distribucija II tipa. Dopunske specifikacije su neophodne.

Prvo, Paretova distribucija govori o distribuciji verovatnoće da neka individua ostvaruje određeni prihod. Ona se odnosi na finalno stanje, odn. ukupnu količinu novca kojom neko raspolaže. U teoriji poverenja, isto kao i u teoriji izgleda, pretpostavljamo da se analiza odlučivanja odnosi na *relativne* događaje dobitaka i gubitaka. U teoriji poverenja koristimo dve reprezentacije: jednu standardnu Paretovu distribucija II tipa, koja sa parametrom $q_{\mathrm{g}}$ opisuje strukturu verovatnoća dobitaka (pozitivnih monetarnih ishoda u ekonomskim interakcijama poput odigravanja nekog rizičnog loza), i drugu, sa parametrom $q_l$, koja opisuje strukturu verovatnoća gubitaka. Dakle, teorija poverenja ne radi sa finalnim stanjima tj. ukupnim količinama vrednosti koje neka individua poseduje, već sa pozitivnim i negativnim priraštajima u tim prihodima.

Dok prvu pretpostavku zadovoljavaju i teorija izgleda i teorija poverenja, druga pretpostavka je karakteristična za teoriju poverenja. Pogledajmo sliku 19. Na njoj je prikazano ponašanje familije standardnih Paretovih distribucija II tipa. U ekonomskim analizama, uobičajeno je da se ove distribucije posmatraju kao distribucije same monetarne vrednosti odn. novca. Drugim rečima, skala na kojoj se nalazi posmatrana varijabla $x$ je skala novca, objektivne vrednosti. U teoriji poverenja, mi pretpostavljamo da donosioca odluka karakteriše određena funkcija korisnosti za ishode, $u : x \to u(x)$, $x, u(x) \in R$, i *pretpostavljamo da je memorijska reprezentacija strukture verovatnoća u relevantnom ekonomskom okruženju data nad korisnostima*, a ne nad vrednostima. Za sistem čiji je zadatak optimizacija ponašanja organizma u odnosu na adaptivne pritiske, relevantan opis distribucije neke vrednosti u sredini je upravo u jedinicima subjektivne korisnosti. Formalno, pretpostavka ne uvodi nikakve komplikacije, pošto teorija poverenja proširuje Viskuzijevu teoriju perspektivne reference za koju već postoji dokaz o postojanju funkcije $u(\cdot)$ nad ishodima (Viscusi, 1989).

Da rezimiramo dosadašnje pretpostavke. Pre svega, memorijska reprezentacija environmentalne distribucije relevantne vrednosti (novca) je takve forme da nju bar aproksimativno može da opiše standardna Paretova distribucija tipa II. Drugim rečima, ova pretpostavka glasi da su kognitivna reprezentacija distribucije u odgovarajućoj okolini i objektivna distribucija u toj okolini izomorfne. Subjekti sempluju distribucije verovatnoća dobitaka i gubitaka u okolini, tako da nema garancija da je ova subjektivna reprezentacija perfektna. Dalje, postoje dve nezavisne memorijske reprezentacije: jedna za verovatnoće dobitaka, i druga za verovatnoće gubitaka. Sledeća pretpostavka jeste da se reprezentacije verovatnoća ishoda prethodnih ekonomskih interakcija razvijaju nad negativnim i pozitivnim priraštajima, a ne nad finalnim stanjima tih interakcija. Konačno, pretpostavka je da se memorijske reprezentacije formiraju na skali korisnosti, a ne na skali vrednosti. Konsekventno, standardne Paretove distribucije tipa II koje koristimo u razvoju teorije poverenja su date nad korisnostima.

Sledeća pretpostavka precizno formuliše informaciju koju kognitivni sistem memorijski reprezentuje. U teoriji poverenja, distribucija environmentalnih verovatnoća dobitaka i gubitaka, kognitivno se reprezentuje u formi *dekumulativne funkcije* $S(x)$ standardne Paretove distribucije II tipa. Opravdanje ove pretpostavke je veoma jednostavno. Prvo, novac - osnovna vrednost na koju primenjujemo teoriju poverenja i druge teorije odlučivanja - je *po svojoj prirodi kumulativna*

*vrednost.* Svako ko ima 100 dinara, ima i 50 dinara. Svako ko ima 200 dinara može za 140 dinara da kupi paklu cigareta i da očekuje kusur od 60 dinara. Svako ko je odigrao neki rizičan loz i osvojio npr. 400 dinara, osvojio je i 300 dinara, osvojio je i 200 dinara, i tako dalje. Prirodna reprezentacija korisnosti novca je, dakle, kumulativne prirode. Međutim, ako bi kognitivni sistem koristio kumulativnu funkciju $F(x)$, susreo bi sa sledećim problemom: naime, na kumulativnim funkcijama, sa priraštajem u vrednosti raste i kumulativna verovatnoća. Ovakav opis ne odgovara svakodnevnoj intuiciji da je teže (tj. da je manje verovatno) osvojiti 400 nego 200 dinara. Dekumulativna funkcija $S(x)$ zadovoljava oba intuitivno opravdana uslova za reprezentaciju monetarnih dobitaka i gubitaka: ona zadržava kumulativnost kao karakteristiku vrednosti novca, i ujedno direktno korespondira intuitivnim sudovima prema kojima su veći dobici manje verovatni od manjih dobitaka, kao što su veći gubici manje verovatni od manjih gubitaka.

Pošto smo sada u potpunosti opisali kognitivnu reprezentaciju strukture verovatnoća monetarnih dobitaka i gubitaka kakvu predviđa teorija poverenja, potrebno je da opišemo mehanizam kojim na osnovu svoje interne funkcije $S(x)$ donosilac odluka formuliše verovanja o *a priori* verovatnoćama $P'$ osvajanja ishoda $X$ u lozu $(X, P)$.

*Formiranje verovanja o rizičnim ishodima.* Dekumulativna funkcija $S(x)$, u idealnom slučaju, sadrži kompletnu informaciju o strukturi verovatnoća svih mogućih monetarnih dobitaka i gubitaka koje donosilac odluka uopšte može da razmatra. Loz $(X, P)$, međutim, predstavlja distribuciju verovatnoća nad samo *nekim* ishodima $X$, koji formiraju loz. Na koji način, polazeći od distribucije verovatnoća nad celim domenom dobitaka ili gubitaka, donosilac odluka formira *a priori* distribuciju verovatnoće samo nad onim ishodima u lozu $(X, P)$? Podsetimo se da je $P$ koja odlikuje loz $(X, P)$ prava distribucija verovatnoće: uvek važi $\sum_{i=1}^{n} p_i = 1$, $p_i \in P$.

Rešenje ovog problema u teoriji poverenja počiva na *Lusovom aksiomu izbora* (Luce, 1959, 1977). Ishodi $X$ na lozu $(X, P)$ jesu neki podskup svih mogućih monetarnih ishoda $M$ čije verovatnoće opisuje funkcija $S(x)$. Dakle, $X \subset M$. Uvodimo sledeću notaciju. Neka je $x \in X$, i $X \subset M$; sa $P_M(x)$ označavamo verovatnoću da će ishod $x$ biti izabran ukoliko je skup svih alternativnih ishoda $M$, i sa $P_X(x)$ verovatnoću da će ishod $x$ biti izabran ukoliko je skup svih alternativnih ishoda $X$. Interesuje nas, dakle, na koji način treba da računamo verovatnoću ishoda

$x$ kada se on posmatra u kontekstu samo onih ishoda koji se nalaze na lozu $(X, P)$, ako znamo koja je njegova verovatnoća u skupu svih mogućih ishoda M. Lusov aksiom izbora predstavlja proširenje standardne aksiomatike teorije verovatnoće koji omogućava da verovatnoću $P_X(x)$ izračunamo kao:

$$P_X(x) = P_M(x|X) \tag{76}$$

odn. kao zavisnu verovatnoću $x$ u $M$, ako znamo da je izabran ishod u $X$, gde je $X \subset M$. Iz ovoga direktno sledi da, ako se ishod $x$ čija nas verovatnoća interesuje, nalazi u nekom podskupu $X$ svih mogućih ishoda $M$, onda uvek možemo da izračunamo njegovu verovatnoću na sledeći način:

$$P_X X = \frac{P_M(x)}{\sum_{y \in X} P_M(y)} \tag{77}$$

Lusov aksiom izbora formalizuje opravdanje sledećeg potpuno intuitivnog suda. Ukoliko imamo neki skup mogućih ishoda $M = \{x, y, z\}$, čijim su svim članovima pridružene verovatnoća izbora, u oznaci $P_M(x)$, $P_M(y)$ i $P_M(z)$, a zatim taj skup izbora svedemo na novi skup $X = \{x, y\}$, *odnosi* verovatnoća izbora ishoda koji su ostali u novom skupu $X \subset M$ će ostati isti kao što su bili u polaznom skupu ishoda. Ako je $M = \{belo\ vino, crno\ vino, konjak\}$, a iz ovog skupa biramo ishode sa verovatnoćama $P_M(belo\ vino)$, $P_M(crno\ vino)$ i $P_M(konjak)$, tako da je $P_M(belo\ vino)+P_M(crno\ vino)+P_M(konjak)=1$, a zatim saznamo da konjaka nema, te je skup izbora sveden na $X = \{belo\ vino, crno\ vino\}$, odnos verovatnoća da će biti izabrano belo ili crno vino ne sme da se promeni. Direktno sledi da $P_X(belo\ vino)+P_X(crno\ vino)=1$. Jednačina (77) garantuje da je odnos verovatnoća invarijantan, upućujući nas da nove verovatnoće izračunamo u odnosu na ukupnu verovatnoću podskupa $X$ u odnosu na polazni skup $M$.

Čitaocu upućenom u diskusije teorija verovatnoće i odlučivanja nije promakao sledeći problem. Lusov aksiom izbora, koji nam omogućava da verovatnoće $P'$ za loz $(X, P)$ izračunamo na osnovu informacije o verovatnoćama svih mogućih ishoda, odnosi se na reprezentaciju putem verovatnoća, *a ne putem dekumulativnih verovatnoća* koje opisuje funkcija $S(x)$. Jedna direktna posledica Lusovog aksioma izbora otklanja ovaj problem. Naime, iz aksioma izbora sledi (Luce, 1977) da uvek postoji pozitivna, realna funkcija $v$ nad $M$, koja je jedinstvena do transformacije multiplikativnom konstantom, tako da za sve elemente $x$ nekog podskupa $X$ skupa

$M$ važi

$$P_X(x) = \frac{v(x)}{\sum_{y \in X} v(y)} \tag{78}$$

Jednačina (77) sada može da se tumači kao primena jednačine (78) u kojoj ulogu funkcije $v$ igra verovatnoća; međutim, direktno je jasno da tu ulogu može da igra i funkcija dekumulativne verovatnoće $S(x)$. Ako su odnosi verovatnoće izbora dva objekta isti bez obzira na izbor podskupa izvornog skupa iz kojeg ih biramo, što je suština Lusovog aksioma izbora, onda i odnosi njihovih kumulativnih i dekumulativnih verovatnoća moraju isto tako ostati isti. Sada je jasno da prethodna verovanja $P'$ o ishodima loza $(X,P)$ dobijamo kao

$$p_{x_i} = \frac{S(x_i)}{\sum_{j \in X} S(x_j)} \tag{79}$$

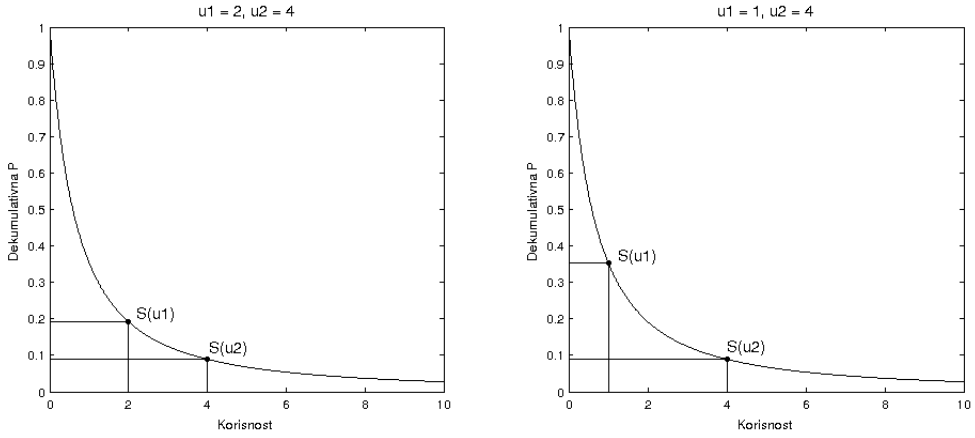
odn. da je vrednost *a priori* verovatnoće $p_x$ za određeni ishod $x$ u skupu $X$ sa loza $(X,P)$ proporcionalna njegovoj relativnoj dekumulativnoj verovatnoći izračunatoj prema jednačini (78), gde funkcija $S(x)$ igra ulogu funkcije $v$ čiju egzistenciju obezbeđuje Lusov aksiom izbora. U analizu sada uključujemo još jednu varijablu za koju će se pokazati da igra odlučujuću ulogu u objašnjenju empirijskih nalaza u odlučivanju u uslovima rizika.

*Neizvesnost rizičnih lozova i stepen poverenja u a priori verovatnoće.* Donosilac odluka koji treba da donese svoj sud o monetarnom ekvivalentu rizičnog loza $(x,p_x;y,p_y)$, prema teoriji poverenja, polazi od *a priori* verovanja $p'_x$ i $p'_y$. Podsetimo se da Viskuzijeva teorija predviđa da će *a posteriori* verovatnoće $p''_x$ i $p''_y$, na osnovu kojih će aposteriorni loz $(X, P'')$ biti evaluiran, dobijaju na osnovu bejzijanske inferencije kroz konjugovanu Dirišleovu i Multinomijalnu distribuciju, pa će te *a posteriori* verovatnoće biti $p''_x = \frac{\alpha''_x}{\alpha''_x+\alpha''_y}$ i $p''_y = \frac{\alpha''_y}{\alpha''_x+\alpha''_y}$ , gde su $\alpha''_x, \alpha''_y$ parametri *a posteriori* Dirišleove distribucije. Mi prvo moramo da odredimo vrednost ovih parametara za *a priori* Dirišleovu distribuciju, i smatramo da nije rano otkriti da će te vrednosti biti proporcionalne prethodno izračunatim *a priori* verovatnoćama $p'_x$ i $p'_y$. Međutim, još jedna, ključna varijabla, mora da se inkorporira u teoriju.

Podsetimo se da u Dirišle-Multinomijalnoj (kao i Beta-Binomijalnoj) bejzijanskoj inferenciji $N = \alpha_x + \alpha_y$ možemo da tumačimo kao broj posmatranja na osnovu kojih je donosilac odluka formirao svoja verovanja. Pogledajmo za trenutak ponovo jednačine (56) i (57). One nam govore da vrednosti odgovarajućih *a posteriori*

verovatnoća, $p''_x$ i $p''_y$, zavise i od vrednosti $N$. Tako „snaga“, odn. mera pozdanosti prethodnih verovanja, $p'_x$ i $p'_y$, učestvuje i u formiranju *a posteriori* verovatnoća. To znači da ukoliko dopustimo da vrednost $N$ varira, obezbeđujemo dodatnu fleksibilnost suda o *a posteriori* verovatnoćama, $p''_x$ i $p''_y$. Može da se postavi pitanje zašto uopšte treba da dozvolimo da parametar $N = \alpha_x + \alpha_y$ ima varijabilnu vrednost, posebno u odnosu na Viskuzijev zaključak (up. dodatak A.1 njegovom radu Viscusi, 1989) da takva mogućnost vodi u komplikacije. Viskuzi u razvoju teorije perspektivne reference nije uvideo da se variranjem vrednosti $N = \alpha_x + \alpha_y$ osvaja značajna eksplanatorna moć. To je, s druge strane, ključan uvid u teoriji poverenja. Jednačine (56) i (57), ponovo, pokazuju odnos *a posteriori* verovatnoća i verovatnoća datih na lozu $(X, P)$, odnos koji je linearan kao što prikazuje slika 17. Forma jednačine (57), kao što smo već diskutovali, pokazuje da član $\frac{100}{N+100}$ kontroliše nagib, a član $\frac{N}{N+100}p'$ intercept linearne funkcije koja povezuje *a posteriori* verovatnoću $p''_x$ sa verovatnoćom $p$ datom na lozu. Pošto ova linearna jednačina igra istu ulogu kao i funkcija ponderisanja verovatnoća u teoriji izgleda, ako teorija dozvoli da vrednost $N$ varira kroz različite lozove, ona dozvoljava da nagib i intercept ove linearne funkcije budu varijabilni, obezbeđujući tako veću mogućnost da se uklope različiti slučajevi odstupanja subjektivnih verovatnoća od onih koje su date na lozu. Verujemo da postoji još jedan razlog zbog kojeg je Viskuzi izbegao analizu mogućnosti da je vrednost $N$ varijabilna. U originalnoj ekspoziciji njegove teorije, koja se minimalno razlikuje od naše, ponder koji donosilac odluka stavlja na *a priori* verovatnoće i ponder koji stavlja na objektivne verovatnoće su oba *slobodni parametri*. U Viskuzijevoj teoriji perspektivne reference ovo za posledicu ima to da vrednosti ova dva parametra ne mogu *simultano* da se ocene: moguće je oceniti samo njihov odnos (Viscusi, 1989). Pošto smo mi vrednost objektivnih verovatnoća fiksirali na 100, pretpostavljajući da je skala procenata prirodna skala na kojoj su izraženi rizični lozovi, problem je rešiv, jer preostaje samo da se oceni vrednost jednog parametra: pondera koji donosilac odluka stavlja na *a priori* verovatnoće. Međutim, zbog navedene mogućnosti da se linearna funkcija ponderisanja verovatnoće učini fleksibilnijom, mi odustajemo i od toga da parametar $N$ tretiramo kao slobodan parametar, već dopuštamo da on *sistematski* varira kroz različite lozove - mogućnost koju je Viskuzi predvideo, ali smatrao komplikovanom. Sada dajemo teorijsko rešenje za problem kako odrediti koju vrednost $N$ treba dati *a priori* verovatnoćama na određenom lozu. Podsetimo se još jednom teorijskog motiva za rešavanje ovog problema: vrednost $N$, u jednačinama (56-57), učestvuje

u kontroli intercepta i nagiba linearne funkcije ponderisanja verovatnoća. Pošto je upravo ta funkcija ključna u objašnjenju odstupanja empirijskih od normativnih sudova i izbora, mogućnost da ta funkcija uzima različite forme sigurno obezbeđuje veću eksplanatornu moć. Slika 20. će poslužiti u objašnjenju našeg teorijskog predloga.




Slika 20. *Uloga neizvesnosti loza u teoriji poverenja.* Levi i desni panel prikazuju istu dekumulativnu funkciju definisanu nad korisnostima, $S(u(x))$, sa vrednošću parametra $q$=1.5. Na levom panelu, *a priori* dekumulativne $S(u_1)$ i $S(u_2)$ verovatnoće za loz sa dva ishoda koji odgovaraju subjektivnim korisnostima $u_1 = 2$ i $u_2 = 4$. Na desnom panelu, *a priori* dekumulativne $S(u_1)$ i $S(u_2)$ verovatnoće za loz sa dva ishoda koji odgovaraju subjektivnim korisnostima $u_1 = 1$ i $u_2 = 4$.

Na slici 20, oba panela prikazuju jednu istu dekumulativnu funkciju $S$. Po pretpostavci teorije poverenja, ovakva funkcija korespondira osobinama memorijske reprezentacije prethodnih ekonomskih interakcija donosioca odluka. Na levom panelu, prikazane su dekumulativne verovatnoće za loz sa dva ishoda koji korespondiraju subjektivnim korisnostima $u_1 = 2$ i $u_2 = 4$; na desnom panelu, isti takav loz ali sa ishodima koji korespondiraju subjektivnim korisnostima $u_1 = 1$ i $u_2 = 4$. Očigledno je da je razlika $S(u_1)$-$S(u_2)$ veća u slučaju loza na desnom nego u slučaju loza na levom panelu. Posle primene Lusovog aksioma izbora po jednačini (79), jasno je da će *a priori* verovatnoća za *istu* korisnost $u_2$= 4 biti *različita* na ova dva loza: ona će biti veća za loz na levom nego za loz na desnom panelu. Vrednost a priori verovatnoće asocirana za ishod određene korisnosti u teoriji poverenja je funkcija konteksta: taj kontekst, zbog primene Lusovog aksioma izbora, čine *svi ishodi* koji se nalaze na lozu oblika $(X, P)$. Generalno, što su manje dekumulativne

verovatnoće koje čine kontekst dekumulativnoj verovatnoći ishoda koji posmatramo, to će *a priori* verovatnoća za taj ishod, posle primene Lusovog aksioma, biti veća za taj posmatrana ishod. Ukoliko su razlike između *a priori* verovatnoća ishoda na nekom lozu u proseku *veće*, to su a priori verovatnoće koje odlikuju ishode na tom lozu za donosioca odluka *informativnije*. Loz sa dva ista ishoda će odlikovati njihove *a priori* verovatnoće u odnosu 50:50. To znači da je distribucija *a priori* verovatnoća za takav loz uniformna. Što se *a priori* verovatnoće više razlikuju, npr. kao na lozu za koji one stoje u odnosu 90:10, ili 25:75, to je distribucija *a priori* verovatnoća za taj loz udaljenija od uniformne. Visina samih *a priori* verovatnoća za loz $(X, P)$, koja u potpunosti zavisi od reprezentovanih dekumulativnih verovatnoća na funkciji $S$ i skupa ishoda $X$ koji odlikuju taj loz, određuje u kojoj meri će *a priori* verovatnoće korigovati sud o verovatnoćama $P$ koje su date na lozu.

Postoji način da se u sud o *a posteriori* verovatnoćama inkorporira i upravo diskutovana informativnost distribucije *a priori* verovatnoća. Svaku distribuciju verovatnoća odlikuje određen stepen *neizvesnosti*, koji se u teoriji verovatnoće meri *informacionom entropijom* te distribucije verovatnoća. Za loz $(X, P)$ koji sadrži $n$ ishoda, za koji smo na osnovu dekumulativne funkcije $S$ primenom Lusovog aksioma izbora izračunali distribuciju *a priori* verovatnoća $P'$, informaciona entropija *a priori* verovatnoća $P'$ je zadata izrazom:

$$H(P') = -\sum_{i=1}^{N} p'_i \cdot log_2(p'_i) \tag{80}$$

prema poznatoj Šenonovoj jednačini informacione entropije (Shannon, 1948). Podsetimo se, ponovo, izraza za *a posteriori* verovatnoće, datog jednačinama (56-57): *a priori* verovatnoće $P'$ i verovatnoće date na lozu $P$ se linearno kombinuju da bi odredile *a posteriori* verovatnoće $P''$. U toj linearnoj kombinaciji, svaka od njih doprinosi vrednosti *a posteriori* verovatnoće $P''$ proporcionalno tome koliko „vrede“ posmatranja na kojima su formirane: to je 100, u slučaju verovatnoća datih na lozu, i $N$, u slučaju *a priori* verovatnoća. Upravo sada formulišemo teorijski predlog o određivanju vrednosti $N$ - stepena u kome *a priori* verovatnoća sa loza $(X, P)$ učestvuju u formiranju vrednosti *a posteriori* verovatnoća na osnovu kojih će loz biti evaluiran. Predlažemo da vrednost $N$ varira između 0 i 1, na skali *relativne informacione entropije a priori verovatnoća* $P'$, te da ta vrednost bude pomnožena sa 100 kako bi se dovela na istu skalu na kojoj se nalazi 100 posmatranja - odn. na skalu procenata - sa koje dolazi stepen u kome verovatnoće date na lozu učestvuju

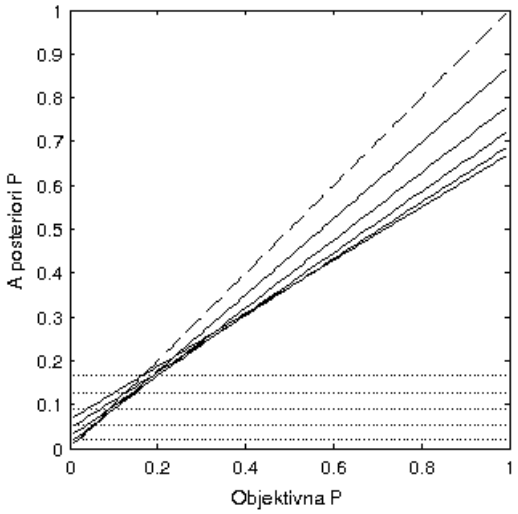
u formiranju *a posteriori* verovatnoća. Relativna informaciona entropija *a priori* verovatnoća $P'$ je data sa

$$H_R(P') = \frac{H(P')}{H_{max}(P')} \tag{81}$$

odn. predstavlja odnos informacione entropije $H(P')$ i maksimalne entropije koju može da dostigne distribucija verovatnoća za loz sa $n$ ishoda: ta maksimalna vrednost odlikuje uniformnu distribuciju verovatnoća i iznosi $-log_2(\frac{1}{n})$. Stepen u kome donosilac odluka koristi *a priori* verovatnoće je onda određen kao $N = 100 \cdot H_{rel}(P')$. Prethodni izraz množenjem sa 100 samo reskalira relativnu informacionu entropiju na skalu procenata. Ovakav pristup izračunavanju $N$ ima očigledne i bitne teorijske posledice. Pre svega, on zadovoljava uslov koji smo postavili, a to je da ova vrednost može da varira u funkciji ishoda koje loz sadrži, omogućavajući tako fleksibilnost linearne funkcije ponderisanja verovatnoća. Druga važna teorijska osobina ovako određenog stepena $N$ je sledeća: ona nikada ne dozvoljava da *a priori* verovatnoće utiču na *a posteriori* verovatnoće *više* nego što na njih utiču verovatnoće date na lozu. Donosilac odluka ipak mora više da uzme u obzir verovatnoće koje loz nosi nego svoja prethodna verovanja o mogućnostima osvajanja ishoda koji se na njemu nalaze. Tako, ukoliko loz sadrži ishode koji se dosta razlikuju, donosilac odluka formira *a priori* verovatnoće koje se dosta razlikuju. Već zahvaljujući tim razlikama među *a priori* verovatnoćama u bejzijanskoj inferenciji dolazi do značajne korekcije verovatnoća datih na lozu. Vrednost $N$, kako mi predlažemo da se ona izračunava prema jednačini (81), balansira ovaj efekat *a priori* verovatnoća u odnosu na efekat verovatnoća datih na lozu, jer što *a priori* verovatnoće *imaju nižu entropiju* (tj. više se razlikuju) *to će vrednost N biti manja.* Loz sa *a priori* verovatnoćama u odnosu 50:50 tako odlikuje maksimalna entropija *a priori* verovatnoća, pa će njen uticaj kroz vrednost $N$ na formiranje *a posteriori* verovatnoća biti maksimalan (isti koliki i uticaj verovatnoća datih na lozu), ali se taj efekat niveliše činjenicom da će *a priori* verovatnoće u odnosu 50:50 po sebi veoma malo korigovati verovatnoće date na lozu. Verovatnoće na lozu sa *a priori* verovatnoćama u odnosu 90:10 će biti značajno korigovane vrednostima samih tih *a priori* verovatnoća, ali će ovaj efekat biti umanjen vrednošću $N$ koja će biti niža, jer verovatnoće u odnosu 90:10 imaju daleko nižu entropiju od maksimalne, dopuštajući tako verovatnoćama na lozu da ipak presudno utiču na evaluaciju loza.

Konačno, u odnosu na linearno ponderisanje verovatnoća kakvo sledi iz

Viskuzijeve bejzijanske teorije perspektivne reference, kakve promene unosi mehanizam formiranja verovanja koji predstavlja teorija poverenja? Uzmimo neposredan primer da bismo ovo ilustrovali. Pretpostavimo da je funkcija korisnosti donosioca odluka linearna za potrebe ovog primera. Neka on reprezentuje svoja verovanja o mogućnostima osvajanja ishoda dekumulativnom funkcijom $S$ sa parametrom $q$=1.5. Neka je dat loz (75,$p$; $y$,1-$p$); neka drugi ishod u lozu, $y$, varira kao $y = 5, 10, 15, 20, 25$, a verovatnoća $p$ od 0 do 1, generišući tako kontinuum lozova sa dva fiksna ishoda. Posmatramo samo koje vrednosti *a posteriori* verovatnoća se dobijaju za veći od dva ishoda, 75, polazeći od date reprezentacije $S$ i varirajući verovatnoću $p$ i drugi ishod na lozu $y$ kao što je opisano. Slika 21. ilustruje linearne odnose između verovatnoća datih na lozu i *a posteriori* verovatnoća kakve predviđa teorija poverenja.




Slika 21. *Linearno ponderisanje verovatnoća u teoriji poverenja* (rešenje za varijabilni efekat a priori verovatnoća u slučaju višestrukih referentnih tačaka u terminologiji Viscusi, 1989). Primer je generisan pretpostavljajući linearnu funkciju korisnosti, varirajući $y$ kao 5, 10, 15, 20, 25, sa vrednošću $q$=1.5 za dekumulativnu funkciju $S$.

Teorija poverenja predviđa da je ponderisanje verovatnoća fenomen koji opisuje *familija linearnih funkcija*. Preciznu formu ponderisanja verovatnoća nekog ishoda na lozu $(X, P)$ određuju $(i)$ vrednost parametra $q$ dekumulativne funkcije $S$ koja predstavlja reprezentaciju prethodnih verovanja donosioca odluka, $(ii)$ nelinearnost funkcije korisnosti (određena parametrom $\rho$ u slučaju stepene funkcije) i $(iii)$ kontekst koji daju *a priori* verovatnoće drugih ishoda na istom lozu. *Fundamentalna predikcija teorije poverenja je da ponderisanje verovatnoća ni u kom slučaju ne može*

*biti samo funkcija verovatnoća datih na lozovima.* Za niz lozova oblika ($x$,$p$;$y$,1-$p$), koji nastaju tako što ishod $x$ držimo fiksnim, a ishod $y$ variramo proizvoljno tako da je uvek $x>y$, držeći verovatnoću $p$ fiksnom, kumulativna teorija izgleda uvek predviđa *iste* pondere odluka. Kumulativna teorija izgleda je model korisnosti zavistan od *ranga* (tj. RDU model); dok je rang određenog ishoda na nekom lozu nepromenjen, ponderisanje verovatnoće na tom ishodu biće nepromenjeno. Nasuprot tome, predikcija teorije poverenja je da će *a posteriori* verovatnoće - koje u teoriji poverenja i teoriji perspektivne reference igraju ulogu pondera odluka Kanemena i Tverskog - biti i funkcija korisnosti ishoda $y$, odn. $u(y)$. Ovo se u teoriji poverenja ogleda u različitim vrednostima za $S(u(y))$, koje imaju efekta u formiranju $p'_y$, i prema mehanizmu teorije koji je prethodno opisan, u formiranju svih *a posteriori* verovatnoća na osnovu kojih će loz biti evaluiran.

*Tretman mešovitih lozova.* Pošto smo u prethodnim redovima predstavili novi predlog za tretman fenomena ponderisanja verovatnoća, sada generalizujemo pristup na problem mešovitih lozova - lozova koji sadrže i dobitke i gubitke. Prema Viskuzijevoj teoriji perspektivne reference, videli smo, ukupna očekivana korisnost loza računa se kao

$$EU(X,P) = \sum_{i=1}^{n} p''_{x_i} u(x_i) \tag{82}$$

što je u suštini ista jednačina očekivane korisnosti koja se u proučavanju odlučivanja u uslovima rizika koristi od Bernulija do savremenih autora. Ipak, ne zaboravimo da je nju u okviru Viskuzijeve teorije moguće dekomponovati u komponentu koja se odnosi na očekivanu korisnost *a priori* i komponentu očekivane korisnosti izračunatu na osnovu verovatnoća datih na lozu (up. jednačine (59-61)). Kakav tretman imaju mešoviti lozovi, lozovi koji sadrže i dobitke i gubitke? Naš predlog prati ideju teorije izgleda Tverskog i Kanemena (Kahneman & Tversky, 1979, Tversky & Kahneman, 1992), da se ukupna očekivana korisnost mešovitog loza dekomponuje u pozitivni i negativni deo :

$$EU(X,P) = EU^{+}(X,P) + EU^{-}(X,P) \tag{83}$$

tako da se očekivana korisnost pozitivnog dela loza izračunava nezavisno od očekivane korisnosti negativnog dela loza. Pošto se evaluira pozitivni i negativni deo loza, ukupna očekivana vrednost mešovitog loza je zbir očekivanih korisnosti

pozitivnog i negativnog dela. Postavlja se pitanje kakva je evaluacija lozova oblika ($x$,$p$;-$y$,1-$p$), lozova koji sadrže samo jedan dobitak i jedan gubitak? Pošto se loz dekomponuje na pozitivni i negativni deo, svaki od njih sadrži samo po jedan ishod, i sledi da bi *a priori* verovatnoća u tom slučaju morala biti jedan. Ovaj problem se rešava tako što se loz oblika ($x$,$p$;-$y$,1-$p$) dekomponuje u pozitivni deo ($x$,$p$;0,1-$p$) i negativni deo (0,$p$;-$y$,1-$p$). Pošto je dekumulativna funkcija $S$ dobro definisana u tački nula - gde uvek ima vrednost jedan - evaluacija pozitivnog i negativnog dela nije problematična. Drugim rečima, pozitivni i negativni deo mešovitog loza koji sadrže samo po jedan ishod se edituju u lozove koji sadrže odgovarajuće ishode i ishod 0 sa preostalom verovatnoćom. Ishod 0, po pretpostavci koja se koncenzusom prihvata u diskusijama teorije izgleda, ima korisnost $u(0)=0$, tako njeno uključivanje u loz nema nikakvog efekta na očekivanu korisnost loza. Interesantno je da Viskuzi diskutuje ovakvu opciju (ne u kontekstu analize mešovitih lozova), ali predlaže drugačije rešenje od našeg, rešenje koje je po našem mišljenju nedovoljno motivisano (up. dodatak A.1 njegovom radu Viscusi, 1989). Viskuzijev predlog počiva na pretpostavci da će donosilac odluka uvek formirati distribuciju *a priori* verovatnoća za odgovarajući loz tako da se te verovatnoće sabiraju do jedan (tj. da čine pravu distribuciju verovatnoća); ovakvo objašnjenje ostaje *ad hoc* pošto Viskuzi nije motivisao ovu pretpostavku teorije perspektivne reference. Primena reprezentacije prethodnih verovanja putem dekumulativne funkcije $S$ i normalizacija koju uvek obezbeđuje Lusov aksiom izbora omogućavaju teoriji poverenja da izbegne ovaj i više sličnih problema - problema koje je već prepoznao i diskutovao Viskuzi u razvoju teorije perspektivne reference.

U odnosu na savremene bihejvioralne teorije odlučivanja poput kumulativne teorije izgleda, teoriji poverenja preostaje posao inkorporacije još jednog robustnog empirijskog fenomena: *averzije prema gubicima*.

*Status averzije prema gubicima.* Inkorporacija parametra averzije prema gubicima u matematički model teorije poverenja nije problematična. Savremene teorije odlučivanja, videli smo, razmatraju tri glavna izvora rizika koji učestvuju u evaluaciji lozova: rizik koji je posledica forme funkcije korisnosti, probabilistički rizik koji se odnosi na subjektivni tretman verovatnoća, i rizik koji potiče iz odnosa prema referentnoj tački odn. rizik vezan za averziju prema gubicima. Sva ova tri izvora rizika su nezavisna; tako, slobodno možemo da inkorporiramo parametar averzije prema gubicima u model teorije poverenja na isti način na koji je on inkorporiran u teoriju izgleda. Sledeće jednačine onda daju funkcije korisnosti za dobitke i gubitke

u teoriji poverenja:

$$u(x) = \begin{cases} x^{\rho} & x \geq 0 \\ -\lambda(-x)^{\rho} & x < 0 \end{cases} \tag{84}$$

i potpuno su identične jednačinama (7-8) za kumulativnu teoriju izgleda.

*Normativni status teorije poverenja.* Teorija poverenja, kao teorija o formiranju verovanja u odlučivanju u uslovima rizika, nema nikakvu obavezu korespondencije prema normativnim uslovima teorije odlučivanja. Proces formiranja verovanja je potpuno izolovan (kognitivno enkapsuliran) od procesa evaluacije lozova tj. odlučivanja samog. Između njih ne postoji nikakva interakcija. Na algoritamskom nivou analize, ova dva procesa operišu serijalno: prvo formiranje verovanja, a zatim evaluacija lozova i donošenje odluka. Uopšte, naša strategija konstrukcije teorije poverenja imala je za cilj da postigne upravo to da se *izvori fenomena ograničene racionalnosti izoluju od elemenata donošenja odluka.* Ako analiziramo strukturu Viskuzijeve teorije, uključujući i aksiomatsku analizu koju on pruža (Viskusi, 1989), u njoj ne nalazimo ništa što bi bilo *prima facie* „ograničeno racionalno“: njena struktura se od strukture teorije fon Nojmana i Morgenšterna razlikuje tek u tehničkim detaljima neophodnim da se inkorporiraju reprezentacija *a priori* verovatnoća i mehanizam bejzijanske inferencije. Samo odlučivanje uzima oblik dobro poznate jednačine očekivane korisnosti. *A posteriori* verovatnoće na osnovu kojih se evaluiraju lozovi su, posle bejzijanske inferencije, prave verovatnoće (tj. podležu aksiomima Kolmogorova), za razliku od najčešće *subaditivnih* kapaciteta odn. pondera odluke koji se koriste u teoriji Tverskog i Kanemena. Konzistentna upotreba mere verovatnoće u teoriji odlučivanja svakako nije osobina koju treba zanemarivati, a činjenica da je teorije izgleda ne obezbeđuje predstavlja možda najslabiju teorijsku tačku u celoj njenoj konstrukciji.

Formiranje verovanja, koje u našem modelu objašnjava isključivo komponenta teorije poverenja, bazira se na formalnim rešenjima za koje mi ne možemo da kažemo da su na bilo koji način deskripcija neke ograničene racionalnosti. Upotreba prethodnih informacija, tamo gde su te prethodne informacije korisne, potpuno je racionalna odluka. Bejzijanski mehanizam kojim se one inkorporiraju u strukturu informacija koju daje sredina predstavlja koncenzusom prihvaćeno normativno rešenje za problem revizije verovanja uopšte. Lusov aksiom, čija primena obezbeđuje da *a priori* verovatnoće formiraju pravu distribuciju verovatnoća u bejzijanskoj

inferenciji, odslikava toliko jednostavnu intuiciju o izbornim verovatnoćama da je teško dovesti u pitanje njegovu normativnu adekvatnost. Upotreba informacije o neizvesnosti (relativnoj entropiji) koju nosi distribucija *a priori* verovatnoća, koja se koristi da bi se odredio stepen u kome te prethodne informacije treba da koriguju informacije koje donosi loz, na prvi pogled predstavlja čisto deskriptivnu intervenciju. Mi smo naglašavali da je njena uloga da omogući veću fleksibilnost linearne funkcije ponderisanja verovatnoća. Uloga ovog mehanizma teorije poverenja postaće jasnija u narednoj sekciji, posle primene njenog modela na eksperimentalne podatke. Raspravu o normativnom statusu ovog mehanizma ostavljamo otvorenom, svesni toga da su mogući i drugačiji teorijski predlozi[84] o ovom mehanizmu koji suštinski određuje *stepen poverenja* koji donosilac odluke ulaže u informacije koje su mu predočene - i po kome cela teorija poveranja nosi ime.

Mi smo u model teorije poverenja inkorporirali parametar averzije prema gubicima, $\lambda$. Averzija prema gubicima se smatra najrobustnijim odstupanjem od normativne racionalnosti u odlučivanju (Wakker, 2010). Podsetimo se našeg primera iz sekcije 7.1 - primera sa odlukom o tome za kojim zebrama treba a za kojima ne treba da potrči lavica u lovu. Lavica, uvideli smo, mora da uračunava određeni trošak u energiji koju će uložiti u trk za nekom zebrom; ukoliko je njena procena karakteristika te zebre takva da ona predstavlja potencijalni gubitak (rasipanje energije bez izgleda za ulov), sledi da je racionalno da je njen sistem odlučivanja kroz parametar averzije prema gubicima upozorava na mogućnost takvog bespotrebnog gubitka energije. Ali, šta ako govorimo o nekoj fiktivnoj lavici koja je neprestano prezasićena, i za koju je lov tako postao više igra nego stvar realnog preživljavanja? Bilo koja osoba koja raspolaže proizvoljno visokim resursima, ili veruje da raspolaže takvim resursima, u tolikoj meri da ne mora da se brine za lokalne, manje dobitke i gubitke, može sebi (lokalno) da dopusti da bude *sklona dobicima*, što bi dovelo do toga da vrednost $\lambda$ bude veća od jedan. U tipičnom eksperimentu merenja monetarnih ekvivalenata rizičnih lozova, na primer, ispitanici ne moraju da se ponašaju prema fiktivnim ishodima onako kako bi se ponašali u realnosti. Oni mogu da pristupe suđenju o monetarnim ekvivalentima „opušteno“, određujući njihove minimalne cene tako da zadovolje svoje stavove prema količini novca koju bi na njima voleli da zarade, iako bi u realnim situacijama verovatno zauzeli konzervativnije stavove. Takvi stavovi u eksperimentalnim situacijama predstavljaju izvor sklonosti prema dobicima. U većini eksperimenata u oblasti odlučivanja u uslovima rizika, ispitanici se plaćaju nasumičnim izborom nekih od lozova koje su procenili ili

odigrali, koji se onda stvarno odigravaju. Ovakva procedura bi trebalo da garantuje da će ispitanik evaluirati lozove kako bi to činio i da oni jesu realne ekonomske interakcije kojima je svakodnevno izložen. Ipak, ni takve procedure ne garantuju da se među ispitanicima neće pojaviti „kockar“ koji je sklon dobicima, ili ima konveksnu funkciju korisnosti, ili potcenjuje male a precenjuje visoke verovatnoće. Na osnovu empirijskih nalaza koje smo diskutovali u sekciji 7.1 i i analize koju smo ovde predstavili, mi smatramo da fenomen averzije prema gubicima nosi jednu u suštini pogrešnu konotaciju, implicirajući da je prirodan, očekivan bihejvioralni nalaz taj da ispitanici pokazuju averziju prema gubicima, iako oni, videli smo, često pokazuju sklonost ka dobicima. Videli smo da i te kako postoje situacije, kao i da one nisu neočekivane u eksperimentima, u kojima ispitanici mogu da demonstriraju upravo efekte obrnute od averzije prema gubicima. Mi predlažemo da ovaj izvor rizika u savremenim teorijama odlučivanja zovemo jednostavno *zavisnošću od referentne tačke*, što je termin koji je neutralan u odnosu na to da li se govori o dobicima ili o gubicima. Kao što smo videli u sekciji 7.1, zavisnost od referentne tačke je prilično nestabilan empirijski fenomen, koji je u nekim situacijama izražen više a u nekim manje, i koji odlikuje visoka intersubjektivna varijabilnost, uključujući tu i mogućnosti manipulacije ovim efektom.

Već je Viskuzi pokazao da teorije perspektivne reference obuhvata tipične empirijske nalaze ograničene racionalnosti poput Aleovog paradoksa i drugih rezultata koji se objašnjavaju ponderisanjem verovatnoća uopšte (up. Viscusi, 1989). Pošto se Viskuzijeva teorija proširi teorijom poverenja, ona je u stanju da obuhvati još širu klasu empirijskih fenomena. Analiza nekih od tih fenomena pod teorijom poverenja predstavljamo u Prilogu A na kraju rada (*Formalna analiza ograničene racionalnosti pod Viskuzijevom teorijom i teorijom poverenja*).

U narednoj sekciji prikazujemo rezultate primene standardne eksperimentalne metodologije u testiranju nove teorije. Poredićemo njenu eksplanatornu moć sa onom kumulativne teorije izgleda Kanemana i Tverskog, ukazujući na strukture eksperimentalnih podataka koje vodeća deskriptivna teorija odlučivanja ne može da uklopi, i otkrivajući načine na koje ih u svoju strukturu uklapa teorija poverenja. Koristimo i dodatne eksperimentalne metode koji služe za proveru bitnih pretpostavki teorije poverenja.

# 11 Selekcija modela odlučivanja

U sledećim sekcijama predstavljamo eksperimentalne testove teorije poverenja. Eksperiment u sekciji 11.1 testira fundamentalnu pretpostavku teorije poverenja: da donosioce odluka odlikuju kognitivne reprezentacije *a priori* verovatnoća monetarnih ishoda. Eksperimenti u sekciji 11.2 koriste metodologiju ocene monetarnih ekvivalenata sa najopsežnijim nacrtom sistematskog merenja monetarnih ekvivalenata rizičnih lozova koji je do sada primenjen. Pored ovih originalnih eksperimenata, u sekciji 11.2 reanaliziramo eksperiment merenja monetarnih ekvivalenata Gonzalesa i Vua iz 1999. godine (Gonsales & Wu, 1999). U sekciji 11.3 koristimo eksperimente izbora između rizičnih lozova u procesu selekcije modela odlučivanja.

## 11.1 Reprezentacija monetarnih vrednosti

Fundamentalna pretpostavka teorije poverenja je da donosioce odluka karakterišu kognitivne reprezentacije distribucija *a priori* verovatnoća za dobitke i gubitke. U eksperimentima 1a i 1b pokušavamo da proverimo ovu hipotezu oslanjajući se na metodologiju merenja sličnosti između monetarnih dobitaka i gubitaka. Prema našem znanju, ova metoda nikada nije korišćena u domenu monetarnih vrednosti. Ideja počiva na sledećem predlogu: sud o sličnosti između dva različita monetarna dobitka (ili gubitka), npr 50 evra i 150 evra, može da bude (a) funkcija objektivne razlike između njih, (b) funkcija razlike između subjektivnih korisnosti, (c) funkcija objektivne razlike između njih *i* razlike u *a priori* verovatnoći da one budu osvojene, ili (d) funkcija razlike između subjektivnih korisnosti *i* razlike u *a priori* verovatnoći da one budu osvojene. Eksperimenti 1a i 1b se izvode po standardnoj metodologiji procene sličnosti, dajući tako podatke iz koje je moguće zaključiti koja od hipoteza (a-d) najbolje objašnjava sudove ispitanika. Od ispitanika se, pored ocena sličnosti, zahteva i procena relativnih frekvencija odgovarajućih dobitaka i gubitaka, čime dobijamo varijablu koja može da posluži u prepoznavanju *a priori* verovatnoća u kognitivnim reprezentacijama dobitaka i gubitaka. U zavisnosti od toga koja hipoteza o reprezentaciji monetarnih dobitaka i gubitaka može da objasni distance dobijene primenom nemetričkog multidimenzionalnog skaliranja na procene sličnosti, donećemo zaključak o tome koja je najverovatnija forma te reprezentacije.

## Eksperimenti 1a i 1b

Eksperimentima 1a i 1b prikupljeni su podaci o subjektivnim procenama različitosti monetarnih dobitaka u evrima.

### Metod

*Ispitanici.* U eksperimentu 1a učestvovalo je trideset i osam studenata, oba pola, I godine psihologije Departmana za psihologiju Fakulteta za medije i komunikacije, Univerzitet Singidunum, u okviru obaveznih vežbi na kursu psihologije; četrdesetoro njihovih kolega, takođe oba pola, takođe I godine studija, učestvovalo je u eksperimentu 1b. Ispitanici su bili naivni u odnosu na poznavanje teorije odlučivanja, ali su neki od njih koncept sličnosti i metode procene sličnosti prethodno proučavali u okviru odgovarajućeg kursa.

*Dizajn i stimulusi.* Stimulusi eksperimentu 1a su parovi monetarnih vrednosti u evrima: parovi dobitaka, i parovi gubitaka. Polazeći od skupa ishoda: 10 *EUR*, 25 *EUR*, 50 *EUR*, 75 *EUR*, 100 *EUR*, 150 *EUR* i *250* EUR, generisano je svih mogućih 21 različitih parova ishoda. Na isti način su generisani parovi gubitaka. Tako je ukupno bilo 42 para monetarnih vrednosti unutar kojih je procenjivana sličnost. Potpuno ista procedura generisanja stimulusa primenjena je u eksperimentu 1b, samo što je polazni skup ishoda bio u dinarskim vrednostima, i to: 50 *RSD*, 100 *RSD*, 200 *RSD*, 500 *RSD*, 1000 *RSD*, 5000 *RSD* i 7000 *RSD*.

*Procedura.* Eksperimentalna procedura je ista u eksperimentima 1a i 1b. je Ispitanici su procenjivali *stepen različitosti* između monetarnih dobitaka i gubitaka. Svaki ispitanik je dobio listu parova dobitaka i gubitaka. Prvo su procenjivane različitosti između dva monetarna dobitka u svakom paru, a zatim između dva monetarna gubitka u odgovarajućim parovima. Korišćena je skala Likertovog tipa sa podeocima od 0 do 100 i korakom od 5; ispitanicima je dato sledeće upustvo: „*Prvo procenjujete sličnost novčanih dobitaka. Zamislite da ste u prilici da slučajno, bez posebnog razloga, dobijete neki od dva ponuđena iznosa na skali. Koliko su za Vas, na skali od 100 stepeni, slična ili različita ova dva dobitka? Manji brojevi na skali odgovaraju većoj sličnosti, veći brojevi na skali - većoj različitosti.*“ Slično uputstvo je ponovljeno pre početka procene različitosti monetarnih gubitaka. Ispitanici su dodatnim uputstvima upućeni na to da koriste sve podeoke skale koji im se čine pogodnim za procenu različitosti u određenom paru, vodeći računa da koriste i ekstremne vrednosti skale kada smatraju da to odgovara njihovom sudu. Korišćena

su dva slučajna rasporeda stimulsa. Položaj monetarnih dobitaka (gubitaka) u odnosu na skalu procene (levo ili desno) bio je balansiran koliko je to bilo moguće sa ovakvim eksperimentalnim nacrtom; svaka upotrebljena vrednost nalazila se levo odn. desno od skale procene aproksimativno podjednak broj puta.

Posle procene različitosti u parovima monetarnih dobitaka i gubitaka, od ispitanika je tražena procena relativnih frekvencija dobitaka i gubitaka koji su korišćeni u eksperimentu 1. Formular je sadržao sledeću instrukciju (navodimo je u formi korišćenoj za eksperiment 1a; isto uputstvo je korišćeno za eksperiment 1b, samo su vrednosti u *RSD* zamenile vrednosti u *EUR*):

„Zamislite sledeću situaciju. Vi vodite neki posao u kome su moguće zarade od:

10 EUR, 25 EUR, 50 EUR, 75 EUR, 100 EUR, 150 EUR, 200 EUR

Pod zaradom ovde ne mislimo na neke sigurne dobitke poput nedeljne ili mesečne plate, ili dnevnice za neki posao. Razmišljate o ovome kao o nekom poslu koji Vi vodite; svaki put kada se uopšte ukaže prilika za poslovanje, ukazuje se i određena šansa da zaradite nešto novca. U zavisnosti od toga kako posao "ide", moguće su različite zarade: nekad veće, nekad manje. Pretpostavite da ste u nekom poslu u kom su moguće samo ove zarade koje su gore navedene: 10 evra, 25 evra, 50 evra, 75 evra, 100 evra, 150 evr i 200 evra. Šta mislite, koliko često može da se zaraditi više od svakog od ovih iznosa? Koliko često u nekom ovakvom poslu biste mogli da zaradite više od 10, koliko često više od 25, koliko često više od 50, koliko često više od 75, koliko često više od 100, koliko često više od 150, a koliko često više od 200 evra? Odgovore dajete popunjavanjem tabele koja je odštampana ispod ovog uputstva.

Za svaki od ovih iznosa treba da navedete broj od 1 do 100 koji govori o tome koliko puta u 100 prilika da se ostvari zarada veća od navedenog iznosa bi se takva zarada zaista i ostvarila.

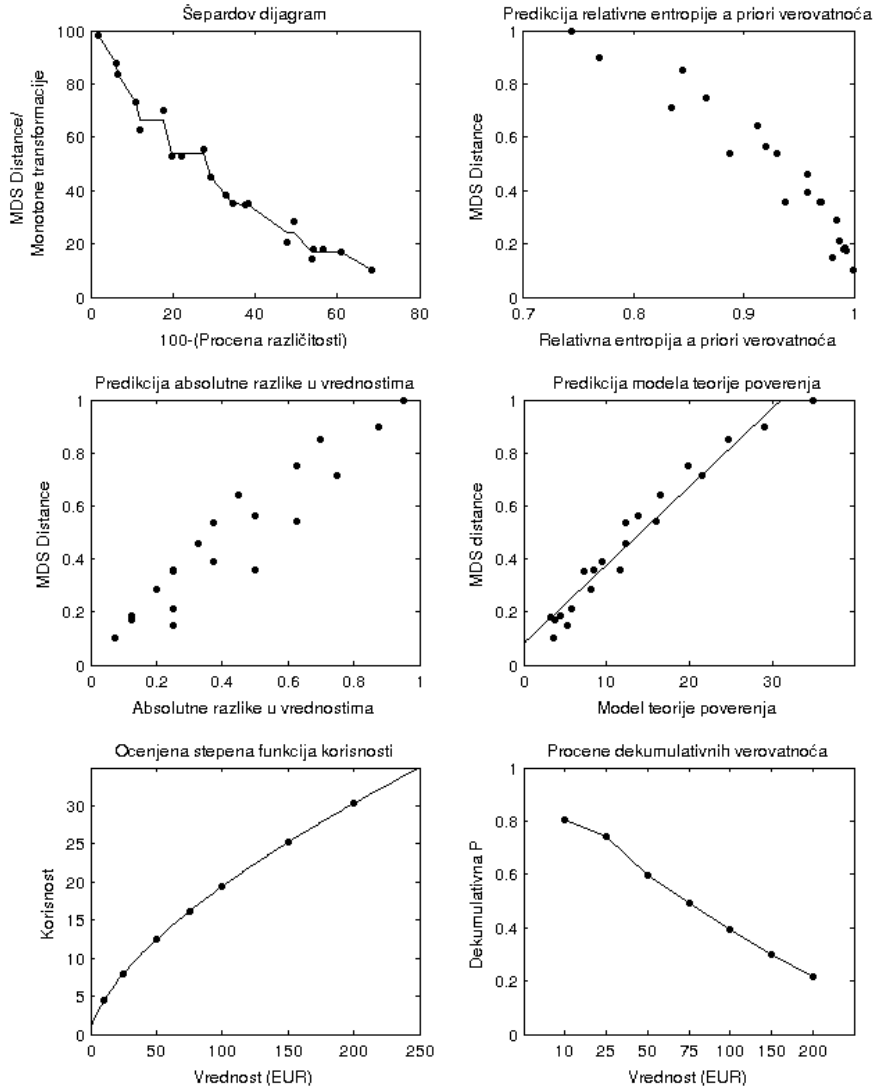
Pokušajte da za svaki iznos u tabeli date Vaše mišljenje o sledećem:

KOLIKO PUTA BI SE OD 100 PRILIKA U KOJIMA SE UKAZUJE ŠANSA ZA ZARADU VEĆU OD ODREĐENOG IZNOSA DOGODILO DA SE TA ZARADA ZAISTA I OSTVARI? Dakle, u svaki red za odgovore unesite broj od 1 do 100 koji označava šta vi mislite koliko puta od 100 prilika za zaradu veću od iznosa u tom redu bi se ta zarada zaista ostvarila.“

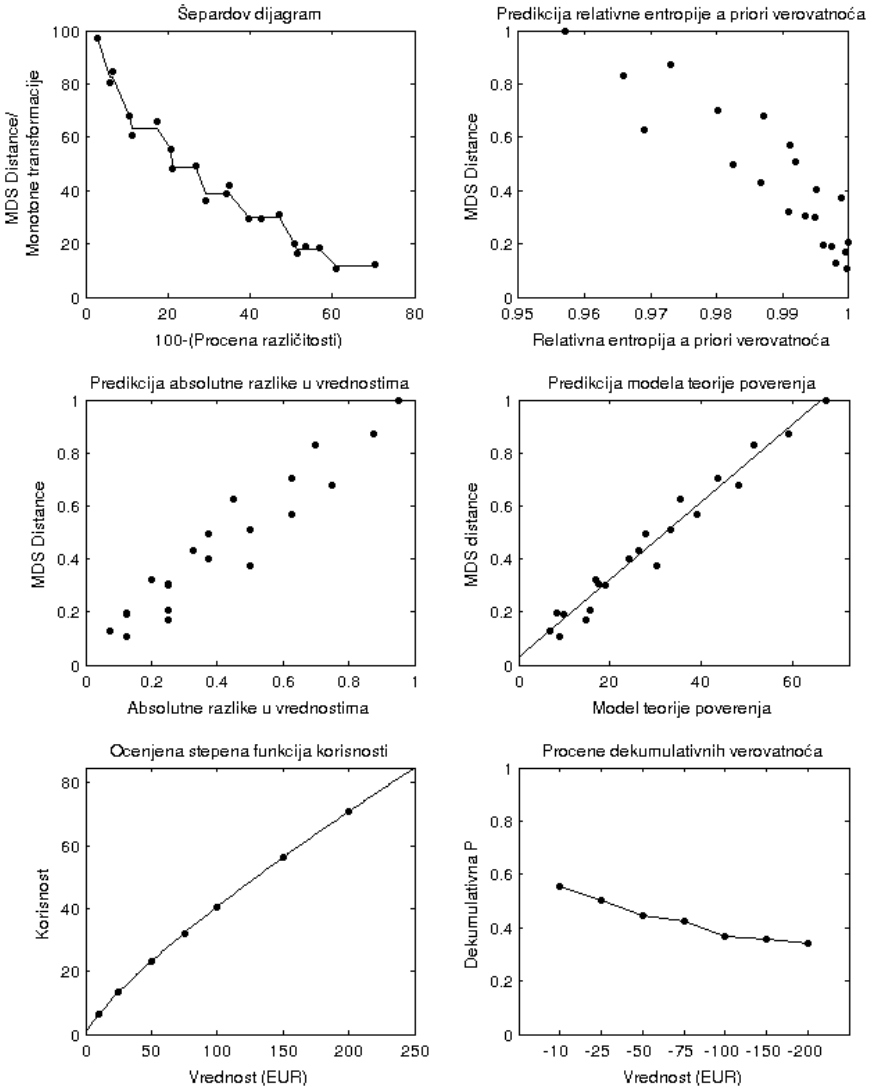
Ispod ovog uputstva je predstavljena tabela čija je prva kolona sadržala odgovarajuće dobitke, a pored nje se nalazila kolona u koju je trebalo upisati celobrojnu procenu od 1 do 100. Posle popunjavanja ove tabele, ispitanici su upućeni na drugu stranu formulara gde su popunjavali istovetnu tabelu kojom se zahtevala procena relativnih frekvencija gubitaka. Procedura merenja relativnih frekvencija u ovom eksperimentu očigledno korespondira sudu o dekumulativnim verovatnoćama odgovarajućih događaja. Ispitanici koji su učestvovali u eksperimentu 1b nisu davali ove procene relativnih frekvencija gubitaka.

### Rezultati

Na osnovu podataka 38 ispitanika koji su uzeli učešća u eksperimentu 1a izračunate su prosečne procene različitosti za po 21 par monetarnih dobitaka i gubitaka. Iz ovih podataka formirane su matrice različitosti koje su zatim podvrgnute nemetričkom multidimenzionalnom skaliranju. Korišćena je minimizacija klasične *stress* mere (Borg & Groenen, 2005). Jednodimenzionalna rešenja pokazuju sledeće karakteristike: $stress = 0.034$ za matricu različitosti dobitaka, i $stress = 0.030$ za matricu različitosti gubitaka. Levi panel u prvom redu slika 22a (dobici) i 22b (gubici) sadrži Šepardov dijagram za odgovarajuću finalnu konfiguraciju u jednoj dimenziji: na slici 22a za dobitke, na slici 22b za gubitke. Ostali paneli na slikama 22a i 22b biće objašnjeni dalje u tekstu.




Slika 22a. *Rezultati modeliranja distanci finalne jednodimenzionalne MDS konfiguracije za matricu različitosti monetarnih dobitaka u eksperimentu 1a.* Objašnjenje u tekstu.




Slika 22b. *Rezultati modeliranja distanci finalne jednodimenzionalne MDS konfiguracije za matricu različitosti monetarnih gubitaka u eksperimentu 1a.* Objašnjenje u tekstu.

Na abscisi Šepardovog dijagrama se nalaze prosečne procene ispitanika (reskalirane u sličnost kao *100-različitost*, samo da bi dijagram uzeo formu u kojoj se najčešće sreće u literaturi), dok na oordinati nalazimo dve mere. Jedna su distance (tačke na dijagramu) na jednoj izdvojenoj dimenziji u MDS rešenju za matricu različitosti, drugo su monotone transformacije podataka (engl. *disparities*), tipične u upotrebi u nemetričkim MDS procedurama. Za nas te monotone transformacije nisu od značaja; uneli smo ih na dijagram radi kompletnosti i konzistentnosti sa

standardnim formama Šepardovog dijagrama. Oba Šepardova dijagrama, za dobitke i gubitke, odlikuje očekivana, blaga nelinearnost.

Desni panel u prvom redu slika 22a i 22b prikazuje odnos između relativne entropije izračunate po jednačini (81) teorije poverenja i distanci iz finalne MDS konfiguracije. Distance su reskalirane tako da maksimalna distanca iznosi 1; ovo reskaliranje nije problematično pošto predstavlja samo promenu merne jedinice (nemetrička MDS rešenja su inače invarijantna do u monotone transformacije). Relativne entropije za svaki par monetarnih dobitaka odn. gubitaka izračunate su na sledeći način. Dekumulativne verovatnoće dobitaka (gubitaka) u paru u kome je procenjena različitost dobijene su na osnovu procena relativnih frekvencija koje su ispitanici davali posle procene različitosti. Za svaki par dobitaka (gubitaka), odgovarajuće dekumulativne verovatnoće su normalizovane primenom Lusovog aksioma izbora. Iz dobijenih verovatnoća izračunata je entropija, i zatim relativna entropija ovih *a priori* verovatnoća, za svaki par monetarnih dobitaka (gubitaka) koji je dat na procenu različitosti. Podsećamo, ispitanici su prvo davali procene različitosti, a zatim procene dekumulativnih verovatnoća; ove druge, dakle, nikako drugačije nisu mogle da imaju efekta na procenu različitosti osim do ako su bile prethodno reprezentovane i upotrebljene u donošenju odgovarajućeg suda. Vidimo da postoji očekivan, aproksimativno monoton, opadajući odnos između relativne entropije *a priori* verovatnoća i distanci između odgovarajućih dobitaka (gubitaka). Što je razlika između dva monetarna dobitka ili gubitka veća, to je, sasvim očekivano, veća i razlika u verovatnoći da se oni realizuju. Sa porastom te razlike između njih, njihova relativna entropija opada.

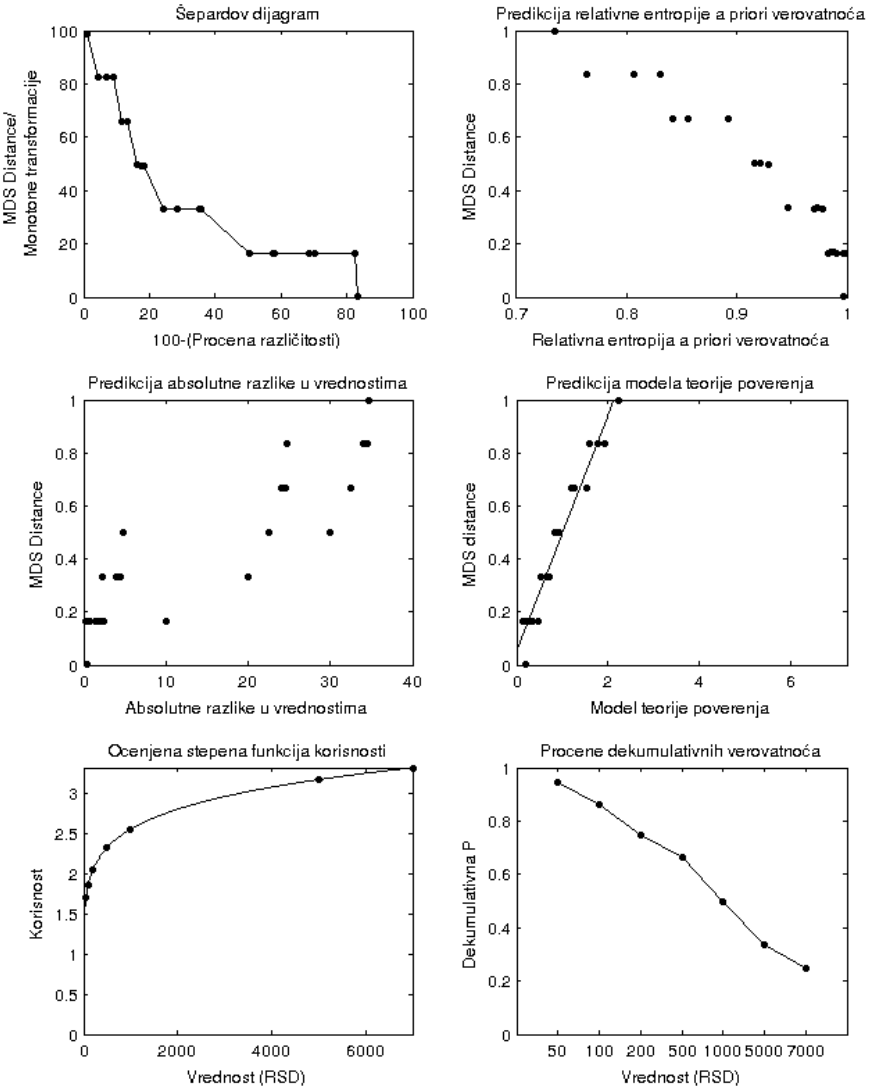
Levi panel u drugom redu slika 22a i 22b prikazuje odnos razlika objektivnih vrednosti monetarnih dobitaka (slika 22a) i gubitka (22b) sa distancama iz finalne MDS konfiguracije. Apsolutne razlike su reskalirane tako da maksimalna iznosi 1. Ponovo očekivano, ovaj odnos je pozitivan. Sa porastom apsolutne vrednosti razlike između dva dobitka (gubitka), raste i procena različitosti između njih. Vidimo da i monetarna vrednost i procena verovatnoće da bi ona mogla da bude realizovana jesu regularno povezani sa distancama između odgovarajućih monetarnih vrednosti u MDS reprezentaciji. Očigledno je da su monetarna vrednost i njena verovatnoća, kako to teorija poverenja i tvrdi, korelirane: niži ishodi imaju više verovatnoće da se realizuju od visokih, bilo da su u pitanju dobici ili gubici. Postavlja se pitanje kako inkorporirati obe ove informacije, jednu o vrednosti ishoda, i drugu o njegovoj verovatnoći, u model koji bi dao bolju predikciju distanci između odgovarajućih

monetarnih vrednosti u MDS reprezentaciji procena različitosti.

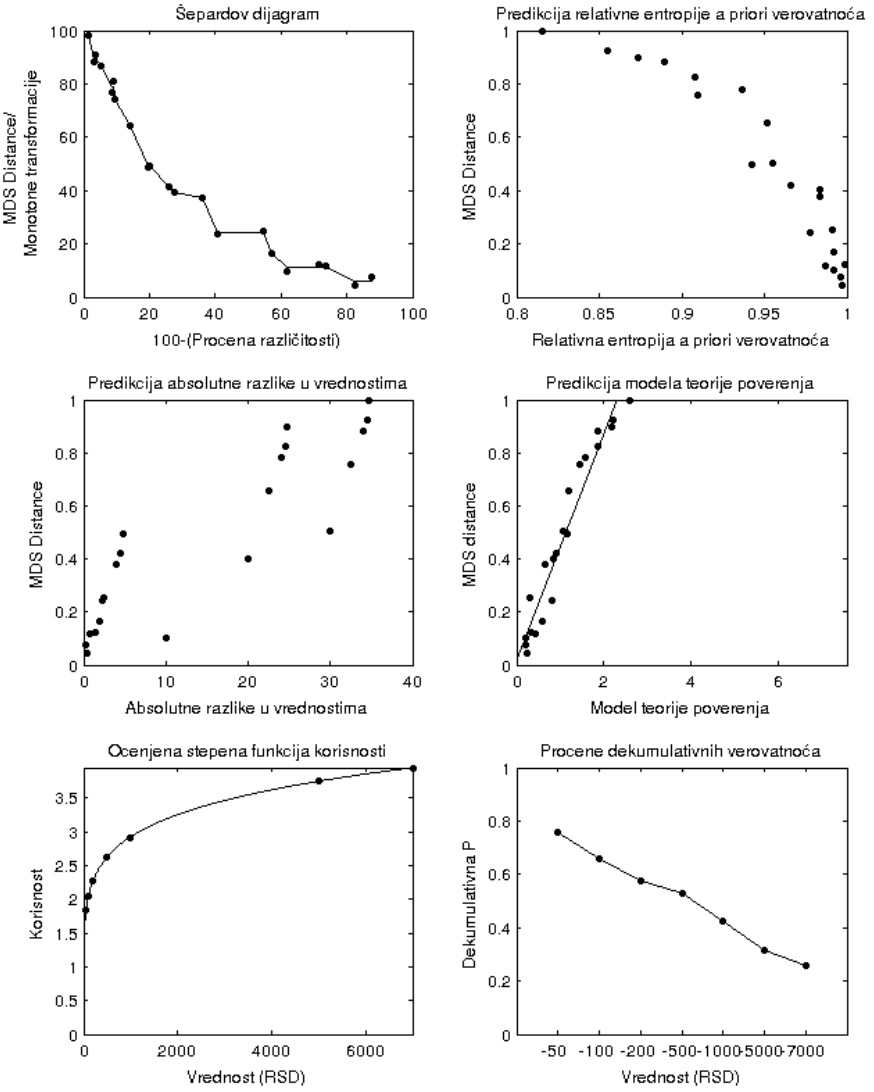
Pokazuje sa da to u okviru teorije poverenja nije teško. Pretpostavljamo da kognitivni sistem ne reprezentuje same monetarne vrednosti, već njihove korisnosti, dakle njihove transformacije odgovarajućom subjektivnom funkcijom. U tom slučaju, distance u reprezentaciji monetarnih vrednosti biće: (a) monotono rastuća funkcija razlike između korisnosti odgovarajućih ishoda, i (b) monotono opadajuća funkcija relativne entropije njihovih *a priori* verovatnoća. Nameće se jednostavan model forme $d_{ij} = f\left(\frac{x_i^\rho - x_j^\rho}{H_{rel}(p_{x_i}, p_{x_j})}\right)$, gde smo sa $d_{ij}$ označili distancu između reprezentacije ishoda $x_i$ i $x_j$; relativna entropija $H_{rel}$ se računa za distribuciju odgovarajućih *a priori* verovatnoća, a $\rho$ se, standardno, odnosi na eksponent stepene funkcije korisnosti. Funkcija $f$ za koju pretpostavljamo da povezuje distance $d_{ij}$ sa ovim modelom je po pretpostavci linearna. Rezultati primene ovog modela u objašnjenju distanci finalnih MDS reprezentacija za dobitke i gubitke nalaze se na desnim panelima u drugom redu slika 22a (dobici) i 22b (gubici). Model je fitovan na distance primenom Nelder-Midovog simpleks algoritma, minimizujući kvadratne razlike između distanci i predikcija modela. Optimalna vrednost eksponenta stepene funkcije korisnosti - jedinog slobodnog parametra u ovom modelu - za dobitke je $\rho_g = .64$, a za dobitke $\rho_l = .80$. Linearna regresiona analiza pokazuje sledeće za predikciju distanci između dobitaka: $R^2 = .95$, $F(1{,}19) = 326.80$, $p < .01$; za za predikciju distanci između gubitaka, $R^2 = .97$, $F(1{,}19) = 563.79$, $p < .01$.

Levi paneli u trećem redu slika 22a i 22b predstavljaju odgovarajuće ocenjene funkcije korisnosti za dobitke (slika 22a) i gubitke (slika 22b). Desni paneli u trećem redu prikazuju profil prosečnih procena dekumulativnih verovatnoća, reskaliranih sa bihejvioralne skale 1-100 na skalu verovatnoće.

Doslovce ista logika modeliranja eksperimentalnih rezultata primenjena je na podacima dobijenim u eksperimentu 1b. Ispitanici u eksperimentu 1b sa dinarskim vrednostima nisu davali procene relativnih frekvencija monetarnih dobitaka i gubitaka. U modeliranju njihovih procena različitosti iskoristili smo procene relativnih frekvencija monetarnih dobitaka i gubitaka koje su dali ispitanici u eksperimentu 2b (v. sekciju Eksperimenti 2a i 2b). Te procene u eksperimentu 2b odnosile su se na tačno iste monetarne vrednosti korišćene u eksperimentu 1b. Ideja je bila sledeća: ukoliko ljudi reprezentuju verovatnoće dobitaka i gubitaka, i ukoliko mi to možemo da pokažemo modeliranjem eksperimentalnih nalaza kao što je to učinjeno za eksperiment 1a, onda bi to moralo da je moguće i pokazati upotrebom *ocena tih verovatnoća sa drugog uzorka iz iste populacije.*




Slika 22c. *Rezultati modeliranja distanci finalne jednodimenzionalne MDS konfiguracije za matricu različitosti monetarnih dobitaka u eksperimentu 1b.* Objašnjenje u tekstu.




Slika 22d. *Rezultati modeliranja distanci finalne jednodimenzionalne MDS konfiguracije za matricu različitosti monetarnih gubitaka u eksperimentu 1b.* Objašnjenje u tekstu.

Modeliranje rezultata eksperimenta 1b pokazuje da je to upravo moguće. Slika 22c (*dobici*) prikazuje za eksperiment 1b iste informacije koje prikazuje slika 22a za eksperiment 1a, i slično tome, slika 22d (*gubici*) za eksperiment 1b prikazuje iste informacije koje slika 22b prikazuje za eksperiment 1a. Reprezentacije u jednodimenzionalnim MDS rešenjima ponovo su bile zadovoljavajuće, sa *stress* = 0.000 za matricu različitosti dobitaka, i *stress* = 0.018 za matricu različitosti gubitaka. Model teorije poverenja je ponovo fitovan primenom Nelder-Midovog simpleks algoritma; optimalna vrednost eksponenta stepene funkcije korisnosti u eksperimentu 1b za dobitke je $\rho_g = .14$, a za dobitke $\rho_l = .16$. Linearna regresiona analiza pokazuje sledeće za predikciju distanci između dobitaka: $R^2 = .96$, $F(1,19) = 417.92$, $p < .01$; za za predikciju distanci između gubitaka, $R^2 = .94$, $F(1,19) = 304.02$, $p < .01$.

Diskusija

Jednostavan model, razvijen na pretpostavkama teorije poverenja, izuzetno uspešno objašnjava distance koje se dobijaju u MDS reprezentaciji matrica različitosti monetarnih dobitaka i gubitaka. Model inkorporira dve različite informacije: (*i*) informaciju o korisnostima samih ishoda čija se različitost ocenjuje, i (*ii*) informaciju o verovatnoći da će se ishod određene visine realizovati. Ove dve mere jesu korelirane, kao što teorija poverenja i predviđa, ali nisu iste: dekumulativne verovatnoće, koje su u eksperimentu 1 direktno procenjivane od strane subjekata, predstavljaju nelinearne opadajuće funkcije odgovarajućih korisnosti u teoriji poverenja. Kombinujući ove dve mere na najjednostavniji način, teorija poverenja pruža izvanrednu predikciju distanci u jednodimenzionalnom reprezentacionom prostoru.

Još jednom, podsetimo se da ispitanici u eksperimentima 1a i 1b donose sudove o različitosti monetarnih ishoda pre nego što od njih zahtevamo ocenu odgovarajućih dekumulativnih verovatnoća (ispitanici u eksperimentu 1b uopšte ni ne donose te sudove, već se koriste ocene dobijene od drugih ispitanika). Eksperimentalna procedura tako garantuje da je ma kakav efekat odgovarajućih verovatnoća realizacije ishoda u proceni različitosti poreklom iz samog reprezentacionog, memorijskog sistema. Drugim rečima, *ljudski kognitivni sistem reprezentuje relevatne verovatnoće realizacije monetarnih ishoda*, ne samo te ishode ili njihovu funkciju korisnosti; takođe, kognitivni sistem koristi ovu probabilističku informaciju u donošenju suda o različitosti odn. sličnosti dva monetarna ishoda.

Zahvaljujući dobrim rezultatima modeliranja eksperimenata 1a i 1b, pokazali smo da postoji još jedna *metoda ocene funkcije korisnosti*: to je, naime, metoda procene različitosti (ili sličnosti) između monetarnih ishoda, praćena procenom odgovarajućih verovatnoća. Za razliku od mukotrpnih eksperimenata merenja monetarnih ekvivalenata koji traju satima, celokupna eksperimentalna sesija eksperimenta poput ovog ne uzima više od 15 do 20 minuta, i za ispitanike sigurno nije naporna. Ocenjene funkcije korisnosti (levi paneli u trećem redu slika 22a-d) su konkavne, pokazujući očekivanu karakteristiku averzije prema riziku. Buduća istraživanja bi trebalo da jasnije odrede odnos ovakvog pristupa merenju funkcije korisnosti prema standardnim metodama, mada je njegov praktičan značaj očigledan.

Ukoliko se pokaže da teorija poverenja može da parira drugim bihejvioralnim teorijama u modeliranju podataka specifičnih za odlučivanje, kao što su sudovi o monetarnim ekvivalentima, rezultati modeliranja eksperimenata 1a i 1b onda pokazuju da je reč o prvoj teoriji koja u jedinstvenom teorijskom okviru pruža analizu procene sličnosti i odlučivanja u uslovima rizika. Naredne dve sekcije testiraju teoriju poverenja upravo u odnosu na njen osnovni predmet: odlučivanje u uslovima rizika.

## 11.2 Sudovi o monetarnim ekvivalenatima

Sledeća dva eksperimenta predstavljaju velike, sistematske skupove ocena monetarnih ekvivalenata u evrima (eksperiment 2a) i dinarima (eksperiment 2b). Oba eksperimenta poštuju isti eksperimentalni dizajn i procedure. Analiza ovih eksperimenata otkriva strukture podataka koje igraju suštinsku ulogu u evaluaciji teorije poverenja.

### Eksperimenti 2a i 2b

Ispitanici su u dva eksperimenta donosili numeričke sudove o monetarnim ekvivalentima rizičnih lozova oblika ($x$,$p$;$y$,$1$-$p$). Dok su ispitanici u eksperimentu 2a donosili sudove o monetarnim ekvivalentima lozova koji su nosili ishode u evrima, ispitanici u eksperimentu 2b su donosili sudove o monetarnim ekvivalentima lozova u dinarima. Pored standardne metodologije procene monetarnih ekvivalenata, ispitanici su dali i procene relativnih frekvencija za sve ishode koji se pojavljuju na lozovima u ova dva eksperimenta.

## Metod

*Ispitanici.* Dvadeset i osam studenata Fakulteta organizacionih nauka Beogradskog univerziteta učestvovalo je u eksperimentu 2a (13 ispitanika) i eksperimentu 2b (15 ispitanika). Svi ispitanici su učešćem u eksperimentima obezbedili kredite za kurs psihologije koji uzimaju u sklopu nastavnog programa. Uzorak je obuhvatio ispitanike oba pola koji prethodno nisu prošli nijedan kurs teorije odlučivanja.

*Dizajn i stimulusi.* 495 rizičnih lozova za eksperiment 2a sa ishodima u evrima generisano je na sledeći način. Skup pozitivnih ishoda (dobitaka) obuhvatao je vrednosti od 0, 25, 50, 75, 100 i 150 evra; skup negativnih ishoda obuhvatao je iste ove vrednosti. Prvo su formirani svi parovi dobitaka, svi parovi gubitaka, i svi parovi dobitaka i gubitaka. Zatim su svi dobijeni parovi upotrebljeni za generisanje lozova tako što je sistematski varirana verovatnoća prvog ishoda u paru kao 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95% i 99%, a verovatnoća drugog loza u paru kao komplementarna do 100% verovatnoći prvog ishoda. Na taj način je generisano po 90 striktno pozitivnih i striktno negativnih lozova (lozova koji sadrže dva dobitka ili dva gubitka, oba različita od nule), po 45 ne-pozitivnih (jedan gubitak i nula) i ne-negativnih (jedan dobitak i nula) lozova, i 225 mešovitih lozova (koji sadrže jedan dobitak i jedan gubitak, oba različita od nule). Dakle, rizični lozovi obuhvataju šest nivoa vrednost i devet nivoa verovatnoće. Lozovi za eksperiment 2b su generisani na isti način pri čemu je umesto skupa vrednosti u evrima korišćen sledeći skup vrednosti u dinarima: 0, 100, 200, 500, 1000, 1500 i 5000. Na dan održavanja eksperimenta, 18. decembra 2011. godine, prema srednjem kursu Narodne banke Srbije jedan evro ($EUR$) je razmenjivan za 101.18 dinara ($RSD$).

Pored ocene monetarnih ekvivalenata rizičnih lozova, ispitanici su dali i ocene relativnih frekvencija dobitaka i gubitaka koji su korišćeni da bi se rizični lozovi formirali. Metodologija procene relativnih frekvencija je bila ista kao ona primenjena u eksperimentu 1, s tim što su ispitanici u eksperimentu 2b procenjivali relativne frekvencije dobitaka i gubitaka u dinarskim vrednostima.

*Procedura.* Procena 495 monetarnih ekvivalenata trajala je aproksimativno između jednog i dva i po časa po ispitaniku. Svakom ispitaniku je data knjižica na kojoj su u odvojenim kolonama navođeni: (1) verovatnoća osvajanja prvog ishoda na lozu izražena u procentima, (2) prvi ishod na lozu, (3) verovatnoća osvajanja drugog ishoda na lozu, (4) drugi ishod na lozu. Na kraju odgovarajućeg reda kojim je kroz

ove četiri kolone definisan svaki od 495 lozova, nalazila se prazna kućica u koju su ispitanici upisivali svoju direktnu numeričku procenu monetarnog ekvivalenta odgovarajućeg loza.

Pre početka sesije procene monetarnih ekvivalenata, eksperimentator je dao svim ispitanicima precizne instrukcije o načinu ocene monetarnih ekvivalenata. Instrukcije su date grupno, kroz prateću prezentaciju na slajdovima koja je sadržala primer svakog tipa loza (pozitivni, negativni, ne-pozitivni, ne-negativni i mešoviti). Lozovi su prikazani vizuelno, standardnim metodom koji loz oblika ($x$,$p$;$y$,$1$-$p$) prikazuje kao dve grane koje polaze od čvora odlučivanja ka odgovarajućim ishodima, a svaka grana je označena verovatnoćom sa kojom se odgovarajući ishod ostvaruje. Ispitanicima je zatim objašnjeno na koji način se lozovi predstavljaju u redovima knjižice koja je sadržala svih 495 lozova. Uputstvo za ocenu monetarnih ekvivalenata pozitivnih lozova glasilo je: „*Vaš zadatak je da odredite minimalan iznos za koji biste prodali ovakav loz*“, za ocenu monetarnih ekvivalenata negativnih lozova: „*Vaš zadatak je da odredite maksimalan iznos koji biste platili da biste izbegli da odigrate ovakav loz*“, i za ocenu monetarnih ekvivalenata mešovitih lozova: „*Vi treba da odlučite da li ćete odrediti minimalan iznos ispod kojeg ne biste prodali loz, ili maksimalan iznos koji biste platili da ne biste morali da ga odigrate*“. Uzimajući u obzir da ispitanicima nije pružena nikakva novčana nadoknada za učešće u ovom eksperimentu, niti su oni mogli da zarade novac na nekim lozovima koje bismo slučajno izvukli i odigrali pred njima, izbegli smo komplikovanje eksperimentalnih instrukcija do kojih bi dovela primena poznate Beker-Degrut-Maršakove procedure (Becker, DeGroot & Marschak, 1964) za koju se veruje da forsira iskrenost u odgovorima ispitanika. Ispitanicima je objašnjeno da bi realno odigravanje loza odgovaralo situaciji u kojoj iz slepe kutije slučajno izvlačimo jednu od sto kuglica, među kojima su pomešane crne i bele kuglice u proporciji verovatnoća (procenata) sa kojima se osvajaju dva ishoda na lozu. Ispitanici su upućeni da koriste celobrojne ocene monetarnih ekvivalenata.

Instrukcijama je naglašeno da negativni lozovi moraju da dobiju negativne monetarne ekvivalente, a pozitivni lozovi pozitivne, kao i da monetarni ekvivalenti moraju da leže u rasponu između najmanjeg i najvećeg ishoda na lozu, tako da je eksperimentalna procedura u tom smislu forsirala odgovore u domen podataka koji su prihvatljivi za analizu standardnih teorija odlučivanja. Ispitanici su odmah posle prezentacije i zadavanja eksperimentalnih instrukcija motivisani da postave ma koje pitanje koje ih je interesovalo u odnosu na metod rada. Takođe, eksperimentator

je bio na raspolaganju za individualna pitanja ispitanika tokom rada, kojih je bilo minimalno i koja su se najčešće odnosila na razjašnjenje i podsećanje na način davanja odgovora kada bi se ispitanik prvi put susreo sa novim tipom loza u knjižici. Zbog dužeg vremena koje je potrebno za ocenu monetarnih ekvivalenata svih 495 lozova, ispitanicima je dozvoljeno da uzmu pauzu bilo kada, uključujući mogućnost da napuste kabinet u kome su radili, prošetaju se, odmore i zatim nastave sesiju.

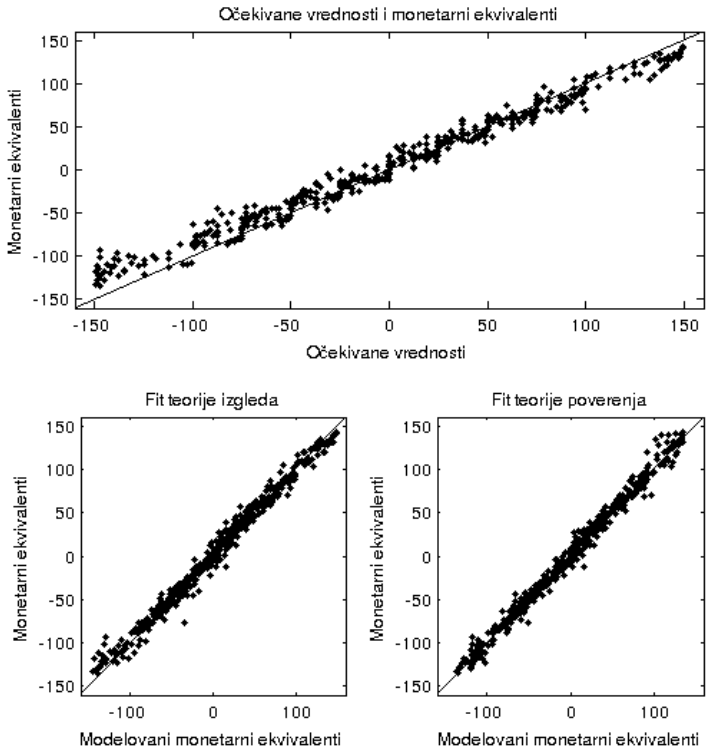
Pre sesije procena monetarnih eksperimenata, ispitanici su popunili formular u kome su davali procene relativnih frekvencija dobitaka i gubitaka koji su korišćeni na lozovima. Procedura i eksperimentalne instrukcije su bili identični kao u eksperimentu 1. Ispitanici u eksperimentu 2b su takođe pružili procene relativnih frekvencija dobitaka i gubitaka, ali je njihov formular sadržao vrednosti u dinarima koje su korišćene na odgovarajućim lozovima.

Rezultati

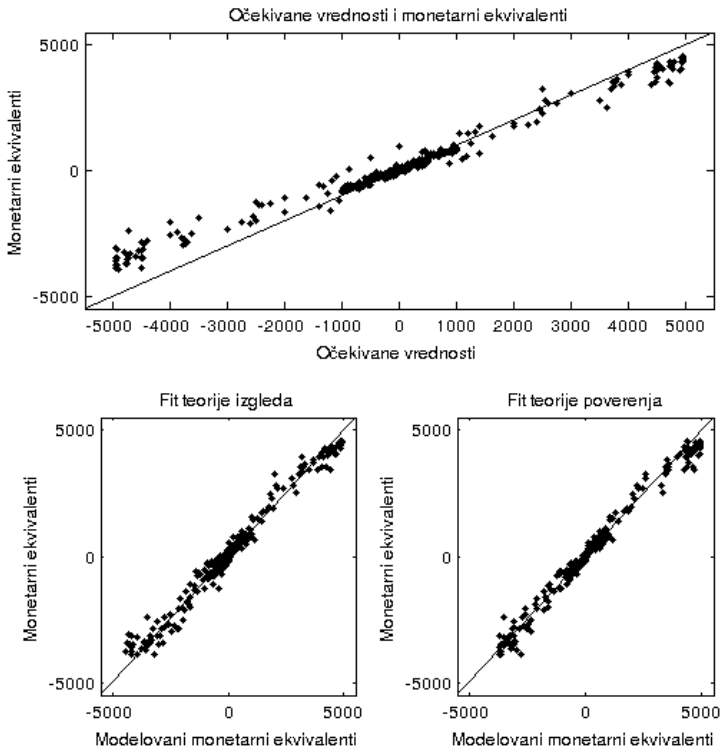
*Provera validnosti i konzistentnosti prikupljenih skupova podataka.* Pre statističkih analiza i modeliranja proverili smo validnost dobijenih podataka u eksperimentima 2a i 2b. Pre svega, provera se odnosila na ispravnost znaka monetarnog ekvivalenta (obavezno pozitivan za pozitivne i negativan za negativne lozove). Druga provera se odnosila na domen monetarnih ekvivalenata, odn. činjenicu da oni za određeni loz moraju da se nalaze između najniže i najviše vrednosti koju taj loz sadrži. Konačno, treća provera validnosti podataka se odnosila na mešovite lozove. Ispitanici, kojima je nekada potrebno i više od dva sata da daju direktne numeričke ocene monetarnih ekvivalenata, skoro sigurno će ponekada pogrešiti u proceduri. Podaci su zato provereni i pročišćeni na sledeći način. Prvo, pretpostavili smo da slučajevi u kojima monetarni ekvivalenti imaju pogrešan znak za striktno pozitivne, striktno negativne, ne-negativne i ne-pozitivne lozove predstavljaju očigledne greške (najčešće „ispuštanje minusa“ u kućici za odgovor) i automatski ispravili znak monetarnog ekvivalenta. Ovo nije moguće konzistentno izvesti za greške koje mogu da se jave kod mešovitih lozova, zbog čega smo se odlučili da podatke analiziramo za svaki procenjeni monetarni ekvivalent u rasponu od -2 do +2 standardne devijacije *kroz procene svih ispitanika.* Uzmimo za primer loz (-50 *EUR*,25%;25 *EUR*,75%). U zavisnosti od karakteristika ispitanika, ovaj loz može da bude prosuđen kao loz sa pozitivnim ili negativnim monetarnim ekvivalentom. Ipak, moguće su i greške („ispuštanje minusa“ pri upisu u kućicu za odgovore). Analizirajući samo monetarne ekvivalente koji se nalaze u rasponu od -2 do +2 standardne devijacije kroz procene svih ispitanika obezbeđuje da,

ukoliko je manji broj ispitanika u eksperimentu napravio grešku u znaku monetarnog ekvivalenta (u odnosu na očekivani, većinski znak odgovora oko kojeg se slaže veći broj ispitanika), njihova procena tog monetarnog ekvivalenta neće biti analizirana. U poslednjim kolonama tabela 6a, 6b, 7a i 7b može se videti koliko od 495 monetarnih ekvivalenata je analizirano za svakog ispitanika u eksperimentima 2a i 2b. Iz tabela se vidi da je visok procenat podataka zadržan u svim analizama. Konačno, ako su monetarni ekvivalenti ležali van granica koji određuju najniža i najviša vrednost u lozu, korigovani su na $min$+1 (negativna granica) ili $max$-1 (pozitivna granica). Broj ovakvih intervencija koje obezbeđuju validnost prikupljenih podataka je bio minimalan. Procedure poput serijskog izbora u merenju monetarnih ekvivalenata, koje se izvode kompjuterski, automatski sprečavaju ispitanike da počine neke od opisanih greški, ali bi u svrhe prikupljanja 495 monetarnih ekvivalenata takvom procedurom bilo potrebno sigurno nekoliko dana.

*Ocena parametera za modele teorija odlučivanja.* Kumulativna teorija izgleda i teorija poverenja su fitovane na procene monetarnih ekvivalenata svih ispitanika kao i na prosečne monetarne ekvivalenete za svaki od dva eksperimenta posebno. Model kumulativne teorije izgleda pratio je prethodno diskutovane jednačine i obuhvatao četiri parametra: eksponent stepene funkcije korisnosti $\rho$, parametre Prelecove jednoparametarske forme funkcije ponderisanja verovatnoća za dobitke $\gamma_g$ i gubitke $\gamma_l$, i parametar averzije prema gubicima $\lambda$; funkcija ponderisanja verovatnoća uzima formu $w(p) = exp(-(-ln(p)^\gamma)$. Model teorije poverenja takođe obuhvata četiri parametra: eksponent stepene funkcije korisnosti $\rho$, parametre dekumulativne funkcije verovatnoća $S$ za dobitke $q_g$ i gubitke $q_l$, i parametar averzije prema gubicima $\lambda$. Vrednosti ovih parametara su određene za svakog ispitanika pojedinačno i za prosečne monetarne ekvivalente u oba ogleda. Posle specifikacija odgovarajućih matematičkih modela za obe teorije, kvadratne razlike predikcija modela i monetarnih ekvivalenata su minimizovane Nelder-Midovom simpleks optimizacijom. Optimizacija Nelder-Midovom simpleks metodom je generalno konvergirala pod uslovima veoma visoke preciznosti u vrednostima parametara i tolerancije na vrednost objektivne funkcije. Pošto je Nelder-Mid heuristička metoda koja ne garantuje konvergenciju u globalni minimum, modeli su optimizovani deset puta za svakog ispitanika, i samo ako je svih deset optimizacija konvergiralo u iste vrednosti parametara te vrednosti su prihvaćene. Ukoliko je makar jedna od deset optimizacija odstupala, modeli su u takvim slučajevima optimizovani sto puta, i prihvaćeno je rešenje sa najmanjom vrednošću sirove kvadratne greške.




Slika 23a. *Prosečni monetarni ekvivalenti u eksperimentu 2a.*




Slika 23b. *Prosečni monetarni ekvivalenti u eksperimentu 2b.*

*Prosečni monetarni ekvivalenti i struktura stavova prema riziku.* Slika 23a. prikazuje (a) odnos očekivanih vrednosti svih 495 lozova i prosečnih monetarnih ekvivalenata (na gornjem panelu), (b) predikcije prosečnih monetarnih ekvivalenata teorije izgleda (donji levi panel) i (c) predikcije prosečnih monetarnih ekvivalenata teorije poverenja (donji desni panel) za eksperiment 2a sa vrednostima ishoda u evrima. Slika 23b. prikazuje iste ove odnose za prosečne monetarne ekvivalente u eksperimentu 2b sa vrednostima ishoda u dinarima. Prvo što možemo videti već sa slika 23a i 23b jeste očekivano visoka korelacija između očekivanih vrednosti monetarnih ekvivalenata i procenjenih monetarnih ekvivalenata. U eksperimentu 2a (vrednosti u evrima), očekivane vrednosti lozova objašnjavaju oko 98% varijanse procenjenih monetarnih ekvivalenata: $R$=0.99, $R^2$=0.98, $F(1,493) = 29484.07$, $p < .01$, $RMSE = 8.74$, slično kao i eksperimentu 2b (dinarske vrednosti): $R$=0.99, $R^2$=0.98, $F(1,493) = 21229.2$, $p < .01$, $RMSE = 234.32$. Naravno, bez obzira na ovako visoke vrednosti koeficijenata determinacije, očekivana korisnost ni iz daleka nije model koji može da objasni finese u eksperimentalnim rezultatima ocene monetarnih ekvivalenata. Pre svega, i teorija izgleda i teorija poverenja, pored skoro nikakvih razlika u vrednostima $R^2$, imaju niže vrednosti $RMSE$ u odnosu na očekivane vrednosti u eksperimentu 2a: $RMSE = 8.15$ za teoriju izgleda i $RMSE = 7.35$ za teoriju poverenja u eksperimentu 2a, dok u eksperimentu 2b, $RMSE = 249.21$ za teoriju izgleda i 227.74 za teoriju poverenja - interesantno je da je $RMSE$ vrednost za teoriju izgleda viša u odnosu na $RMSE$ za očekivane vrednosti u eksperimentu 2b.

Tabela 6a. sadrži (1) optimalne vrednosti parametara, (2) vrednost sirove kvadratne greške ($SSE$), (3) srednje kvadratne greške penalizovane za broj slobodnih parametara ($RMSE$), (4) vrednosti $R$ i $R^2$ za regresionu analizu odnosa predikcija monetarnih ekvivalenata na osnovu optimalnih vrednosti parametara i ocena monetarnih ekvivalenata dobijenih od ispitanika za kumulativnu teoriju izgleda u eksperimentu 2a, i (5) broj podataka na kojima je bazirana ocena parametara. Sve navedene vrednosti $R$ i $R^2$ u ovoj i narednim tabelema značajne su na nivou .01. U poslednjem redu Tabele 6a. nalaze se iste ove informacije za prosečne monetarne ekvivalente ($M_{ME}$) iz eksperimenta 1a. Tabela 6b. sadrži iste informacije za teoriju poverenja u eksperimentu 1a. Tabele 7a i 7b sadrže iste informacije za kumulativnu teoriju izgleda i teoriju poverenja za eksperiment 2b sa vrednostima u dinarima.

Dalje, strukturu prosečnih procenjenih monetarnih ekvivalenata odlikuju osobine koje ni očekivane vrednosti ni teorija očekivane korisnosti ne mogu da objasne.

Poznati nalaz o četvoročlanoj strukturi stavova prema riziku (up. tabelu 4. u Tversky & Kahneman, 1992) je delimično repliciran u eksperimentima 2a i 2b. Ovaj eksperimentalni nalaz se odnosi na sledeću strukturu stavova prema riziku: na malim verovatnoćama dobitaka, donosioci odluka su najčešće skloni riziku (tj. monetarni ekvivalenti su viši od očekivanih vrednosti lozova), dok na većim verovatnoćama dobitaka pokazuju averziju prema riziku (tj. monetarni ekvivalenti su niži od očekivanih vrednosti lozova); za gubitke, struktura je upravo obrnuta: na malim verovatnoćama gubitaka, donosioci odluka pokazuju averziju prema riziku, dok na većim verovatnoćama gubitaka pokazuju sklonost ka riziku. Tabele 5a i 5b prikazuju strukturu procenata monetarnih ekvivalenata koji su veći od odgovarajućih očekivanih vrednosti (dakle, strukturu odgovora koji pokazuju sklonost ka riziku) u eksperimentima 2a i 2b za sve ne-negativne i ne-pozitivne lozove.

Vidimo iz tabele 5a da monetarni ekvivalenti iz eksperimenta 2a ne pokazuju savršenu četvoročlanu strukturu stavova prema riziku, što se pre svega ogleda u rezultatima pet ispitanika čijih 100% monetarnih ekvivalenata koji donose dobitke sa visokom verovatnoćom odlikuje sklonost ka riziku, i četiri ispitanika čijih 100% monetarnih ekvivalenata koji donose gubitke sa visokom verovatnoćom odlikuje averzija prema riziku. Međutim, kvalitativno posmatrano, četvoročlana struktura stavova je prisutna u podacima dobijenim eksperimentom 2a. To se ne može tvrditi za rezultate eksperimenta 2b: tek monetarni ekvivalenti dva ispitanika u tabeli 5b pokazuju savršenu četvoročlanu strukturu stavova prema riziku (ispitanici 7 i 13), dok je kod ostalih ona narušena. Za lozove koji nose gubitke, u eksperimentu 2b je kvalitativno dobijena predviđena struktura, ali vidimo da je za lozove koji nose dobitke dominantna sklonost ka riziku bez obzira da li se radi o nižim ili višim verovatnoćama dobitaka. Ipak, za teoriju očekivane korisnosti, koja predviđa konzistentnu averziju prema riziku (i konzistentnu sklonost ka riziku, u domenu gubitaka), ili model očekivane vrednosti koji ništa ne govori o averziji niti sklonosti ka riziku, strukture podataka dobijene eksperimentima 2a i 2b predstavljaju neobjašnjive nalaze. Tek kada budemo ušli u analizu subjektivnog odnosa prema verovatnoćama u ovim eksperimentima, videćemo koliko su strukture podataka u odlučivanju zaista složene. Tek teorije koje obezbeđuju fleksibilnost u upotrebi verovatnoća i uključuju parametre zavisnosti od referentne tačke, poput teorije izgleda i teorije poverenja, mogu sa uspehom da objašnjavaju ovakve strukture podataka.

Tabela 5a. Četvoročlana struktura stavova prema riziku za eksperiment 2a (EUR).

| | **dobici** | | **gubici** | |
|---|---|---|---|---|
| **Subjekat** | **p≤.1** | **p>.5** | **p≤.1** | **p>.5** |
| 1 | 100 | 100 | 0 | 100 |
| 2 | 100 | 100 | 0 | 100 |
| 3 | 100 | 0 | 0 | 0 |
| 4 | 100 | 0 | 0 | 100 |
| 5 | 0 | 0 | 0 | 0 |
| 6 | 100 | 100 | 0 | 100 |
| 7 | 100 | 100 | 0 | 100 |
| 8 | 100 | 0 | 0 | 0 |
| 9 | 100 | 0 | 0 | 100 |
| 10 | 100 | 0 | 0 | 0 |
| 11 | 100 | 0 | 0 | 100 |
| 12 | 100 | 100 | 0 | 100 |
| 13 | 100 | 0 | 0 | 100 |

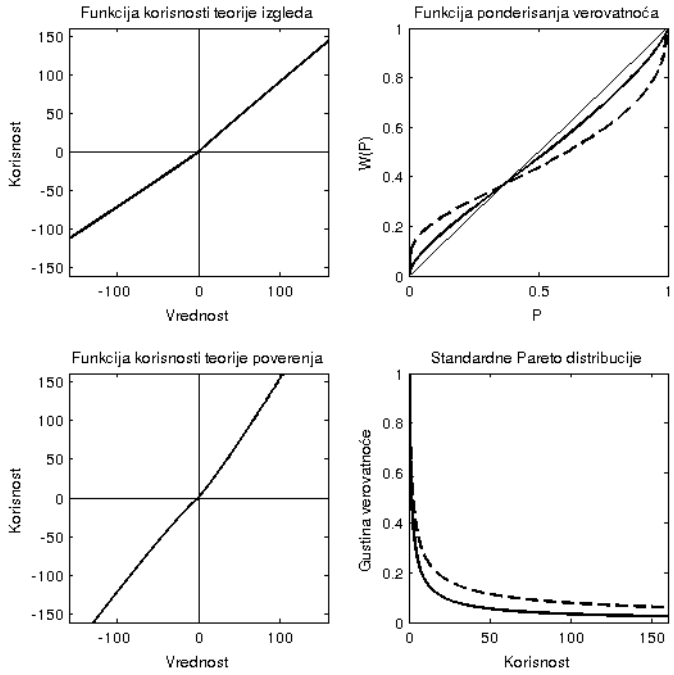
Tabela 5b. Četvoročlana struktura stavova prema riziku za eksperiment 2b (RSD).

| | **dobici** | | **gubici** | |
|---|---|---|---|---|
| **Subjekat** | **p≤.1** | **p>.5** | **p≤.1** | **p>.5** |
| 1 | 100 | 100 | 0 | 100 |
| 2 | 100 | 100 | 0 | 100 |
| 3 | 100 | 100 | 0 | 100 |
| 4 | 100 | 100 | 0 | 100 |
| 5 | 100 | 0 | 0 | 0 |
| 6 | 100 | 100 | 0 | 100 |
| 7 | 100 | 0 | 0 | 100 |
| 8 | 100 | 100 | 0 | 100 |
| 9 | 100 | 0 | 0 | 0 |
| 10 | 100 | 100 | 0 | 0 |
| 11 | 100 | 0 | 0 | 100 |
| 12 | 100 | 100 | 0 | 100 |
| 13 | 100 | 0 | 0 | 100 |
| 14 | 100 | 100 | 0 | 0 |
| 15 | 100 | 0 | 0 | 0 |

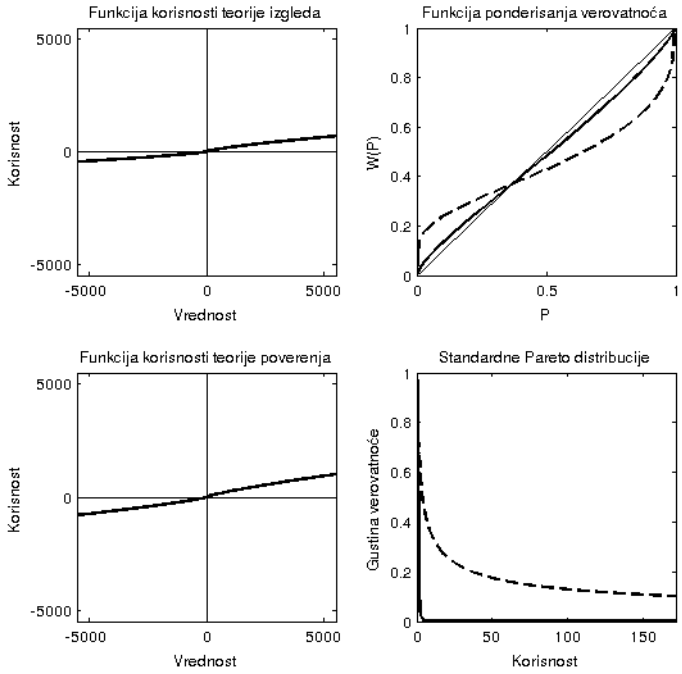
*Rezultati modeliranja modelima teorije izgleda i teorije poverenja.* Posvetićemo sada pažnju rezultatima modeliranja ovih eksperimentalnih nalaza modelima teorije izgleda i teorije poverenja. Slike 24a i 24b prikazuju funkcije ovih teorija sa optimalnim vrednostima parametara dobijenim fitovanjem prosečnih monetarnih ekvivalenata iz eksperimenata 2a i 2b. Gornja dva panela obe slike se odnose na funkcije teorije izgleda; donja dva, na funkcije teorije poverenja. Funkcije ponderisanja verovatnoća na ovim slikama su iscrtane punim linijama za dobitke i isprekidanim za gubitke. Isto tako, dekumulativne funkcije teorije poverenja za reprezentacije *a priori* verovatnoća su iscrtane punim linijama za dobitke, i

isprekidanim za gubitke. Iz tabela 6a i 6b za eksperiment 2a (*EUR*) vidimo da su eksponenti funkcija korisnosti blizu 1, odn. da su te funkcije skoro linearne (teorija poverenja predviđa blago konveksnu funkciju za prosečne monetarne ekvivalente iz ovog eksperimenta). Obe funkcije imaju eksponente manje od jedan za prosečne monetarne ekvivalente u eksperimentu 2b (*RSD*) i pokazuju očekivanu osobinu averzije prema riziku.

Funkcije ponderisanja verovatnoća teorije izgleda su očekivanog inverznog-S oblika u oba eksperimenta, podjednako za dobitke i gubitke. Na slici 24b vidimo da je funkcija ponderisanja verovatnoća za dobitke skoro linearna u eksperimentu 2b (RSD), mada za oba eksperimenta važi da teorija izgleda predviđa značajnije ponderisanje verovatnoća za gubitke nego za dobitke. Dekumulativna funkcija verovatnoće $S$, koja reprezentuje informacije o *a priori* verovatnoćama donosioca odluka u teoriji poverenja, u oba eksperimenta se za gubitke nalazi na višim verovatnoćama nego za gubitke. Ova funkcija za dobitke u eksperimentu 2b (slika 23b) sa ishodima u dinarskim vrednostima je praktično „priljubljena“ uz abscisu: sa ovakvom dekumulativnom funkcijom, donosilac odluka bi praktično sve *a priori* verovatnoće tretirao kao jednake (i veoma bliske nuli), a parametar $N$ bi imao istu vrednost (veoma blisku 1) za sve lozove. Ovakva dekumulativna funkcija u teoriji poverenja obezbeđuje veoma malo ponderisanja verovatnoća uopšte, i dobro korespondira sa rezultatom dobijenim modeliranjem teorijom izgleda koja za eksperiment 2b otkriva skoro linearnu funkciju ponderisanja verovatnoća.




Slika 24a. *Funkcije kumulativne teorije izgleda i teorije poverenja za optimalne vrednosti parametara na prosečnim monetarnim ekvivalentima u eksperimentu 2a (EUR).* Objašnjenje u tekstu.




Slika 24b. *Funkcije kumulativne teorije izgleda i teorije poverenja za optimalne vrednosti parametara na prosečnim monetarnim ekvivalentima u eksperimentu 2b (RSD).* Objašnjenje u tekstu.

Tabela 6a. Fitovi kumulativne teorije izgleda za eksperiment 2a (EUR).

| **Subjekat** | $\rho$ | $\gamma_g$ | $\gamma_l$ | $\lambda$ | **SSE** | **RMSE** | **R** | $\mathbf{R^2}$ | **Podaci** |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.08 | 1.26 | 0.10 | 0.40 | 129285.12 | 16.78 | 0.97 | 0.94 | 463 |
| 2 | 0.73 | 0.89 | 0.04 | 0.89 | 187979.97 | 21.36 | 0.94 | 0.88 | 416 |
| 3 | 0.75 | 0.34 | 0.11 | 0.60 | 164320.75 | 21.05 | 0.91 | 0.83 | 375 |
| 4 | 1.14 | 0.70 | 0.65 | 0.70 | 92815.89 | 13.91 | 0.98 | 0.97 | 484 |
| 5 | 0.85 | 1.42 | 1.34 | 0.86 | 66444.52 | 11.64 | 0.99 | 0.98 | 494 |
| 6 | 0.82 | 0.61 | 0.50 | 0.90 | 128980.48 | 16.29 | 0.97 | 0.94 | 490 |
| 7 | 0.91 | 0.64 | 0.80 | 0.86 | 130702.58 | 16.42 | 0.97 | 0.95 | 489 |
| 8 | 0.99 | 0.92 | 1.04 | 0.82 | 59924.44 | 11.08 | 0.99 | 0.98 | 492 |
| 9 | 0.86 | 1.75 | 0.79 | 0.94 | 95154.50 | 13.99 | 0.98 | 0.97 | 490 |
| 10 | 0.95 | 0.54 | 0.37 | 0.51 | 196147.80 | 20.39 | 0.95 | 0.90 | 476 |
| 11 | 1.05 | 0.95 | 1.11 | 1.15 | 69216.46 | 11.97 | 0.99 | 0.98 | 487 |
| 12 | 1.13 | 0.56 | 0.39 | 0.83 | 149869.10 | 17.80 | 0.97 | 0.94 | 477 |
| 13 | 1.31 | 1.05 | 0.93 | 0.70 | 128805.22 | 16.59 | 0.98 | 0.96 | 472 |
| $\mathbf{M_{ME}}$ | **0.98** | **0.79** | **0.51** | **0.79** | **32675.56** | **8.16** | **0.99** | **0.99** | **495** |

Tabela 6b. Fitovi teorije poverenja za eksperiment 2a (EUR).

| **Subjekat** | $\rho$ | $q_g$ | $q_l$ | $\lambda$ | **SSE** | **RMSE** | **R** | $\mathbf{R^2}$ | **Podaci** |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.03 | 2.83 | 0.29 | 0.41 | 136435.46 | 17.24 | 0.97 | 0.94 | 463 |
| 2 | 0.50 | 18.46 | 0.51 | 1.23 | 217901.05 | 23.00 | 0.93 | 0.87 | 416 |
| 3 | 0.62 | 0.29 | 1.43 | 0.34 | 176630.83 | 21.82 | 0.91 | 0.83 | 375 |
| 4 | 1.34 | 0.55 | 0.53 | 0.61 | 111405.69 | 15.23 | 0.98 | 0.96 | 484 |
| 5 | 0.99 | 13.95 | 10.55 | 0.85 | 75610.60 | 12.42 | 0.99 | 0.97 | 494 |
| 6 | 0.90 | 0.68 | 0.68 | 0.81 | 111441.65 | 15.14 | 0.97 | 0.94 | 490 |
| 7 | 1.15 | 12.58 | 0.63 | 0.82 | 116618.84 | 15.51 | 0.97 | 0.94 | 489 |
| 8 | 0.99 | 6.98 | 11.90 | 0.84 | 59841.39 | 11.07 | 0.99 | 0.98 | 492 |
| 9 | 0.94 | 4.69 | 3.06 | 0.81 | 91115.57 | 13.69 | 0.98 | 0.97 | 490 |
| 10 | 1.15 | 12.58 | 0.63 | 0.82 | 116618.84 | 15.72 | 0.94 | 0.88 | 476 |
| 11 | 1.07 | 5.25 | 7.59 | 1.18 | 68410.13 | 11.90 | 0.99 | 0.98 | 487 |
| 12 | 1.13 | 0.00 | 0.51 | 1.27 | 147483.55 | 17.66 | 0.97 | 0.94 | 477 |
| 13 | 1.36 | 10.85 | 0.86 | 0.80 | 120580.19 | 16.05 | 0.98 | 0.96 | 472 |
| $\mathbf{M_{ME}}$ | **1.09** | **0.74** | **0.55** | **0.80** | **26541.27** | **7.35** | **0.99** | **0.99** | **475** |

Tabela 7a. Fitovi kumulativne teorije izgleda za eksperiment 2b (RSD).

| **Subjekat** | $\rho$ | $\gamma_g$ | $\gamma_l$ | $\lambda$ | **SSE** | **RMSE** | **R** | $\mathbf{R^2}$ | **Podaci** |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.49 | 3.06 | 0.21 | 0.50 | 190074436.58 | 661.79 | 0.92 | 0.84 | 438 |
| 2 | 0.58 | 0.56 | 0.57 | 1.28 | 116226938.94 | 495.70 | 0.94 | 0.89 | 477 |
| 3 | 0.93 | 0.74 | 0.38 | 0.63 | 88007059.21 | 434.58 | 0.97 | 0.94 | 470 |
| 4 | 0.64 | 2.85 | 0.60 | 0.49 | 205636339.91 | 706.48 | 0.92 | 0.85 | 416 |
| 5 | 0.73 | 0.68 | 0.12 | 0.32 | 103051151.85 | 468.25 | 0.94 | 0.89 | 474 |
| 6 | 0.57 | 0.74 | 0.02 | 0.35 | 87880196.96 | 435.20 | 0.94 | 0.88 | 468 |
| 7 | 0.76 | 0.78 | 1.15 | 1.18 | 58364421.56 | 348.70 | 0.98 | 0.96 | 484 |
| 8 | 0.84 | 0.71 | 0.75 | 1.73 | 79800926.28 | 406.05 | 0.97 | 0.95 | 488 |
| 9 | 0.81 | 0.93 | 0.80 | 0.85 | 71191093.45 | 382.34 | 0.98 | 0.96 | 491 |
| 10 | 0.87 | 0.61 | 0.31 | 0.37 | 56556777.26 | 345.79 | 0.98 | 0.95 | 477 |
| 11 | 0.85 | 1.18 | 0.78 | 0.41 | 57062170.38 | 341.25 | 0.98 | 0.97 | 494 |
| 12 | 0.85 | 1.26 | 1.58 | 0.52 | 56272352.53 | 342.39 | 0.99 | 0.97 | 484 |
| 13 | 0.55 | 0.35 | 0.13 | 2.05 | 82546600.86 | 440.71 | 0.87 | 0.76 | 429 |
| 14 | 1.06 | 0.69 | 0.87 | 0.27 | 70897348.29 | 387.15 | 0.98 | 0.96 | 477 |
| 15 | 0.83 | 0.75 | 0.63 | 0.85 | 60068678.72 | 350.84 | 0.98 | 0.96 | 492 |
| $\mathbf{M_{ME}}$ | **0.76** | **0.85** | **0.44** | **0.65** | **30493858.65** | **249.21** | **0.99** | **0.98** | **495** |

Tabela 7b. Fitovi teorije poverenja za eksperiment 2b (RSD).

| **Subjekat** | $\rho$ | $q_g$ | $q_l$ | $\lambda$ | **SSE** | **RMSE** | **R** | $\mathbf{R^2}$ | **Podaci** |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.57 | 11.26 | 0.58 | 0.49 | 218098620.05 | 708.89 | 0.90 | 0.81 | 438 |
| 2 | 0.78 | 0.40 | 0.36 | 1.73 | 119011856.85 | 501.61 | 0.94 | 0.89 | 477 |
| 3 | 1.04 | 1.94 | 0.26 | 0.73 | 83160863.29 | 422.44 | 0.97 | 0.94 | 470 |
| 4 | 0.78 | 57.19 | 0.91 | 0.32 | 204802490.64 | 705.05 | 0.92 | 0.85 | 416 |
| 5 | 0.57 | 4.71 | 0.37 | 0.35 | 121959768.42 | 509.40 | 0.94 | 0.88 | 474 |
| 6 | 0.37 | 4.07 | 1.07 | 0.21 | 110885614.03 | 488.85 | 0.92 | 0.84 | 468 |
| 7 | 0.76 | 2.78 | 2.80 | 1.33 | 56897243.14 | 344.29 | 0.98 | 0.96 | 484 |
| 8 | 0.80 | 3.24 | 1.52 | 2.15 | 79423012.19 | 405.09 | 0.97 | 0.95 | 488 |
| 9 | 0.87 | 3.54 | 0.61 | 1.03 | 64430238.71 | 363.73 | 0.98 | 0.96 | 491 |
| 10 | 1.05 | 0.35 | 0.26 | 0.17 | 63279239.61 | 365.76 | 0.97 | 0.95 | 477 |
| 11 | 0.84 | 10.51 | 4.41 | 0.38 | 60864759.15 | 352.44 | 0.98 | 0.97 | 494 |
| 12 | 0.95 | 3.27 | 3.96 | 0.65 | 65714920.23 | 370.01 | 0.98 | 0.97 | 484 |
| 13 | 0.39 | 0.99 | 0.59 | 1.76 | 79813874.39 | 433.36 | 0.88 | 0.77 | 429 |
| 14 | 1.08 | 2.61 | 1.36 | 0.14 | 74875290.69 | 397.87 | 0.98 | 0.96 | 477 |
| 15 | 0.93 | 2.03 | 0.42 | 1.21 | 67222756.34 | 371.15 | 0.98 | 0.95 | 492 |
| $\mathbf{M_{ME}}$ | **0.80** | **2.95** | **0.44** | **0.80** | **24361049.82** | **222.74** | **0.99** | **0.98** | **495** |

Pređimo sada na trenutak sa modela prosečnih monetarnih ekvivalenata na diskusiju fitova monetarnih ekvivalenata individualnih rezultata. Generalno, fitovi su visokog kvaliteta, što se jasno vidi u kolonama $R^2$ i $RMSE$ u tabelama 6a, 6b, 7a i 7b. U slučaju procena pojedinačnih ispitanika, svakako, očekujemo veće eksperimentalne greške nego u slučaju prosečnih monetarnih ekvivalenata. Pet ispitanika u eksperimentu 2a ($EUR$) odlikuje konveksna funkcija korisnosti (sklonost ka riziku) prema teoriji izgleda; prema ocenama teorije poverenja, čak sedam od trinest ispitanika odlikuje takva funkcija korisnosti u eksperimentu 2a. Generalno, teorija poverenja predviđa više vrednosti eksponenata stepene funkcije korisnosti, $\rho$, nego teorija izgleda. Slična je situacija i u slučaju eksperimenta 2b ($RSD$ vrednosti, up. tabele 7a i 7b): ocene teorije izgleda otkrivaju samo jednog ispitanika sa konveksnom funkcijom korisnosti, dok ocene teorije poverenja otkrivaju tri takva ispitanika u eksperimentu u kome su funkcije korisnosti mahom konkavne.

Oba eksperimenta odlikuje *sklonost ka dobicima*: ocenjeni parametri $\lambda$ za prosečne monetarne ekvivalente su manji od 1 u oba eksperimenta, prema ocenama oba modela. To znači da su naši ispitanici mahom pokazivali sklonost ka dobicima, a ne averziju prema gubicima. Ocene teorije izgleda nalaze samo jednog ispitanika koji pokazuje averziju prema gubicima u eksperimentu 2a ($EUR$), dok ocene teorije poverenja nalaze tri takva ispitanika. U eksperimentu sa dinarskim vrednostima 2b, teorija izgleda otkriva četiri ispitanika koji pokazuju averziju prema gubicima, a teorija poverenja pet takvih ispitanika. U oba skupa eksperimentalnih podataka, samo dva ispitanika ukupno odlikuje vrednost averzije prema gubicima oko 2, vrednost koja je neopravdano stekla „kultni status“ od kako su je Kaneman i Tverski prepoznali u procenama monetarnih ekvivalenata na kojima su izvedene prve ocene kumulativne teorije izgleda[85] (Tversky & Kahneman, 1992).

Na prosečnim monetarnim ekvivalentima, teorija poverenja ima nižu vrednost srednje kvadratne greške ($RMSE$) od teorije izgleda. Međutim, potrebno je da uđemo dublje u strukturu dobijenih podataka da bismo otkrili informacije koje će omogućiti jasniju selekciju adekvatnog modela odlučivanja u uslovima rizika. Te strukture podataka ćemo otkriti upravo na terenu subjektivnog tretmana verovatnoća u donošenju odluka, ključnom fenomenu čijim je objašnjenjem (kumulativna) teorija izgleda stekla status najznačajnije deskriptivne teorije odlučivanja.

*Subjektivni tretman verovatnoća u odlučivanju u uslovima rizika.* Analizu koju sada predstavljamo prvi su koristili Tverski i Kaneman u radu o kumulativnoj teoriji

izgleda iz 1992. godine (Tversky & Kahneman, 1992). Tverski i Keler su, u analizi odlučivanja u uslovima neizvesnosti, modifikovali ovu analizu u njen zapravo jedini prihvatljiv oblik, pošto, videćemo, način na koji je ona korišćena u radu Tverskog i Kanemana iz 1992. godine nije teorijski plauzibilan (Tversky & Koehler, 1994; Tverski i Keler ne daju nikakvo objašnjenje za način na koji je analiza o kojoj je reč prvi put korišćena 1992. godine u Tversky & Kahneman, 1992). Posmatrajmo ne-negativan loz oblika (*x*,*p*;*0*,*1-p*), loz koji sa verovatnoćom $p$ daje ishod $x$ i ništa u suprotnom. Prema teoriji izgleda važi sledeće:

$$w(p_x) \cdot x^{\rho} = [MEq(x,p)]^{\rho} \tag{85}$$

gde sa *MEq*(*x*,*p*) označavamo monetarni ekvivalent ne-negativnog loza ($x$,$p_{\mathrm{x}}$); u zapisu potiskujemo nebitnu vrednost nula. Prethodno tvrđenje bi trebalo da bude potpuno očigledno. Sa leve strane jednačine (85) nalazimo očekivanu korisnost loza koji sadrži jedan dobitak i nulu, i koja ne zavisi od korisnosti nule (koja je u teoriji izgleda uvek nula), već je u potpunosti određena korisnošću jednog dobitka $x$ na lozu koju množi odgovarajući ponder odluke verovatnoće $p_{\mathrm{x}}$ asocirane sa tim ishodom. Sa desne strane izraza nalazimo korisnost monetarnog ekvivalenta tog loza. Iz ovoga direktno sledi

$$w(p_x) = \frac{[MEq(x,p)]^{\rho}}{x^{\rho}} \tag{86}$$

odn. da ako znamo monetarni ekvivalent loza, a njega smo u eksperimentima 2a i 2b dobili kroz procene ispitanika, i eksponent odgovarajuće stepene funkcije korisnosti $\rho$, možemo da ocenimo ponder odluke, $w(p_x)$, koji je korišćen da bi se izračunala očekivana korisnost loza. Tverski i Kaneman su u radu iz 1992. godine (up. slike 1 i 2, Tversky & Kahneman, 1992) izveli sličnu analizu, ali ne oslanjajući se na eksponent stepene funkcije korisnosti, već ocenjujući ponder odluke kroz odnos $\frac{MEq(x,p)}{x}$. Ovaj pristup sasvim sigurno ne vodi oceni pondera odluke $w(p_x)$. Tverski i Keler su 1994. godine ocenu pondera odluke izračunali ispravno, upotrebom jednačine (86) (mada nisu imali ocenu eksponenta stepene funkcije za podatke koje analiziraju, pa su se oslonili na vrednost dobijenu u prethodnom istraživanju Tverskog i Kanemana iz 1992, up. Tversky & Koehler, 1994). Za lozove koji sadrže dva dobitka ili dva gubitka, odn. striktno pozitivne ili striktno negativne lozove oblika ($x$,$p_x$;$y$,$p_y$), moguće je oceniti ponder odluke za jedan od dva ishoda na lozu

na sledeći način:

$$w(p_x) = \frac{[MEq(x, p_x; y, p_y)]^\rho - y^\rho}{x^\rho - y^\rho} \tag{87}$$

Ovakva analiza nije moguća za mešovite lozove. Međutim, veliki broj striktno pozitivnih i negativnih, te ne-negativnih i ne-pozitivnih lozova u našim eksperimentima 2a i 2b omogućа nam da ocenimo veliki broj pondera odluka, tako da možemo direktno empirijski da testiramo formu koju uzima subjektivni tretman verovatnoća. Dok se kumulativna teorija izgleda oslanja na inverzne-S funkcije poput Prelecove jednoparametarske forme koju mi koristimo, teorija poverenja predviđa linearne funkcije koje opisuju odnos između verovatnoća datih na lozu i *a posteriori* verovatnoća na osnovu kojih se evaluira vrednost loza. Teorija poverenja dalje omogućava da te linearne funkcije uzimaju varijablne nagibe i intercepte kao posledice interakcija funkcije korisnosti, dekumulativne funkcije *a priori* verovatnoća i vrednosti samih ishoda koje se nalaze na lozovima. U teoriji poverenja, analiza ponderisanja verovatnoća se za ne-negativne i ne-pozitivne lozove oblika (*x,p;0,1-p*) izvodi slično kao i za teoriju izgleda, po jednačini

$$p''_x = \frac{[MEq(x, p)]^\rho}{x^\rho} \tag{88}$$

dok se analiza za striktno pozitivne i striktno negativne lozove oblika ($x$,$p_x$;$y$,$p_y$) izvodi prema

$$p''_x = \frac{[MEq(x, p_x; y, p_y)]^\rho - y^\rho}{x^\rho - y^\rho} \tag{89}$$

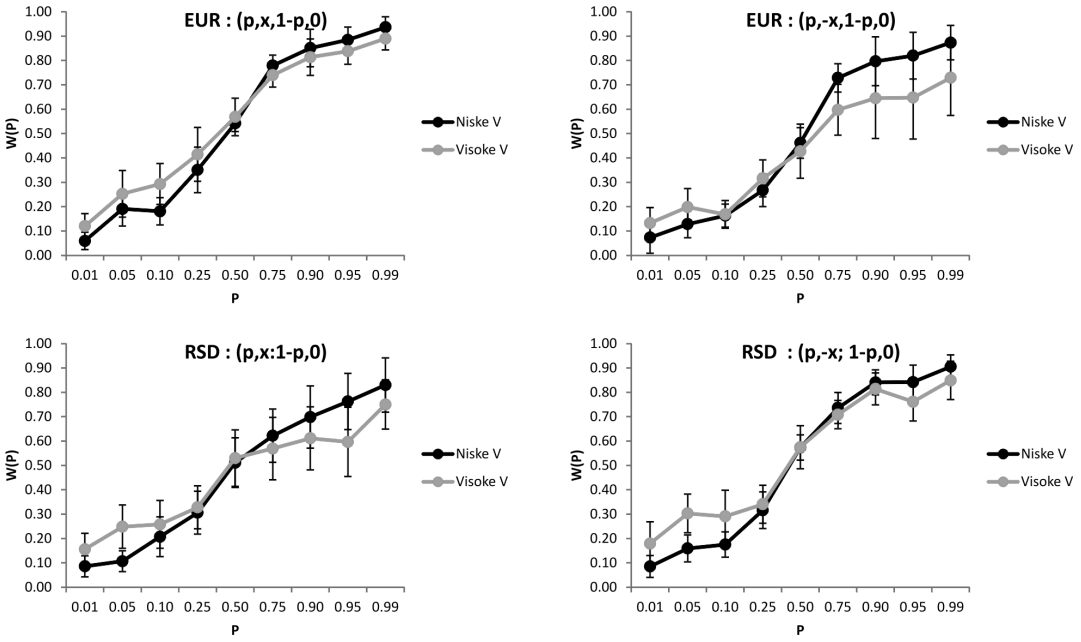
Veoma značajna pretpostavka teorije izgleda koja se neposredno odnosi na diskusiju subjektivnog tretmana verovatnoća jeste *uslov homogenosti preferencija*. Uslov je u suštini jednostavan i glasi ovako: homogenost preferencija je zadovoljena ako za sve lozove oblika $(X, P)$, ako je $c$ monetarni ekvivalent loza, onda, za sve pozitivne realne brojeve $k$, važi da je monetarni ekvivalent loza $(kX, P)$ jednak $kc$. Na primer, ukoliko je monetarni ekvivalent loza (50 $EUR$,$\frac{1}{4}$;0 $EUR$,$\frac{3}{4}$) jednak 10 evra, onda je monetarni ekvivalent loza (4×50 $EUR$,$\frac{1}{4}$;0 $EUR$,$\frac{3}{4}$) jednak 4×10 evra. Tverski je još u radu iz 1967. godine, u kome koristi skoro potpuno isti metod direktne numeričke procene monetarnih ekvivalenata koji je primenjen u eksperimentima 2a i 2b, pokazao da je *homogenost preferencija nužan i dovoljan*

*uslov za egzistenciju stepene funkcije korisnosti* (Tversky, 1967), što je činjenica koju rekapituliraju Tverski i Keler 1994 (Tversky & Koehler, 1994). Ovu vezu između uslova homogenosti preferencija i stepene funkcije korisnosti ponovo otkrivaju Bredli, al-Novaii i Dami u kratkom radu iz 2008. godine, dokazujući da iz homogenosti preferencija *nužno sledi* (*i*) da je funkcija korisnosti upravo stepena funkcija, (*ii*) da je ta funkcija ista za dobitke i gubitke, (*iii*) da su funkcije ponderisanja verovatnoća iste za dobitke i gubitke, i (*iv*) da averzija prema gubicima mora da ima vrednost veću od jedan (Bradley, al-Nowaihi & Dhami, 2008). Ovi autori pogrešno navode da su navedene karakteristike empirijski relevatne, pošto u vreme objavljivanja njihovog rada postoji već sijaset radova koji pokazuju da vrednosti parametara odgovarajućih funkcija nikako nisu konzistentne sa klasičnom skicom ograničeno racionalnog donosioca odluka kakvu je nudila kumulativna teorija izgleda još 1992. godine, ali zato uopšte ne citiraju rad Tverskog iz 1967. koji na mnogo jednostavniji način od onoga koji oni nude dokazuje da je homogenost preferencija nužan i dovoljan uslov za stepenu funkciju korisnosti (up. Bradley, al-Nowaihi & Dhami, 2008)[86]! Na primer, Koberlingova i Vaker su još 2005. demonstrirali paradoks koji sledi ako se stepene funkcije korisnosti za dobitke i gubitke razlikuju (diskutovali smo ga u sekciji 7.1, up. Köbberling & Wakker, 2005); u preglednom radu Foksa i Poldreka (up. tabelu 11.3, Fox, & Poldrack, 2009) uočljivo je da su procene funkcija korisnosti za dobitke i gubitke uglavnom slične, ali ne bez odstupanja, dok se funkcije ponderisanja verovatnoća za dobitke i gubitke razlikuju kao po pravilu. Zbog problema vezanog za stepenu funkciju korisnosti koji su primetili Koberlingova i Vaker, mi smo odlučili da koristimo jednu istu funkciju korisnosti za dobitke u gubitke u modelima koje ovde analiziramo.

Međutim, možda najvažnija posledica homogenosti preferencija za kumulativnu teoriju izgleda jeste sledeća. Ako je homogenost preferencija zadovoljena, onda očigledno sledi da će ponderi odluka za verovatnoće na nekom lozu $(p_1,x_1;p_2,x_2;...;p_n,x_n)$ *ostati isti* ako se svi ishodi na tom lozu pomnože nekom konstantom $k$, pa dobijemo loz oblika $(p_1,kx_1;p_2,kx_2;...;p_n,kx_n)$. Ovo je direktna posledica homogenosti preferencija i formi jednačina (86-87) za ocenu vrednosti pondera odluke. Ali, da li je uslov homogenosti preferencija zadovoljen u eksperimentalnim podacima?

Na osnovu jednačina (86-87) izračunali smo pondere odluka iz prosečnih monetarnih ekvivalenata za sve lozove (osim mešovitih, za koje ova analiza nije moguća) u eksperimentima 2a i 2b. Istu analizu smo ponovili za sve lozove na

osnovu jednačina (88-89), izračunavajući vrednosti *a posteriori* verovatnoća koje su morale biti korišćene u evaluaciji lozova prema teoriji poverenja. Na osnovu ocene parametara modela teorije izgleda na prosečnim monetarnim ekvivalentima, znamo kako izgledaju njene funkcije ponderisanja verovatnoća za dobitke i gubitke. Tako, možemo direktno da testiramo koliko funkcije ponderisanja verovatnoća fituju empirijske vrednosti pondera odluka dobijenih na osnovu jednačina (86-87). Slično, na osnovu parametara ocenjenih na prosečnim monetarnim ekvivalentima za teoriju poverenja, možemo da testiramo u kojoj meri linearne funkcije ponderisanja verovatnoće kakve ona predviđa dobro objašnjavaju empirijske vrednosti *a posteriori* verovatnoća dobijenih iz jednačina (88-89). Međutim, pre ovih analiza, sledili smo analizu Tverskog i Kanemana iz 1992, koja u originalnom radu nije podržana statističkim zaključivanjem. Za eksperimente 2a i 2b, izračunali smo empirjske vrednosti pondera odluka prema kumulativnoj teoriji izgleda za sve ne-negativne i ne-pozitivne lozove i za sve ispitanike koji su učestvovali u ovim eksperimentima. Prosečne vrednosti ovako izračunatih empirijskih pondera odluka date su na slici 25.




Slika 25. *Analiza homogenosti preferencija prema kumulativnoj teoriji izgleda za lozove u eksperimentima 2a i 2b.* Objašnjenje u tekstu.

Gornja dva panela slike 25. se odnose na eksperiment 2a (*EUR*), donja dva na eksperiment 2b (*RSD*); levi paneli na slici se odnose na lozove koji donose jedan dobitak ili ništa (ne-negativni), desni paneli na lozove koji donose jedan gubitak

ili ništa (ne-pozitivni). Na abscisi se nalaze verovatnoće sa kojima je odgovarajući dobitak ili gubitak ponuđen na lozu, a na oordinati empirijske vrednosti pondera odluka $w(p)$ izračunate na osnovu jednačine (86). Pošto smo u eksperimentima 2a i 2b koristili po pet različitih vrednosti ishoda za dobitke i gubitke (ne računajući ishod 0), posmatranja smo podelili na lozove sa visokim ishodima, uključujući sve lozove sa najviše dve vrednosti (100 *EUR* i 150 *EUR* za eksperiment 2a, odn. 1000 *RSD* i 5000 *RSD* za eksperiment 2b), i lozove sa niskim ishodima, uključujući sve lozove sa najniže dve vrednosti (25 *EUR* i 50 *EUR* za eksperiment 2a, odn. 100 *RSD* i 200 *RSD* za eksperiment 2b). Da bismo obezbedili podjednak broj posmatranja u ove dve situacije, koje su na slici 25. predstavljene crnom (niske vrednosti) i sivom (visoke vrednosti) linijom, iz analize za oba eksperimenta smo izbacili „srednji“ nivo vrednosti: to su lozovi koji sadrže ishod 75 *EUR* u eksperimentu 2a, i lozovi koji sadrže ishod 500 *RSD* u eksperimentu 2b. Bez obzira na ovo izbacivanje „umerenih“ ishoda iz analize, ako je uslov homogenosti preferencija zadovoljen, sve tačke na crnim i sivim linijama na odgovarajućim (istim) nivoima verovatnoće na slici 25. *moraju da se poklope*, jer iz homogenosti preferencija sledi da ponderi odluka moraju da budu isti za lozove sa niskim i visokim ishodima (trivijalno: visoki ishodi su izvesno jednaki proizvodu neke konstante proprocionalnosti $k$ sa niskim ishodima). To očigledno nije slučaj.

Izveli smo četiri poptuno ponovljene analize varijanse na podacima koje prikazuje slika 25. Prvi faktor u svakoj analizi je verovatnoća sa kojom je dobitak ili gubitak dat na lozu, sa devet nivoa, dok je drugi faktor visina ishoda, i ima dva nivoa (visoki i niski ishodi). Važi za sve rezultate analiza varijanse koje sada navodimo: ako je Mošlijev test sfericiteta značajan, navedena vrednost $F$-testa koju saopštavamo je korigovana Grinhaus-Gajserovom korekcijom.

(I) Za analizu lozova koji sadrže dobitke u eksperimentu 2a (gornji levi panel slike 25), osnovni efekat visine ishoda nije značajan, osnovni efekat verovatnoće jeste: $F(2.18,\ 21.82) = 157.97$, $p < .01$, a značajne je i dvofaktorska interakcija: $F(3.37,\ 33.74) = 3.87$, $p < .05$.

(II) Za analizu lozova koji sadrže gubitke u eksperimentu 2a (gornji desni panel slike 25), osnovni efekat visine ishoda je na granici statističke značajnosti ($p = .084$), osnovni efekat verovatnoće je značajan: $F(1.72,\ 18.97) = 51.05$, $p < .01$, i dvofaktorska interakcija je značajna: $F(4.18,\ 45.98) = 5.51$, $p < .01$.

(III) Za analizu lozova koji sadrže dobitke u eksperimentu 2b (donji levi panel slike 25), osnovni efekat visine ishoda nije statistički značajan, osnovni efekat verovatnoće jeste: $F(2.62, 31.41) = 179.46$, $p < .01$, kao i dvofaktorska interakcija: $F(8, 96) = 8.41$, $p < .01$.

(IV) Za analizu lozova koji sadrže gubitke u eksperimentu 2b (donji desni panel slike 25), osnovni efekat visine ishoda je na granici statističke značajnosti ($p = .07$), osnovni efekat verovatnoće je značajan: $F(2.59, 36.32) = 5.59$, $p < .01$, i dvofaktorska interakcija je značajna: $F(8, 112) = 3.95$, $p < .01$.

Dok je u ovim analizama varijanse osnovni efekat verovatnoće potpuno očekivan, uslov homogenosti preferencija je prekršen bilo da su značajni (a) osnovni efekat faktora visine ishoda ili (b) dvofaktorska interakcija visine ishoda i verovatnoće. Videli smo da efekti faktora visine ishoda tek ponekad dostižu marginalnu statističku značajnost, ali su zato sve dvofaktorske interakcije statistički značajne. Interakcije ne samo da su statistički značajne, već na osnovu grafikona prikazanih na slici 25. vidimo da su one i sistematske prirode: na lozovima sa niskim verovatnoćama dobitka i gubitaka, ponderi odluka teže da budu manji za lozove sa niskim ishodima nego za lozove sa visokim ishodima, i obrnuto za lozove sa visokim verovatnoćama dobitaka ili gubitaka. Rezultati ovih analiza već predstavljaju jak empirijski dokaz za fundamentalnu tvrdnju teorije poverenja: *fenomen ponderisanja verovatnoća nije nezavistan od visine ishoda koji se ponderišu.* Ovaj nalaz je u potpunoj kontradikciji sa standardnom formom modela kumulativne teorije izgleda. Ako homogenost preferencija nije zadovoljena, iz rezultata do kojih su došli Tverski, te Bredli, al-Novaii i Dami, sledi da *stepena funkcija izvesno ne odgovara opisu empirijske funkcije korisnosti.* Mi ćemo nastaviti da koristimo stepenu funkciju, ali od ovog trenutka svesni činjenice da *ona može da predstavlja samo aproksimaciju empirijske funkcije korisnosti.* Dalje, iz rezultata do Bredlija, al-Novaiia i Damia sledi da je nužno da se funkcije korisnosti poklapaju za dobitke i gubitke, kao i funkcije ponderisanja verovatnoća; takođe, averzija prema gubicima je nužno veća od jedan, što su sve činjenice koje se očigledno ne poklapaju sa eksperimentalnim rezultatima koje upravo diskutujemo. Konačno, pošto homogenost preferencija nije zadovoljena, *funkcija ponderisanja verovatnoća kumulativne teorije izgleda nije jedinstvena*, što se nalazi u dubokoj kontradikciji sa aksiomatskom strukturom ove teorije iz koje sledi dokaz o jedinstvenosti ove funkcije (up. Wakker & Tversky, 1993). Vidimo kako su četiri jednostavne analize varijanse već temeljno zatresle strukturu najbolje prihvaćene deskriptivne teorije odlučivanja.

U narednim analizama za striktno pozitivne i striktno negativne lozove oblika (*x,p;0,1-p*) važi konvencija da se ponder odluke (u teoriji izgleda) odn. *a posteriori* verovatnoća (u teoriji poverenja) uvek *izračunava za veći od dva ishoda na lozu*. Ovo je, dakle, čisto stvar konvencije: ovakvu analizu bismo svejedno mogli da izvedemo birajući bilo koji od dva ishoda na striktno pozitivnim ili striktno negativnim lozovima kao njen predmet. Na prosečnim monetarnim ekvivalentima iz eksperimenata 2a i 2b smo izračunali (a) empirijske pondere odluka prema jednačinama (86-87) za kumulativnu teoriju izgleda, i (b) empirijske *a posteriori* verovatnoće prema kojima su lozovi morali da budu evaluirani pod teorijom poverenja, prema jednačinama (88-89). Konstatovali smo već da je na osnovu parametara modela ocenjenih na prosečnim monetarnim ekvivalentima moguće direktno testirati u kojoj meri njihove teorijske funkcije uspešno objašnjavaju ovako empirijski izračunate vrednosti pondera odluka i *a posteriori* verovatnoća.

Slike 26-29 predstavljaju rezultate analize subjektivnog tretmana verovatnoća pod dva teorijska modela koje ovde diskutujemo. Pošto su nalazi složeni, a analiza zahteva dosta grafikona da bi bila predstavljena u potpunosti, grafikoni na slikama 26-29 se odnose samo na eksperiment 2a gde su korišćeni iznosi u evrima. Slike su organizovane sistematski i naredni redovi treba da omoguće lakšu interpretaciju naših analiza.

Pet panela na slici 26a prikazuju funkciju ponderisanja verovatnoća za kumulativnu teoriju izgleda, za sve ne-negativne lozove oblika (*x,p;0,1-p*) u eksperimentu 2a. Podsetimo se da se ponderisanje verovatnoća uvek odnosi na veći od dva ishoda u lozu. Paneli, dakle, redom prikazuju odnos između vrednosti $w(p)$ koja je izračunata „empirijski“, dakle na osnovu jednačine (86). Isprekidana linija na dijagonali je linija identiteta: nalaziće se na svim ovakvim grafikonima. Kada uopšte ne bi bilo ponderisanja verovatnoća, sve tačke bi morale da leže na liniji identiteta. Na abscisama grafikona na slici 26a nalaze se „objektivne“ verovatnoće, tj. verovatnoće sa kojima je odgovarajući ishod naveden na lozovima oblika (*x,p;0,1-p*) u kojima je učestvovao. Pošto teorija izgleda predviđa jedinstvenu funkciju ponderisanja verovatnoća, funkcija na svih pet grafikona slici 26a je potpuno ista. Ona predstavlja Prelecovu jednoparametarsku formu ponderisanja verovatnoća sa odgovarajućom vrednošću parametra $\gamma_g$ za dobitke. Grafikoni na slici 26a nam omogućavaju da sagledamo odnos između empirijskih vrednosti $w(p)$, i vrednosti $w(p)$ koje predviđa model teorije izgleda.

Slika 26b prikazuje linearne regresione funkcije za vrednost $w(p)$ koju predviđa model teorije izgleda (funkcije ponderisanja verovatnoća na slici 26a) kao prediktor, i empirijske vrednosti $w(p)$ kao zavisnu varijablu. Interesantno: bez obzira na veoma jasna odstupanja koja se vide na pet grafikona na slici 26a, regresione linije su izuzetno dobre. Rezultati regresionih analiza koje se odnose na sliku 26b dati su u tabeli 8a.

Slika 26c prikazuje funkcije ponderisanja verovatnoća teorije izgleda za striktno pozitivne lozove oblika ($x$,$p$;$y$,*1-p*). Ponderisanje verovatnoća se posmatra uvek u odnosu na veći ishod na lozu. Naslovi iznad grafikona jasno ukazuju na to koji loz je u pitanju, a na abscisi pratimo porast u verovatnoći sa kojom je dat veći od dva ishoda kroz sve moguće lozove koje sadrže ta dva ishoda na lozu. Na oordinati se ponovo nalaze empirijske vrednosti $w(p)$, izračunate na osnovu jednačine (87). Slika 26d onda predstavlja odgovarajuće linearne regresione linije i daje odgovor na pitanje u kojoj meri predikcije modela teorije izgleda sa parametrima ocenjenim na prosečnim monetarnim ekvivalentima odgovaraju empirijskim vrednostima $w(p)$. Rezultati regresionih analiza koji se odnose na sliku 26d dati su u tabeli 8b.

Slika 27a i 27b prikazuju isti tip rezultata kao slike 26a i 26b, ali ovaj put za sve ne-pozitivne lozove oblika (*-x,p;0,1-p*) u eksperimentu 2a; funkcije ponderisanja verovatnoća su dobijene na osnovu parametra $\gamma_l$ za gubitke koji je ocenjen na prosečnim monetarnim ekvivalentima za odgovarajuće lozove. Rezultati regresionih analiza koji se odnose na sliku 27a dati su u tabeli 8c.

Slike 27c i 27d nose iste informacije kao slike 26c i 26d, ali za striktno negativne lozove oblika (*-x,p;-y,1-p*). Organizacija grafikona je ista kao u slučaju striktno pozitivnih lozova na slici 26c. Rezultati regresionih analiza koji se odnose na sliku 27d nalaze se u tabeli 8d, a za sliku 27b u tabeli 8c.

Slike 28 i 29 se odnose na analizu *a posteriori* verovatnoća pod teorijom poverenja. Slika 28a se odnosi na sve ne-negativne lozove oblika (*x,p;0,1-p*) u eksperimentu 2a. Isprekidana linija na dijagonali predstavlja liniju identiteta na kojoj bi se nalazile sve tačke da nema nikakve subjektivne korekcije verovatnoća datih na lozu. Na abscisi svih grafikona varira vrednost verovatnoće sa kojom je na lozu prikazana odgovarajuća vrednost $x$ - navedena iznad svakog od pet grafikona. Na oordinati se nalaze „empirijske“ vrednosti *a posteriori* verovatnoća izračunate na osnovu jednačine (88). Tanja crna linija predstavlja najbolju linearnu regresionu liniju koja opisuje odnos između „objektivnih“ verovatnoća na abscisama (verovatnoćama kakve su navedene na lozovima) i empirijskih *a posteriori*

verovatnoća. Podebljana crna linija predstavlja linearnu funkciju ponderisanja verovatnoća kakvu predviđa model teorije poverenja na osnovu vrednosti parametara ocenjenih na prosečnim monetarnim ekvivalentima. Da su predikcije teorije poverenja perfektne, ove dve linije bi se poklapale.

Slika 28b predstavlja linearne regresione linije između *a posteriori* verovatnoća kakve predviđa model teorije poverenja i empirijskih *a posteriori* verovatnoća za sve analize na slici 28a. Tabela 9a sadrži rezultate regresionih analiza koje se odnose na sliku 28b.

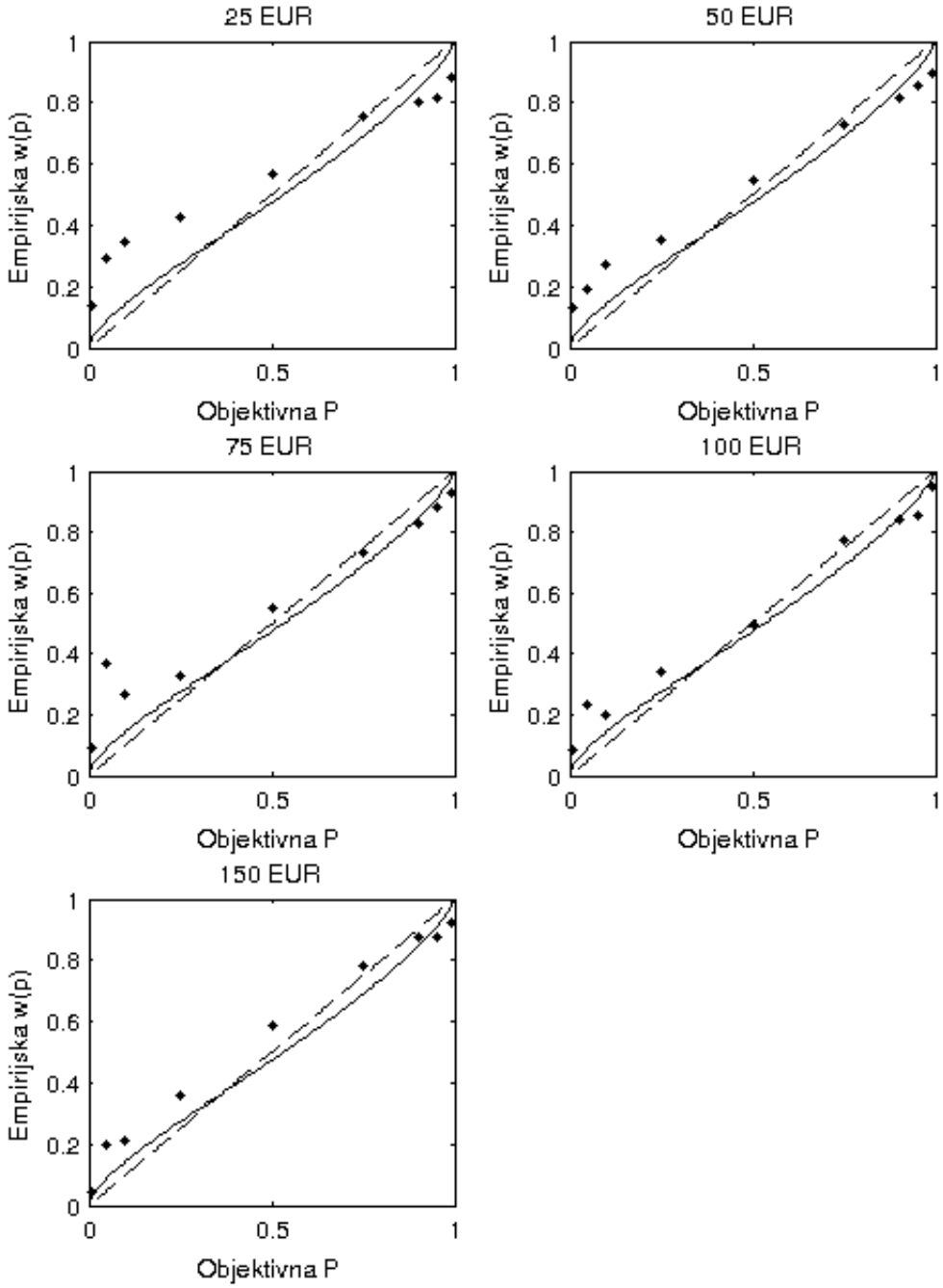
Slika 28c predstavlja istu ovu analizu pod teorijom poverenja ali za striktno pozitivne lozove; slika 28d predstavlja linearne regresione funkcije između *a posteriori* verovatnoća iz modela i empirijskih *a posteriori* verovatnoća izračunatih na osnovu jednačine (89), a rezultati regresionih analiza koji se odnose na sliku 28d dati su u tabeli 9b.

Konačno, slike 29a-29d sadrže iste informacije kao i slike 28a-28d, redom za ne-pozitivne i striktno negativne lozove u eksperimentu 2a; rezultati odgovarajućih regresionih analiza dati su u tabelama 9c i 9d.

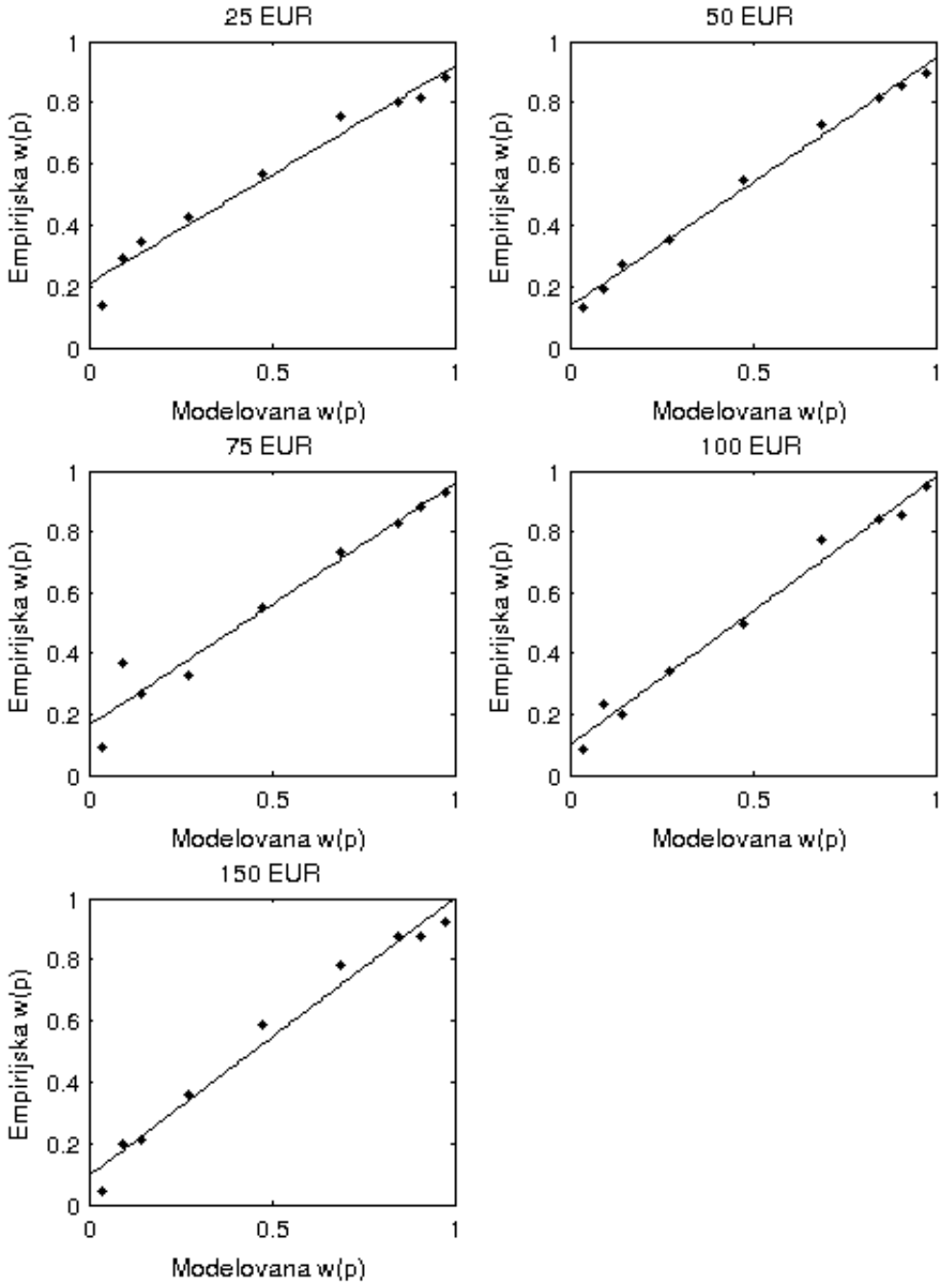
Rezultati naših analiza su donekle začuđujući. Prvo, vidimo da i funkcija ponderisanja verovatnoća teorije izgleda i linearna funkcija ponderisanja verovatnoća teorije izgleda omogućavaju predikciju empirijskih vrednosti $w(p)$ odn. $p''$. Rezultati linearnih regresionih analiza su odlični, u svim analizama, za obe funkcije: niti jedna jedina vrednost $R^2$ nije ispod .90. Ono što je čudno je činjenica da funkcija ponderisanja verovatnoća teorije izgleda, uprkos tome što podaci ne zadovoljavaju uslov homogenosti preferencija, iz čega sledi da ih *sigurno* ne opisuje jedinstvena funkcija ove forme, uspešno objašnjava rasporede $w(p)$ koji se i grubim pregledom grafikona očigledno razlikuju od loza do loza. Vratićemo se ovom pitanju uskoro.

Presudni empirijski nalaz u selekciji modela odlučivanja za merenje monetarnih ekvivalenata u eksperimentima 2a i 2b zahteva da strukturu podataka sagledamo iz još jedne perspektive. Na grafikonima koji prikazuju ponderisanje verovatnoća pod teorijom poverenja nalaze se, pored linearne funkcije kakvu predviđa teorija, i najbolje regresione linije koje opisuju odnos objektivnih verovatnoća (datih na lozovima) i empirijskih *a posteriori* verovatnoća izračunatih na osnovu jednačina (88-89). Prethodne diskusije Viskuzijevog modela pokazale su da njegovu linearnu formu ponderisanja verovatnoća kontrolišu parametar $N$ i vrednost *a priori* verovatnoće koja je formirana pre evaluacije odgovarajućeg loza (jednačine (56-57)). Posle uvođenja teorije poverenja, videli smo da ona predviđa mogućnost da je

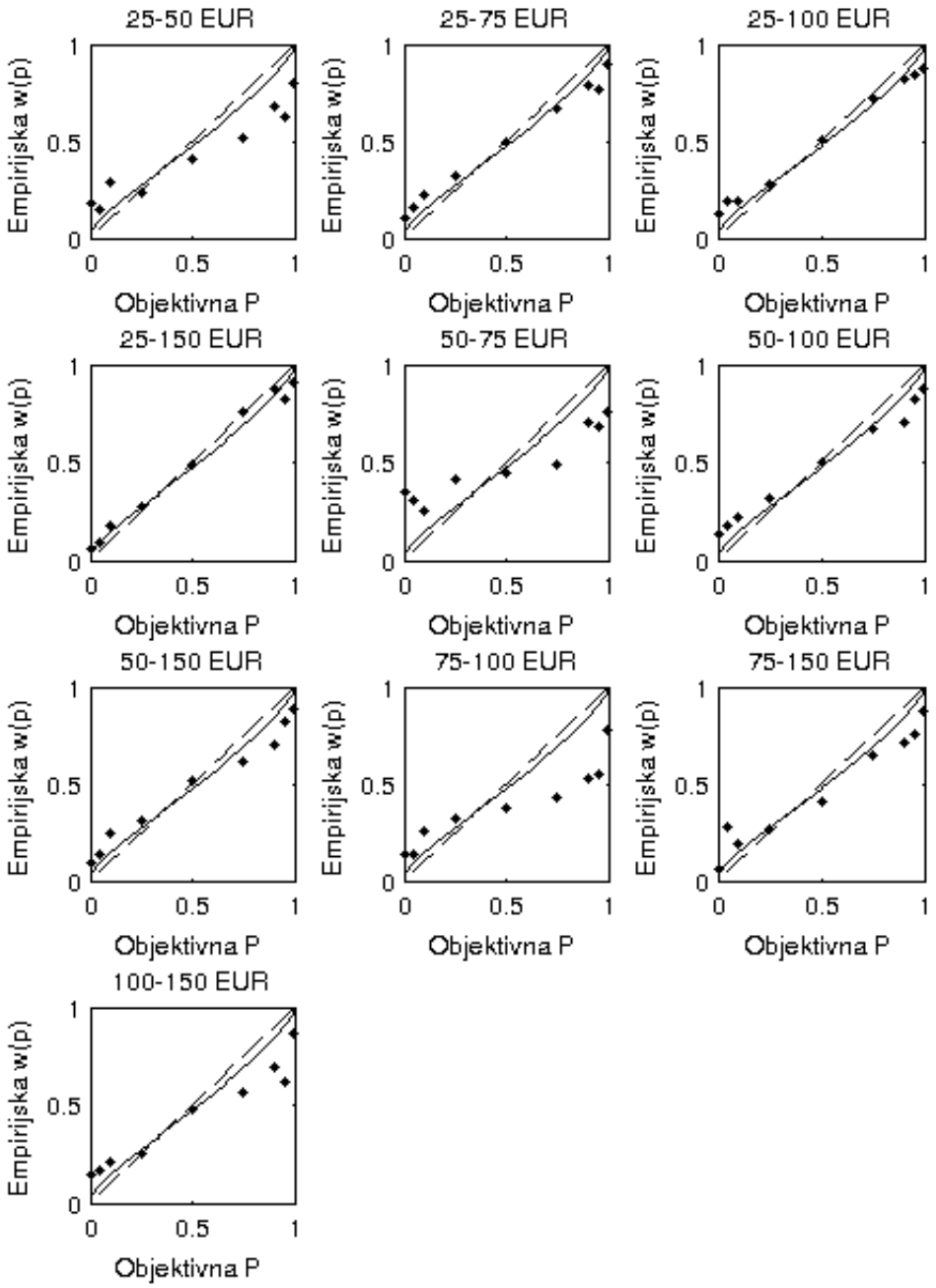
ponderisanje verovatnoća opisivo *familijom linearnih funkcija*, čiji nagibi i intercepti zavise od interakcije osobina dekumulativne funkcije $S$ (koja određuje *a priori* verovatnoće), funkcije korisnosti i konteksta koji čine svi ishodi na određenom lozu (i koji presudno utiče na izračunavanje vrednosti parametra $N$ kao relativne entropije a priori verovatnoća na lozu, jednačina (81)). Posle ocene parametara teorije poverenja na prosečnim monetarnim ekvivalentima u eksperimentu 2a, mi smo tako u stanju da na osnovu njihovih vrednosti *predvidimo intercepte i nagibe linearnih funkcija* koje odlikuju veze između objektivnih verovatnoća i *a posteriori* verovatnoća koje su korišćene u evaluaciji lozova. Tako, možemo da uporedimo nagibe i intercepte najboljih regresionih linija između objektivnih i a posteriori verovatnoća na slikama 28a, 28c, 29a i 29c sa predikcijama nagiba i intercepte koje nam daje teorijski model.



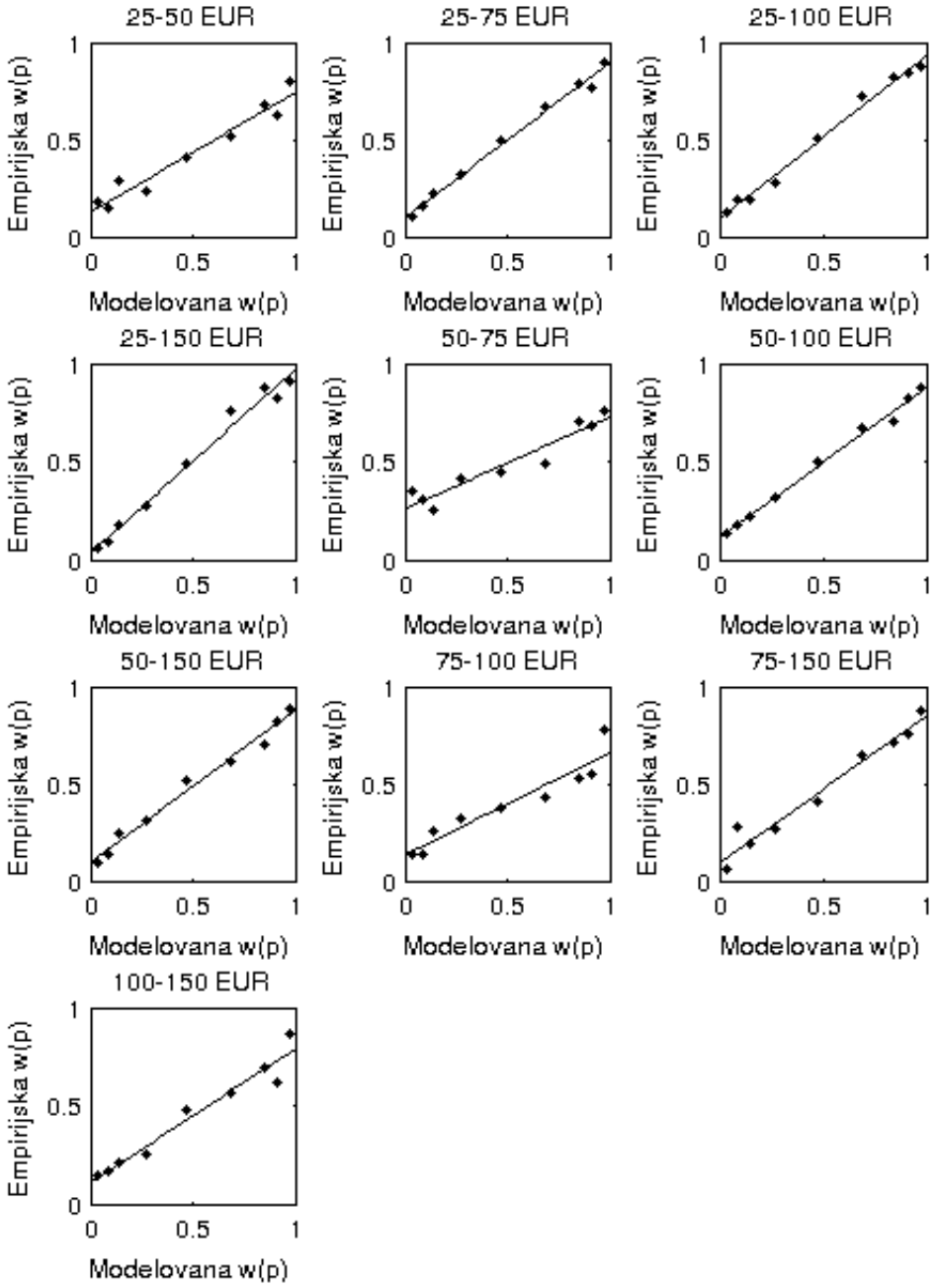

Slika 26a. *Funkcija ponderisanja verovatnoća za kumulativnu teoriju izgleda: nenegativni lozovi oblika (x,p;0,1-p) u eksperimentu 2a.*



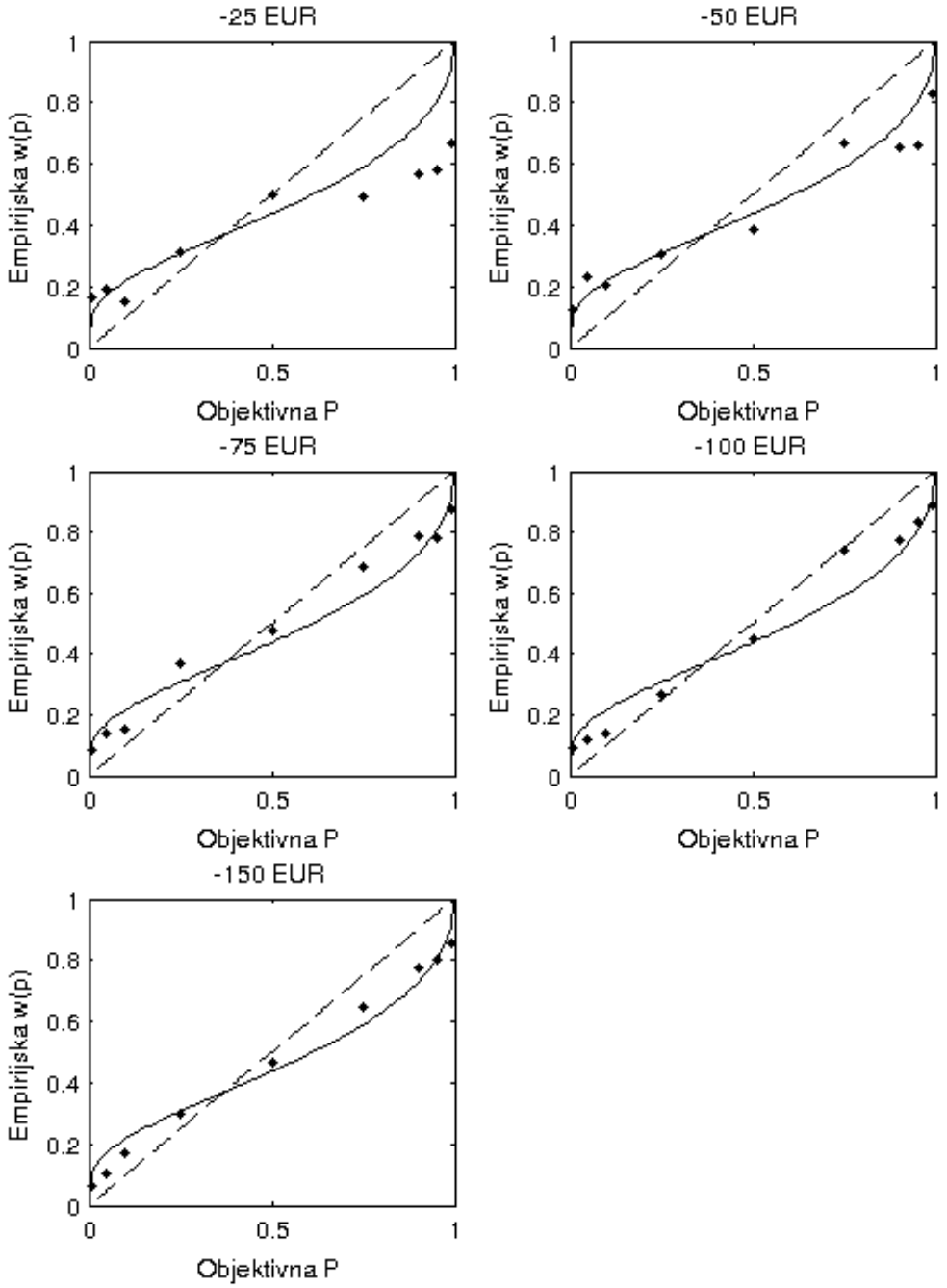

Slika 26b. *Linearne regresije funkcije ponderisanja verovatnoća za kumulativnu teoriju izgleda: ne-negativni lozovi oblika (x,p;0,1-p) u eksperimentu 2a.*



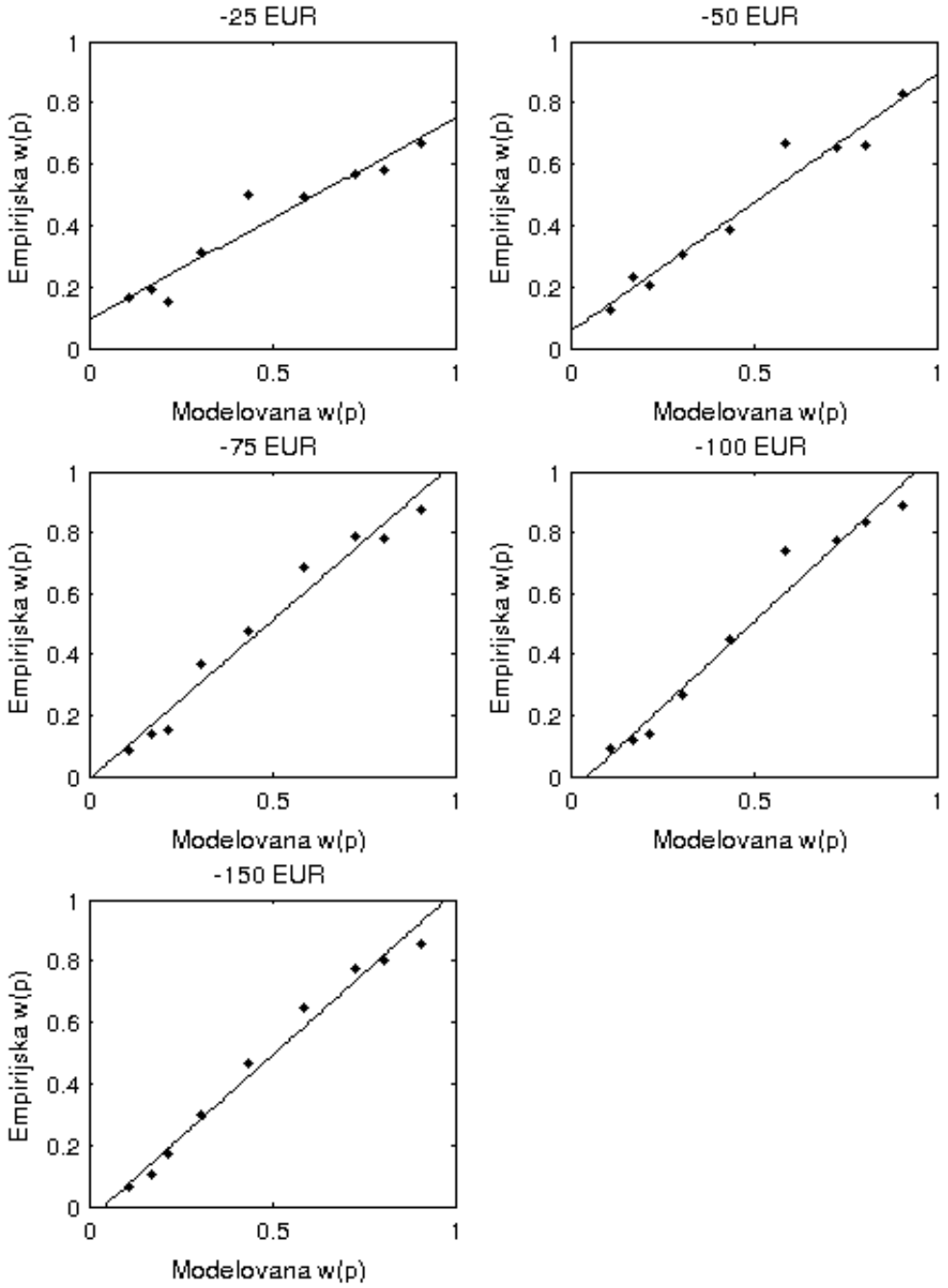

Slika 26c. *Funkcije ponderisanja verovatnoća za kumulativnu teoriju izgleda: striktno pozitivni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*



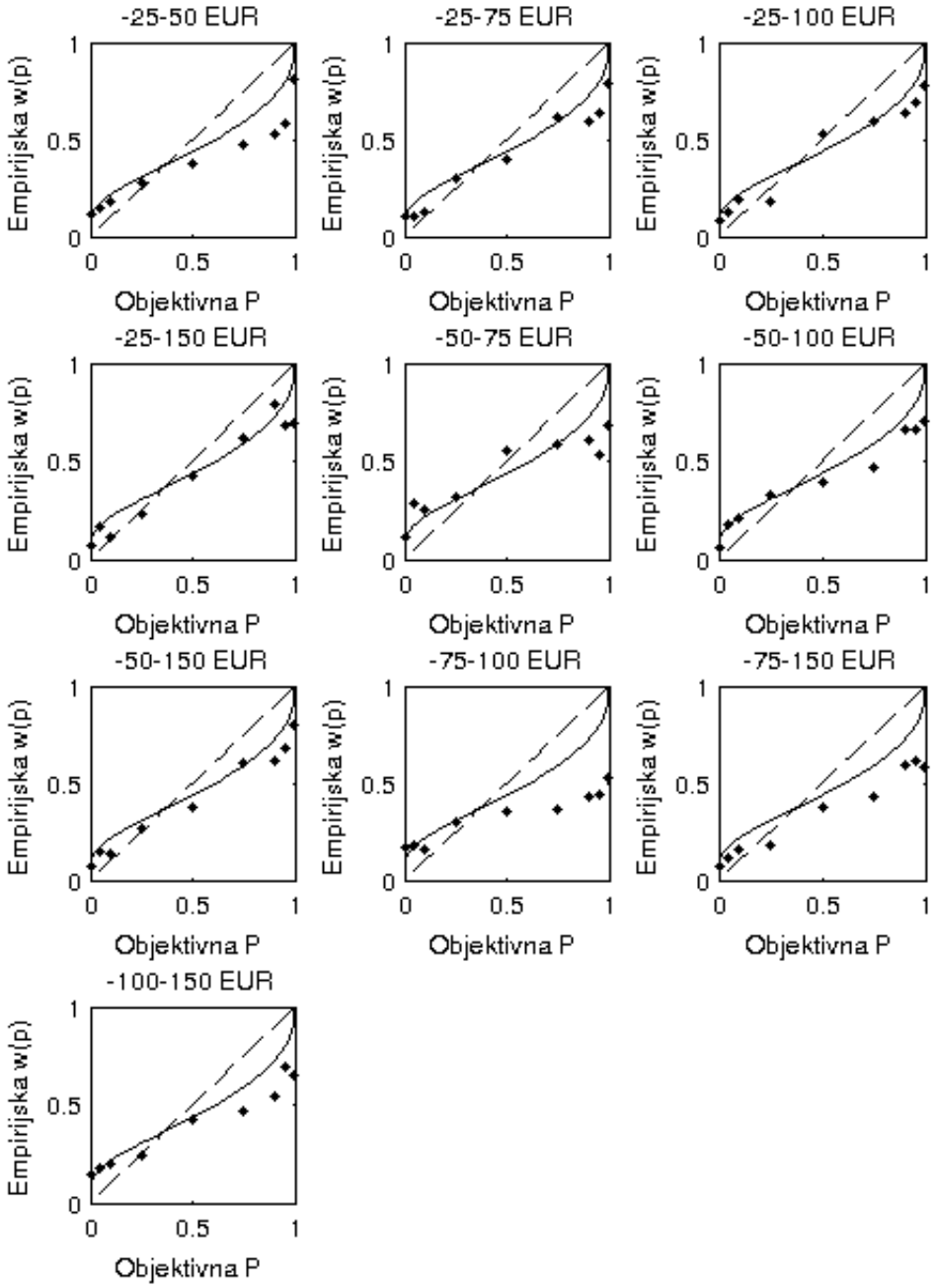

Slika 26d. *Linearne regresije funkcije ponderisanja verovatnoća za kumulativnu teoriju izgleda: striktno pozitivni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*



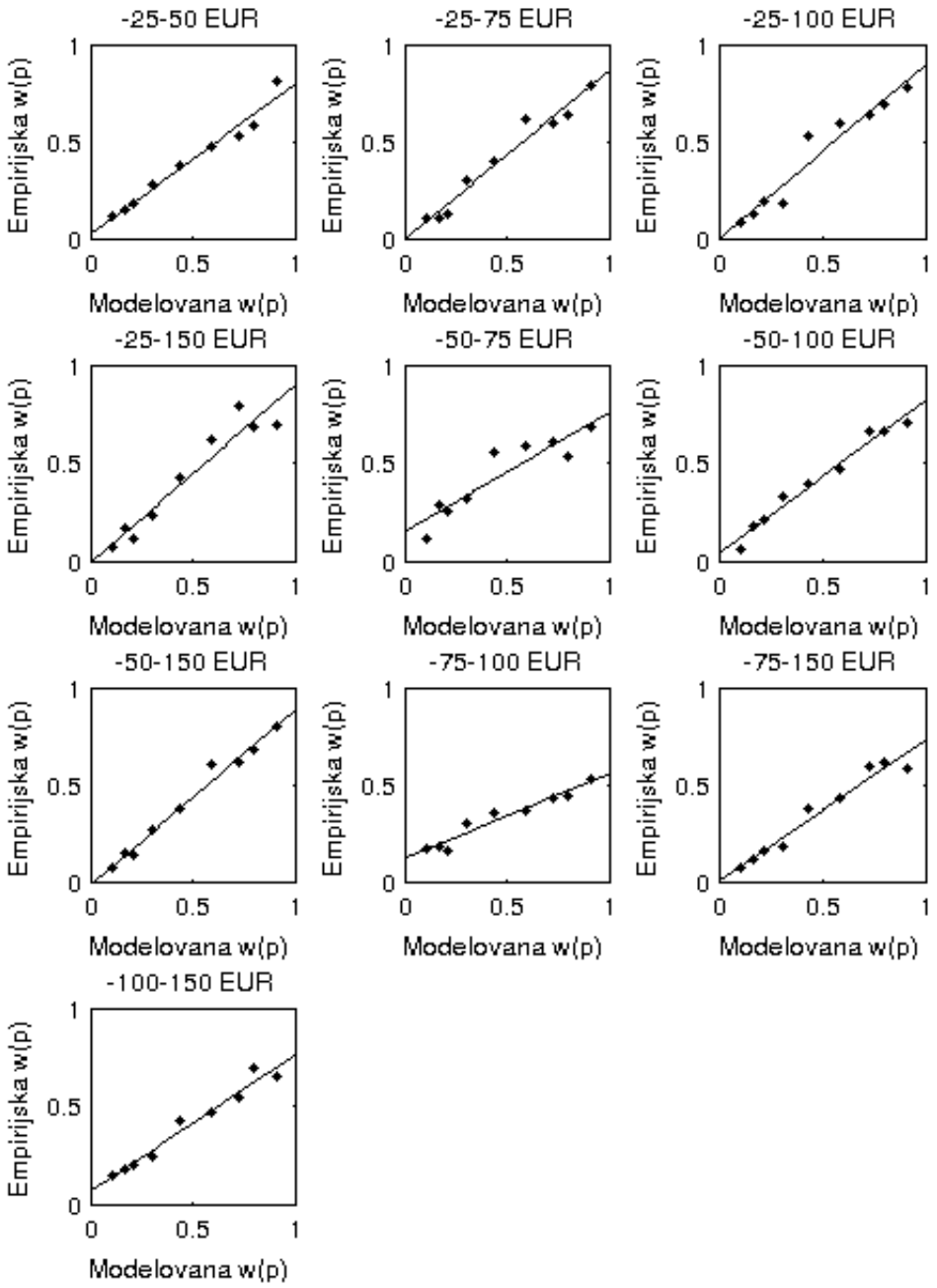

Slika 27a. *Funkcije ponderisanja verovatnoća za kumulativnu teoriju izgleda: nepozitivni lozovi oblika (-x,p;0,1-p) u eksperimentu 2a.*



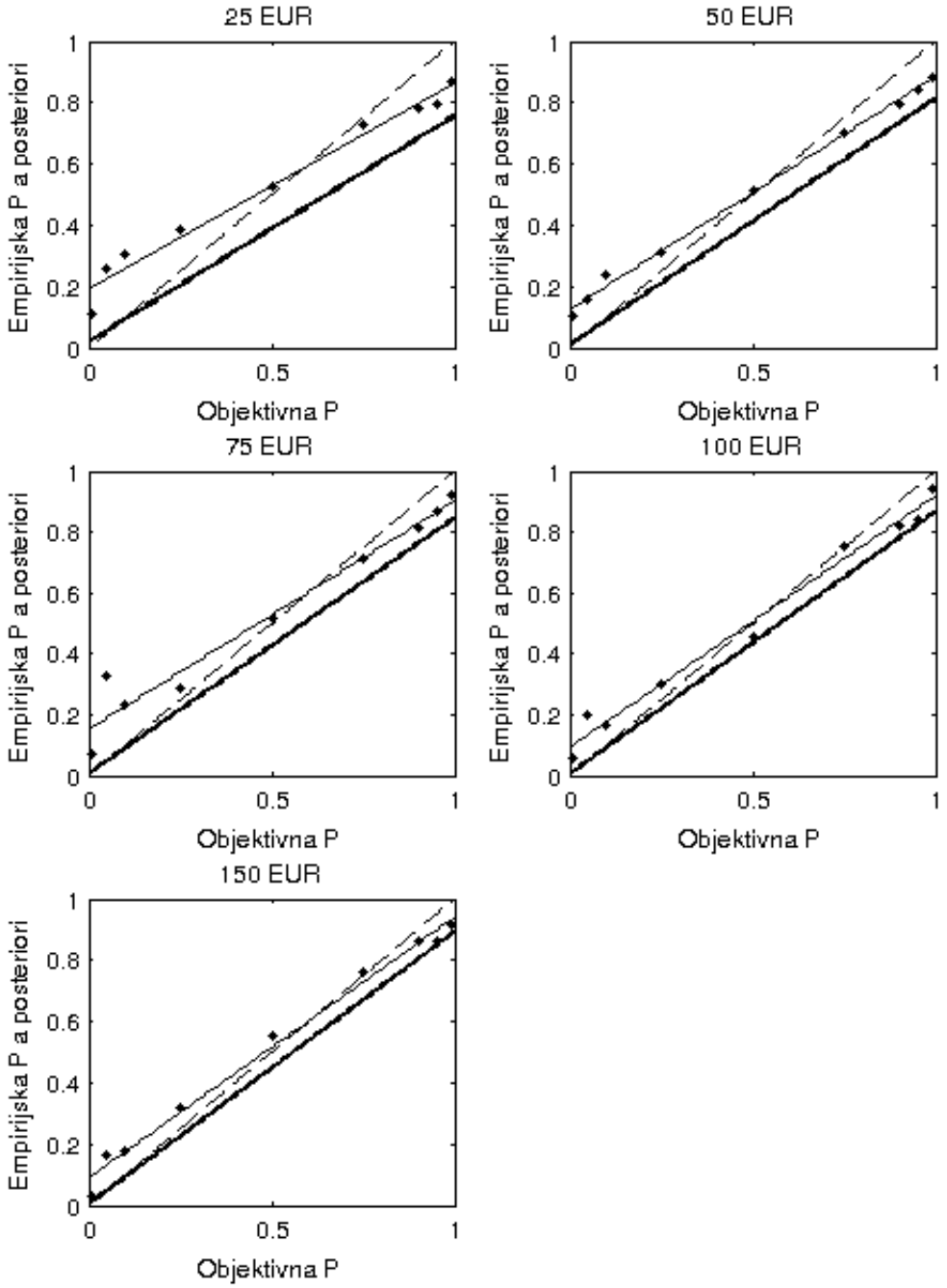

Slika 27b. *Linearne regresije funkcije ponderisanja verovatnoća za kumulativnu teoriju izgleda: ne-pozitivni lozovi oblika (-x,p;0,1-p) u eksperimentu 2a.*



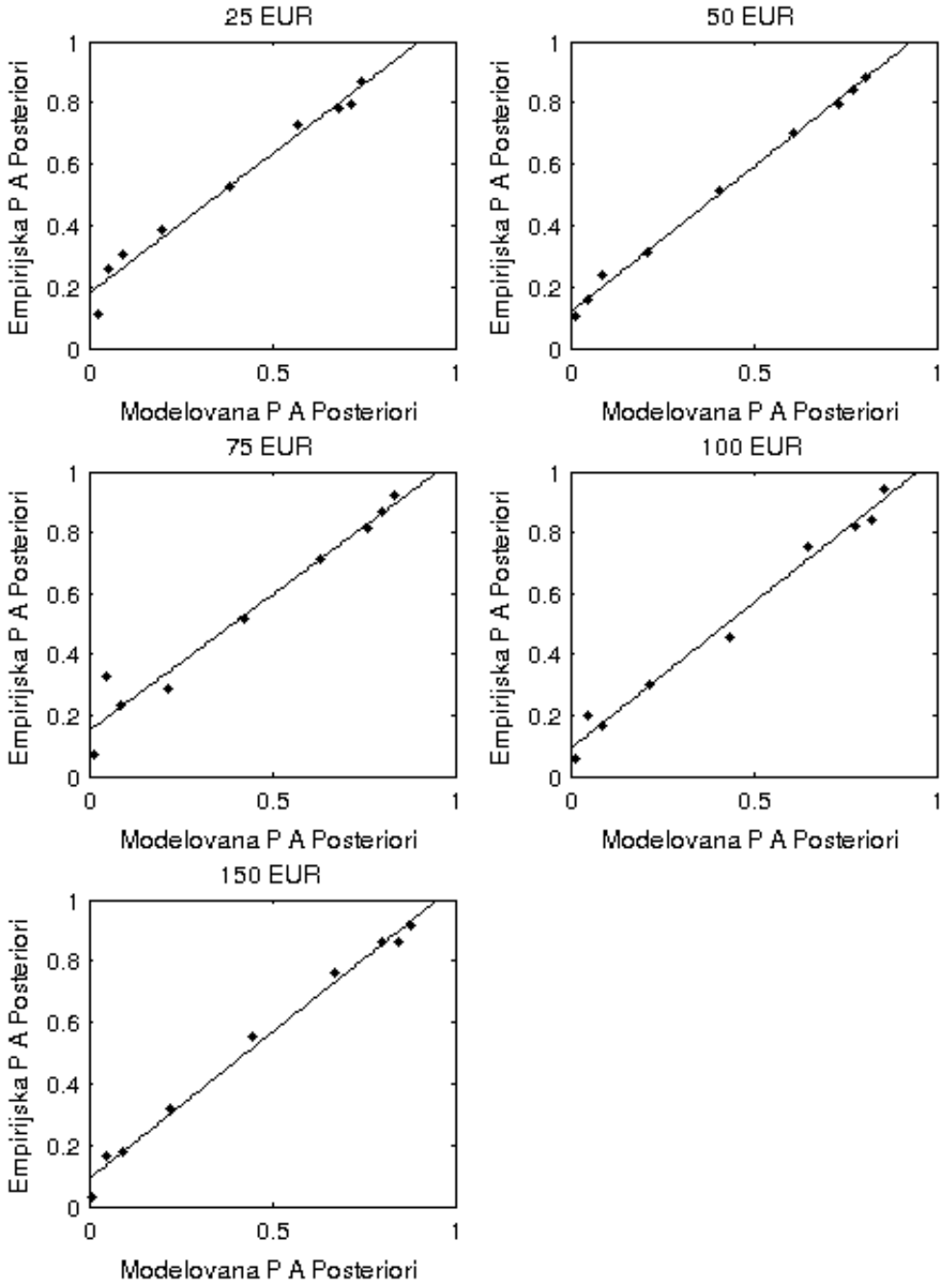

Slika 27c. *Funkcije ponderisanja verovatnoća za kumulativnu teoriju izgleda: striktno pozitivni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*



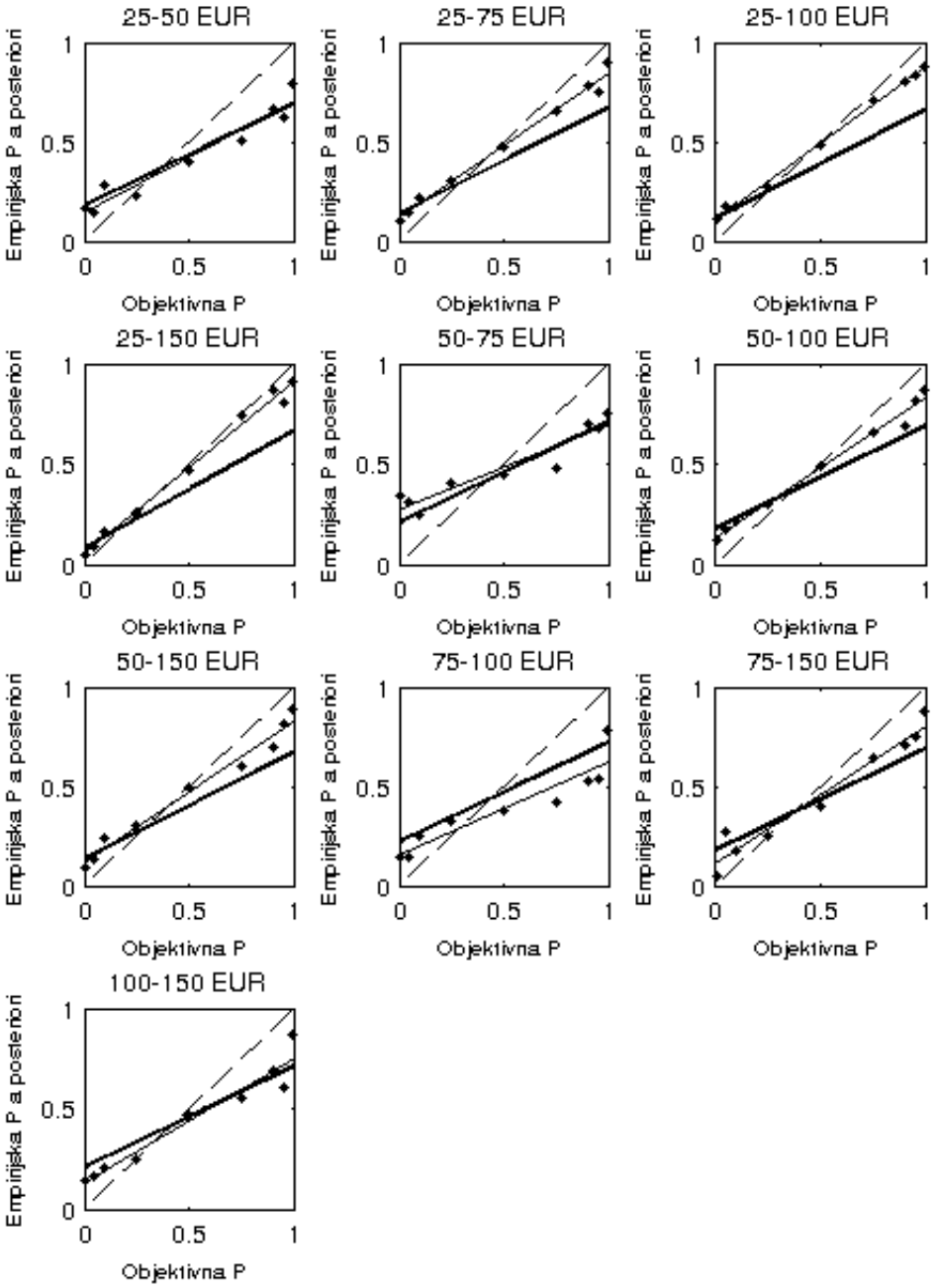

Slika 27d. *Linearne regresije funkcije ponderisanja verovatnoća za kumulativnu teoriju izgleda: striktno pozitivni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*



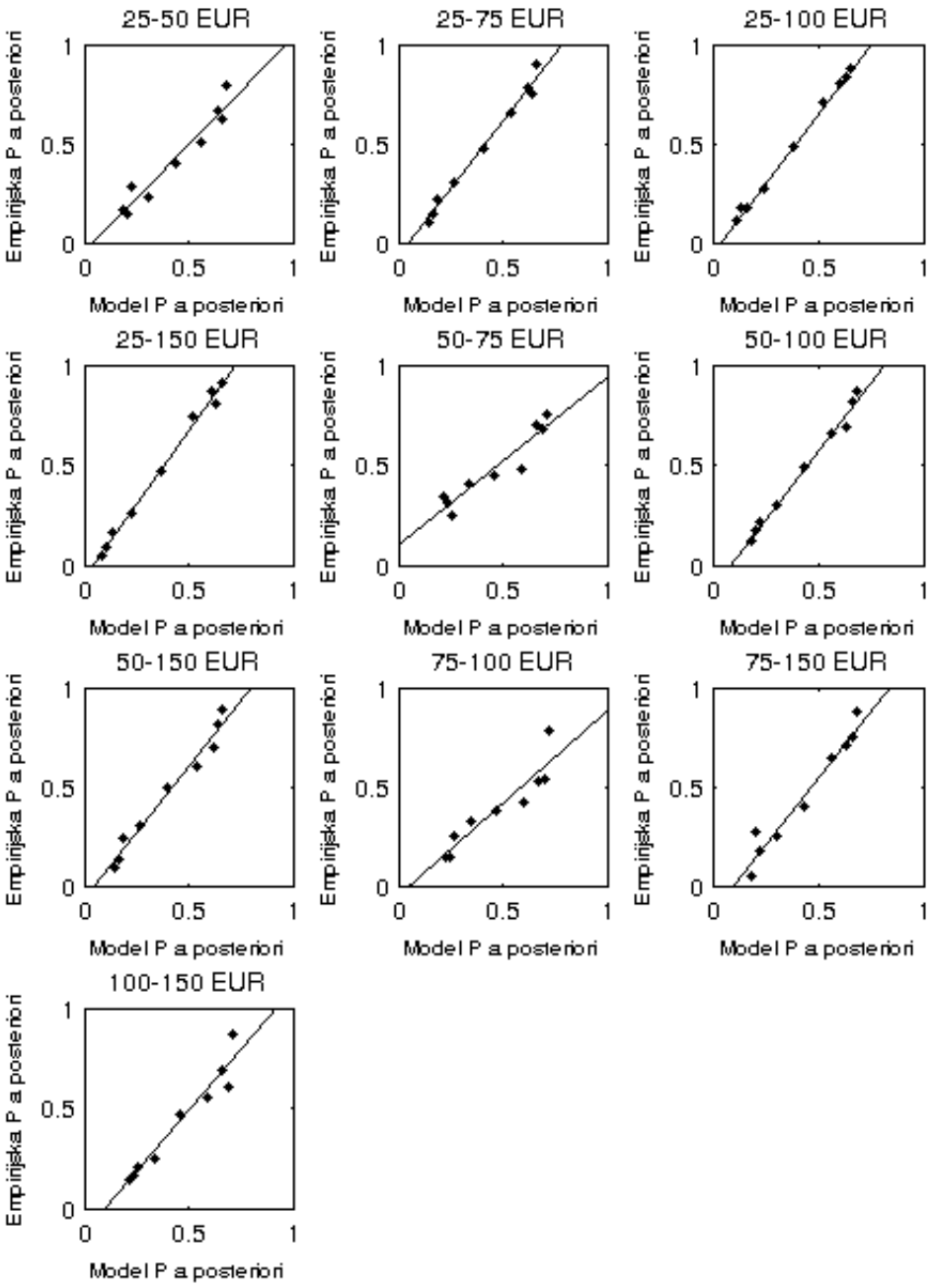

Slika 28a. *Funkcije ponderisanja verovatnoća za teoriju poverenja: ne-negativni lozovi oblika (x,p;0,1-p) u eksperimentu 2a.*



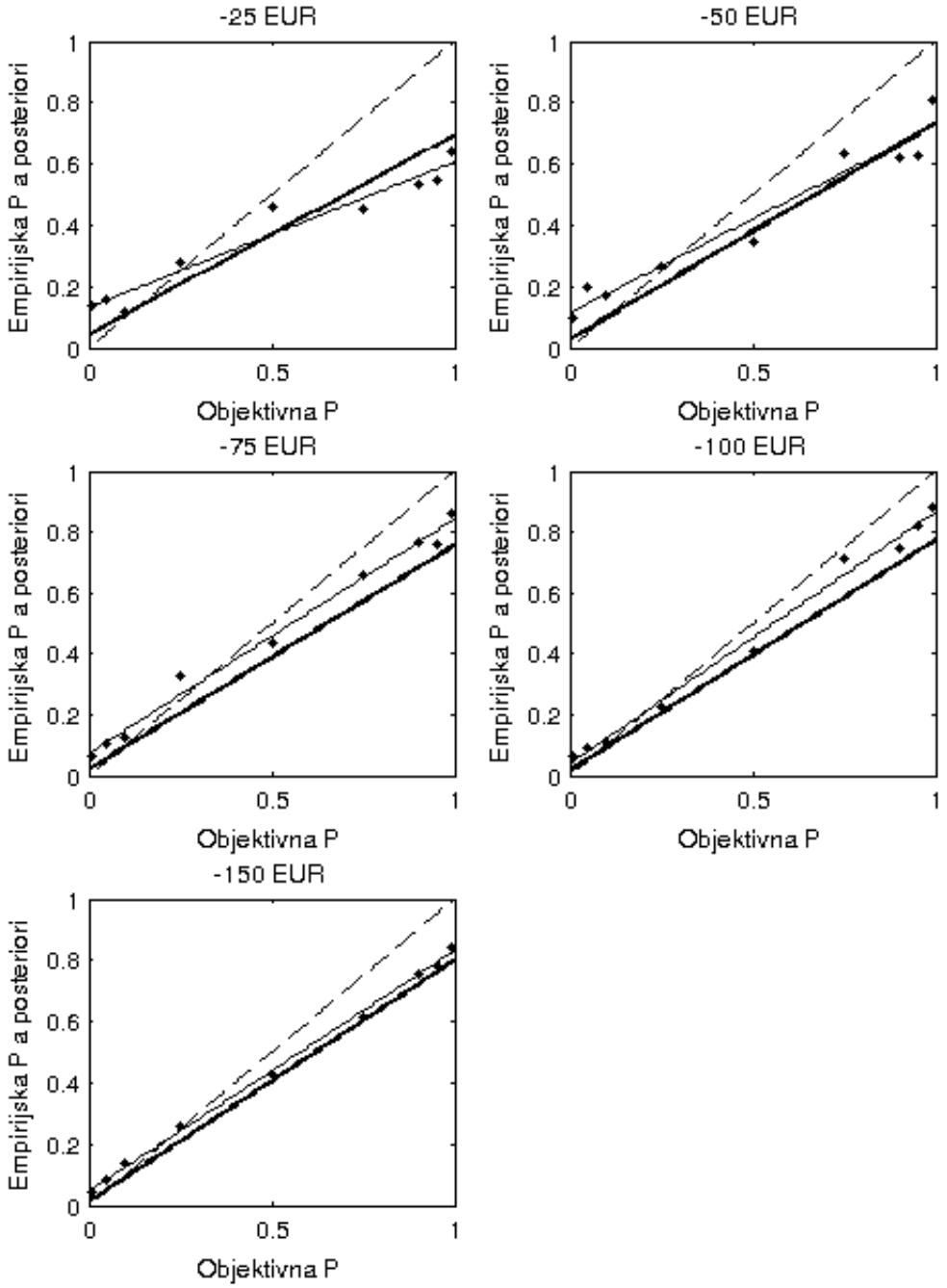

Slika 28b. *Linearne regresije funkcija ponderisanja verovatnoća za teoriju poverenja: ne-negativni lozovi oblika (x,p;0,1-p) u eksperimentu 2a.*



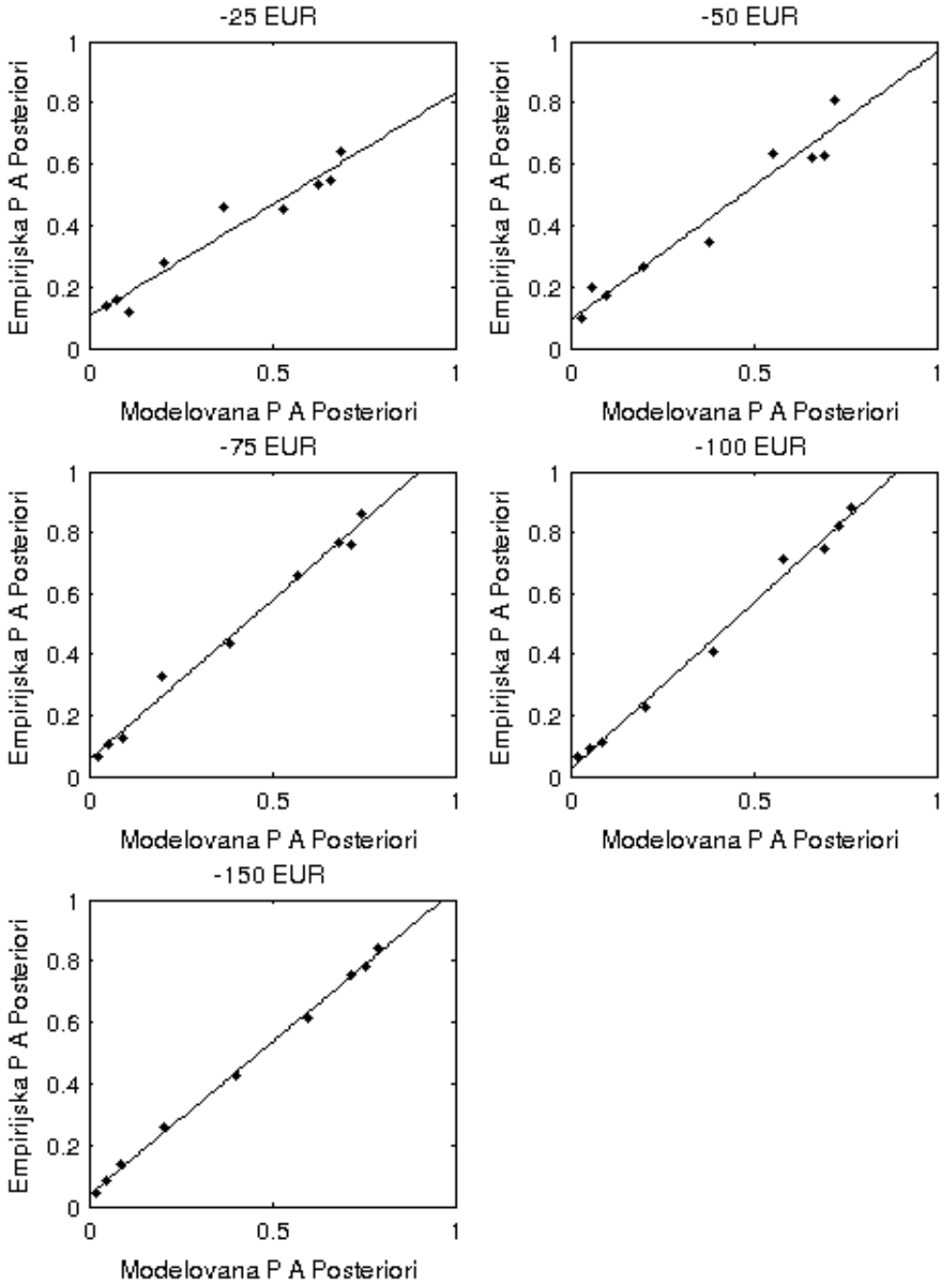

Slika 28c. *Funkcije ponderisanja verovatnoća za teoriju poverenja: striktno pozitivni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*



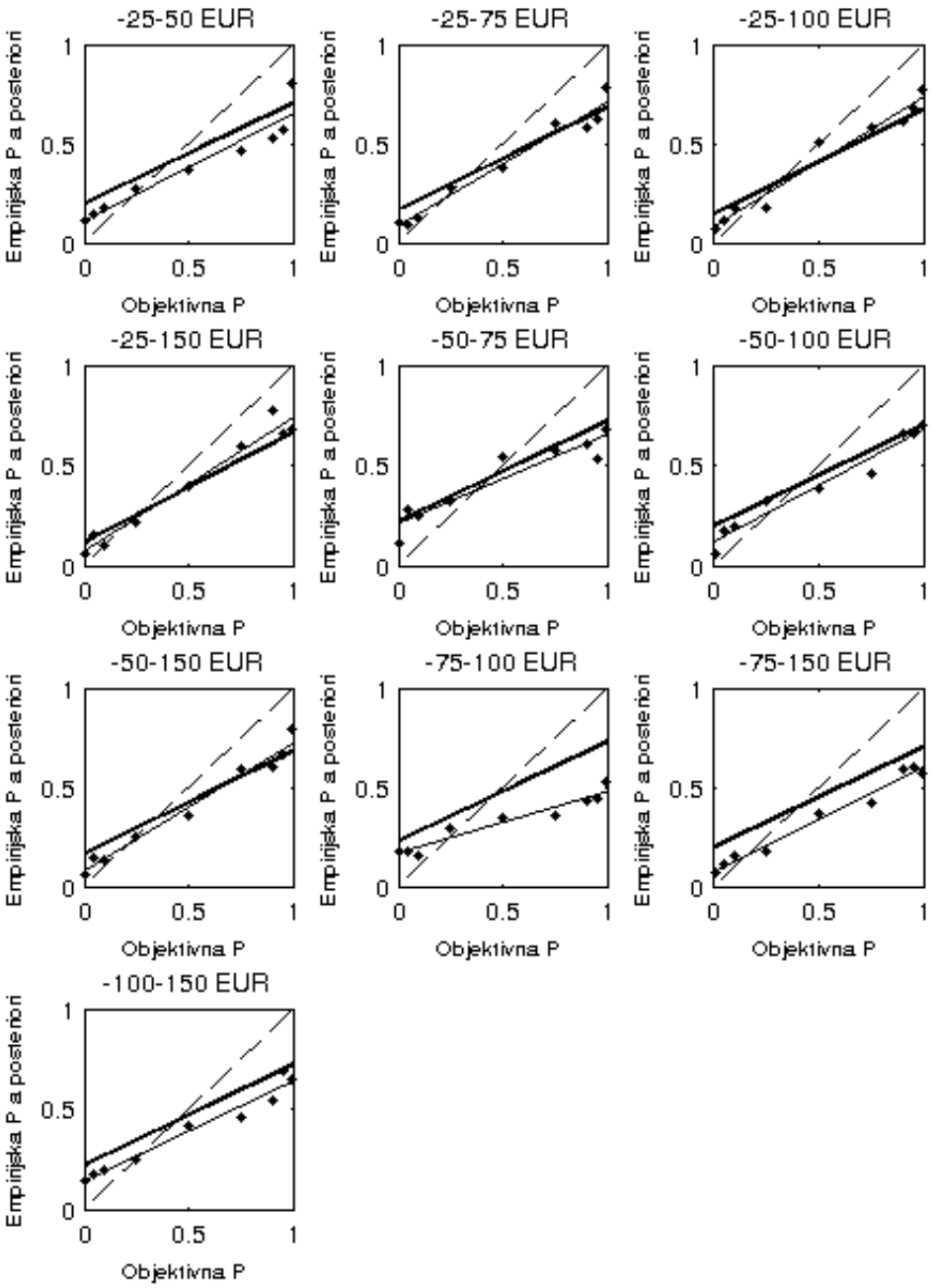

Slika 28d. *Linearne regresije funkcija ponderisanja verovatnoća za teoriju poverenja: striktno pozitivni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*



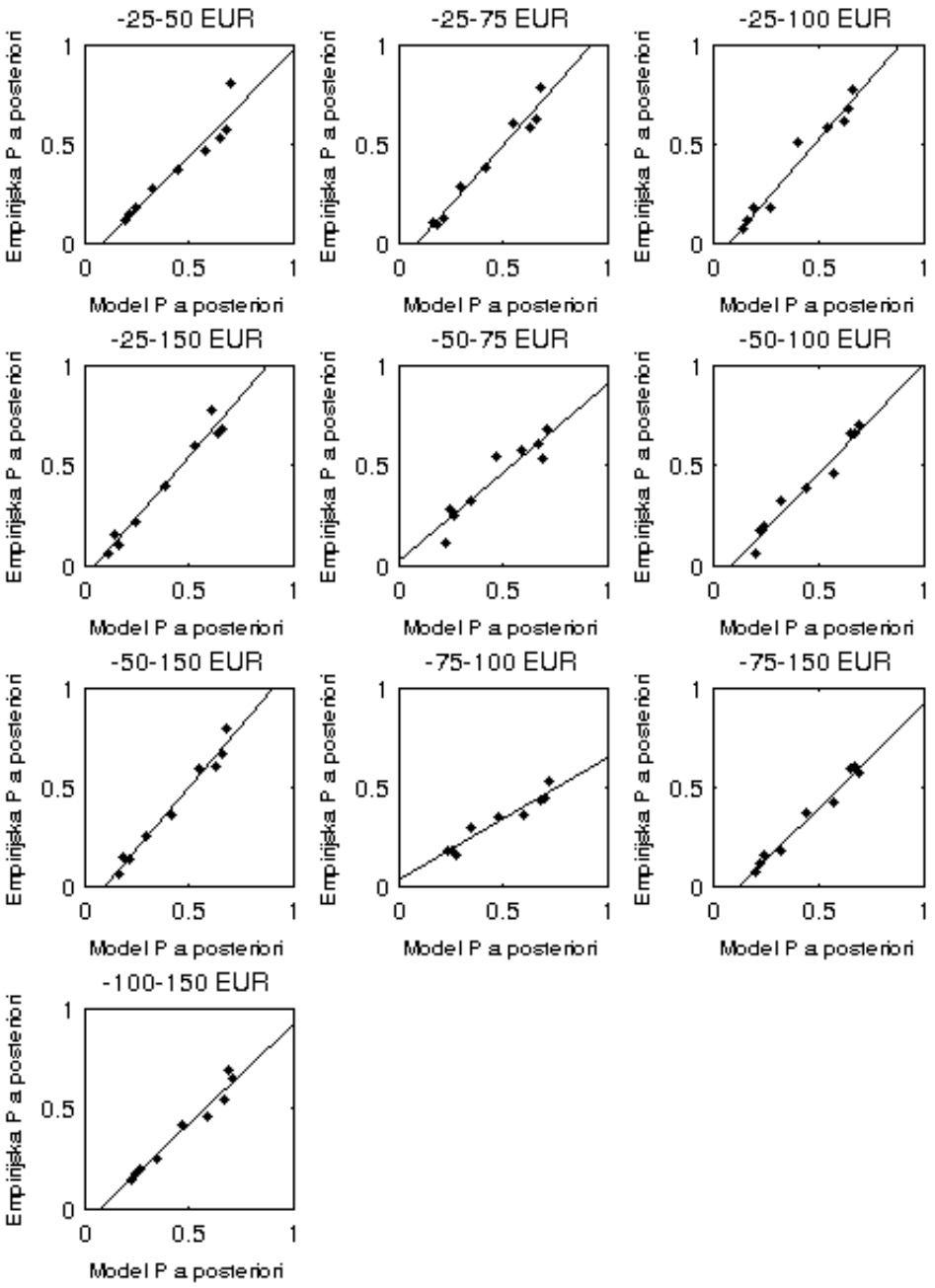

Slika 29a. *Funkcija ponderisanja verovatnoća za teoriju poverenja: ne-negativni lozovi oblika (x,p;0,1-p) u eksperimentu 2a.*




Slika 29b. *Linearne regresije funkcija ponderisanja verovatnoća za teoriju poverenja: ne-negativni lozovi oblika (x,p;0,1-p) u eksperimentu 2a.*




Slika 29c. *Funkcije ponderisanja verovatnoća za teoriju poverenja: striktno negativni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*




Slika 29d. *Linearne regresije funkcija ponderisanja verovatnoća za teoriju poverenja: striktno negativni lozovi oblika (x,p;y,1-p) u eksperimentu 2a.*

Tabela 8a. Teorija izgleda: regresione analize između predikcija modela za $w(p)$ i empirijske $w(p)$ (up. sliku 26b). Analize se odnose na sve ne-negativne lozove oblika ($x$,$p$;0,1-$p$) u eksperimentu 2a.

| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **25 EUR** | 0.71 | 0.21 | 0.97 | 228.38 | .01 | 0.13 |
| **50 EUR** | 0.81 | 0.13 | 0.99 | 980.72 | .01 | 0.08 |
| **75 EUR** | 0.80 | 0.16 | 0.96 | 159.89 | .01 | 0.11 |
| **100 EUR** | 0.88 | 0.10 | 0.99 | 464.58 | .01 | 0.07 |
| **150 EUR** | 0.91 | 0.09 | 0.98 | 300.23 | .01 | 0.07 |

Tabela 8b. Teorija izgleda: regresione analize između predikcija modela za $w(p)$ i empirijske $w(p)$ (up. sliku 26d). Analize se odnose na sve striktno pozitivne lozove oblika ($x$,$p$;$y$,1-$p$) u eksperimentu 2a.

| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **25-50 EUR** | 0.62 | 0.12 | 0.95 | 135.15 | .01 | 0.16 |
| **25-75 EUR** | 0.80 | 0.10 | 0.99 | 832.15 | .01 | 0.07 |
| **25-100 EUR** | 0.84 | 0.09 | 0.99 | 864.66 | .01 | 0.06 |
| **25-150 EUR** | 0.94 | 0.03 | 0.99 | 469.94 | .01 | 0.05 |
| **50-75 EUR** | 0.47 | 0.26 | 0.91 | 71.14 | .01 | 0.19 |
| **50-100 EUR** | 0.77 | 0.11 | 0.99 | 856.74 | .01 | 0.08 |
| **50-150 EUR** | 0.79 | 0.09 | 0.99 | 467.04 | .01 | 0.08 |
| **75-100 EUR** | 0.53 | 0.13 | 0.90 | 59.96 | .01 | 0.20 |
| **75-150 EUR** | 0.76 | 0.09 | 0.96 | 173.15 | .01 | 0.10 |
| **100-150 EUR** | 0.68 | 0.11 | 0.95 | 133.53 | .01 | 0.13 |

Tabela 8c. Teorija izgleda: regresione analize između predikcija modela za $w(p)$ i empirijske $w(p)$ (up. sliku 27b). Analize se odnose na sve ne-pozitivne lozove oblika (-$x$,$p$;0,1-$p$) u eksperimentu 2a.

| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **-25 EUR** | 0.65 | 0.09 | 0.92 | 82.95 | .01 | 0.13 |
| **-50 EUR** | 0.84 | 0.05 | 0.96 | 150.12 | .01 | 0.07 |
| **-75 EUR** | 1.04 | -0.01 | 0.97 | 212.14 | .01 | 0.05 |
| **-100 EUR** | 1.12 | -0.05 | 0.97 | 219.36 | .01 | 0.06 |
| **-150 EUR** | 1.07 | -0.04 | 0.98 | 397.61 | .01 | 0.05 |

Tabela 8d. Teorija izgleda: regresione analize između predikcija modela za $w(p)$ i empirijske $w(p)$ (up. sliku 27d). Analize se odnose na sve striktno negativne lozove oblika (-$x$,$p$;-$y$,1-$p$) u eksperimentu 2a.

| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **-25-50 EUR** | 0.78 | 0.02 | 0.96 | 192.69 | .01 | 0.12 |
| **-25-75 EUR** | 0.87 | -0.01 | 0.96 | 158.64 | .01 | 0.09 |
| **-25-100 EUR** | 0.90 | -0.01 | 0.94 | 115.61 | .01 | 0.09 |
| **-25-150 EUR** | 0.91 | -0.01 | 0.91 | 74.65 | .01 | 0.10 |
| **-50-75 EUR** | 0.61 | 0.15 | 0.85 | 38.64 | .01 | 0.14 |
| **-50-100 EUR** | 0.79 | 0.03 | 0.97 | 229.44 | .01 | 0.10 |
| **-50-150 EUR** | 0.91 | -0.02 | 0.98 | 351.05 | .01 | 0.08 |
| **-75-100 EUR** | 0.44 | 0.12 | 0.94 | 118.18 | .01 | 0.22 |
| **-75-150 EUR** | 0.73 | 0.00 | 0.95 | 147.38 | .01 | 0.16 |
| **-100-150 EUR** | 0.70 | 0.06 | 0.97 | 206.98 | .01 | 0.12 |

Tabela 9a. Teorija poverenja: regresione analize između predikcija modela za *p"* i empirijske *p"* (up. sliku 28b). Analize se odnose na sve ne-negativne lozove oblika ($x$,$p$;0,1-$p$) u eksperimentu 2a.

| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **25 EUR** | 0.91 | 0.18 | 0.98 | 282.51 | .01 | 0.55 |
| **50 EUR** | 0.95 | 0.12 | 1.00 | 2117.42 | .01 | 0.61 |
| **75 EUR** | 0.89 | 0.15 | 0.96 | 170.35 | .01 | 0.62 |
| **100 EUR** | 0.96 | 0.09 | 0.99 | 511.76 | .01 | 0.65 |
| **150 EUR** | 0.96 | 0.09 | 0.99 | 632.52 | .01 | 0.67 |

Tabela 9b. Teorija poverenja: regresione analize između predikcija modela za *p"* i empirijske *p"* (up. sliku 28d). Analize se odnose na sve striktno pozitivne lozove oblika ($x$,$p$;$y$,1-$p$) u eksperimentu 2a.

| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **25-50 EUR** | 1.09 | -0.05 | 0.94 | 106.06 | .01 | 0.06 |
| **25-75 EUR** | 1.37 | -0.07 | 0.99 | 630.76 | .01 | 0.11 |
| **25-100 EUR** | 1.40 | -0.05 | 1.00 | 2081.27 | .01 | 0.14 |
| **25-150 EUR** | 1.48 | -0.07 | 0.99 | 848.59 | .01 | 0.16 |
| **50-75 EUR** | 0.84 | 0.10 | 0.90 | 61.35 | .01 | 0.07 |
| **50-100 EUR** | 1.37 | -0.11 | 0.99 | 628.21 | .01 | 0.09 |
| **50-150 EUR** | 1.34 | -0.07 | 0.98 | 293.36 | .01 | 0.11 |
| **75-100 EUR** | 0.94 | -0.06 | 0.87 | 45.79 | .01 | 0.11 |
| **75-150 EUR** | 1.35 | -0.13 | 0.95 | 139.37 | .01 | 0.09 |
| **100-150 EUR** | 1.23 | -0.13 | 0.94 | 114.14 | .01 | 0.08 |

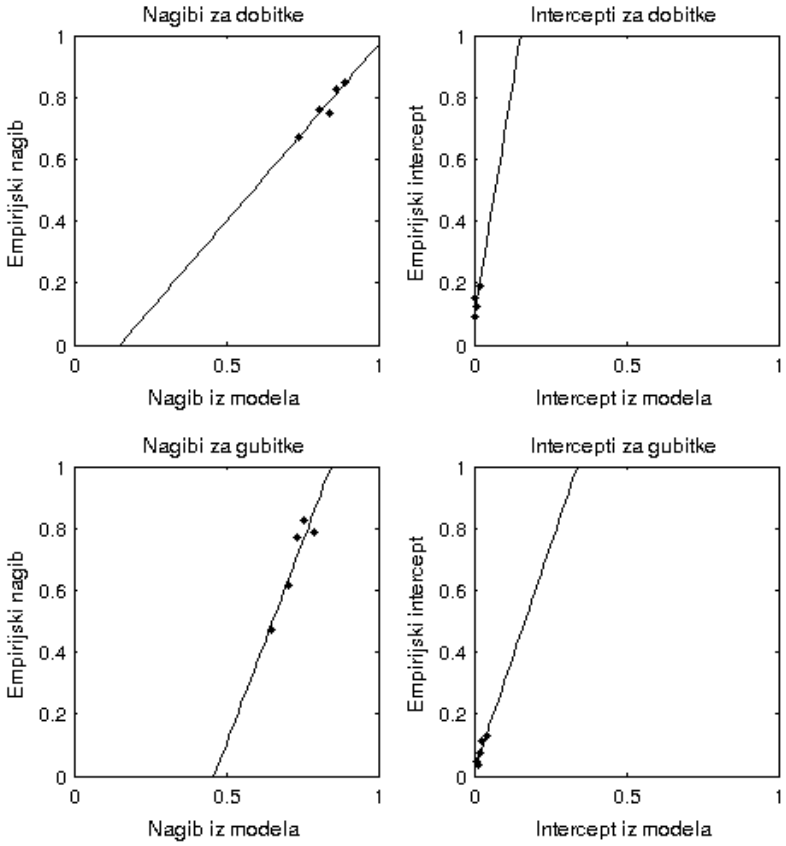
Tabela 9c. Teorija poverenja: regresione analize između predikcija modela za *p"* i empirijske *p"* (up. sliku 29b). Analize se odnose na sve ne-pozitivne lozove oblika (-$x$,$p$;0,1-$p$) u eksperimentu 2a.

| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **-25 EUR** | 0.73 | 0.10 | 0.94 | 118.41 | .01 | 0.51 |
| **-50 EUR** | 0.87 | 0.09 | 0.95 | 128.80 | .01 | 0.55 |
| **-75 EUR** | 1.05 | 0.05 | 0.99 | 640.34 | .01 | 0.60 |
| **-100 EUR** | 1.09 | 0.02 | 0.99 | 805.38 | .01 | 0.63 |
| **-150 EUR** | 1.00 | 0.04 | 1.00 | 3840.80 | .01 | 0.63 |

Tabela 9d. Teorija poverenja: regresione analize između predikcija modela za *p"* i empirijske *p"* (up. sliku 29d). Analize se odnose na sve striktno negativne lozove oblika (-$x$,$p$;-$y$,1-$p$) u eksperimentu 2a.

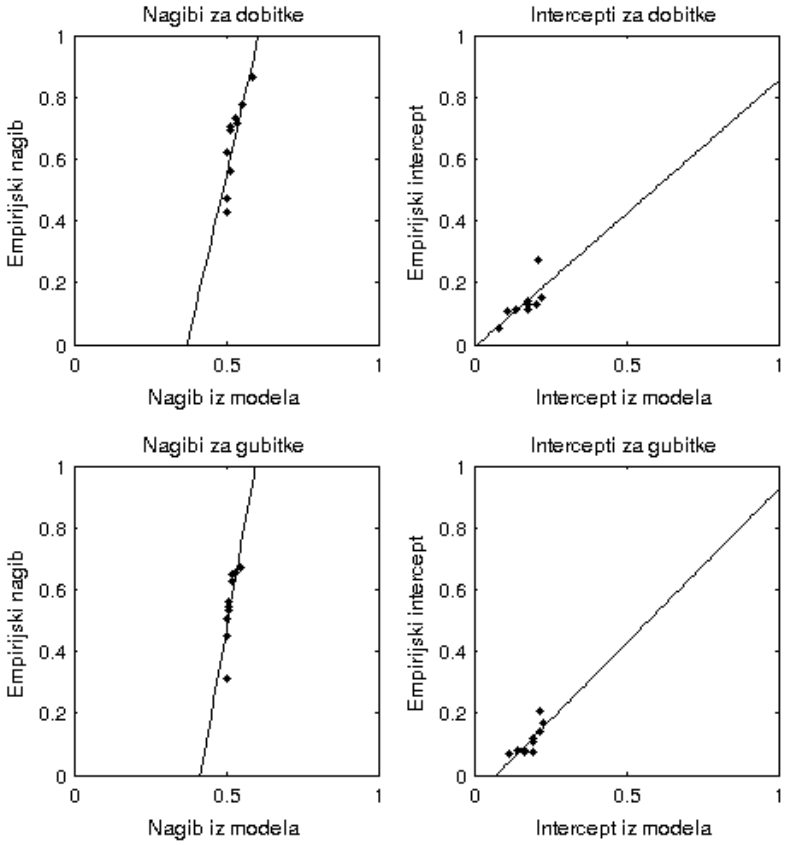
| dobici | nagib | intercept | $R^2$ | F | p < | RMSE |
|---|---|---|---|---|---|---|
| **-25-50 EUR** | 1.07 | -0.10 | 0.91 | 73.79 | .01 | 0.10 |
| **-25-75 EUR** | 1.21 | -0.12 | 0.96 | 188.43 | .01 | 0.07 |
| **-25-100 EUR** | 1.24 | -0.10 | 0.96 | 191.39 | .01 | 0.07 |
| **-25-150 EUR** | 1.22 | -0.07 | 0.96 | 190.88 | .01 | 0.07 |
| **-50-75 EUR** | 0.89 | 0.01 | 0.88 | 50.34 | .01 | 0.08 |
| **-50-100 EUR** | 1.11 | -0.10 | 0.96 | 162.93 | .01 | 0.07 |
| **-50-150 EUR** | 1.25 | -0.13 | 0.97 | 264.63 | .01 | 0.07 |
| **-75-100 EUR** | 0.62 | 0.03 | 0.93 | 94.80 | .01 | 0.17 |
| **-75-150 EUR** | 1.05 | -0.13 | 0.98 | 329.91 | .01 | 0.11 |
| **-100-150 EUR** | 1.00 | -0.08 | 0.96 | 188.46 | .01 | 0.09 |

Slika 30. prikazuje predikcije teorije poverenja za nagibe i intercepte linearnih regresionih funkcija za odnos objektivnih verovatnoća i *a posteriori* verovatnoća za ne-negativne i ne-pozitivne lozove u eksperimentu 2a. Pojasnimo: na oordinatama se nalaze nagibi i intercepti linearnih regresionih funkcija (tanke linije na slikama 28a i 28b) koje opisuju odnos objektivnih verovatnoća i empirijskih *a posteriori* verovatnoća na slikama 28a (dobici, gornji panel na slici 30) i 28c (gubici, donji panel na slici 30), dok se na abscisama nalaze predikcije teorije izgleda. Teorija poverenja je u stanju da pruži dobru predikciju nagiba i intercepta ovih linearnih funkcija: za predikciju nagiba za ne-negativne lozove, $R^2 = .91$, $F(1,3) = 29.45$, $p < .05$, za predikciju intercepta za ne-negativne lozove: $R^2 = .77$, $F(1,3) = 10.22$, $p < .05$; za predikciju nagiba za ne-pozitivne lozove, $R^2 = .87$, $F(1,3) = 20.67$, $p < .05$; za predikciju intercepta za ne-pozitivne lozove, $R^2 = .85$, $F(1,3) = 17.37$, $p < .05$. Kumulativna teorija izgleda nema eksplanatorne mehanizme kojima bi mogla da uklopi upravo diskutovani nalaz.




Slika 30. *Predikcije nagiba i intercepta linearnih regresionih funkcija između verovatnoća datih na lozovima i empirijskih a posteriori verovatnoća za ne-negativne (gornji panel) i ne-pozitivne (donji panel) lozove u eksperimentu 2a.*

Slika 31. prikazuje istu analizu kao i slika 30. ali za striktno pozitivne i striktno negativne lozove - nagibi i intercepti koje sada pokušavamo da predvidimo odlikuju linearne regresione funkcije (tanke linije) na slikama 29a i 29c.

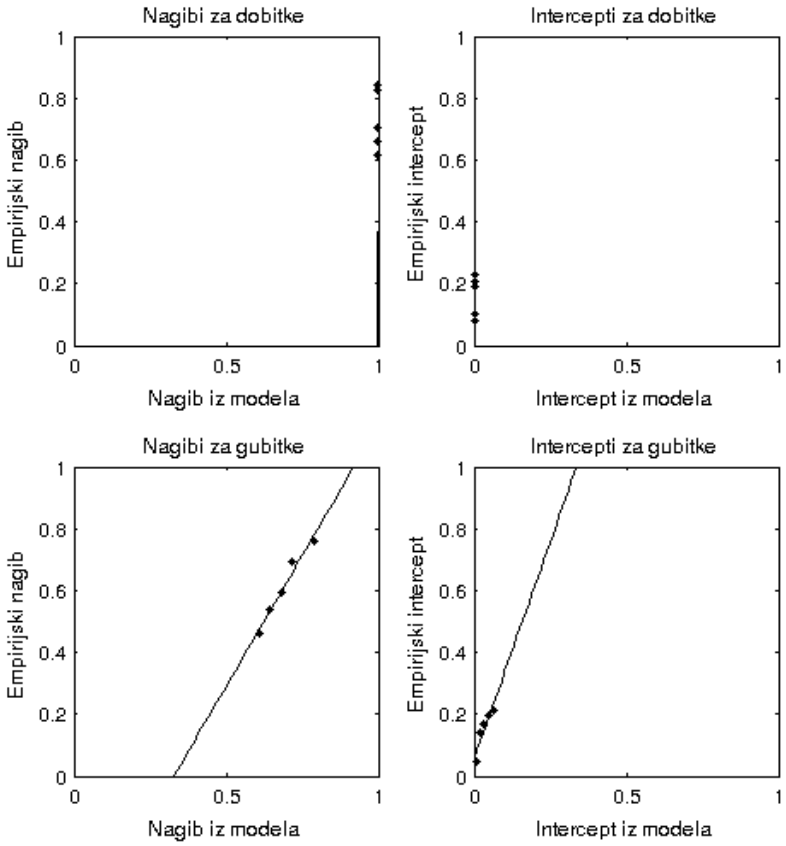



Slika 31. *Predikcije nagiba i intercepta linearnih regresionih funkcija između verovatnoća datih na lozovima i empirijskih a posteriori verovatnoća za striktno pozitivne (gornji panel) i striktno negativne (donji panel) lozove u eksperimentu 2a.*

Rezultati regresionih analiza za sliku 31. pokazuju sledeće: za predikciju nagiba za striktno pozitivne lozove, $R^2 = .68$, $F(1,8) = 16.89$, $p < .01$; za predikciju intercepta za striktno pozitivne lozove: $R^2 = .49$, $F(1,8) = 7.68$, $p < .05$; za predikciju nagiba za striktno negativne lozove, $R^2 = .58$, $F(1,8) = 11.15$, $p < .01$; za predikciju intercepta za striktno negativne lozove, $R^2 = .59$, $F(1,8) = 11.70$, $p < .01$. Kao što vidimo, predikcije su daleko od perfektnih, ali je generalni linearni trend predvidljiv.

Sve prethodno diskutovane analize ponderisanja verovatnoća ponovljene su za eksperiment 2b sa ishodima u dinarskim vrednostima. Samo iz želje da izbegnemo dopunskih dvadeset strana grafikona i tabela, ne prikazujemo ove analize detaljno[87].

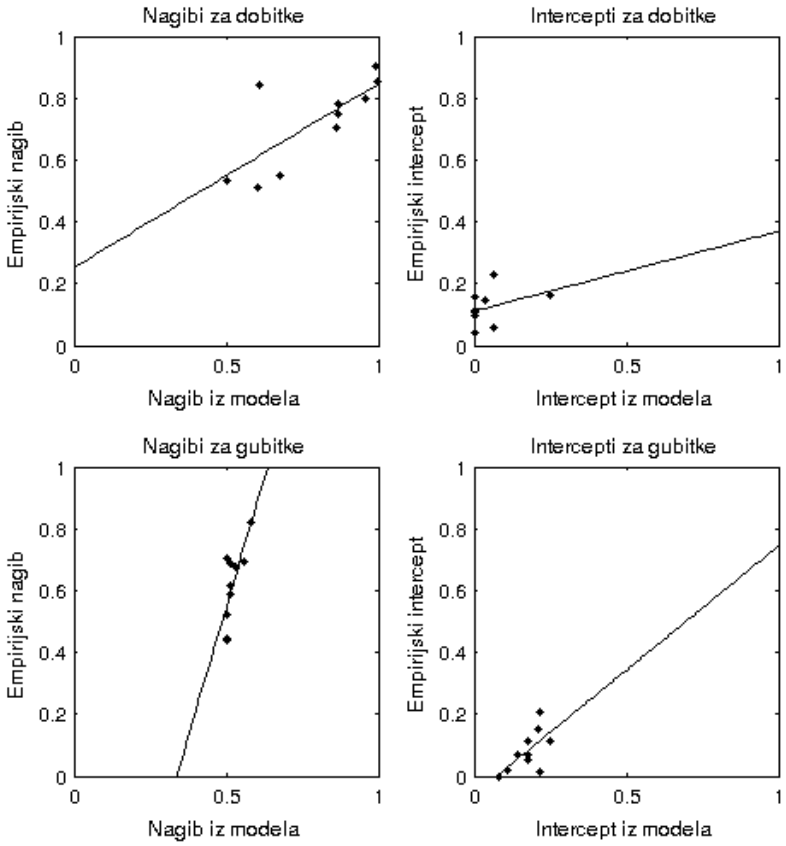
Njihovi rezultati su, na kvalitativnom nivou, poptuni isti kao rezultati koje smo ovde predstavili za eksperiment 2a. Slike 32. i 33. zato prikazuju samo predikcije nagiba i intercepta za eksperiment 2b - predikcije koje smo upravo diskutovali za podatke iz eksperimenta 2a.




Slika 32. *Predikcije nagiba i intercepta linearnih regresionih funkcija između verovatnoća datih na lozovima i empirijskih a posteriori verovatnoća za ne-negativne (gornji panel) i ne-pozitivne (donji panel) lozove u eksperimentu 2b.*

Rezultati regresionih analiza za sliku 32, eksperiment 2b, pokazuju sledeće: za predikciju nagiba za ne-negativne lozove, $R^2 = .08$, $F(1,3) = .27$, nije značajno; za predikciju intercepta za ne-negativne lozove: $R^2 = .07$, $F(1,3) = .22$, nije značajno; za predikciju nagiba za ne-pozitivne lozove, $R^2 = .96$, $F(1,3) = 63.61$, $p < .01$; za predikciju intercepta za ne-pozitivne lozove, $R^2 = .81$, $F(1,3) = 13.19$. $p < .05$. Vidimo da teorija poverenja ne predviđa dobro nagibe i intercepte linearnih regresionih funkcija između objektivnih i *a posteriori* verovatnoća za ne-negativne lozove u eksperimentu 2b. U tabeli 7.b nalazimo da je parametar dekumulativne funkcije $S$ za dobitke u eksperimentu 2b $q_{\mathrm{g}} = 2.95$ na prosečnim

monetarnim ekvivalentima, a slika 24b pokazuje da je funkcija $S$ sa ovom vrednošću parametra „priljubljena“ uz abscisu. Posledica je to da su sve *a priori* verovatnoće za dobitke skoro identične, što implicira veoma sličnu formu ponderisanja verovatnoća za ma koje lozove. Prema ocenama teorije poverenja, u proseku, ispitanici su procenama monetarnih ekvivalenata dobitaka u eksperimentu 2b pristupali kao da su njihova verovanja o odgovarajućim *a priori* verovatnoćama predstavljene uniformnim distribucijama.




Slika 33. *Predikcije nagiba i intercepta linearnih regresionih funkcija između verovatnoća datih na lozovima i empirijskih a posteriori verovatnoća za striktno pozitivne (gornji panel) i striktno negativne (donji panel) lozove u eksperimentu 2b.*

Rezultati regresionih analiza za sliku 33. koja se odnosi na striktno pozitivne i striktno negativne lozove u eksperimentu 2b pokazuju sledeće: za predikciju nagiba za striktno pozitivne lozove, $R^2 = .55$, $F(1,8) = 9.95$, $p < .05$; za predikciju intercepta za striktno pozitivne lozove: $R^2 = .13$, $F(1,8) = 1.21$, nije značajno; za predikciju nagiba za striktno negativne lozove, $R^2 = .54$, $F(1,8) = 9.25$, $p < .05$; za predikciju intercepta za striktno negativne lozove, $R^2 = .39$, $F(1,8) =$

5.15, marginalno značajno sa $p < .053$. Kao što vidimo, predikcije su slabije od prethodnih, predikcija intercepta za strikno pozitivne lozove nije ni dostigla nivo statističke značajnosti, ali je generalni linearni trend u ostalim analizama ponovo donekle predvidljiv.

Diskusija

Analiza eksperimenata 2a i 2b pokazala je, na prvom mestu, koliko su složene strukture podataka koje moramo da posmatramo da bismo prepoznali sve finese u odnosu različitih teorijskih modela prema empirijskim procenama monetarnih ekvivalenata. Merenje monetarnih ekvivalenata koje smo mi izveli u eksperimentima 2a i 2b predstavlja ubedljivo najopsežniju primenu ove metode od kada je ona u upotrebi. To ne znači da naš metod nije podložan kritikama.

Trenutno želimo da diskutujemo samo neke od njih, pošto ćemo uskoro potvrditi sve empirijske nalaze koje smo demonstrirali u eksperimentima 2a i 2b na skupu monetarnih ekvivalenata prikupljenih drugačijom metodom u studiji drugih istraživača. Prva bitna kritika može da se odnosi na našu odluku da analiziramo samo monetarne ekvivalente u rasponu od -2 do +2 standardne devijacije. Naveli smo već da je naša motivacija za ovu odluku vezana više za činjenicu da nam ona omogućava uklanjanje nekih banalnih grešaka koje su ispitanici mogli da naprave u ocenama monetarnih ekvivalenata mešovitih lozova, a koje nije moguće drugačije prepoznati i ukloniti tokom pregleda prikupljenih podataka. Naš odgovor na kritiku analize samo ograničenog skupa monetarnih ekvivalenata je sledeći: sve ovde prikazane analize i zaključci prvobitno su izvedeni na kompletnom skupu monetarnih ekvivalenata, odn. analizirano je svih 495 monetarnih ekvivalenata svakog ispitanika. Ne postoje nikakve robustne, kvalitativne razlike između rezultata koje smo naveli i rezultata dobijenih u toj prvoj analizi.

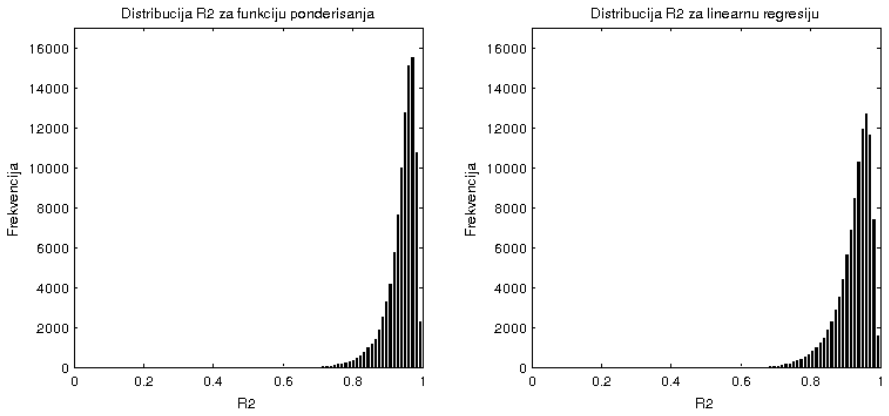
Druga bitna kritika može da se odnosi na analizu homogenosti preferencija koju smo izveli (slika 25. i rezultati analiza varijanse koji joj slede). Podelili smo, naizgled arbitrarno, ne-negativne i ne-pozitivne lozove na one koji sadrže više i one koji sadrže niže ishode, izbacujući čitav jedan nivo vrednosti ishoda (srednji) iz analize. Da li je to odgovarajući način da se analizira pretpostavka o homogenosti preferencija? Naš odgovor je: nedvosmisleno da. Bez obzira na izbacivanje srednjeg nivoa vrednosti, što je učinjeno samo da bi se nacrt analize varijanse izbalansirao na odgovarajući način, sve tačke na istim nivoima verovatnoće na sva četiri grafikona na slici 25. moraju da se poklapaju ako je ova pretpostavka zadovoljena. Pored analize koju smo mi prikazali na slici 25. i analiza varijanse kojima je testirano da

li četiri podskupa monetarnih lozova zadovoljavaju uslov homogenosti preferencija, iste takve analize koje uključuju svih pet nivoa vrednosti u eksperimentima 2a i 2b su izvedene, takođe dajući rezultate koji nedvosmisleno upućuju na to da homogenost preferencija nije zadovoljena. Izbegli smo da ih prikažemo pošto je demonstracija na slici 25. jednostavnija. Isti princip podele lozova na lozove sa višim i nižim ishodima koristili su Tverski i Kaneman u studiji u kojoj uvode kumulativnu teoriju izgleda (Tversky & Kahneman, 1992).

Analiza eksperimenata 2a i 2b pokazala je da bitna pretpostavka o homogenosti preferencija - pretpostavka o kojoj zavisi standardni parametarski model teorije izgleda kako se najčešće javlja u literaturi - nije zadovoljena za širok skup monetarnih ekvivalenata rizičnih lozova. Pored toga, analiza je pokazala da je ova pretpostavka narušena na sistematski način, tj. da to što nije zadovoljena nije posledica tek prisustva grešaka u eksperimentalnim podacima. U analizi ponderisanja verovatnoća, videli smo da postoji sistematsko variranje subjektivnih ocena verovatnoća - pondera odluka u teoriji izgleda i *a posteriori* verovatnoća u teoriji poverenja - koje kumulativna teorija ni u principu ne može da objasni. Za razliku od nje, teorija poverenja, čiji su statistički fitovi na prosečnim monetarnim ekvivalentima dosledno postigli nižu RMSE vrednost od kumulativne teorije izgleda, upravo predviđa takva sistematska variranja. Statistički posmatrano, sve do nivoa analize sistematskih variranja u subjektivnim ocenama verovatnoća do nivoa pojedinačnih lozova, teoriju izgleda i teoriju poverenja nije moguće porediti, pošto su njihove performanse praktično iste. Tek na dubljem nivou eksperimentalne analize monetarnih ekvivalenata otkrivamo strukture podatka koje jedna teorija objašnjava, a druga ne.

Postavlja se pitanje kako je moguće da funkcija ponderisanja verovatnoća koja važi pod kumulativnom teorijom izgleda, i za koju posle analize homogenosti preferencija *de facto* znamo da ne može da bude jedinstvena, ipak postiže izuzetne statističke fitove subjektivnih pondera odluka koji sistematski variraju sa vrednostima ishoda na lozovima? Razmišljanje o odgovoru na ovo pitanje navelo nas je na ideju da izvedemo sledeću stohastičku simulaciju. Zadali smo sledeći vektor od jedanest verovatnoća: .01, .05, .1, .25, .40, .50, .60, .75, .90, .95 i .99. Zatim smo izvukli 100000 uzoraka veličine $n = 11$ slučajnih brojeva sa uniformne distribucije na rasponu od nula do jedan. Pošto je funkcija ponderisanja verovatnoća $w$ funkcije izgleda po definiciji monotono rastuća, kao i odgovarajuća linearna funkcija teorije poverenja, sortirali smo u rastući redosled jedanest dobijenih slučajnih brojeva za

svaki od 100000 uzoraka. Ti brojevi u ovoj simulaciji igraju ulogu slučajnih pondera odluka koji su mogli biti dodeljeni verovatnoćama u definisanom vektoru, i koji zadovoljavaju uslov da ponderi odluke budu monotono rastuća funkcija verovatnoće. Za svih 100000 slučajnih uzoraka odredili smo parametar najbolje fitujuće Prelecove jednoparametarske funkcije ponderisanja koju smo koristili u svim prethodnim analizama. Zatim su izračunate predikcije ponderisanja verovatnoća na osnovu tako dobijene funkcije za sve vrednosti u prethodno definisanom „nacrtu“ odn. vektoru verovatnoća od .01 do .99. Za svaki od 100000 slučajnih uzoraka, regresionom analizom je ustanovljena $R^2$ vrednost linearne regresije vektora od .01 do .99 i slučajno dobijenih monotono rastućih brojeva između nula i jedan. Simulacija pokazuje sledeće rezultate: prosečan $R^2$ kroz 100000 slučajnih uzoraka za Prelecovu funkciju ponderisanja verovatnoća iznosi .9382, dok je ista prosečna vrednost za linearne regresione funkcije .9243! Drugim rečima, na skali verovatnoće, funkcija ponderisanja verovatnoća fituje *slučajne monotono rastuće vrednosti* donekle bolje od linearnih funkcija dobijenih linearnom regresijom. Slika 34. prikazuje distribucije vrednosti $R^2$ dobijene ovom simulacijom za funkciju ponderisanja verovatnoća i za linearnu regresiju: golim okom je vidljiva prednost funkcije ponderisanja verovatnoća na uzorcima slučajnih, monotono rastućih vrednosti između nula i jedan.



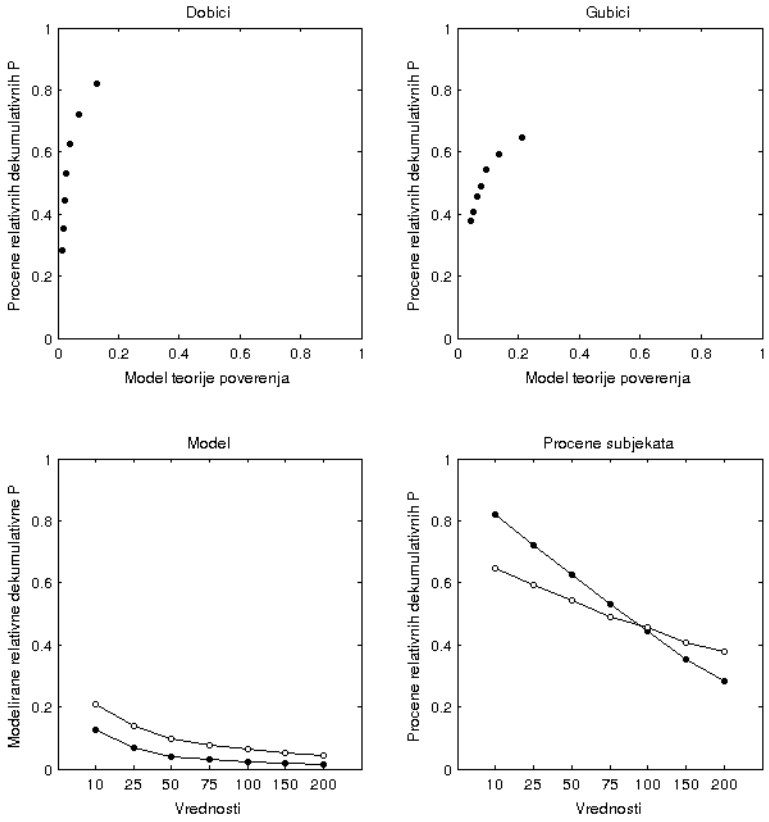

Slika 34. *Rezultati simulacije pondera odlučivanja: distribucije $R^2$ za funkciju ponderisanja verovatnoća (levi panel) i linearnu regresionu funkciju (desni panel).*

Rezultate ove *Monte Carlo* simulacije treba uzeti samo u svrhu ilustracije. Očigledno je sledeće: na intervalu (0,1), monotono rastuće vrednosti, čak i kada predstavljaju slučajne brojeve sa uniformne distribucije, nije teško fitovati sa visokom vrednošću $R^2$. Funkcija ponderisanja verovatnoća, sa izraženom nelinearnom formom i smenom konveksnih i konkavnih regiona, u prednosti je

u odnosu na linearne funkcije kakve predviđa teorija poverenja. Naš utisak, posle svega, je da funkcija ponderisanja verovatnoća u teoriji izgleda *nije dovoljno restriktivna*. Po svemu sudeći, ta čudna forma koju je odlikuje - forma za koju je Prelec, podsetimo se, primetio da je teško da bi je neko nacrtao bez da je na to primoran jakim empirijskim razlozima - joj omogućava da fituje praktično *bilo šta* u svom domenu. Analiza ponderisanja verovatnoća u eksperimentima 2a i 2b pokazala je koliko su ti empirijski razlozi zapravo nedovoljno motivisani.

Podsetimo se da su ispitanici u eksperimentima 2a i 2b, pored procena monetarnih ekvivalenata, davali i procene relativnih frekvencija monetarnih dobitaka i gubitaka. Pošto je osnovna pretpostavka teorije poverenja da kognitivni sistem reprezentuje svoja prethodna iskustva u ekonomskim interakcijama i oslanja se na te reprezentacije u odlučivanju, pokušali smo da ovom metodom proverimo bar osnovni „aksiom“ teorije poverenja: da su verovatnoće pozitivnih i negativnih ishoda monotono opadajuće funkcije njihovih apsolutnih vrednosti. Rezultati dobijeni procenom relativnih frekvencija monetarnih dobitaka i gubitaka potvrđuju ovu pretpostavku. Na procenama relativnih frekvencija za dobitke i gubitke svih ispitanika u ogledu 2a izvedena je potpuno ponovoljena analiza varijanse sa faktorima znaka (dobitak ili gubitak) i visine ishoda. Analiza varijanse otkriva da efekat znaka (dobitak ili gubitak) nije statistički značajan, za razliku od efekta visine ishoda, $F(1.9,14.23) = 17.40$, $p < .01$; dvofaktorska interakcija nije dostigla statističku značajanost. Slični rezultati su dobijeni za procene ispitanika u eksperimentu 2b: efekat znaka (dobitak ili gubitak) je na ivici statističke značajnosti, $F(1,14) = 3.60$, $p < .08$; efekta visine ishoda je značajan, $F(1.42,19.92) = 39.38$, $p < .01$; dvofaktorska interakcija opet ne dostiže nivo statističke značajanosti. Slika 35. prikazuje rezultate ovih procena za eksperiment 2a. Gornji levi panel prikazuje odnos predikcije dekumulativnih verovatnoća dobitaka iz teorije poverenja sa parametrima ocenjenim na prosečnim monetarnim ekvivalentima, a gornji desni panel isti odnos za dekumulativne verovatnoće gubitaka. Procene ispitanika su reskalirane na raspon (0,1). Predikcije modela predstavljaju *a priori* verovatnoće prema odgovarajućim funkcijama $S$. Očigledno, procene ispitanika i modelirane verovatnoće se ne nalaze na istoj skali. Donji levi panel slike 35. prikazuje predikcije modela teorije poverenja, a donji desni panel prosečne procene ispitanika reskalirane na raspon (0,1). Podsetimo da nema statistički značajnih razlika između procena za dobitke i gubitke, kao ni interakcije između faktora znaka (dobitak ili gubitak) i visine ishoda. Rezultati ove analize za eksperiment 2b su veoma slični, osim što zbog vrednosti parametra

funkcije $S$ za dobitke, $q_g$, sigurno nije moguće predvideti porast u dekumulativnoj verovatnoći odgovarajućih vrednosti ishoda.




Slika 35. *Procene relativnih frekvencija monetarnih dobitaka i gubitaka ispitanika iz eksperimenta 2a.* Objašnjenje u tekstu.

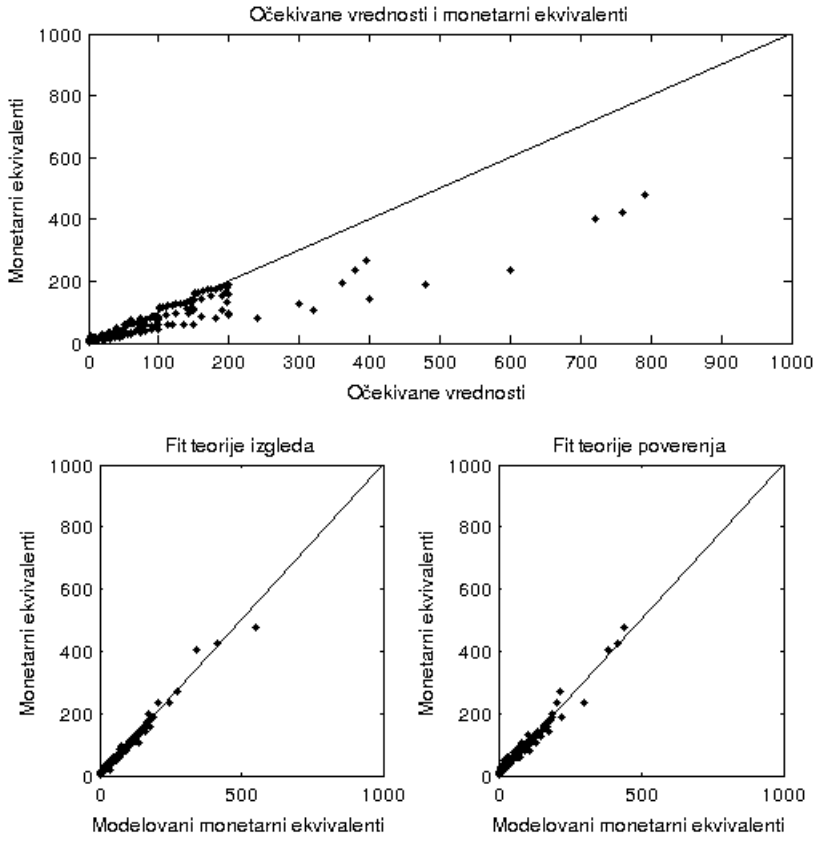
Analize eksperimenata 2a i 2b favorizuju teoriju poverenja nad kumulativnom teorijom izgleda. Našoj metodi je moguće uputiti više kritika, sigurno. Jedna od njih je da direktne numeričke ocene monetarnih ekvivalenata možda nisu najpreciznji način merenja; studije poput Tverskog i Kanemana iz 1992 (Tversky & Kahneman, 1992), ili Gonzalesa i Vua iz 1999 (Gonzales & Wu, 1999), koriste metode koje su *prima facie* bolje kontrolisane od metode koju smo koristili mi. Naši ispitanici nisu bili plaćeni za učešće u eksperimentima 2a i 2b, niti su mogli da osvoje ma kakav realan novac odigravajući neki podskup lozova koje smo im prikazali. U narednim redovima otklanjamo sve sumnje da je naš metodološki pristup, na neki način, izvor strukture eksperimentalnih podataka koja favorizuje teoriju poverenja nad teorijom izgleda.

## Gonzalesa i Vua, 1999.

Gonzales i Vu su 1999. godine objavili studiju merenja monetarnih ekvivalenata u kojoj su koristili istu proceduru procene koja je korišćena u studiji Tverskog i Kanemana 1992. godine (Gonzales & Wu, 1999, Tversky & Kahneman, 1992). Procedura se ne bazira na direktnoj numeričkoj oceni monetarnog ekvivalenta kakva je primenjena u eksperimentima 2a i 2b. U eksperimentu je učestvovalo 10 ispitanika. Stimuluse je predstavljalo 15 osnovnih lozova, ne-negativnih i striktno pozitivnih - dakle, studija nije uopšte razmatrala gubitke niti averziju prema gubicima - koji su kombinovani sa jedanest nivoa verovatnoće na isti način na koji je to izvedeno u dizajnu eksperimenata 2a i 2b. Ishodi na lozovima su dati u dolarima, a petnest osnovnih parova ishoda su bili: 25–0, 50–0, 75–0, 100–0, 150–0, 200–0, 400–0, 800–0, 50–25, 75–50, 100–50, 150–50, 150–100, 200–100, i 200–150. Jedanest nivoa verovatnoće koji su pridruženi jednom ishodu na lozu (dok je drugi uvek dobijao 1-$p$) bili su sledeći: .01, .05, .10, .25, .40, .50, .60, .75, .90, .95, i .99. Na taj način je proizvedeno ukupno 165 rizičnih lozova.

Procedura merenja monetarnih ekvivalenata u ovoj studiji je sledeća. Ispitanicima se prvo prikazuje loz koji prati tabela sigurnih ishoda. Zadatak ispitanika je da za svaki od ponuđenih sigurnih ishoda označi da li bi pre prihvatio da odigra rizični loz koji mu je prikazan, ili bi pre prihvatio siguran ishod. Raspon sigurnih ishoda koji je ponuđen u ovim izborima odgovarao je rasponu ishoda koje je sadržao loz. Eksperimentalni softver je onemogućavao ispitanike da daju nekonzistentne odgovore (na primer, da odbiju siguran ishod od 75$ pošto su već prethodno prihvatili siguran ishod od 50$ za isti loz). Pošto bi ispitanici tako doneli svoje odluke o tome koje iznose prihvataju a koje ne u odnosu na mogućnost odigravanja rizičnog loza, softver bi odredio raspon u kome mora da se nalazi monetarni ekvivalent loza kao onaj između maksimalnog sigurnog ishoda koji je odbijen i minimalnog sigurnog ishoda koji je prihvaćen. Procedura se ponavlja, ali se u sledećem koraku raspon sigurnih ishoda nalazi između dve prethodno određene vrednosti. Na isti način, procedura se ponavlja dok se monetarni ekvivalent ishoda ne odredi sa preciznošću od 1$. Kaneman i Tverski su 1992. godine naveli da je prednost ove procedure što se bazira na izborima ispitanika, a ne na direktnoj numeričkoj proceni. Iako ova tvrdnja zahteva složeniju elaboraciju, niko od istraživača je nije pružio.

Skup monetarnih ekvivalenata koji su prikupili Gonzales i Vu u originalnoj studiji je korišćen u razvoju metode neparametrijske ocene funkcija teorije izgleda. Mi smo procene monetarnih ekvivalenata deset ispitanika iz ove studije iskoristili da izvedemo iste analize koje smo izveli i diskutovali za eksperimente 2a i 2b. Model kumulativne teorije izgleda za ovaj eksperiment uključuje samo dva parametra: eksponent stepene funkcije korisnosti, $\rho$, i parametar funkcije ponderisanja verovatnoća, $\gamma$. Teorija poverenja takođe je zadata kroz samo dva parametra: eksponent stepene funkcije korisnosti, $\rho$, i parametar dekumulativne funkcije $S$, $q$. Modeli obe teorije su fitovani na procene monetarnih ekvivalenata pojedinačnih ispitanika i na prosečne procene monetarnih ekvivalenata istom metodom korišćenom u eksperimentima 2a i 2b. Slika 36. prikazuje odnos očekivanih vrednosti i prosečnih monetarnih ekvivalenata, te predikcije teorije izgleda i teorije poverenja.




Slika 36. *Prosečni monetarni ekvivalenti i njihove očekivane vrednosti u studiji Gonzalesa i Vua (gornji panel); fitovi kumlativne teorije izgleda (dole levo) i teorije poverenja (dole desno) dobijeni na osnovu optimalnih vrednosti parametara za prosečne monetarne ekvivalente.*

Tabela 10a sadrži optimalne vrednosti parametara za sve ispitanike i prosečne monetarne ekvivalente, sirovu kvadratnu grešku ($SSE$), srednju kvadratnu grešku ($RMSE$), vrednosti $R$ i $R^2$ za model teorije izgleda. Tabela 10b sadrži iste ove informacije za model teorije poverenja. Fitovanje modela otkriva očekivane funkcije korisnosti sa osobinom averzije prema riziku, kao i tipično ponderisanje verovatnoća - potcenjivanje visokih i precenjivanje visokih verovatnoća navedenih na lozovima. Već na gornjem panelu slike 36. vidimo da je averzija prema riziku očekivano zastupljena u prosečnim procenama monetarnih ekvivalenata. Kvalitet statističkih fitova teorije izgleda i teorije poverenja opet nije moguće uporediti, jer su oba skoro perfektna. Teorija poverenja ponovo otkriva manje konkavnu funkciju korisnosti u prosečnim monetarnim ekvivalentima u odnosu na ocenu kumulativne teorije izgleda. Slika 37. prikazuje teorijske funkcije modela teorije poverenja i teorije izgleda sa optimalnim vrednostima parametara za procene prosečnih monetarnih ekvivalenata.

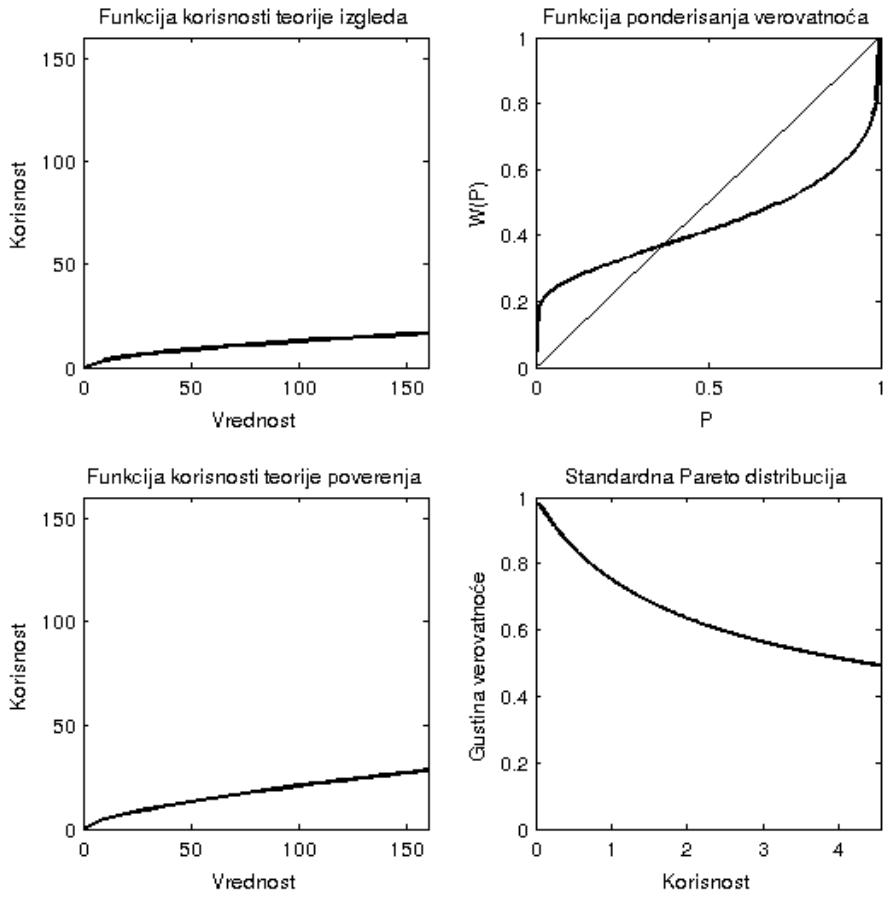
Tabela 10a. Fitovi teorije izgleda za monetarne eksperimente iz studije Gonzalesa i Vua, 1999.

| **Subjekat** | $\rho$ | $\gamma$ | **SSE** | **RMSE** | **R** | $\mathbf{R^2}$ |
|---|---|---|---|---|---|---|
| 1 | 0.51 | 0.32 | 67565.99 | 20.36 | 0.96 | 0.92 |
| 2 | 0.372 | 0.631 | 144909.20 | 29.82 | 0.96 | 0.91 |
| 3 | 0.990 | 0.446 | 103961.22 | 25.25 | 0.97 | 0.95 |
| 4 | 0.325 | 0.084 | 31940.18 | 14.00 | 0.97 | 0.94 |
| 5 | 0.568 | 0.275 | 171482.59 | 32.44 | 0.91 | 0.83 |
| 6 | 0.843 | 0.944 | 16721.06 | 10.13 | 1.00 | 0.99 |
| 7 | 0.429 | 0.151 | 40536.83 | 15.77 | 0.96 | 0.92 |
| 8 | 0.281 | 0.271 | 95013.71 | 24.14 | 0.91 | 0.84 |
| 9 | 0.509 | 0.834 | 46094.81 | 16.82 | 0.99 | 0.98 |
| 10 | 0.519 | 0.473 | 115104.97 | 26.57 | 0.95 | 0.91 |
| $\mathbf{M_{ME}}$ | **0.55** | **0.34** | **17191.58** | **10.27** | **0.99** | **0.98** |

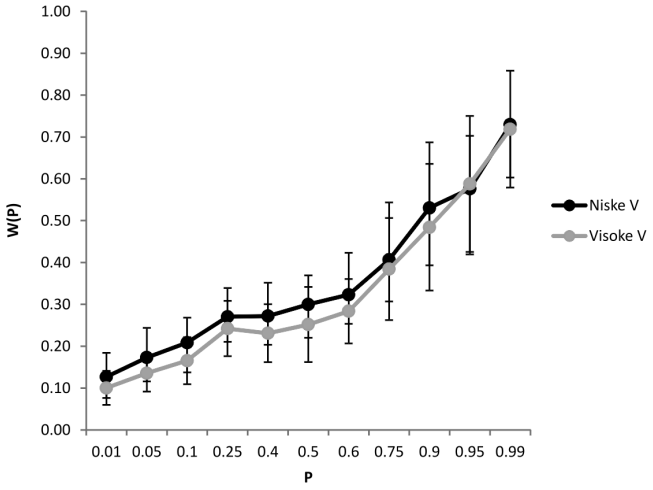
Tabela 10b. Fitovi teorije poverenja za monetarne eksperimente iz studije Gonzalesa i Vua, 1999.

| **Subjekat** | $\rho$ | $q$ | **SSE** | **RMSE** | **R** | $\mathbf{R^2}$ |
|---|---|---|---|---|---|---|
| 1 | 0.42 | 0.00 | 93392.03 | 23.94 | 0.94 | 0.89 |
| 2 | 0.18 | 486.35 | 173183.48 | 32.60 | 0.96 | 0.91 |
| 3 | 1.32 | 0.23 | 115762.97 | 26.65 | 0.97 | 0.94 |
| 4 | 0.21 | 0.64 | 57048.51 | 18.71 | 0.96 | 0.92 |
| 5 | 0.62 | 0.29 | 127819.60 | 28.00 | 0.93 | 0.87 |
| 6 | 0.83 | 128.03 | 17489.80 | 10.36 | 1.00 | 0.99 |
| 7 | 0.32 | 0.27 | 55576.41 | 18.47 | 0.95 | 0.91 |
| 8 | 0.05 | 6.54 | 111764.29 | 26.19 | 0.91 | 0.83 |
| 9 | 0.50 | 1.73 | 48907.46 | 17.32 | 0.99 | 0.98 |
| 10 | 0.48 | 1.16 | 101865.99 | 25.00 | 0.96 | 0.93 |
| $\mathbf{M_{ME}}$ | **0.66** | **0.41** | **24858.46** | **12.35** | **0.99** | **0.97** |

Slika 38. predstavlja empirijske vrednosti $w(p)$ izračunate prema jednačini (86); analiza homogenosti preferencija je ista kao što je izvedena za eksperimente 2a i 2b. Svi ne-negativni lozovi oblika ($x$,$p$;0,1-$p$) su podeljeni u dve grupe, grupu sa visokim ishodima (150$, 200$, 400$ i 800$) i grupu sa niskim ishodima (25$, 50$, 75$ i 100$). Izvedena je potpuno ponovoljena analiza varijanse sa faktorom visine ishoda (dva nivoa) i faktorom verovatnoće sa kojim je određeni ishod naveden na lozu (jedanest nivoa). Rezultati analize varijanse otkrivaju značajan efekat faktora visine ishoda, $F(1,9) = 19.40$, $p < .01$, značajan efekat faktora verovatnoće, $F(10,90) = 30.52$, $p < .01$, i značajan efekat dvofaktorske interakcije, $F(10,90) = 2.37$, $p < .05$. Dakle, ni u skupu monetarnih eksperimenata iz studije Gonzalesa i Vua nije zadovoljen uslov homogenosti preferencija.




Slika 37. *Teorijske funkcije modela teorije izgleda i teorije poverenja u studiji Gonzalesa i Vua (1999).* Vrednosti parametera ovih funkcija su optimalne vrednosti za prosečne monetarne ekvivalente.




Slika 38. *Analiza homogenosti preferencija za studiju Gonzalesa i Vua (1999).* Analiza obuhvata sve ne-negativne lozovi oblika (x,p;0,1-p).

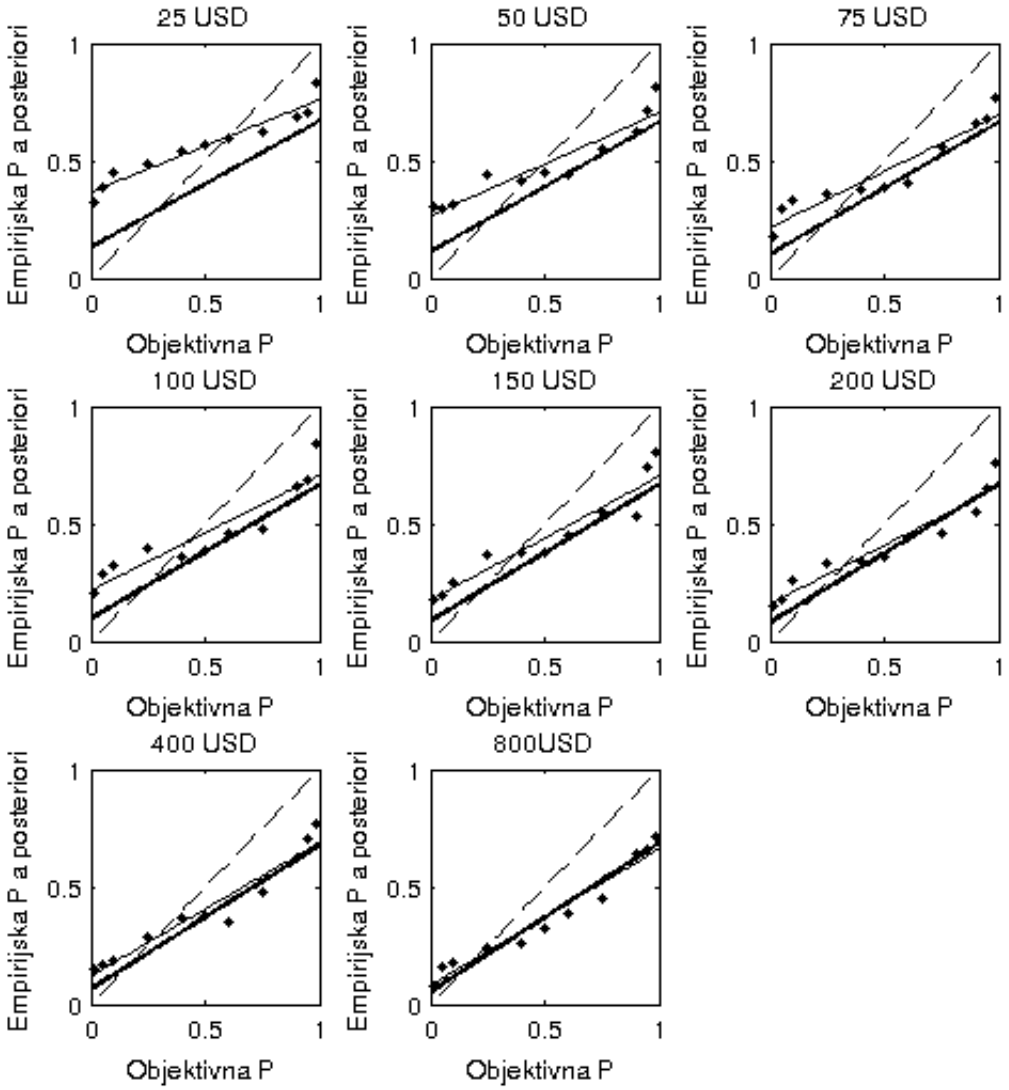
Slika 39a prikazuje analizu ponderisanja verovatnoća pod teorijom poverenja za ne-negativne lozove iz ove studije, a sledeća slika, 39b, prikazuje linearne regresije između modeliranih i empirijskih *a posteriori* verovatnoća. Slike 40a i 40b prikazuju iste odnose za striktno pozitivne lozove iz ove studije. Generalno, nije teško uvideti da se sistematska variranja subjektivnih verovatnoća kakva smo prepoznali u analizama eksperimenata 2a i 2b javljaju i u studiji Gonzalesa i Vua koja je koristila drugačiji metod procene monetarnih ekvivalenata.

Slike 41a i 41b prikazuju predikciju nagiba i intercepta za odgovarajuće linearne regresione funkcije između objektivnih verovatnoća i empirijskih a posteriori verovatnoća za ne-negativne (slika 41a) i striktno pozitivne (slika 41b) lozove. Regresione analize koje se odnose na sliku 41a pokazuju sledeće: za predikciju nagiba linearnih regresionih funkcija za ne-negativne lozove, $R^2 = .90$, $F(1,6) = 57.44$, $p < .01$; za predikciju intercepta za ne-negativne lozove, $R^2 = .97$, $F(1,6) = 200.00$, $p < .01$. Za regresione analize koje se odnose na sliku 40b: za predikciju nagiba linearnih regresionih funkcija za striktno pozitivne lozove, $R^2 = .41$, $F(1,5) = 3.59$, nije značajno; za predikciju intercepta za striktno negativne lozove, $R^2 = .65$, $F(1,5) = 9.48$, $p < .05$. Iako ne perfektno u svim slučajevima, ponovo vidimo da teorija poverenja predviđa pozitivan linearan trend u ocenama nagiba i intercepta empirijski najboljih linearnih regresionih funkcija koje opisuju odnos objektivnih i *a posteriori* verovatnoća.

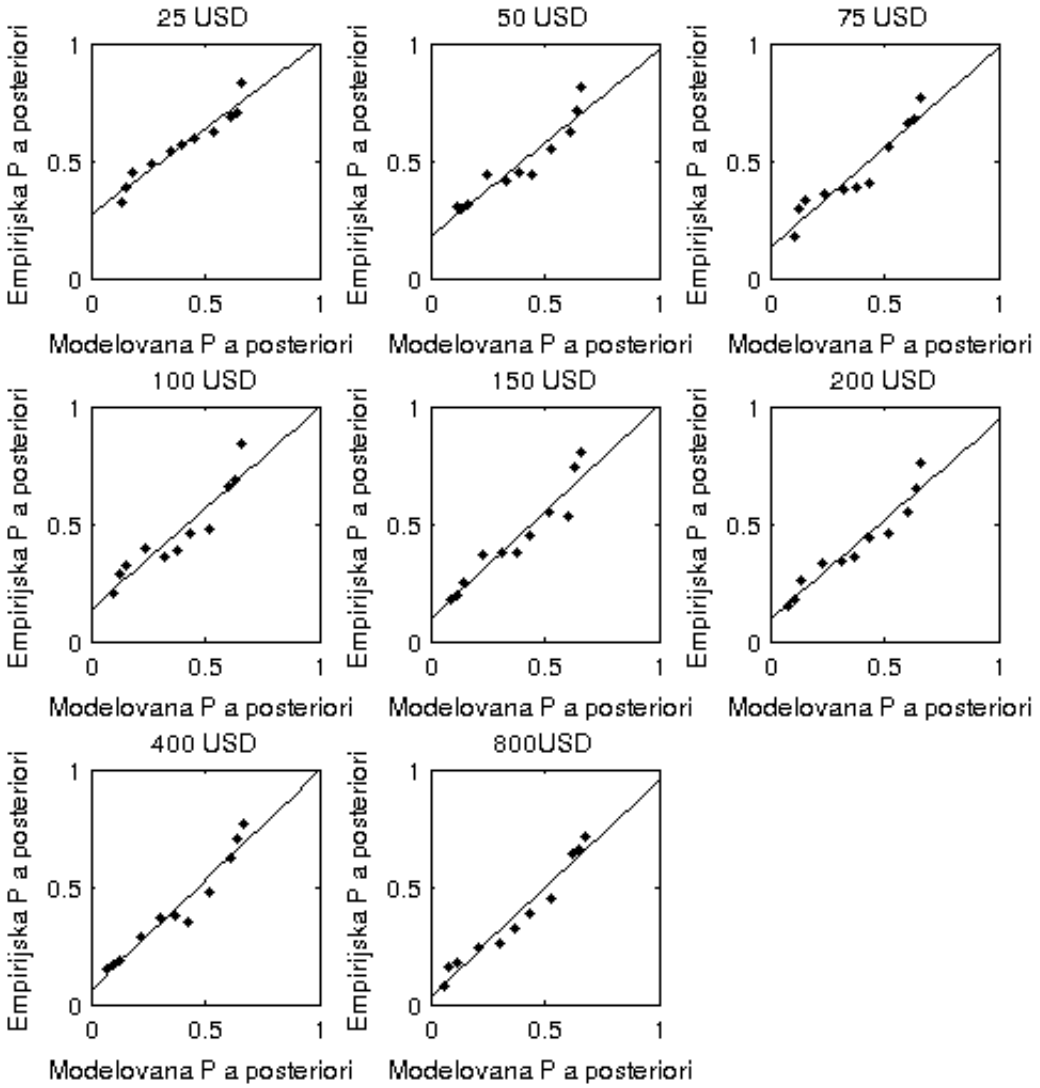
Upravo kao i analiza eksperimenata 2a i 2b, analiza procena monetarnih ekvivalenata koje su prikupili Gonzales i Vu u studiji iz 1999. godine ukazuje

na krupne probleme teorije izgleda. Homogenost preferencija nije zadovoljena; podsetimo se, to znači da standardni model teorije izgleda važi samo kao aproksimacija. Teorija poverenja, s druge strane, fituje eksperimentalne podatke dobro koliko i teorija izgleda, ali je pored toga u stanju da objasni odnose u dubljim strukturama eksperimentalnih podataka za koje teorija izgleda nema adekvatne teorijske mehanizme.

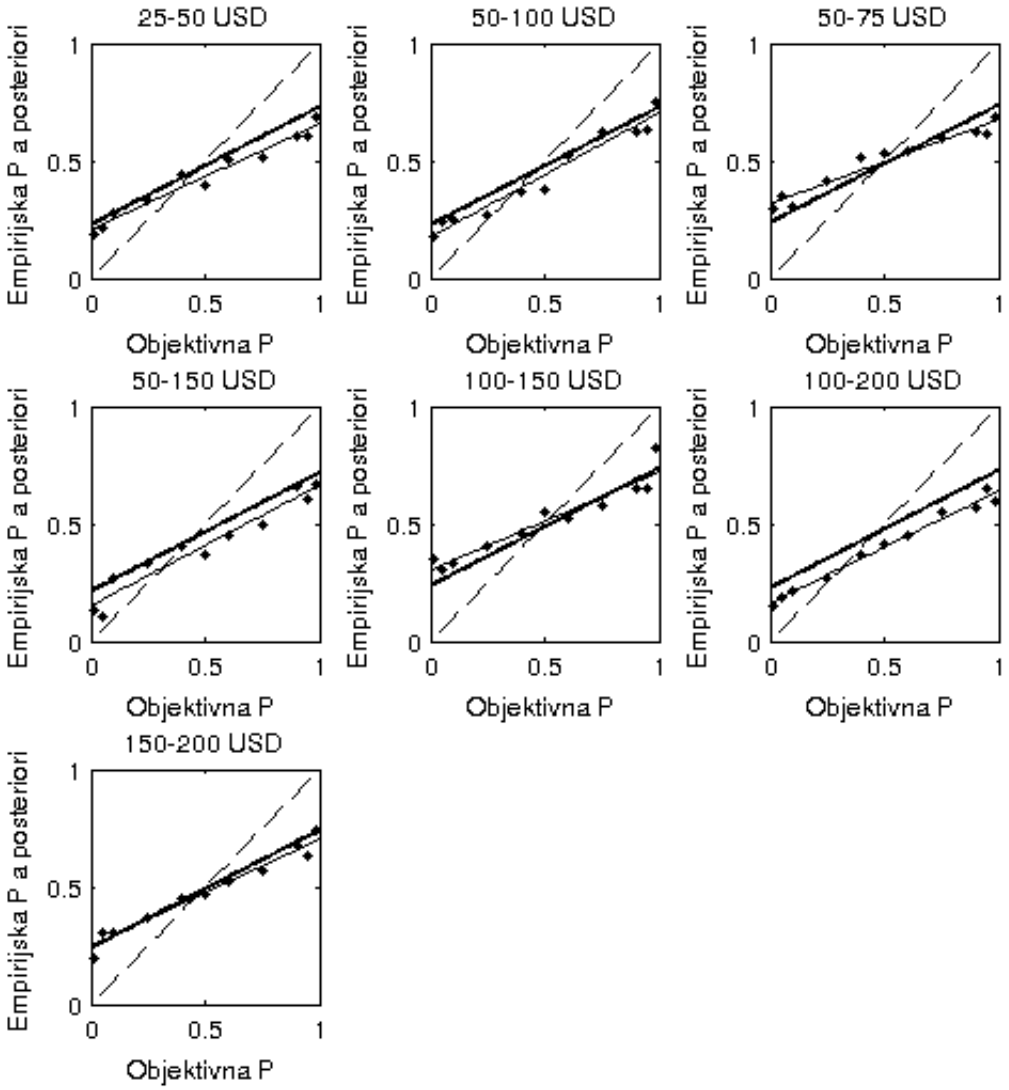
Očigledno, te strukture podataka nisu posledica metode merenja monetarnih ekvivalenata koju smo mi primenili u eksperimentima 2a i 2b. Homogenost preferencija, zgodna teorijska pretpostavka koja, začudo, do danas nije statistički analizirana ni u jednom poznatom skupu procena monetarnih ekvivalenata u ovoj oblasti, neodrživa je. Bilo koja deskriptivna teorija odlučivanja koja pokušava da obuhvati sve eksperimentalne strukture podataka moraće da počiva na strukturi u kojoj subjektivni tretman verovatnoća nije tek nezavistan od rangova odgovarajućih ishoda, već nije nezavistan od samih *visina* ishoda na lozovima koji se evaluiraju. Predlažemo da ovakve teorije - od kojih je teorija poverenja za sada jedina razvijena - nazovemo *modelima subjektivne verovatnoće zavisnim od korisnosti* (engl. *Utility-Dependent Subjective Probability*, skr. *UDSP* model).



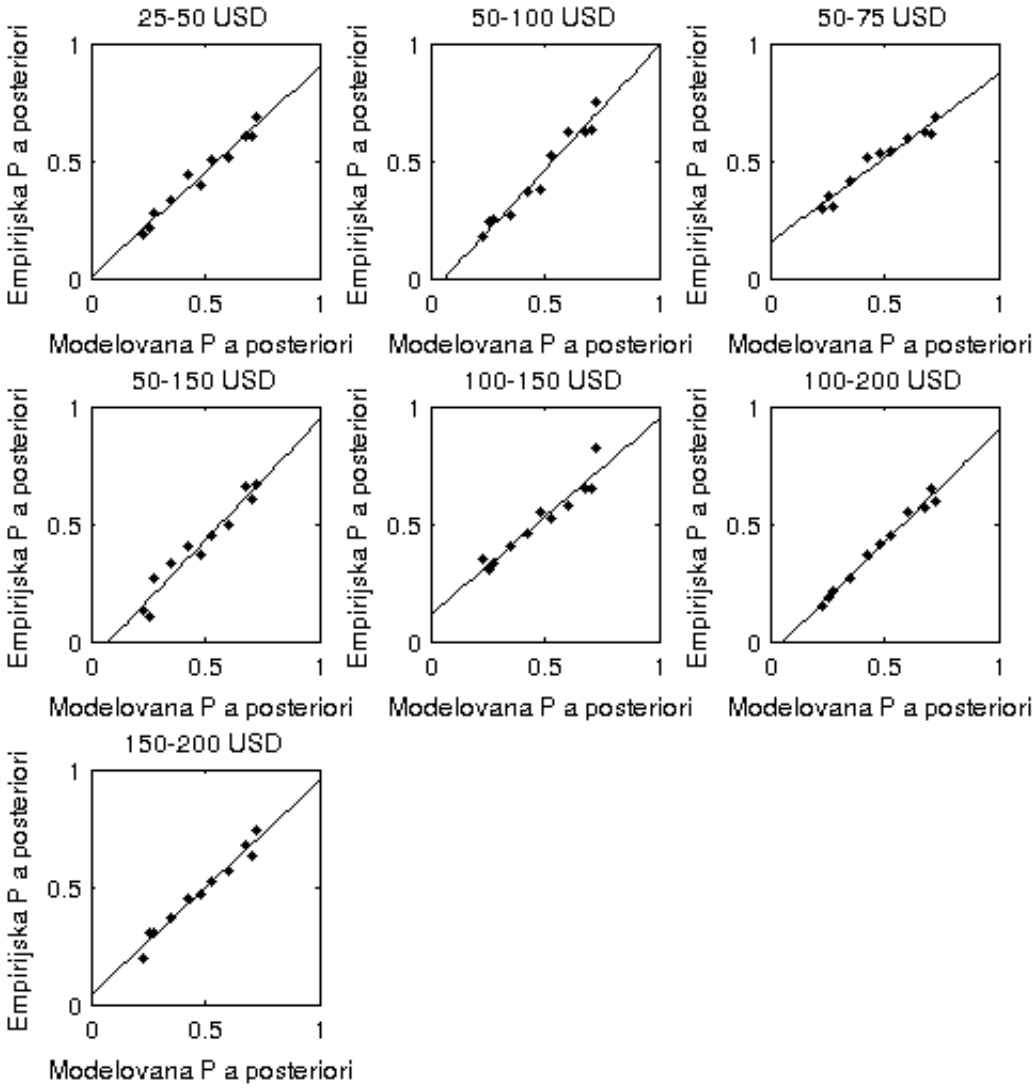

Slika 39a. *Analiza ponderisanja verovatnoća za studiju Gonzalesa i Vua (1999) za sve ne-negativne lozove oblika (x,p;0,1-p).* Tanja linija predstavlja najbolju regresionu funkciju koja povezuje objektivne i a posteriori verovatnoće, deblja teorijsku predikciju teorije poverenja.



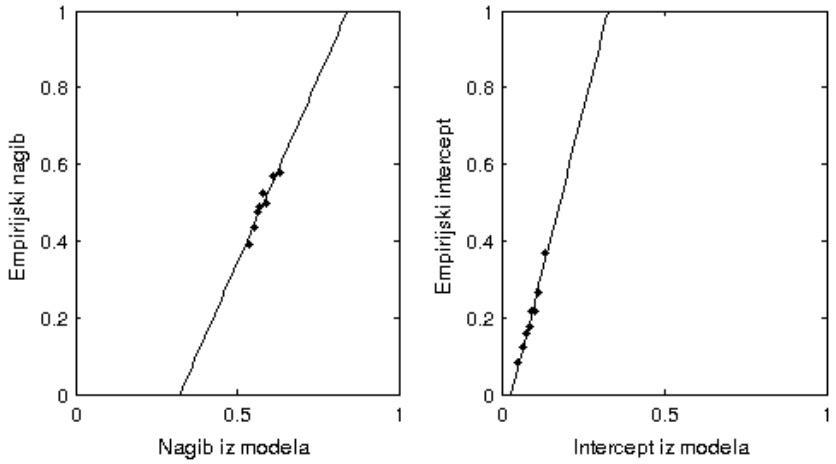

Slika 39b. *Regresione analize odnosa modeliranih a posteriori verovatnoća i empirijskih a posteriori verovatnoća za studiju Gonzalesa i Vua (1999) za sve ne-negativne lozove oblika (x,p;0,1-p).*



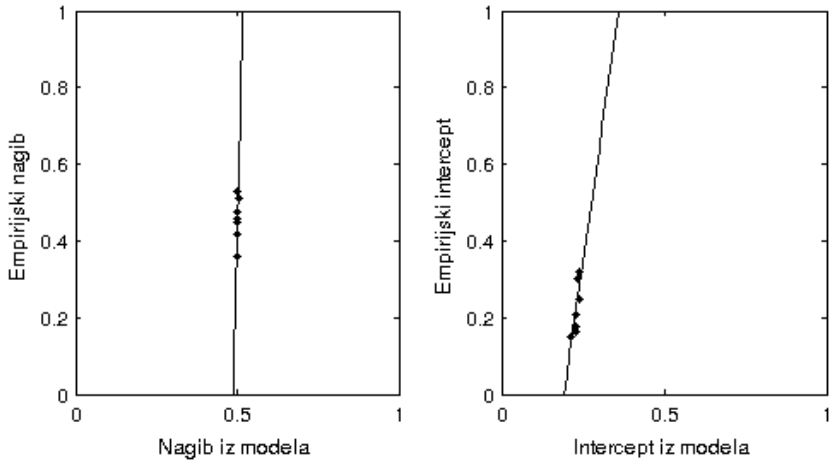

Slika 40a. *Analiza ponderisanja verovatnoća za studiju Gonzalesa i Vua (1999) za sve striktno pozitivne lozove oblika (x,p;y,1-p).* Tanja linija predstavlja najbolju regresionu funkciju koja povezuje objektivne i a posteriori verovatnoće, deblja teorijsku predikciju teorije poverenja.




Slika 40b. *Regresione analize odnosa modeliranih a posteriori verovatnoća i empirijskih a posteriori verovatnoća za studiju Gonzalesa i Vua (1999) za sve striktno pozitivne lozove oblika (x,p;y,1-p).*




Slika 41a. *Predikcija nagiba i intercepta linearnih regresionih funkcija odnosa objektivnih verovatnoća i empirijskih a posteriori verovatnoća za studiju Gonzalesa i Vua (1999) za sve ne-negativne lozove oblika (x,p;0,1-p).*




Slika 41b. *Predikcija nagiba i intercepta linearnih regresionih funkcija odnosa objektivnih verovatnoća i empirijskih a posteriori verovatnoća za studiju Gonzalesa i Vua (1999) za sve striktno pozitivne lozove oblika (x,p;y,1-p).*

## 11.3 Eksperimenti izbora

U eksperimentima izbora ispitanicima prikazujema dva loza i od njih zahtevamo da između njih izaberu onaj koji bi radije odigrali. U literaturi u oblasti kognitivne psihologije i eksperimentalne ekonomije ne postoji veliki broj ovakvih studija. Dizajn eksperimenata izbora predstavlja notorno težak metodološki zadatak. Ne postoji koncenzus o tome koji tip eksperimentalnog nacrta najviše odgovara procesu selekcije bihejvioralnog modela odlučivanja. Modeliranje ovakvih eksperimenata,

zahvaljujući još nerešenim problemima koji se odnose na izbor odgovarajućeg psihometrijskog (odn. ekonometrijskog) modela, modela pod kojim bi adekvatno bile merene odgovarajuće teorijske funkcije (up. diskusiju u Sekciji 7.1), tek predstavlja izazov. Činjenica da se u ovoj metodologiji modeliraju isključivo izbori pojedinačnih ispitanika dalje komplikuje problematiku uvodeći probleme vezane za efekte individualnih razlika. Ovde predstavljamo dizajn dva originalna eksperimenta izbora čijim smo izvođenjem došli do skupova podataka na kojima smo testirali teoriju poverenja i kumulativnu teoriju izgleda.

## Eksperiment 3a

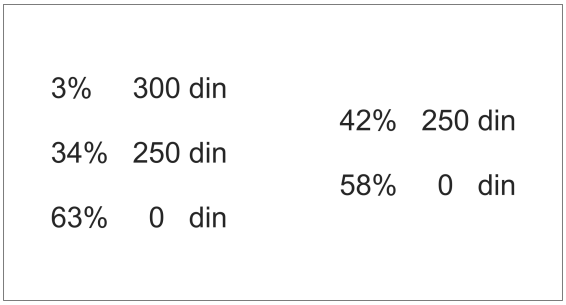
Dizajn eksperimenta 3a blisko prati dizajn koji je prvi put korišćen u studiji Vua i Gonzalesa iz 1996 (Wu & Gonzalez, 1996) sa standardnom metodologijom, i kasnije modifikovan u eksperimentima sa odlučivanjem u motornim akcijama (Wu, Delgado & Maloney, 2009). Eksperiment obuhvata 90 odluka između lozova sa tri i dva ishoda. Predstavljamo eksperimentalni dizajn samo u onoj meri koja je dovoljna za precizan opis eksperimenta; za detalje vezane za motivaciju ovakvog eksperimentalnog nacrta čitaoca upućujemo na studiju Vua i Gonzalesa (Wu & Gonzalez, 1996) i dodatno objašnjenje eksperimentalnog nacrta koji daju Vu, Delgado i Maloni (Wu, Delgado & Maloney, 2009). Dok studija Vua i Gonzalesa iz 1996. koristi ovaj nacrt motivisano, u cilju testiranja specifičnih hipoteza, naša studija, poput studije Vua, Delgada i Malonija iz 2009, upotrebljava ovaj nacrt kao prigodnu shemu za generisanje većeg broja rizičnih lozova u eksperimentu izbora.

### Metod

*Ispitanici.* Dvadeset i pet studenata, oba pola, I godine studija psihologije, Odeljenja za psihologiju Filozofskog fakulteta, Univerzitet u Beogradu, uzelo je učešća u eksperimentu ispunjavajući uslove za kurs psihologije u okviru programa nastave. Niko od ispitanika nije prethodno pohađao kurs teorije odlučivanja.

*Dizajn.* Lozovi korišćeni u eksperimentu 3a razvijeni su polazeći od pet osnovnih parova ishoda u dinarima: 10-60, 250-300, 470-520, 1330-1380 i 4940-4990. Razlika između dva ishoda u ovim parovima uvek iznosi 50 RSD. Na osnovu ovih parova ishoda formirana su dva loza za poređenje, od kojih prvi loz uvek sadrži tri ishoda, a drugi uvek dva. Prvi loz uključuje oba ishoda u paru i nulu, dok drugi loz uključuje samo niži od dva ishoda i nulu. Ovakva struktura „3-2 lozova“ varirana je kroz 18 različitih shema dodele verovatnoća ishodima na lozovima. Primer loza dat

je na sledećoj slici, a tabela ispod nje prikazuje način variranja verovatnoća kroz ishode na dva loza u paru - prvi sa tri, drugi sa dva ishoda. Pet osnovnih parova ishoda korišćeni su kao viši i niži ishod u shemi produkcije lozova; na taj način je konstruisano 90 parova lozova, 18 puta ponavljajući shemu dodele verovatnoća na pet osnovnih planova za par lozova. Lozovi su ispitanicima prikazani kao na slici 42.




Slika 42. *Primer stimulusa u eksperimentu 3a.* Jedan loz u paru uvek ima tri ishoda (250 RSD, 300 RSD i 0 u primeru na slici), a drugi sadrži samo niži od dva pozitivna ishoda sa prvog loza (250 RSD) i nulu. Shema variranja verovatnoća kroz lozove data je u tabeli 11.

Tabela 11. Shema variranja verovatnoće kroz ishode na lozovima u eksperimentu 3a. Osamnest shema dodele verovatnoća ishodima na lozovima je prikazano za I loz (tri ishoda) i II loz (dva ishoda). Brojevi u ćelijama tabele odgovaraju procentima sa kojima je dat odgovarajući ishod na lozu.

| **I loz** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
|---|---|---|---|---|---|---|---|---|---|
| Veći ishod | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| Manji ishod | 4 | 9 | 14 | 19 | 24 | 29 | 34 | 39 | 44 |
| 0 | 93 | 88 | 83 | 78 | 73 | 68 | 63 | 58 | 53 |
| **II loz** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| Manji ishod | 12 | 17 | 22 | 27 | 32 | 37 | 42 | 47 | 52 |
| 0 | 88 | 83 | 78 | 73 | 68 | 63 | 58 | 53 | 48 |

| **I loz** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** |
|---|---|---|---|---|---|---|---|---|---|
| Veći ishod | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| Manji ishod | 49 | 54 | 59 | 64 | 69 | 74 | 79 | 84 | 89 |
| 0 | 48 | 43 | 38 | 33 | 28 | 23 | 18 | 13 | 8 |
| **II loz** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** |
| Manji ishod | 57 | 62 | 67 | 72 | 77 | 82 | 87 | 92 | 97 |
| 0 | 43 | 38 | 33 | 28 | 23 | 18 | 13 | 8 | 3 |

*Procedura.* Lozovi su prikazivani ispitanicima na ekranu kompjutera kao na slici 42. Ispitanici su pritiskom na odgovarajuće tastere birali „*levi*“ ili „*desni*“ loz, označavajući koji bi od ponuđenih lozova u paru radije odigrali. Nije postojalo

nikakvo vremensko ograničenje da se odluka donese. Prikazivanje lozova u paru na levoj i desnoj poziciji na ekranu bilo je balansirano kroz ispitanike, a redosled prikazivanja 90 lozova potpuno randomizovan za svakog ispitanika. Jedna sesija eksperimenta 3a trajala je između 5 i 15 minuta.

Rezultati

Skup podataka do kojeg smo došli eksperimentom 3a čini 25 pojedinačnih skupova binarnih odluka između 90 parova lozova. Uz postavljanje odgovarajućeg psihometrijskog (ekonometrijskog) modela, moguće je fitovanje modela odlučivanja na individualne odluke ispitanika. Prvo su formulisani odgovarajući parametarski modeli kumulativne teorije izgleda i teorije poverenja. Model kumulativne teorije izgleda karakterišu dva parametra: eksponent stepene funkcije korisnosti, $\rho$, i parametar funkcije ponderisanja verovatnoća, $\gamma$. Model teorije poverenja definisan je kroz parametar stepene funkcije korisnosti, $\rho$, i parametar dekumulativne funkcije verovatnoća $S$, u oznaci $q$.

Pored ova dva parametra, oba modela uključuju još po jedan, koji je karakteristika odgovarajućeg psihometrijskog odn. ekonometrijskog modela koji koristimo. Osnovni problem u fitovanju modela teorija odlučivanja na binarne odluke ispitanika je u tome što su sve te teorije suštinski *deterministične*. Ovo je već diskutovano u sekciji 7.1. Za određenu kombinaciju vrednosti parametara, na primer, kumulativna teorija izgleda i teorija poverenja izračunavaju očekivane korisnosti lozova u skladu sa svojim jednačinama, i daju jednu, jedinstvenu predikciju odluke, tj. izbor onog loza koji odlikuje viša očekivana korisnosti. Empirijska realnost odlučivanja je, svakako, drugačija. Očekivati da će donosioci odluka deterministički konzistentno donositi odluke kao što to predviđaju teorije nije realno. Centralna pretpostavka u oceni modela bihejvioralnih teorija odlučivanja je onda sledeća: pretpostavljamo da kognitivni sistem poredi dva signala iz dva loza, čiji je „intenzitet“ određen njihovom očekivanom korisnošću, i pretpostavljamo da izračunavanje tog signala onda odlikuje određena greška $\epsilon$. U eksperimentu 3a, na primer, kognitivni sistem poredi očekivanu korisnost jednog loza sa tri ishoda (od kojih je treći uvek nula) i jedan sa dva ishoda (od kojih je drugi uvek nula). Uzmimo za primer ocenu teorije poverenja. Neka je očekivana korisnost prvog loza $EU(L_1) = p''_x \cdot u(x) + p''_y \cdot u(y)$, pošto ishod 0 ne utiče na očekivanu korisnost loza, i očekivana korisnost drugog loza $EU(L_1) = p''_x \cdot u(x)$. Pretpostavljamo da kognitivni sistem formira varijable $\psi_{L1} = EU(L_1) + \epsilon$ i $\psi_{L2} = EU(L_2) + \epsilon$, tako da odluku donosi na osnovu varijable $\Delta = \psi_{L1} - \psi_{L2}$: ako je $\Delta > 0$ bira loz $L_1$, a ako je $\Delta < 0$,

loz $L_2$. Pretpostavimo da je greška $\epsilon \sim N(0, \delta)$, odn. distribuirana normalno sa prosekom 0 i standardnom devijacijom $\delta$. Onda je

$$p_{L1} = 1 - \Phi(\Delta; 0, \delta) \tag{90}$$

gde je $\Phi$ kumulativna funkcija normalne distribucije, odn. verovatnoća da je $\Delta \leq 0$, a model dobija jedan dodatni parametar, standardnu devijaciju distribucije greške $\delta$. Onda funkciju verodostojnosti za model u celini možemo da formiramo na sledeći način:

$$L(R; \rho, q, \delta) = \prod_{i=1}^{n} \left( p_{L1}^{r_i} \cdot (1 - p_{L1})^{1-r_i} \right) \tag{91}$$

gde $R$ predstavlja $n$ odgovora subjekta: $r_i = 1$ ako je izabran loz $L_1$, $r_i = 0$ ako je izabran loz $L_2$, a vektor $[\rho, q, \delta]$ čine dva odgovarajuća parametra teorije poverenja i parametar psihometrijskog modela, $\delta$. Logaritmovanjem jednačine (91) dolazimo do klasične forme logaritmaske funkcije verodostojnosti koju onda maksimizujemo[88] Nelder-Midovom simpleks metodom. Ocena parametera za kumulativnu teoriju izgleda je potpuno ista, sa tom razlikom što se koriste odgovarajući parametri za izračunavanje očekivane korisnosti lozova: $[\rho, \gamma, \delta]$ . U modeliranju odlučivanja, pristup koji ovde biramo se naziva klasičnim ili fehnerijanskim; razne njegove modifikacije su u upotrebi (za prvu primenu ovog modela, up. Hey & Orme, 1994; za modifikaciju u kojoj je varijansa greške proporcionalna očekivanoj korisnosti, up. Wu, Delgado & Maloney, 2009; za matematički sličan pristup modelu odlučivanja u vizuelnoj percepciji razlika, up. Maloney & Yang, 2003, Knoblauch & Maloney, 2008). Čitalac može da razmišlja o ovom pristupu na sledeći način: deterministički model odlučivanja (model teorije poverenja ili model kumulativne teorije izgleda) čini „unutrašnji model“ čiji se parametri ocenjuju; ocena se vrši inkorporacijom determinističkog, sržnog modela u „spoljašnji“, stohastički model, koji je identičan modelu teorije detekcije signala sa jednakim varijansama i koji igra ulogu modela greške. Zbog sličnosti u terminologiji, upozoravamo da model koji predstavljamo *nije* stohastički model odlučivanja: stohatistički modeli odlučivanja pretpostavljaju distribucije verovatnoća nad vrednostima parametara samih modela odlučivanja; mi ovde ne razmatramo takve modele.

Tabela 12. Rezultati eksperimenta 3a. Tabela prikazuje ocene maksimalne verodostojnosti za parametre odgovarajućih modela, minimalnu vrednost negativne logaritamske verodostojnosti (engl. *negative log-likelihood*) za oba modela, vrednost *Voungovog testa* verodostojnosti za neugnježdene modele (objašnjenje u tekstu) i statističku značajnost ovog testa ('*' ukazuje na značajnost na nivou .05, '-' ukazuje na odsustvo značajnosti).

| | Kumulativna teorija izgleda | | | | Teorija poverenja | | | | Vuongov test | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sub | $\rho$ | $\gamma$ | $\delta$ | $min(nll)$ | $\rho$ | $q$ | $\delta$ | $min(nll)$ | *Vuong z* | p<.05 |
| 1 | 0 | 0 | 21.37 | 62.38 | 1.50 | 0.13 | 1375.38 | 49.04 | 2.58 | * |
| 2 | 0 | 0 | 5.59 | 62.38 | 1.16 | 0.17 | 65.45 | 38.01 | 3.49 | * |
| 3 | 0 | 1.90 | 0.08 | 49.47 | 0 | 4.19 | 0.06 | 50.63 | 0.76 | - |
| 4 | 0 | 0 | 5.44 | 62.38 | 1.31 | 0.15 | 601.93 | 53.76 | 2.08 | * |
| 5 | 0 | 0 | 0 | 60.72 | 0 | 2.83 | 0.04 | 60.63 | 0.21 | - |
| 6 | 0 | 0 | 6.90 | 62.38 | 1.26 | 0.14 | 137.30 | 32.72 | 3.85 | * |
| 7 | 0 | 2.96 | 0.40 | 61.34 | 1.35 | 0.02 | 13534.04 | 61.15 | 0.31 | - |
| 8 | 0 | 2.60 | 0.54 | 61.88 | 0.00 | 3.86 | 0.43 | 62.17 | 0.38 | - |
| 9 | 0 | 0 | 0 | 49.74 | 0 | 2.79 | 0.01 | 33.61 | 2.84 | * |
| 10 | 0 | 1.22 | 0.06 | 42.03 | 0 | 3.22 | 0.02 | 38.20 | 1.38 | - |
| 11 | 0 | 0.68 | 0.38 | 62.00 | 0 | 2.66 | 0.01 | 54.17 | 1.98 | * |
| 12 | 0 | 0 | 8.26 | 62.38 | 0.67 | 0.12 | 10.58 | 50.20 | 2.47 | * |
| 13 | 0 | 0 | 18.99 | 62.38 | 0.76 | 0.24 | 5.05 | 39.72 | 3.37 | * |
| 14 | 0 | 1.41 | 0.11 | 55.45 | 0 | 10.31 | 0.10 | 55.80 | 0.42 | - |
| 15 | 1.02 | 0 | 13.19 | 19.29 | 0 | 1.61 | 0.02 | 18.96 | 0.41 | - |
| 16 | 1.35 | 0.01 | 220.81 | 37.83 | 0.93 | 0.20 | 13.69 | 36.08 | 0.94 | - |
| 17 | 0 | 64.50 | 1.97 | 62.18 | 1.00 | 0.18 | 103.96 | 58.41 | 1.37 | - |
| 18 | 0 | 0 | 18.13 | 62.38 | 1.86 | 0.10 | 3328.52 | 26.02 | 4.27 | * |
| 19 | 0 | 1.36 | 0.20 | 60.30 | 0 | 87.30 | 0.20 | 60.57 | 0.37 | - |
| 20 | 0 | 0 | 7.30 | 62.38 | 1.31 | 0.14 | 158.69 | 31.70 | 3.92 | * |
| 21 | 0 | 0 | 11.89 | 62.38 | 1.34 | 0.13 | 238.17 | 34.26 | 3.75 | * |
| 22 | 1.04 | 0.01 | 19.14 | 33.06 | 0 | 2.04 | 0.02 | 31.75 | 0.81 | - |
| 23 | 0 | 0.67 | 0.15 | 59.87 | 0 | 8.70 | 0.18 | 60.14 | 0.37 | - |
| 24 | 1.10 | 0 | 50.83 | 47.43 | 0.25 | 0 | 0.42 | 46.57 | 0.66 | - |
| 25 | 1.42 | 0.02 | 257.70 | 27.68 | 0.92 | 0.21 | 9.57 | 28.85 | 0.76 | - |

Metodom maksimalne verodostojnosti ocenili smo parametre modela kumulativne teorije izgleda i teorije poverenja. Modeli su optimizovani za odgovore svakog ispitanika upotrebom Nelder-Midovog simpleks algoritma sa po 100 optimizacija za svakog ispitanika; odabrano je rešenje sa najnižom vrednošću negativne logaritamske verodostojnosti. Odlučili smo se za ovako veliki broj optimizacija pošto funkcije verodostojnosti modela odlučivanja predstavljaju u najmanju ruku teške optimizacione probleme. Funkcije verodostojnosti ovakvih modela uzimaju oblik „stepeničastih" funkcija čije površi odlikuje više „platoa" koji mogu da navedu konvergenciju optimizacionih procedura u tek lokalne minimume (up. Carbone & Hey, 2000). Simulacijama smo proverili funkcije verodostojnosti za modele i podatke eksperimenta 3a i uverili se da one uzimaju tu formu. Rešenja koja prikazujemo

smo odabrali tek pošto smo se uverili da je najveći broj od po 100 optimizacija za svakog ispitanika konvergirao u istim vrednostima parametara. Sva rešenja su zatim proverena kroz deset novih optimizacija algoritmom *simuliranog kaljenja* (engl. *simulated annealing*) koji garantuje konvergenciju u globalni minimum (ali je veoma zahtevan po pitanju kompjutacionog vremena) i prihvaćena tek posle verifikacije ovom procedurom. Tabela 12 sadrži rezultate, redom: ocene maksimalne verodostojnosti parametara kumulativne teorije izgleda i $min(nll)$ - minimalnu negativnu logaritamsku verodostojnost - za ovaj model, ocene maksimalne verodostojnosti parametara teorije poverenja i $min(nll)$, vrednost Vuongovog testa za neugnježdene modele i značajnost Vuonogovog testa. Pošto trenutno nemamo nikakvih uvida u to da li teorija poverenja generalizuje teoriju izgleda, ili obrnuto, nismo mogli da koristimo standardan test odnosa logaritamske verodostojnosti da bismo doneli odluku o tome koji model bolje fituje podatke. Vuongov test odnosa verodostojnosti se primenjuje i na neugnježdene modele (Vuong, 1989, prema Loomes, Moffat, & Sugden, 2002).

Činjenica da su vrednosti Vuongovog testa pozitivne za sve subjekte u eksperimentu 3a svedoči o tome da model teorije poverenja konsekventno fituje empirijske izbore bolje od modela kumulativne teorije izgleda; ipak, za samo 11 od 25 (odn. 44%) ispitanika je razlika između fita modela teorije poverenja i fita kumulativne teorije izgleda statistički značajna. Čak osam od 25 ispitanika, pod modelom teorije poverenja, odlikuju konveksne funkcije korisnosti u eksperimentu 3a; pod modelom kumulativne teorije izgleda, sklonost prema riziku odlikuje pet ispitanika, dok odluke ostalih najbolje opisuju funkcije korisnosti sa vrednošću eksponenta od nula - funkcije koje praktično ne prave nikakvu razliku između subjektivne ocene različitih vrednosti u nacrtu eksperimenta 3a. Teorija izgleda i teorija poverenja, dakle, otkrivaju sasvim različite karakteristike donosioca odluka u ovom eksperimentu.

### Diskusija

Ponovićemo još jednom da je modeliranje eksperimenata izbora relativno nova oblast proučavanja u kojoj još ni iz daleka ne postoji koncenzus oko toga koji eksperimentalni dizajn treba koristiti u procesu selekcije modela, niti koji psihometrijski (tj. ekonometrijski) model najbolje opisuje termin greške koji je neophodno inkorporirati u determinističke modele odlučivanja. Ovo ne treba da zakloni činjenicu da je, prema našoj analizi sprovedenoj upotrebom najjednostavnijeg i najčešće korišćenog modela greške, teorija poverenja fitovala

bolje podatke 44% ispitanika (11 od 25) u eksperimentu 3a od modela kumulativne teorije izgleda.

Ne treba da iznenađuju pojave vrednosti eksponenta stepene funkcije korisnosti od nula: ova vrednost ukazuje na to da je ispitanik donosio odluke tretirajući korisnosti svih alternativa kao podjednake, i ova pojava nije uopšte retkost u eksperimentima izbora (up. Blavatsky, 2011; pri tom treba držati na umu da je $u(0)=0$ po definiciji za oba modela koja ovde razmatramo). Vrednost od nula standardne devijacije $\delta$ modela greške ukazuje na determinističko donošenje odluka kod ispitanika koga odlikuje ova vrednost. Vrednost od nula za parametar funkcije ponderisanja verovatnoća ukazuje na ekstremnu formu precenjivanja niskih i potcenjivanja visokih verovatnoća - do mere u kojoj se one poistovećuju na jednom istom nivou. Sve ovde diskutovane vrednosti parametara - koje smo predstavili kao da imaju vrednosti nula - su zapravo *bliske* nuli u rezultatima; njihove realne razlike su minimalne, a praktične nikakve, pa smo dugačke eksponencijalne zapise zamenili nulama u tabeli 12.

## Eksperiment 3b

Eksperiment 3b je tipičan eksperiment izbora kao i eksperiment 3a, sa drugačijim eksperimentalnim dizajnom koji ovde predstavljamo. Kao i u slučaju eksperimenta 3a, teorijska motivacija za upotrebu eksperimentalnog dizajna je dublja i nećemo je diskutovati; eksperiment sličnog dizajna predstavljen je i motivisan u studiji sa odlukama motornog sistema Vua, Delgada i Malonija iz 2009 (up. dopunske materijale za eksperiment 2 u Wu, Delgado & Maloney, 2009). Eksperimentalni dizajn se bazira na principu da pri konstantnoj proporciji visine pozitivnih ishoda na dva ne-negativna loza ponuđena u izbornom paru, sa povećanjem visine pozitivnog ishoda na lozu, opada verovatnoća sa kojim je on ponuđen. Tako su svi stimulusi specifični za test paradoksa zajedničke proporcije koji su prvi sistematski testirali Kaneman i Tverski (Kahneman & Tversky, 1979).

### Metod

*Ispitanici.* Dvadeset i pet studenata, oba pola, I godine studija psihologije, Odeljenja za psihologiju Filozofskog fakulteta, Univerzitet u Beogradu, uzelo je učešća u eksperimentu ispunjavajući uslove za kurs psihologije u okviru programa nastave. Niko od ispitanika nije prethodno pohađao kurs teorije odlučivanja.

*Dizajn.* Svi lozovi korišćeni u eksperimentu 3b bili su oblika (*x,p;0,1-p*), odn. sadržali su jedan pozitivan ishod i nulu. Korišćene su dinarske vrednosti. Neka je loz $L_1$:(*x,p;0,1-p*), a loz $L_2$:(*y,q;0,1-q*). U eksperimentu 3b lozovi su generisani tako što je odnos u kome stoje visina ishoda $x$ na lozu $L_1$, i visina ishoda y na lozu $L_2$, $x/y$, variran sistematski kao 1.2, 1.6 i 2, i praćen obrnutim variranjem u odnosu verovatnoća osvajanja odgovarajućih ishoda, odn. $q/p$: 1.2, 1.6 i 2. Ako pustimo da $x$ varira kao: 100, 300, 500, 700, 900 i 1100, dolazimo do odgovarajućih vrednosti ishoda y pri odnosu $x/y$ = 1.2: 120, 360, 600, 840, 1080 i 1320. Ako pustimo sada $q$, verovatnoću osvajanja ishoda $y$ na lozu $L_2$, da varira kao: .15, .30, .45, .60, .75 i .90, sa odnosom $q/p$ = 1.2 dobijamo sledeće vrednosti za $p$, verovatnoću osvajanja ishoda na lozu $L_1$: .125, .25, .375, .5, .625 i .75. Princip generisanja lozova je taj da kako visina pozitivnog ishoda na jednom lozu raste u odnosu na drugi loz, tako opada verovatnoća sa kojom je on ponuđen. Dakle, polazeći od osnovnih vrednosti za ishod $x$ koje smo naveli, variranjem verovatnoća za osvajanje drugog ishoda, $y$, takođe kao što smo naveli, i prolazeći kroz tri nivoa variranja odnosa ishoda i verovatnoća (1.2, 1.6 i 2), moguće je generisati 108 različitih parova lozova. Svih 108 takvih parova lozova su prikazani ispitanicima u proceduri binarnog izbora. Prethodnim simulacijama pod eksperimentalnim dizajnom eksperimenta 3b, za veliki raspon parametara kumulativne teorije izgleda i teorije poverenja, uverili smo se da je pod ovakvim nacrtom moguća konzistentna ocena parametara modela koje testiramo.

*Procedura.* Eksperimentalna procedura je ista kao u eksperimentu 3a.

### Rezultati

Procedura selekcije modela je istovetna kao u eksperimentu 3a, uz tek neophodno prilagođavanje modela formi lozova koji su korišćeni u eksperimentu 3b. Tabela 13 predstavlja rezultate modeliranja eksperimenta 3b modelima kumulativne teorije izgleda i teorije poverenja.

### Diskusija

Vuongov test pokazuje da je retko kad moguće statistički razlikovati fitove kumulativne teorije izgleda od fitova teorije poverenja u ovom eksperimentu. Odluke tek četiri od 25 ispitanika (odn. 16%) bolje fituje teorija poverenja od kumulativne teorije izgleda (na nivou statističke značajnosti od .05), uprkos činjenici da pozitivna vrednost Vuongovog testa za sve ispitanike ukazuje na to da je model teorije poverenja konsekventno bliži empirijskim izborima. Dok su ocene stepene funkcije

korisnosti skoro perfektno linearne za sve ispitanike pod teorijom poverenja, ocene eksponenta ove funkcije su pod modelom kumulativne teorije izgleda nula, ili bliske nuli, za 12 od 25 ispitanika.

Tabela 13. Rezultati eksperimenta 3b. Tabela prikazuje ocene maksimalne verodostojnosti za parametre odgovarajućih modela, minimalnu vrednost negativne logaritamske verodostojnosti za oba modela, vrednost *Voungovog testa* verodostojnosti za neugnježdene modele i statističku značajnost ovog testa.

| | Kumulativna teorija izgleda | | | | Teorija poverenja | | | | Vuongov test | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sub | $\rho$ | $\gamma$ | $\delta$ | $min(nll)$ | $\rho$ | $q$ | $\delta$ | $min(nll)$ | *Vuong z* | p<.05 |
| 1 | 0 | 0 | 0 | 27.88 | 1.00 | 5.23 | 0 | 28.13 | 0.35 | - |
| 2 | 0 | 0 | 0 | 61.06 | 0.94 | 0.52 | 0.99 | 57.02 | 1.42 | - |
| 3 | 0 | 0 | 0 | 71.24 | 1.00 | 1.28 | 0.52 | 64.28 | 1.87 | * |
| 4 | 0.37 | 0 | 0.43 | 29.40 | 0.94 | 0.42 | 1.19 | 29.03 | 0.43 | - |
| 5 | 1.00 | 1.00 | 0 | 73.10 | 1.00 | 1.86 | 0.08 | 69.08 | 1.42 | - |
| 6 | 1.00 | 1.00 | 0 | 58.86 | 1.00 | 2.02 | 0.05 | 56.00 | 1.20 | - |
| 7 | 0 | 0 | 0 | 33.33 | 1.00 | 2.12 | 0.00 | 31.71 | 0.90 | - |
| 8 | 1.00 | 1.00 | 0 | 63.75 | 1.00 | 0.81 | 5.66 | 62.91 | 0.65 | - |
| 9 | 1.00 | 1.00 | 0 | 57.84 | 0.99 | 0.89 | 1.71 | 59.25 | 0.84 | - |
| 10 | 0 | 0 | 0 | 59.01 | 0.92 | 0.37 | 2.75 | 60.19 | 0.77 | - |
| 11 | 0 | 0 | 0 | 68.76 | 0.94 | 0.52 | 1.71 | 66.75 | 1.00 | - |
| 12 | 0 | 0 | 0 | 64.69 | 1.00 | 3.59 | 0 | 53.31 | 2.39 | * |
| 13 | 1.00 | 1.00 | 0 | 66.23 | 1.00 | 1.25 | 1.07 | 68.44 | 1.05 | - |
| 14 | 1.00 | 1.00 | 0 | 43.93 | 1.00 | 1.53 | 0.10 | 36.72 | 1.90 | * |
| 15 | 0.05 | 0.00 | 0.01 | 54.18 | 1.00 | 5.74 | 0 | 53.66 | 0.51 | - |
| 16 | 1.00 | 1.00 | 0 | 64.94 | 1.00 | 1.69 | 0.17 | 63.57 | 0.83 | - |
| 17 | 0 | 0 | 0 | 35.95 | 1.00 | 3.28 | 0 | 34.66 | 0.80 | - |
| 18 | 0 | 0 | 0 | 49.64 | 0.93 | 0.45 | 1.05 | 48.58 | 0.73 | - |
| 19 | 0.01 | 0 | 0.00 | 45.28 | 1.00 | 3.77 | 0 | 41.54 | 1.37 | - |
| 20 | 1.00 | 1.00 | 0 | 68.49 | 0.95 | 0.64 | 2.72 | 71.08 | 1.14 | - |
| 21 | 0 | 0 | 0 | 66.37 | 1.00 | 2.03 | 0.03 | 63.90 | 1.11 | - |
| 22 | 4.62 | 12.15 | 2.64e+15 | 74.44 | 1.00 | 6.41 | 0 | 73.69 | 0.61 | - |
| 23 | 0.22 | 0 | 0.18 | 51.79 | 0.95 | 0.39 | 3.51 | 51.87 | 0.20 | - |
| 24 | 0 | 0 | 0 | 24.77 | 1.00 | 5.08 | 0 | 15.92 | 2.10 | * |
| 25 | 0.11 | 0 | 0.04 | 43.51 | 0.93 | 0.38 | 1.73 | 43.53 | 0.10 | 0 |

Rezultati eksperimenata izbora koje smo ovde predstavili nisu konkluzivni u odnosu na pitanje selekcije adekvatnog deskriptivnog modela odlučivanja u uslovima rizika. Činjenica je da ne postoje izbori niti jednog ispitanika za kog teorija poverenja pokazuje lošiji fit od kumulativne teorije izgleda. S druge strane, u objašnjenju empirijskih izbora određenih ispitanika, teorija poverenja i kumulativna teorija izgleda grubo divergiraju u ocenama stepenih funkcija korisnosti. Ovo nije bio slučaj u analizi sudova o monetarnim ekvivalentima. Kada je moguće ustanoviti statistički značajno različit kvalitet fita odluka u eksperimentima 3a i 3b, prednost je isključivo na strani teorije poverenja, ali samo u 44% slučajeva

u eksperimentu 3a i tek 16% slučajeva u eksperimentu 3b. Sve ovde navedene rezultate treba uzeti uz dozu opreza: ponovimo, ne postoji koncenzus oko izbora „pravog“ modela greške za ocenu modela odlučivanja, još uvek nije jasno ni da li izvor neizvesnosti u diskretnim izborima treba uopšte tražiti u slučajnim greškama ili je on inherentan parametrima određenih modela (ili oba), i konačno, tek je otvoreno skoro uznemiravajuće pitanje o tome da li uopšte postoji reprezentativan subjekt samo jednog određenog modela odlučivanja - ili su empirijski izbori proizvod više strategija odlučivanja koje kognitivni sistem koristi uporedo.

# 12 Eksplanatorne strategije teorija odlučivanja

Prethodne analize u ovom poglavlju pokazale su da je moguća formulacija deskriptivne teorije odlučivanja u uslovima rizika sa sledećim osobinama: (*i*) njena teorijska konstrukcija je konstrukcija tipične racionalne, bejzijanske teorije koja odlikuje skoro sve modele razvijene u tradiciji racionalne analize u prethodnih dvadesetak godina; (*ii*) njena eksplanatorna moć je, ocenjena u tipičnim eksperimentalnim paradigmama, prema rezultatima standardnih statističkih testova, najmanje onolika kolika je eksplanatorna moć kumulativne teorije izgleda, koja je skoro koncenzusom prihvaćena kao deskriptivno najmoćnija teorija odlučivanja; (*iii*) prema našim rezultatima, teorija poverenja, kao sistematsko proširenje Viskuzijeve teorije perspektivne reference, može bar delimično da modelira prethodno nepoznate efekte u tipičnim eksperimentalnim paradigmama - efekte za koje kumulativna teorija izgleda nema eksplanatorne mehanizme. Ove zaključke treba prihvatiti ne gubeći iz vida sledeća ograničenja. Prvo, parametarska forma kumulativne teorije izgleda koja se najčešće koristi, i koja je korišćena u prethodnim eksperimentalnim analizama u ovom radu, izvesno nije odgovarajući model odlučivanja. Činjenica da eksperimentalni podaci otkrivaju sistematsko kršenje uslova homogenosti preferencija ima više posledica po ovaj popularan model. Najvažnija od njih se sigurno odnosi na stepenu funkciju korisnosti: nju je moguće tretirati samo kao aproksimaciju prave funkcije korisnosti. Drugo, aksiomatika kumulativne teorije izgleda ne obavezuje na upotrebu parametarske forme koja se najčešće koristi, koju smo mi koristili, i koju smo empirijski pobili. Međutim, iz aksiomatskog okvira kumulativne teorije izgleda nužno sledi egzistencija *jedne i jedinstvene funkcije ponderisanja verovatnoća.* Ovo je u gruboj suprotnosti sa našim eksperimentalnim nalazima koji ukazuju na to da je subjektivni tretman

verovatnoća fenomen koji najbolje opisuje familija linearnih funkcija (karakteristična za teoriju poverenja). Posle naših analiza, čini se da ova činjenica predstavlja najteži teorijski problem za kumulativnu teoriju izgleda; nju, verujemo, može empirijski da „pokrije“ samo fleksibilna forma inverzne-S funkcije ponderisanja verovatnoća, koja predstavlja *ad hoc* izbor i nije motivisana aksiomatskim okvirom same kumulativne teorije izgleda (npr. Prelecova forma *jeste* aksiomatizovana, ali zahvaljujući uvođenju dodatnih aksioma, up. Prelec, 1998). Treće, odnos prema referentnoj tački, odn. fenomen poznat kao averzija prema gubicima (za koji se sve više pokazuje da neopravdano nosi ovo ime), nema adekvatan aksiomatski tretman u Viskuzijevoj teoriji perspektivne reference, pa tako ni u teoriji poverenja. Parametar odnosa prema referentnoj tački je u model teorije poverenja ugrađen *ad hoc*, bez prethodnog aksiomatskog tretmana i analize, polazeći od danas široko prihvaćenog stava da je u pitanju nezavisna komponenta rizika. Zbog toga, model teorije poverenja koji smo koristili obuhvata funkciju korisnosti za dobitke i gubitke koja je potpuno identična funkciji korisnosti koja se koristi u modelu kumulativne teorije izgleda. Jedino što razlikuje dva modela koje smo poredili u eksperimentalnim testovima jeste pristup subjektivnom tretmanu verovatnoća.

Prethodna diskusija eksperimentalnih rezultata u našem pokušaju selekcije odgovarajućeg bihejvioralnog modela odlučivanja za mnoge bi verovatno predstavljala tek *tour de force* primenjene matematike u eksperimentalnoj psihologiji. To je neophodno. Diskutovati stanje u jednoj disciplini koja pretenduje na status prirodne nauke je nezamislivo bez diskusije primene matematičkog aparata u toj disciplini. Međutim, sama primena matematičkih modela u oblastima kognitivne psihologije koje su fokus naše rasprave nije ključna za našu diskusiju u celini; ključne su teorijske posledice koje proizlaze iz raznih formi uspeha i neuspeha primene odgovarajućih matematičkih modela. Matematičke, formalne teorije, počivaju na određenim teorijskim pretpostavkama. Posle njihove primene na bihejvioralne podatke, mi saznajemo više o prirodi suštinskih teorijskih koncepata, kao i o prirodi metodološkog, eksperimentalnog aparata koji koristimo. Diskusija teorijskih koncepata i načina na koji oni objašnjavaju eksperimentalne podatke je zato suštinska; ne sama primena matematičkog aparata, *već analiza posledica njegove primene*, rasvetljava najinteresantnije probleme u analizi racionalnosti saznanja.

## 12.1 Dispozicioni i reprezentacioni pojmovi u teorijskom objašnjenju odlučivanja

Teorija poverenja i kumulativna teorija izgleda koriste dve bitno različite eksplanatorne strategije u odnosu na problem odlučivanja u uslovima rizika i neizvesnosti. Ovo ćemo lakše razumeti uvođenjem dihotomije između (*i*) teorijskih eksplanatornih koncepata koje nazivamo *dispozicionim*, i (*ii*) teorijskih eksplanatornih koncepata koje ćemo nazvati *reprezentacionim*. Ove dve vrste eksplanatornih koncepata nalazimo u svim teorijama odlučivanja kao i u teorijama drugih kognitivnih fenomena. Iako su dispozicioni teorijski pojmovi dobro poznati i temeljno diskutovani u istoriji psihologije (Ryle, 1949/1975, Radonjić, 1967/94), mi ćemo precizno i eksplicitno objasniti tačno koje pojmove ćemo smatrati dispozicionim u kontekstu naše rasprave. Uzmimo za primer način na koji teorija izgleda objašnjava subjektivni tretman verovatnoća. Empirijski nalazi u odlučivanju uzimaju formu koja se objašnjava tek pretpostavkom da je subjektivni tretman verovatnoća drugačiji od onog koji implicira njihova objektivna reprezentacija. Subjektivni tretman verovatnoća je centralni fenomen koji motiviše razvoj deskriptivnih teorija odlučivanja. Kumulativna teorija izgleda postulira funkciju ponderisanja verovatnoća, sa osobinama (*i*) striktno monotonog rasta na intervalu od 0 do 1 i (*ii*) vrednostima 0 i 1 na odgovarajućim ekstremima skale kumulativnih verovatnoća. Parametarski model ove teorije koji se najčešće razmatra implicira inverznu-S funkciju u kojoj se smenjuju konveksni i konkavni regioni. Ako odaberemo neku funkcionalnu formu ponderisanja verovatnoća, kao što naša i mnoge druge studije biraju Prelecovu funkciju, na raspolaganju dobijamo slobodan parametar (ili dva, u zavisnosti od izbora funkcije), čija vrednost onda opisuje subjektivni tretman verovatnoća nekog određenog donosioca odluka. Postavimo sledeće pitanje: *zašto* funkcija ponderisanja verovatnoća uzima oblik koji uzima, i zašto određeni donosilac odluka vrši izbore u skladu sa određenom vrednošću parametra $\gamma$ ove funkcije? Kumulativna teorija izgleda zapravo nema odgovore na ova pitanja. Inverzni-S oblik ponderisanja verovatnoća nije ništa drugo do korisna empirijska generalizacija. Oblik inverzne-S funkcije ponderisanja verovatnoće jednostavno odgovara subjektivnom tretmanu verovatnoća koji može da inkorporira bitne eksperimentalne nalaze: Aleov paradoks, paradoks zajedničke proporcije, precenjivanje niskih i potcenjivanje visokih verovatnoća, efekat izvesnosti i efekat mogućnosti. Teorijsko, psihološko objašnjenje ovih fenomena u kumulativnoj teoriji

izgleda izostaje: kažemo da ljudi donose odluke u skladu sa inverznom-S funkcijom ponderisanja verovatnoća jer takva funkcija najbolje opisuje njihove empirijske izbore. Individue odlikuju izbori koji ih odlikuju zato što su njihove odluke najbolje objašnjene funkcijom ponderisanja verovatnoće sa određenom vrednošću parametra $\gamma$: ta vrednost je, onda, suštinska odlika nekog kognitivnog sistema. Očigledno, naučno objašnjenje odlučivanja koje odstupa od normativnih standarda se u kumulativnoj teoriji izgleda oslanja na *suštinske karakteristike samog donosioca odluka*: same osobine njegovog kognitivnog sistema su te koje analizi odlučivanja nameću funkcije poput inverzne-S. Smatraćemo *dispozicionim* sve teorijske koncepte koji objašnjavaju neko određeno ponašanje kao *posledicu inherentnih karakteristika kognitivnog sistema.* Upozoravamo da način na koji koristimo termin „dispozicioni pojmovi“ sada nema nikakve veze sa Karnapovom upotrebom ovog termina koja je diskutovana u II delu ove rasprave. Koncept averzije prema gubicima - posmatran van konteksta racionalne analize koju smo mi ponudili u ovom (V) delu naše rasprave - takođe je dispozicioni teorijski koncept. Koncept averzije prema riziku je dispozicioni koncept i u kumulativnoj teoriji izgleda i u teoriji poverenja (i svakako u teoriji očekivane korisnosti koja ga uvodi u savremene rasprave odlučivanja). Koncept neuroticizma u psihologiji uopšte je dispozicioni koncept u upravo opisanom smislu, kao što bi to bile i sve druge crte ličnosti. Teorija izgleda, kao teorija ograničene racionalnosti, u potpunosti predstavlja dispozicionu teorijsku strukturu u objašnjenju odlučivanja u uslovima rizika i neizvesnosti: njena eksplanatorna strategija se sastoji u lociranju eksplanatorne moći u *osobine samog donosioca odluka*, koje onda više nisu objašnjenje ničime do činjenicom da korespondiraju sa empirijskim nalazima, što određuje deskriptivni, ne normativni, karakter teorije. Jedini mogući način da se odstrani cirkularnost iz odnosa dispozicionog teorijskog koncepta i ponašanja na koje on nosi referencu, jeste da se *dispozicioni koncept razume kao esencijalna, kauzalno relevantna odlika samog sistema koji se proučava.* Inverzna-S funkcija je, dakle, nešto što je suštinski deo ljudske prirode.

Objašnjenje ponderisanja verovatnoća u teoriji poverenja, koje u ovoj teoriji samo uslovno nazivamo „ponderisanjem“ jer ona subjektivni tretman verovatnoća objašnjava kao vid bejzijanske inferencije, nije dispoziciono objašnjenje. Razlika je fina i suštinska: u teoriji poverenja, donosilac odluka jeste karakterisan subjektivnim, njemu svojstvenim parametrom koji izdvaja tačno jednu dekumulativnu funkciju verovatnoće osvajanja monetarnih ishoda *S*. Taj parametar je interna, subjektivna varijabla koji predstavlja opis jedne karakteristike datog kognitivnog sistema.

Međutim, taj parametar suštinski nije deo deskripcije procesa odlučivanja, već je *osobina reprezentacije okoline* koju je taj donosilac odluka izgradio kroz svoje prethodne interakcije u njoj. Teorijske koncepte koji objašnjavaju određeno ponašanje tako što ga dovode u vezu sa prethodno reprezentovanim karakteristikama problema adaptacije koji kognitivni sistem rešava nazvaćemo *reprezentacionim teorijskim konceptima.* Koncept dekumulativne funkcije verovatnoća realizacije monetarnih ishoda u teoriji poverenja je, dakle, reprezentacioni koncept. Koncept averzije prema gubicima posle racionalne analize koju smo predstavili u ovom poglavlju - koja pokazuje da on nužno sledi iz pretpostavljene strukture elementarnih ekonomskih interakcija - predstavlja reprezentacioni koncept. Dakle, teorijski koncept koji se na bihejvioralnoj ravni idenitifikuje kao jedinstven - poput averzije prema gubicima ili averzije prema riziku, jer se oba demonstriraju svojstvenim, jedinstvenim skupovima eksperimentalnih nacrta i procedura - može da ima i dispozicionu, i reprezentacionu intepretaciju. Ovaj mogući dualitet u interpretaciji bihejvioralno jedinstvenog fenomena je od najvećeg značaja za diskusije koje slede. Da ilustrujemo jednim poznatim primerom u psihologiji uopšte, teorijski koncept kompleksa u klasičnoj psihoanalizi, emotivno-kognitivni konglomerat koji okuplja oko sebe nesvesne asocijacije i u čijem semantičkom jezgru se nalazi potisnuta reprezentacija nekog prethodnog događaja iz lične istorije individue, predstavlja reprezentacioni teorijski koncept (iako, pod uslovom da je psihoanalitička teorija tačna, taj koncept posreduje u određenju mnogih drugih konstrukata koji bi svojoj prirodi bili dispozicioni).

Pod kauzalnom analizom ponašanja, reprezentacioni koncepti imaju direktnu interpetaciju kao kauzalni uzročnici, koja se čini više opravdanom nego u slučaju dispozicionih koncepata. Zašto ispitanici subjektivno tretiraju niske verovatnoće kao više nego što jesu, a visoke kao niže nego što jesu? Teorija poverenja, sledeći Viskuzijev bejzijanski formalizam, odgovara da je takav empirijski fenomen direktna posledica primene normativne bejzijanske inferencije polazeći od prethodno reprezentovanih verovantoća. Sled je očigledno kauzalan: osmotreno ponašanje je takvo kakvo je, jer ga kognitivni sistem planira optimalno u odnosu na relevantne reprezentacije takve kakve su. U slučaju kumulativne teorije izgleda, kao i ma kog drugog dispozicionog objašnjenja, stvari nisu tako jasne. Osmotreno ponašanje koje ukazuje na subjektivni tretman verovatnoća različit od objektivnog jeste kauzalna posledica primene funkcije ponderisanja verovatnoća, ali nije jasno čega je kauzalna posledica upravo ta funkcija. Zašto su visoke verovatnoće potcenjene,

a niske precenjene? Zato što funkcija ponderisanja verovatnoća precenjuje niske, a potcenjuje visoke verovatnoće. Zašto funkcija ponderisanja uzima ovaj inverzni-S oblik, onda? Zato što je to suštinska osobina empirijskog, realnog donosioca odluka. Zašto ljudi pridaju višu (negativnu) korisnost gubicima nego dobicima odgovarajuće vrednosti? Zato što su ljudi averzivni prema gubicima. Zašto bi ljudi pokazivali averziju prema gubicima, što je normativno neopravdano? Zato što je to odlika njihovog kognitivnog sistema, na šta ukazuju robusni empirijski nalazi; da bi se izbegla cirkularnost, onda, averzija prema gubicima mora da postane esencijalna, kauzalno relevantna odlika kognitivnog sistema. Međutim, averzija prema gubicima, videli smo, može biti i posledica sasvim racionalnog odnosa prema distribuciji resursa koji su organizmu na raspolaganju u ekonomskim interakcijama.

Ostavimo za sada po strani strukture podataka koje teorija poverenja očigledno može da inkorporira, a kumulativna teorija izgleda ne; vrlo verovatno je moguće konzistentno teorijsko proširenje kumulativne teorije izgleda koje bi joj omogućilo da uklopi efekte koji nastaju kao posledica kršenja uslova homogenosti preferencija. Naša eksperimentalna analiza pokazala je da, u većini slučajeva, statistički ne možemo da razlikujemo opis donosioca odluka pod teorijom poverenja i pod kumulativnom teorijom izgleda. Teorijska analiza koju smo sada predstavili pokazuje da ove dve teorije koriste različite, skoro simetrične, eksplanatorne strategije u objašnjenju bitnih fenomena odlučivanja u uslovima rizika. Da li je moguće da postoje dva, teorijski različita, modela koji podjednako uspešno objašnjavaju fenomene sa *istim* bihejvioralnim referencama? Upravo smo postavili jedno od dva suštinska pitanja za našu analizu racionalnosti saznanja; odgovor predstavljamo u VI delu naše rasprave.

## 12.2 Environmentalna i ekološka racionalnost odlučivanja

Teorija poverenja, posmatrana kao proširenje Viskuzijeve bejzijanske teorije perspektivne reference, omogućava nam da odlučivanje u uslovima rizika posmatramo kao proces environmentalne i ekološke racionalnosti. Prethodno smo specifikovali precizno kada ćemo za neku teoriju reći da je teorija environmentalne racionalnosti: očigledno, teorija koju ovde diskutujemo ima precizno onu formu na koju smo mislili. Racionalna bejzijanska inferencija bazirana na *a priori* verovatnoćama poreklom iz interakcija sa strukturom objektivnog okruženja predstavlja „kičmu“ teorije poverenja. Samo u fazi formiranja verovanja - fazi koju specifikuje upravo teorija poverenja - proces odstupa od potpuno tipične slike

racionalne analize. U toj fazi, videli smo, neophodno je postulirati (*i*) određena skaliranja u percepciji verovatnoće (primena Lusovog aksioma izbora), (*ii*) procese izračunavanja određenih, čisto subjektivnih parametara (relativne entropije *a priori* distribucija za relevantne lozove), te (*iii*) specifičnu formu reprezentacije okoline. Sve što u teoriji poverenja preostaje posle ovih intervencija u teorijskoj strukturi jeste struktura jedne kognitivne teorije environmentalne racionalnosti. U procesu formiranja verovanja nema ničega što bi moglo biti okarakterisano kao iracionalno ili ograničeno racionalno; razlika u odnosu na sve diskusije racionalnih modela do sada je u tome što nedostaju normativni kriterijumi za proces formiranja verovanja kakav teorija opisuje. Mi se ovde nećemo upuštati u analizu normativne adekvatnosti procesa formiranja verovanja u teoriji poverenja; za sada, jasna je deskriptivna validnost ovog procesa, dok s druge strane nije jasno šta bi u njemu bilo moguće kritikovati kao normativno neadekvatno (i u odnosu na koje standarde).

Teorija poverenja zadovoljava i Gigerencerov koncept ekološke racionalnosti. Njena relativno složena formalna struktura ne otkriva odmah ovu činjenicu: nisu li Gigerencerove ideje o ekološkoj racionalnosti bazirane na tvrdnji da kognitivni sistem ima na raspolaganju jednostavne, ne nužno tačne ili optimalne, heuristike kojima rešava probleme odlučivanja i slične? Dok se struktura teorije poverenja ne oslanja ni na jednu heuristiku eksplicitno, ono što je čini sličnom teorijama ekološke racionalnosti jeste *kompjutaciona jednostavnost* koja se odnosi na centralne procese koje ona postulira. Ekspozicija teorije odlučivanja u uslovima rizika uvek obiluje formalizmima i tehničkim detaljima; to za cilj ima održavanje jasne teorijsko-konceptualne veze sa prethodnim razvojima u ovoj, za društvene nauke, solidno matematički proučenoj problematici. Ipak, kao posledica formalne strukture teorije poverenja javlja se bejzijanski kognitivni mehanizam baziran na Dirišle-multinomijalnoj inferenciji koja je u svojoj osnovi *jednostavan proces aditivne prirode*: teško je zamisliti jednostavniju formu bejzijanske inferencije od one koju pretpostavlja teorija poverenja. Praktično svi kompjutacioni resursi neophodni za izračunavanje prema teoriji poverenja svode se na sabiranje i normalizaciju (primena Lusovog aksioma); izračunavanje relativne entropije je još najsloženiji kompjutacioni proces koji model obuhvata. U tom smislu, teorija poverenja možda predstavlja najelegantnije ponuđeno matematičko rešenje u nizu bihejvioralnih modela odlučivanja: (*i*) teorijskom konstrukcijom se, kroz Viskuzijevu teoriju perspektivne reference, u potpunosti uklapa u kontinuitet koji je povezuje sa radom Sevidža, fon Nojmana i Morgnešterna, dok (*ii*) sa druge strane nastavlja tradiciju

racionalne analize u psihologiji - ali tako što počiva na jednostavnim kompjutacionim mehanizmima.

U narednom delu naše rasprave razvijamo argumente koji vode ka našim konačnim zaključcima o problemu racionalnosti saznanja. Da li se naša kritika Andersonove racionalne teorije pamćenja (up. 7.3) odnosi i na strukturu teorije poverenja, teorije koja putem reprezentacionih teorijskih koncepata objašnjava mnoštvo bihejvioralnih fenomena odlučivanja u uslovima rizika, ne pretpostavljajući kompjutaciono suviše složene kognitivne procese? Pokazaćemo, u narednim redovima, da se u određenom smislu takva kritika - kritika prema kojoj se eksplanatorni, kauzalni lanac neke kognitivne teorije zatvara u krug koji povezuje sredinu i kognitivni sistem koji saznaje u njoj - neizbežno odnosi na *svaki* pokušaj racionalne analize kognitivnih funkcija. Međutim, otićićemo korak dalje, i pokazati kako iz odnosa stukture racionalnih teorija, podložnih takvoj kritici, i konceptualne strukture teorija ograničene racionalnosti - poput teorije izgleda - zapravo sledi jedan dublji uvid u našu mogućnost izgradnje kognitivne teorije uopšte, bila ona racionalna ili ne. Metateorijski koncepti dispozicionih i reprezentacionih pojmova u strukturi psiholoških teorija koje smo predstavili u ovom delu rasprave biće od ključnog značaja u formulisanju tog argumenta. S druge strane, diskusija paradigme racionalne analize, koju planiramo da sprovodemo primenom koncepata *teorije igara* - i to upravo kako se oni koriste u evolucionoj biologiji i bihejvioralnoj ekologiji, discplinama čiji koncepti svejedno motivišu racionalne analize u psihologiji - otvoriće nam uvid u ograničenja analize racionalnosti saznanja iz još jedne, verujemo, ključne perspektive.

# Deo VI

# Racionalnost uma i racionalnost psihološke teorije uma

U sekciji 13, predstavljamo tri argumenta koja grade naš opšti, konačni zaključak o naučnom statusu koncepta racionalnosti saznanja: *racionalnost saznanja nije naučni pojam*, i kao takav treba da bude isključen iz naučne diskusije kognitivnih fenomena.

Eksplanatorni mehanizmi kognitivnih teorija mogu da budu bazirani na reprezentacionim ili dispozicionim pojmovima - razliku između ove dve vrste teorijskih pojmova u psihološkom objašnjenju uveli smo u prethodnoj sekciji. Tako, kognitivne teorije mogu da budu reprezentacione ili dispozicione prema karakteru objašnjenja koje nude - ili da u svojoj eksplanatornoj strukturi sadrže obe vrste pojmova. Teorije onda mogu biti čisto reprezentacione prirode, čisto dispozicione prirode, ili mešovite u odnosu na upotrebu dve vrste teorijskih pojmova. Takođe, ukoliko određeni koncept neke teorije objašnjava precizno određen domen bihejvioralnih podataka, kao što npr. koncept funkcije ponderisanja verovatnoće objašnjava tačno određen skup fenomena u teoriji izgleda, tu teoriju nazivamo dispozicionom u odnosu na taj domen bihejvioralnih podataka; slično

razumemo i reprezentacione teorije. Naš prvi argument tvrdi sledeće: ako postoji skup bihejvioralnih fenomena u kome je nemoguće razlikovati reprezentacionu od odgovarajuće dispozicione kognitivne teorije, a prethodna eksperimentalna analiza u V delu rasprave i formalna analiza u Prilogu A pokazuju da takav skup bihejvioralnih fenomena izvesno postoji - bar dok je reč o odlučivanju u uslovima rizika - *sledi da postoji skup relevantnih bihejvioralnih fenomena u kome nije moguće razlikovati teorije environmentalne racionalnosti od teorija ograničene racionalnosti.* Implikaciju podržava činjenica da teorije ograničene racionalnosti nužno moraju da sadrže dispozicione teorijske pojmove - jer nešto što nije poreklom u stimulusima (sredini) mora da objašnjava odstupanje od optimalnih sudova i odluka - dok teorije environmentalne racionalnosti moraju da sadrže reprezentacione pojmove, jer nešto mora da pruži organizmu prethodne informacije o sredini na osnovu kojih će on odrediti optimalan bejzijanski odgovor na tu sredinu. Ovo za direktnu posledicu ima to da pojam racionalnosti saznanja gubi naučno relevantno značenje: kroz dve teorije, semantički različite, cirkuliše ista informacija, što za posledicu ima samo dislociranje eksplanatorne moći kroz različite eksplanatorne strukture - ako te strukture u krajnjoj liniji uvek imaju iste bihejvioralne reference. Pitanje racionalnosti saznanja gubi na značaju i postaje samo pitanje arbitrarne odluke o deskriptivnom jeziku koji ćemo izabrati da govorimo o empirijskim fenomenima koje raspravljamo.

Prethodne diskusije već su nam omogućile da razumemo da kognitivni sistem u odnosu na neka jedinstvena stanja svoje okoline sebi može da postavi više ciljeva izračunavanja. Tih ciljeva izračuvanja, teorijski, može da bude i beskonačno mnogo: baš kao što je statistik bilo šta što je izračunato iz podataka (iako nije svaki statistik podjednako koristan za ostvarenje nekog cilja), bilo koje kognitivno izračunavanje iz istih podataka ostvaruje neku kognitivnu funkciju (iako neće svaka ovako definisana kognitivna funkcija biti korisna za ostvarenje *relevantnih* ciljeva). Međutim, kognitivni sistem koji, suočen sa određenim podacima iz svoje okoline, ne zna tačno koji kognitivni cilj sebi treba da postavi u odnosu na te podatke, opravdan je u pokušaju da svojim kognitivnim izračunavanjima zadovolji uporedo više ciljeva, pripremajući tako sebe za adaptaciju na upravo optimalan način. Ovakav pristup analizi problema kognitivne adaptacije vodi nas u formulisanje koncepta kognitivne strategije, koji se formalizuje kroz pojam mešovite strategije teorije igara. U mešovitim strategijama, igrač bira svoje akcije u datom okruženju samo sa određenom verovatnoćom: njega ne odlikuje izbor karakteristične akcije u

nekoj situaciji, već tek distribucija verovatnoće izbora određenih akcija. Naš drugi argument onda pokazuje da ako kognitivni sistemi zaista koriste mešovite strategije u izboru kompjutacionih ciljeva koje teže da zadovolje u učenju, odlučivanju, rezonovanju, suđenju i drugim zadacima, fenomeni ograničene racionalnosti mogu da se jave kao posledice racionalnosti koja odlikuje ma kog racionalnog donosioca odluka. Iz ovoga direktno sledi da ma koja analiza racionalnosti saznanja koja nije uzela u obzir ovu mogućnost - a nama nije poznata nijedna koja jeste - nije mogla da donese zaključak o racionalnosti kognitivne funkcije na koju se odnosila. Međutim, dalji razvoj našeg argumenta otkriva još ozbiljniji problem u analizi racionalnosti saznanja: faktori koji nisu kognitivne prirode, poput subjektivnih funkcija korisnosti, mogu da na suštinski način utiču na formu koju uzimaju kognitivne funkcije ako ih kognitivni sistem koristi u rešavanju problema okoline koju odlikuje fundamentalna nepredviljivost upravo one prirode koja odlikuje strategijske interakcije kao predmet teorije igara.

Konvencionalnost je osnovna osobina simboličkih sistema koje proučavaju kognitivna psihologija, semotika i lingvistika. *Intrepretacija* značenja određenog znaka je takođe uvek arbitrarna u nekoj meri, odn. *zavisna od konvencije pod kojom* (ili konteksta u kome) *se odvija.* Postoje ubedljivi empirijski argumenti o tzv. *kvazi-regularnoj prirodi* (Rogers & McClelland, 2004) organizacije našeg konceptualnog sistema. Naš treći argument polazi od činjenice da elementarne simboličke kognitivne funkcije - poput funkcije interpretacije znaka u kategorizaciji i učenju kategorija - fundamentalno jesu konvencionalne prirode. Postavljajući problem interpretacije znaka formalno, pokazuje se da njegovo rešenje može da leži u domenu teorije igara, odn. analize konvencionalnosti kakvu je predložio još Dejvid Luis (Lewis, 1969), ali po ceni da faktori koji nisu kognitivne prirode određuju formu interpretacije svih znakova koje suštinski ne odlikuje jednoznačnost - odn. svih zaista konvencionalnih znakova uopšte. Naš argument onda pokazuje da *nije moguće postaviti kriterijume normativne adekvatnosti za ovaj skup kognitivnih funkcija*, što dodatno dovodi u pitanje naučni status koncepta racionalnosti saznanja.

# 13 Kritika koncepta racionalnosti saznanja

Redom diskutujemo (13.1) ekvivalenciju dispozicionih i reprezentacionih kognitivnih teorija, (13.2) koncept kognitivnih strategija, njihovu stabilnost i relativnost u značenju racionalnosti saznanja koja se javlja posle primene

ovih koncepata u racionalnoj analizi i (13.3) pitanje racionalnosti elementarnih simboličkih funkcija. Konačno, posle opsežne kritike koncepta racionalnosti saznanja, u VII delu pružamo skicu šire slike o kognitivnoj psihologiji *kao prirodnoj nauci* u kojoj nema mesta pojmovima kao što je racionalnost saznanja.

## 13.1 Argument I: Posledice ekvivalencije dispozicionih i reprezentacionih kognitivnih teorija

U V delu naše rasprave demonstrirali smo mogućnost jedne teorije odlučivanja koja po svojoj strukturi odgovara teorijama iz tradicije racionalne bejzijanske analize. Teorija poverenja, kako smo pokazali u V delu i elaborirali u Prilogu A, objašnjava više robusnih empirijskih nalaza koji se interpretiraju kao znaci ograničene racionalnosti u donošenju odluka. Takođe, pokazali smo da najčešće nije moguće statistički razlikovati objašnjenja bihejvioralnih podataka (npr. ocena monetarnih ekvivalenata, eksperimenti 2a i 2b) koja pružaju teorija poverenja, kao racionalna teorija, i kumulativna teorija izgleda, kao teorija ograničene racionalnosti. Sada je vreme da izvedemo sve teorijske konsekvence koje slede iz ovog empirijskog nalaza.

*Ekvivalencija dispozicionih i reprezentacionih kognitivnih teorija.* Ekvivalencija teorije poverenja i kumulativne teorije izgleda je, kao što pokazujemo u Prilogu A, veoma složeno matematičko pitanje. Dva modela odlučivanja mogu da budu ekvivalentna, ali egzaktni uslovi pod kojima oni daju iste evaluacije rizičnih lozova su veoma kompleksni i zavise od vrednosti većeg broja parametara oba modela i distribucije verovatnoća na lozovima koji se posmatraju. Naš eksperimentalni rad u V delu je pokazao da ova dva modela *praktično* nije moguće razlikovati u standardnim bihejvioralnim testovima. Teorija poverenja je, doduše, u stanju da makar delom objasni neke strukture podataka koje kumulativna teorija izgleda trenutno ne objašnjava, ali smo ostavili otvorenim pitanje o tome da li bi minimalna modifikacija kumulativne teorije izgleda (odn. uvođenje više funkcija ponderisanja verovatnoća) dovela do toga da i ona obuhvati te strukture podataka. Robusni empirijski efekti koji su korišćeni u debati o racionalnosti u oblasti odlučivanja mogu da se objasne kako jednim tako i drugim razmatranim modelom. Neposredan zaključak je sledeći: *verovatno nije moguća eksperimentalna diferencijacija racionalnog modela odlučivanja i modela odlučivanja ograničene racionalnosti.* Pošto semantika dve diskutovane teorije otkriva dve potpuno različite

slike o prirodi odlučivanja kao kognitivne funkcije, otvara se pitanje o tome *kako, uopšte, donosimo zaključak o racionalnosti kognitivnih funkcija u situacijama poput ove*? Pitanje motiviše sledeću egzaktniju analizu.

Vraćamo se ponovo centralnoj tvrdnji o racionalnosti saznanja: *kognitivno je racionalan subjekt* S *čije ponašanje* $B$ *konzistentno svedoči o tome da on dela u skladu sa svojim verovanjima* $\psi$, *kako bi ostvario svoje ciljeve* $G$ *u nekoj sredini* $E$. Formalno dajemo prethodnu tvrdnju zapisom $S(\psi|G,E)\rightarrow B$, koji čitamo kao: „$S$ primenjuje kognitivne funkcije u skladu sa svojim verovanjima $\psi$, *ako su dati ciljevi* $G$ *i sredina* $E$, da bi proizveo ponašanje $B$"; kažemo da $B$ *zadovoljava* $G$ *u* $E$ ako ponašanje $B$ vodi ka ispunjenju ciljeva $G$ u sredini $E$. Kognitivno je racionalan, onda, svaki kognitivni akter $S$, koji posle primene kogntivnih funkcija na verovanja $\psi$ proizvodi ponašanje $B$, $S(\psi|G,E)\rightarrow B$, tako da $B$ zadovoljava $G$ u $E$. U slučaju odlučivanja u uslovima rizika, tvrdnja da $B$ zadovoljava $G$ u $E$ je jednoznačno određena konsekvencama po donosioca odluka koji odigrava rizične lozove koje je odabrao u skladu sa onim modelom odlučivanja koji za njega važi: objektivno, nezavisno merilo za to da li njegovi opservabilni izbori zadovoljavaju $G$ u $E$ jeste vrednost koju on osvaja tokom ekonomskih interakcija izraženih u formi izbora između rizičnih lozova. Formalizam koji smo uveli omogućava nam da napravimo egzaktnu razliku između racionalnih bejzijanskih teorija koje odlikuje environmentalna racionalnost, poput teorije poverenja, i teorija ograničene racionalnosti, poput kumulativne teorije izgleda. Diskutujemo prvo formu teorija ograničene racionalnosti.

U izrazu $S(\psi|G,E)\rightarrow B$, primetimo da su verovanja kognitivnog aktera, $\psi$, izvesno neka funkcija sredine u kojoj njegovo ponašanje $B$ treba da bude delotvorno u odnosu na ciljeve $G$, odn: $\psi = f_D(\phi), \phi \in E$; ovaj izraz formalizuje sve funkcije koje povezuju parametre sredine $E$ sa određenim subjektivnim verovanjima o njima. Oznaka $f_D$ za ovakve funkcije se dakle odnosi na sve funkcije kao što su psihofizičke, ili funkcije korisnosti, ili funkcije ponderisanja verovantoća u teorijama odlučivanja. Slovo D u oznaci ove funkcije služi tome da je označi kao funkciju neke dispozicione kognitivne teorije; slično, uskoro ćemo upoznati funkciju $f_R$ u reprezentacionim teorijama. Pošto ovde govorimo o strukturi teorija ograničene racionalnosti, funkcija $f_D$ predstavlja neku subjektivnu transformaciju parametara sredine $E$ koja odslikava karakteristike samog kognitivnog aktera $S$: na primer, stepen njegovog pesimizma ili optimizma, što se ponekad uzima kao semantička osnova za interpretaciju funkcije ponderisanja verovatnoće u kumulativnoj teoriji izgleda (Wakker, 2010).

Za razliku od strukture teorija ograničene racionalnosti, u strukturi bejzijanskih teorija, koje odlikuje environmentalna racionalnost, funkcija koja povezuje subjektivna verovanja $\psi$ sa stanjima okoline $E$ uzima sledeću formu: $\psi = f_R(\phi, \phi'), \phi \in E, \phi \in E'$; indeks R u imenu funkcije $f_R$ sada govori o tome da se radi o funkciji određene reprezentacione kognitivne teorije. Funkcija $f_R$ preslikava *dve klase* objekata, ($a$) objektivna stanja sredine $E$ u kojoj se $S$ nalazi, $\phi$, i (b) subjektivna verovanja koja se odnose na opažanje sredine $E$, u oznaci $E'$, od strane $S$: $\phi'$, na aktualna subjektivna verovanja $\psi$. Na primer, u teoriji poverenja, *a posteriori* verovatnoće koje donosilac odluka koristi u evaluaciji rizičnih lozova ($\psi$) su funkcija objektivnih stanja sredine tj. verovatnoća datih na lozovima ($\phi$) i prethodno formiranih verovanja o tim verovatnoćama tj. *a priori* verovatnoća ($\phi'$).

U teorijama ograničene racionalnosti, ponašanje $B$ se označava kao neadekvatno („ograničeno racionalno“) iz očiglednog razloga što funkcija koja povezuje subjektivna verovanja i odgovarajuća stanja sredine *nije funkcija identiteta*, odn. $\psi = f_D(\phi), \psi \neq \phi$; pošto funkcija $f_D$ tako dobija semantiku *subjektivne interpretacije* sredine $E$, koja je opet funkcija nekih internih osobina $S$, jasno je da proizvedeno $B$ ne može da zadovoljava $G$ u $E$. U teorijama ograničene racionalnosti subjektivna interpretacija sredine je nužno onaj faktor koji u funkciji *inherentnih osobina* aktera $S$ dovodi do neadekvatnosti ponašanja i daje dispozicioni karakter celoj teoriji. Racionalne teorije bazirane na strukturi Bejzove teoreme i environmentalnoj racionalnosti, umesto da direktno koriguju verovanja aktera $S$ u funkciji njegovih inherentnih osobina, koriguju samu sredinu $E$ u funkciji njenih prethodnih stanja koja su bila dostupna i jesu reprezentovana od strane aktera $S$. Funkcija $\psi = f_R(\phi, \phi'), \phi \in E, \phi \in E'$ ukazuje na to da adekvatnost ponašanja $B$ mora da se prosuđuje u odnosu na neku novu, *subjektivno konstruisanu sredinu* $E''$, koja sledi posle korekcije sredine $E$ verovanjima o nekim njenim drugim (prethodnim) stanjima $E'$. Teorije ograničene racionalnosti i racionalne teorije tako uvode dve različite ali simetrične korekcije u semantici fenomena koje pokušavaju da objasne: dok teorije ograničene racionalnosti „fiksiraju“ sredinu $E$, ne dopuštajući da akter $S$ dela u odnosu na neku konstruisanu, subjektivnu sredinu $E''$, ali variraju njegovu subjektivnu ocenu parametara te sredine $E$, racionalni bejzijanski modeli „variraju“ sredinu $E$, forsirajući aktera $S$ da dela u odnosu na subjektivnu, konstruisanu sredinu $E''$, ali „fiksiraju“ racionalni proces formiranja verovanja subjekta $S$, održavajući racionalnost njegovih kognitivnih funkcija u odnosu na efekte njegovog ponašanja $B$ u odnosu na $G$ u subjektivnoj $E''$.

Uzmimo sada da je ponašanje $B$ proizvedeno po modelu neke dispozicione teorije ekvivalentno ponašanju proizvedenom po modelu neke reprezentacione teorije. U interpretaciji jedne teorije, ono će ukazivati na ograničenu racionalnost; u interpretaciji druge, na environmentalnu racionalnost. Očigledno je da ako govorimo o istom ponašanju, a prethodni eksperimentalni rad i proces selekcije modela (V deo) ukazuju na to da zaista govorimo o ponašanju koje podjednako mogu da generišu jedna i druga vrsta teorije, mora biti da su funkcije $f_D$ i $f_R$ *jedna ista funkcija.* To znači da je $f_R(\phi, \phi') = f_D(\phi), \phi \in E, \phi \in E'$. Iz ovoga direktno slede dva zaključka. Prvi je da funkcija formiranja verovanja u dispozicionoj teoriji, $\psi = f_D(\phi)$, mora da bude *složenija* od odgovarajuće funkcije u reprezentacionoj teoriji. Ovo sledi iz toga što funkcija formiranja verovanja u reprezentacionoj teoriji, kao funkcija dve varijable, *koristi više informacija od odgovarajuće funkcije u dispozicionoj teoriji.* Ako razmislimo sada o razlikama između linearne funkcije ponderisanja verovatnoća u teoriji poverenja i Viskuzijevoj teoriji, s jedne, i nelinearnoj funkciji ponderisanja rangova verovatnoća u kumulativnoj teoriji izgleda i drugim modelima zavisnim od ranga, videćemo da je prethodno izneto tvrđenje zadovoljeno. Možda ovo, tek, objašnjava začuđenost nekih istraživača nad neobičnim, složenim oblikom funkcije ponderisanja verovatnoća u kumulativnoj teoriji izgleda. Drugi zaključak koji direktno sledi iz ekvivalencije dve klase funkcija formiranja verovanja jeste da, u reprezentacionim teorijama, verovanja $\phi$ o aktualnoj sredini $E$, verovanja $\phi'$ o nekim drugim, prethodnim stanjima sredine, u oznaci $E'$*, moraju da budu regularno povezana.* Ovo opet sledi iz ekvivalencije dve funkcije formiranja verovanja, $f_D$ i $f_R$ : ako funkcija dispozicione teorije nosi iste informacije koje nosi funkcija dve promenljive reprezentacione teorije, mora da postoji regularnost koja povezuje promenljive nad kojima je definisana funkcija reprezentacione teorije - jer drugačije funkcija dispozicione teorije ne bi mogla da kompresuje tu informaciju upotrebom jedne varijable manje[89]. Ono što sledi iz formalne analize koje smo upravo predstavili jesu *formalni uslovi pod kojima će odgovarajuće dispozicione i reprezentacione teorije biti izjednačene u svojim bihejvioralnim referencama*: dispoziciona teorija mora da koristi složeniju funkciju formiranja verovanja o relevantnim stimulusima, a reprezentaciona teorija mora da koristi funkciju formiranja verovanja koja počiva na regularnoj vezi između onoga što bi mogli biti aktualni stimulusi i onoga što se pretpostavlja da bi stimulusi mogli biti uopšte. Tek ako su oba formalna uslova zadovoljena, dispoziciona teorija ima funkciju formiranja verovanja koja može da kompresuje (i tako reprezentuje) višak informacija koji nosi reprezentaciona

teorija environmentalne racionalnosti; u suprotnom, dispoziciona teorija ni pod kakvim uslovima ne može da opiše ponašanje aktera uspešno koliko odgovarajuća reprezentaciona teorija - jer, jednostavno, ova druga koristi više informacija. Ako se vratimo sada našem kritičkom pregledu debate o racionalnosti, videćemo da su ovi uslovi zadovoljeni makar za najpoznatije slučajeve racionalne analize i/ili teorije ograničene racionalnosti. Funkcija ponderisanja verovatnoća u teoriji izgleda daleko je složenije forme od linearne koju uzima ponderisanje verovatnoća u teoriji poverenja. U teoriji poverenja, verovatnoća određenog ishoda regularno je povezana sa *a priori* verovanjem o toj verovatnoći - upravo preko visine samog ishoda. Da ovi uslovi ne važe, dve teorije nikada ne bi mogle da konvergiraju u isti domen opservabilnog ponašanja. Andersonova racionalna analiza pamćenja polazi od pretpostavke da će verovatnoća pojave neke reči biti regularno povezana sa verovatnoćom njene prethodne upotrebe. Ako su formalni uslovi koje diskutujemo - i za koje priznajemo da su postavljeni kao široki, labavi uslovi da bi mogli da obuhvate rezonovanje o najrazličitijim kognitivnim teorijama - uvek zadovoljeni, *onda je uvek moguće konstruisati odgovarajuću reprezentativnu teoriju za datu dispozicionu teoriju i vice versa.* Prethodnim se ne implicira se da je proces konstrukcije jednostavan, ali se implicira da je bar u principu moguć.

Dok dve vrste psihološkog objašnjenja, dispoziciono („ograničeno racionalno“) i reprezentaciono („environmentalno racionalno“), objašnjavaju istu klasu bihejvioralnih fenomena, pitanje o tome da li je kognitivni sistem (ili podskup njegovih funkcija koje se objašnjavaju) racionalan ili ograničeno racionalan *nema smisla kao naučno pitanje* - zbog toga što *odgovor na njega zavisi samo od izbora deskriptivnog jezika koji ćemo koristiti.* Pokazali smo tačno kako se kreće eksplanatorna moć kroz jednu i kroz drugu eksplanatornu, teorijsku strukturu: dok je teorije ograničene racionalnosti lociraju u nestandardne procese formiranja subjektivnih verovanja i odstupanja od normativnog objašnjavaju na taj način, teorije environmentalne racionalnosti je dislociraju tako što nameću promenu normativnog kriterijuma, tako da inače racionalan proces formiranja verovanja vodi ka ponašanjima koja su prema novom kriterijumu adekvatna. Informacija ostaje konstantna i ista - jer oba objašnjenja konvergiraju u isti domen bihejvioralnih podataka; dva deskriptivna jezika, jezik ograničene racionalnosti i jezik environmentalne racionalnosti, razmenljiva su u objašnjenju kognitivnih funkcija svaki put kada je moguće konstruisati dispozicionu i reprezentacionu teoriju sa istim bihejvioralnim referencama u relevantnom domenu.

*Posledice po strukturu kognitivne teorije: ponovo o stimulusima, odgovorima i neopservabilnim subjektivnim stanjima.* U sledećoj digresiji uz našu prvu kritiku pojma racionalnosti saznanja želimo da skrenemo pažnju na određene promene u standardnoj psihološkoj analizi odnosa između stimulusa, reakcija i subjektivnih stanja koje se nameću posle ustanovljavanja ekvivalencije dispozicionih i reprezentacionih kognitivnih teorija. Generalnost argumenata koji slede zavisi od generalnosti same mogućnosti da se za odgovarajuću dispozicionu teoriju uvek konstruiše bihejvioralno ekvivalentna reprezentaciona teorija - mogućnosti koju diskutujemo odmah posle ove digresije.

Vratimo se na sam početak savremene kognitivne psihologije, u 1959. godinu, kada je objavljena čuvena kritika Skinerovog programa proučavanja verbalnog ponašanja od strane Noama Čomskog (Chomsky, 1959/1967). Kritika Čomskog uzima se u istoriji kognitivne psihologije kao markantan trenutak u kome je izvršena metodološka „provala“ kognitivističkih hipotetičkih konstrukata i subjektivnih stanja kroz kapiju skoro hermetički zatvorene eksperimentalne analize ponašanja svojstvene američkom bihjeviorizmu. U kritici koju je Čomski ponudio 1959. godine, izdvajaju se, grubo, dva argumenta kroz koje on napada bihejvioristički program objašnjenja simboličkih kognitivnih funkcija - konkretno, učenja, razumevanja i upotrebe jezika. Prvi argument, koji kritikuje Skinerove pokušaje egzaktnog određenja stimulusa i reakcija, karakterističan je za ranija od jedanest poglavlja koliko rad sadrži. Drugi argument, daleko više diskutovan u istoriji kognitivne psihologije, sastoji se u uvidu Čomskog o tome da statističke regularnosti i sličnost - na koje svodi sve pokušaje bihejviorističke teorije da objasni simbolička, verbalna ponašanja - ne mogu da objasne razumevanje složenih rečeničkih struktura kao što to može da učini interna kognitivna rekonstrukcija procesa generisanja određene rečenice. Prvi, metodološki deo kritike Čomskog je onaj koji nas ovde interesuje. U suštini, ta kritika je usmerena ka Skinerovom shvatanju da se stimulusi, kao delovi fizičkog sveta koji, prema bihejviorističkoj teoriji, kontrolišu ponašanje, i odgovori, kao delovi ponašanja, izdvajaju zahvaljujući *zakonomernim odnosima* u kojima se nalaze. Ova kritika se ne odnosi isključivo na simboličke funkcije i jezik (iako su svi primeri koje Čomski koristi vezani za analizu verbalnih stimulusa i reakcija). U II poglavlju kritike iz 1959, Čomski piše:

„*Pojmovi stimulusa, odgovora i potkrepljenja su relativno dobro definisani u odnosu na eksperimente sa pritiskanjem poluge i druge, restriktivne na sličan način. Pre nego što možemo da ih proširimo na ponašanje u realnom životu, moramo*

*da se suočimo sa određenim poteškoćama. Pre svega, moramo da odlučimo da li ćemo svaki fizički događaj na koji organizam može da reaguje zvati stimulusom u određenoj prilici, ili samo one na koje organizam zaista i reaguje; slično, moramo da odlučimo da li ćemo svaki deo ponašanja nazvati odgovorom, ili samo onaj koji je na zakonomeran način povezan sa stimulusima.*„

i dalje,

„*Mi možemo* [...] *da nastavimo da održavamo zakonomernost odnosa između stimulusa i odgovora samo ako ih lišimo njihovog objektivnog karaktera. Tipičan primer kontrole* [*ponašanja*, prim.aut] *stimulusa za Skinera bio bi odgovor na muzičko delo izrazom Mocart ili na sliku odgovorom holandsko. Za ove odgovore se tvrdi da su „pod kontrolom ekstremno suptilnih osobina" fizičkog objekta ili događaja.* [...] *Skiner bi mogao samo da kaže da je svaki od ovih odgovora pod kontrolom neke druge stimulusne osobine fizičkog objekta. Ako pogledamo crvenu stolicu i kažemo crveno, odgovor je pod kontrolom crvenila stimulusa; ako kažemo stolica, on je pod kontrolom kolekcije osobina* [...] *i slično za ma koji odgovor.* [...] *Pošto su osobine na raspologanju za naša pitanja (imamo ih onoliko koliko nesinonimnih deskriptivnih izraza u našem jeziku, šta god to značilo tačno), možemo da objasnimo široku klasu odgovora u terminima Skinerove funkcionalne analize identifikacijom sa kontrolišućim stimulusima. Ali reč stimulus je izgubila svu svoju objektivnost u ovakvoj upotrebi. Stimulusi više nisu deo spoljašnjeg fizičkog sveta; oni su odgurani nazad u organizam. Mi identifikujemo stimulus onda kada čujemo odgovor.*"

(citirano prema Chomsky, 1958/1967; naš prevod). Čomski kroz više primera u radu nastavlja da demonstrira koliko Skinerova ideja o „zakonomernoj vezi" između stimulusa i odgovora nije dovoljno restriktivna, te da je objektivno značenje stimulusa potisnuto činjenicom da saznajemo šta je stimulus tek kada shvatimo šta je bio odgovor na njega i *vice versa*, povezujući ih regularno tek *post hoc*. Iz naše prethodne analize ekvivalencije reprezentacionih i dispozicionih kognitivnih teorija, međutim, sledi da je Čomski izvesno grešio kritikujući Skinerovo shvatanje da stimuluse i odgovore (reakcije) možemo da izdvojimo kao jedinice u eksperimentalnoj analizi ponašanja samo ako ustanovimo da su oni regularno povezani - jer uopšte ne postoji drugi način da ustanovimo šta su, u određenom kontekstu, stimulusi, a šta njima korespondentne reakcije. Vratimo se ekvivalenciji reprezentacionih i dispozicionih kognitivnih teorija iz koje sledi $f_R(\phi, \phi') = f_D(\phi), \phi \in E, \phi \in E'$, odn. da su funkcije formiranja verovanja reprezentacionih i dispozicionih teorija iste

funkcije (ako i kada dva tipa teorija objašnjavaju iste podatke). U reprezentacionim teorijama, interpretacija „*nekog dela okoline*“ (fizičkog sveta, kod Skinera) $\phi$, kao stimulusa, *uopšte nije tačna interpretacija*; tačna interpretacija je ona koja uzima u obzir korekciju tog $\phi$ kroz neko $\phi'$ odn. kroz funkciju $f_R(\phi, \phi')$, dakle tek pošto neka već postojeća subjektivna stanja intervenišu u formiranju verovanja. Iz svega prethodno diskutovanog, još od kako smo u ovu raspravu uveli prve Remzijeve ideje, sledi da ocenu subjektivnih verovanja možemo da dobijemo kroz analizu ponašanja uz određene minimalne teorijske pretpostavke; dakle, ono što je *zaista bilo stimulus*, odn. subjektivno verovanje na osnovu kojeg je donet neki sud i formirano određeno ponašanje, saznajemo tek na osnovu analize ponašanja koje osmatramo. Simetrično, karakter ponašanja - kao „ograničeno racionalnog“ ili „environmentalno racionalnog“ - videli smo, određuje deskriptivni jezik koji biramo, a samim tim *definicija stimulusa koju smo odabrali*: ukoliko biramo jezik neke dispozicione teorije, stimulus je *de facto* ono što je dato eksperimentalnom procedurom (deo Skinerovog „fizičkog sveta“), dok ako biramo jezik neke reprezentacione teorije, stimulus postaje subjektivno uverenje koje je funkcija „objektivnog stimulusa“ i drugih subjektivnih verovanja (što je ekvivalentno stimulusu koji je Čomski opisao kao da je vraćen *u organizam*). Drugim rečima: ako promenimo naše pretpostavke o načinu na koji su stimulusi i odgovori regularno povezani, menjamo osobine onoga što određujemo kao stimuluse i reakcije u kontekstu eksperimentalne analize ponašanja.

Analize kao što je prethodna ukazuju na teorijski zaključak koji se, najverovatnije, neće mnogo dopasti pristalicama establišmenta savremene kognitivne psihologije: najstrože govoreći, jedina moguća forma kognitivne teorije uopšte je *konstruktivistička*, teorija koja poput autopoietičke teorije Maturane i Varele u teorijskoj biologiji pravi samo relativnu razliku između organizama i njihovih sredina, a forme njihovih interakcija stavlja u prvi plan, kao prve i suštinske objekte analize. Ako je moguće da postoji više regularnih povezanosti (modela) između *de facto* istih („*fizički istih*“, „*objektivno istih*“) stimulusa i reakcija, zahvaljujući, kao što smo videli, fleksibilnim načinima da se objasni formiranje subjektivnih verovanja u kontekstu analize ponašanja, onda je zapravo jedina kompletna tvrdnja o kognitivnim funkcijama koje se proučavaju u određenoj eksperimentalnoj situaciji ona koja tvrdi *klasu ekvivalencije svih takvih modela.* Međutim, ovakav pristup problemu analize kognitivnih fenomena zahteva potpunu promenu načina na koji se u kognitivnoj psihologiji razmišljalo u XX veku i na koji se razmišlja danas. Implicitna pretpostavka svake analize kognitivnih funkcija jeste da postoji *tačno*

*jedan, određen model* koji povezuje određene stimuluse sa određenim ponašanjima (ili klase stimulusa sa klasama ponašanja); mi smo pokazali da to nije tačno, i uskoro ćemo pokušati da pokažemo da pluralizam teorijskih modela odnosa *istih* stimulusa i reakcija ima još opštije važenje. Ako je to tako, kognitivna psihologija bi trebalo da prihvati *pluralizam mogućih subjektivističkih objašnjenja ponašanja*, i da kao formu *naučnog objašnjenja kognitivnog sistema ponudi upravo formulisanje klase ekvivalencije modela koji nose iste bihejvioralne reference.* Da je Čomski 1959. uopšte pomislio da bi to mogla biti odlika kognitivne psihologije kao nauke, umesto što je, po svemu sudeći, bio implicitno predan modelu „jednog objašnjenja ponašanja", verovatno ne bi organizovao svoju kritiku Skinera oko tvrđenja koja smo diskutovali. S druge strane, Skiner sam verovatno nije pomišljao na nešto drugo do na ideju da stimuluse i reakcije povezuju *jedinstvene* regularne veze. Bernštajnov problem, problem inverznog inženjeringa predeterminisanog sistema, sistema sa velikim brojem redundantnih stepeni slobode, sistema koji jedinstvene, složene probleme može da reši na više načina i koji nam u bihejvioralnoj analizi ne daje dovoljno informacija da identifikujemo način na koji je je ciljni problem rešio u nekom konkretnom kontekstu, problem je *iste forme* kao problem koji nastaje kada shvatimo da postoje različiti, ekvivalentni modeli odnosa *de facto* istih stimulusa i reakcija.

*O generalnosti ekvivalencije teorija ograničene i teorija environmentalne racionalnosti.* Kada kažemo da su neka dispoziciona teorija, poput kumulativne teorije izgleda, i neka reprezentaciona teorija, poput teorije poverenja, ekvivalentne, mi iznosimo jednu veoma jaku tvrdnju. U najstrožem smislu reči, dve teorije odlučivanja će biti ekvivalente ako i samo ako one predviđaju isto uređenje relacije preferencije za ma koji odabrani skup lozova. Dakle, ukoliko njihove funkcije očekivane korisnosti lozova - funkcije koje mere na intervalnim ili racio-skalama - rekonstruišu istu ordinalnu skalu relacije preferencije, teorije su ekvivalentne. U Prilogu A pokazujemo da su kumulativna teorija izgleda i teorija poverenja izvesno ekvivalentne u određenom domenu lozova u kome obe realizuju teoriju očekivane korisnosti, tj. u domenu u kome se njihove predikcije ne razlikuju od predikcija te teorije. Posledice ekvivalencije dva tipa teorija po analizu racionalnosti saznanja koje smo mi upravo izneli nemaju mnogo smisla ako se ekvivalencija dispozicionih i reprezentacionih ustanovljava na takvom domenu. Ekvivalencija ovakve dve teorije postaje interesantna samo ako se ona ustanovljava u domenu koji već sadrži bihejvioralne fenomene koje očekivana korisnost ne može da objasni, jer

tek u tom domenu teorija poverenja i kumulativna teorija izgleda objašnjavaju iste podatke ali pod potpuno različitim interpretacijama ograničene i environmentalne racionalnosti. Prilog A takođe pokazuje da je u domenu van onog u kome se realizuje teorija očekivane korisnosti ekvivalencija dve teorije moguća, ali pod veoma složenim uslovima. S druge strane, naše eksperimentalne analize u V delu pokazuju da na podacima koje odlikuju odstupanja od predikcija teorije očekivane korisnosti statistički nije moguće razlikovati objašnjenje koje nudi teorija poverenja od objašnjenja koje nudi kumulativna teorija izgleda. U kojoj meri je naše prethodne zaključke moguće generalizovati?

Dispoziciona i reprezentaciona teorija *ne moraju da budu striktno ekvivalentne na celom domenu podataka* koji je interesantan za donošenje zaključaka poput naših da bi oni važili. Sasvim je dovoljno da se ustanovi njihova ekvivalencija na makar nekom podskupu tog domena (dakle, van domena koji već objašnjava teorija očekivane korisnosti). Naravno, od suštinske je važnosti izvestan nivo konzistencije u javljanju ekvivalencije između dva različita modela. Ovo kažemo svesni činjenice da ona može da bude i posledica učešća termina greške koji je nemoguće ukloniti iz ma kog bihejvioralnog modela. Činjenica da je statistički jedva moguće razlikovati predikcije modela dve teorije koje smo uporedili u V delu - kako u analizi podataka pojedinačnih ispitanika, tako i u analizi uprosečenih rezultata - bi trebalo da predstavlja dovoljno ubedljiv argument.

Postavljamo sada sledeće, značajnije pitanje: *da li je uvek moguće konstruisati ekvivalentnu reprezentacionu teoriju za datu dispozicionu teoriju, i obrnuto*? Postoji relativno trivijalan i jednostavan argument koji pokazuje da je to uvek moguće. Pošto u kompjutacionoj kognitivnoj psihologiji ne postoji jasan kriterijum selekcije slobodnih parametara koji se odnose na subjektivne, neopservabilne konstrukte, čak i poštujući princip parsimoničnosti, mi možemo da proširimo skup parametara nekog modela proizvoljno - dok ne izjednačimo domen njegovih predikcija sa domenom predikcija nekog drugog modela. Koliki je stepen interpretabilnosti novouvedenih parametara zavisiće praktično sam od mašte teoretičara; za mali broj parametara koji predstavljaju neko nužno proširenje u ovakvoj operaciji izjednačavanja modela, verovatno ćemo uvek biti u stanju da pronađemo interpretaciju. U svakom slučaju, ako zapostavimo semantički aspekt (interpretabilnost) teorije, tako ćemo uvek biti u stanju da jednu teoriju redukujemo na drugu i *vica versa*, pa zaključak o mogućoj ekvivalenciji dispozicionih i reprezentacionih teorija dobija status *a priori* tvrdnje. Sa uvođenjem složenih hipotetičkih konstrukata i subjektivnih verovanja

u teoriju kognitivne psihologije, savremena KKP je na neki način zaista otvorila Pandorinu kutiju - zlo koje je usledilo će principijelno, uvek onemogućavati proces selekcije teorija različitih formi i interpretacija zahvaljujući nedovoljno restriktivnom prostoru neopservabilnih parametara koji su na raspolaganju u modeliranju. U drugim oblastima kognitivne psihologije ovaj problem je izražen podjednako kao i u odlučivanju, oblasti u kojoj mi vodimo ovu diskusiju. Lari Barsalou je tako još 1990. pokazao da modele primeraka i prototipova u teoriji konceptualnog sistema nije moguće striktno razlikovati - na osnovu ma kakve empirijske evidencije (Barsalou, 1990).

Restriktivniji način da se ovo pitanje diskutuje jeste taj da se zahteva da dve teorije - dispoziciona i reprezentaciona - rade se istim brojem slobodnih parametara. To je slučaj u analizi odnosa kumulativne teorije izgleda i teorije poverenja. U slučaju ove restriktivnije analize situacije se komplikuje zbog toga što sada znamo da su funkcije formiranja verovanja u dispozicionim teorijama ($\psi = f_D(\phi)$) izvesno algoritamski složenije od funkcija formiranja verovanja u njima ekvivalentnim reprezentacionim teorijama ($\psi = f_R(\phi, \phi')$). Međutim, dokle god postoji regularan odnos između objektivnih stanja sredine koja specifikuje problem na koji se teorije odnose, $\phi$, i subjektivnih stanja na koja se oslanja reprezentaciona teorija, $\phi'$, ta regularnost će omogućavati funkciji $f_D(\phi)$ dispozicione teorije da kompresuje višak informacija sadržanih u $f_R(\phi, \phi')$ kroz odgovarajuću matematičku formu i tako je efektivno simulira; cena te simulacije, naravno, izražava se brojem slobodnih parametara u modelu dispozicione teorije. Verujemo da je kriterijum ekvivalencije koji zahteva da dispoziciona teorija menja samo formu funkcije formiranja verovanja, a zadržava isti broj slobodnih parametara, dovoljno fer kriterijum da bi mogao da se zahteva od svih budućih analiza ovog tipa; ovaj kriterijum treba ublažiti samo posle egzaktnog dokaza da nijedna funkcija sa istim brojem slobodnih parametara neće biti u stanju da u dispozicionom modelu simulira eksplanatorne mehanizme odgovarajućeg reprezentacionog modela.

## 13.2 Argument II: Kognitivne strategije i njihova stabilnost

U sekciji 8. detaljno smo razmatrali jednu mogućnost koja u paradigmi racionalne analize - i debati o racionalnosti, uopšte - nije prethodno diskutovana. Sugerisali smo da se kognitivni sistem, shvaćen kao prirodni kompjutacioni sistem, *u nekim problemima adaptacije, pred istim environmentalnim stanjima suočava*

*sa izračunavanjima koja treba da optimizuju ponašanje u odnosu na više ciljeva paralelno.* Sada je naš cilj da ovu mogućnost teorijski i formalno egzaktnije proučimo i ukažemo na sve, verujemo, veoma važne, posledice koje ona implicira.

*Tretman različitih kompjutacionih ciljeva u odnosu na ista stanja sredine.* Vratićemo se diskusiji dva kognitivna problema koja smo iskoristili kao ilustracije za mogućnost paralelnog zadovoljavanja više kognitivnih ciljeva u sekciji 8, problemu kauzalnog učenja i odlučivanja u uslovima rizika.

Podsetimo se prvo strukture modela kauzalnog učenja koji Perales i Šenks nazivaju hibridnim pseudonormativnim modelom: $HPN = \gamma\Delta P' + (1-\gamma)p'_c$, (Perales & Shanks, 2007). Ovaj složeni model kauzalnog učenja, očigledno, predstavlja ponderisanu sumu doprinosa oceni asocijativne ili kauzalne veze koju izračunavaju dva podmodela tj. dve komponente: model probabilističkog kontrasta, $\Delta P$, i model kauzalne moći, $p_c$. Perales i Šenks definišu mešoviti statistički model nad ponderisanim verzijama modela probabilističkog kontrasta i kauzalne moći kao što je objašnjeno u sekciji 7.2. Kao što se vidi iz forme modela, doprinos jedne ili druge komponente određena je vrednošću parametra $\gamma$, tako da se doprinosi pondera $\gamma$ i $1-\gamma$ sabiraju do jedan, pa se ponderi u modelu ponašaju kao verovatnoće. Podsetimo se karaktera ovog mešovitog modela: s jedne strane, u njemu učestvuje probabilistički kontrast, koji je normativna mera intenziteta kovarijacije, tj. asocijativnog odnosa, i model kauzalne moći, koji je normativna mera ocene intenziteta kauzalnog odnosa pod teorijom kauzalnih modela. Svaki sistem koji je izračunao kauzalnu moć izračunao je i probabilistički kontrast, tako da uvođenje hibridnog pseudonormativnog modela ne podiže mnogo pritisak na kompjutacione resurse sistema; suština modela je da objasni sud o intenzitetu odnosa između varijabli u eksperimentu kauzalnog učenja koji bi doneo kognitivni sistem kao *kompozitni sud* koji u određenoj proporciji potiče iz rešavanja jednog, i u određenoj proporciji iz rešavanja drugog kompjutacionog problema. Ne zaboravimo: i ocena intenziteta asocijativnog odnosa, i ocena kauzalne moći, donose se na osnovu potpuno istih podataka - tj. u situaciji u kojoj kognitivni sistem koristi *de facto* ista environmentalna stanja za ocenu različitih parametara sredine kojoj se prilagođava. Statistički testovi pokazuju da je eksplanatorna moć ovog mešovitog modela veoma visoka u poređenju sa drugim modelima u oblasti kauzalnog i asocijativnog učenja. To ne sme da čudi: model kombinuje eksplanatornu moć dva dobro poznata i proverena modela relevantnih kognitivnih funkcija, i sasvim je očekivano da će imati eksplanatornu moć veću od pojedinačnih eksplanatornih moći svojih komponenata.

Vratimo se sada na kratko diskusiji odlučivanja u uslovima rizika. U V delu, diskutujući razvoj teorije poverenja, u jednačinama (59-61) za koje smo naglasili da su od posebnog značaja, pokazali smo da taj bejzijanski model u krajnoj oceni očekivane korisnosti rizičnog loza uzima sledeću formu (up. jednačinu (61)):

$$EU(x, p_x; y, p_y) = \frac{N}{N+100} EU'(x, p'_x; y, p'_y) + \frac{100}{N+100} EU(x, p_x; y, p_y) \qquad (92)$$

Već je poznato značenje oznake $N$ - stepena poverenja, u teoriji poverenja, koji je proporcionalan relativnoj entropiji *a priori* verovantoća ishoda koje nosi rizičan loz. Poznato je i poreklo konstante sa vrednošću 100: ona uvodi korespondenciju skale procenata u opisu rizičnih lozova sa „veličinom uzorka“ na kojoj se bazira proces bejzijanske inferencije. Naglasimo, još jednom, da je fiksiranje skale procenata jedino što omogućava jednoznačnu ocenu parametara ovakvog modela. Vreme je da to precizno objasnimo, pošto se sada u raspravi, iz teorijskih razloga, vraćamo originalnoj formi evaluacije lozova koju je predstavio Viskuzi 1989:

$$EU(x, p_x; y, p_y) = \frac{\gamma}{\gamma+\xi} EU'(x, p'_x; y, p'_y) + \frac{\xi}{\gamma+\xi} EU(x, p_x; y, p_y) \qquad (93)$$

gde su $\gamma, \xi$ *slobodni parametri teorije perspektivne reference*, koji redom odgovaraju stepenu u kome sudu o očekivanoj korisnosti lozova doprinosi komponenta očekivane korisnosti izračunata na osnovu *a priori* verovatnoća, $EU'(x, p'_x; y, p'_y)$, i stepenu u kome tom sudu doprinosi komponenta očekivane korisnosti izračunata na osnovu objektivnih verovatnoća, kakve su date na lozu, odn. $EU(x, p_x; y, p_y)$. Vidimo da se efekat parametara $\gamma, \xi$ takođe ponaša kao verovatnoća u Viskuzijevom modelu, upravo kao što se efekat izraza $\frac{N}{N+100}$ i $\frac{100}{N+100}$ ponaša u teoriji poverenja. Suštinski doprinos teorije poverenja je specifikacija načina da se odrede vrednosti parametara $\gamma, \xi$ u Viskuzijevoj teoriji perspektivne reference, i zato teoriju poverenja treba posmatrati pre kao teoriju formiranja verovanja nego kao teoriju odlučivanja (jer se ona u komponenti odlučivanja u potpunosti oslanja na racionalnu konstrukciju Viskuzijevog modela). Još jednom, parametri$\gamma, \xi$ u Viskuzijevom modelu *ne mogu simultano da se ocene* (moguće je na osnovu podataka oceniti samo njihov odnos, $\frac{\gamma}{\xi}$, up. Viscusi, 1989); parametre teorije je moguće oceniti samo ako se jedna od vrednosti fiksira, kao što to činimo u teoriji poverenja gde je doprinos očekivane korisnosti na osnovu objektivnih, datih verovatnoća $\xi = 100$ za

sve lozove. Važnije od svega, već na prvi pogled je jasno *da modeli Viskuzijeve teorije i teorije poverenja imaju upravo formu mešovitih statističkih modela* - oni linearno kombinuju doprinos koji u odlučivanju daje klasičan model očekivane korisnosti sa modelom očekivane korisnosti izračunate iz *a priori* verovatnoća. Podsetimo se i da su mešoviti model - doduše, sa komponentama teorije očekivane korisnosti i teorije izgleda - iskoristili Harison i Rutstromova, pokazujući da nije moguće tvrditi egzistenciju reprezentativnog subjekta jedne ili druge teorije odlučivanja (tj. ispitanika čije bi sve odluke bile dosledno bolje modelirane jednim ili drugim modelom, Harrison & Rutström, 2008). Pošto *a priori* verovatnoće karakterišu subjektivna verovanja donosioca odluka, a objektivne verovatnoće stanja sredine, jasno je da i model teorije perspektivne reference i model teorije poverenja rešavaju problem tretmana više različitih kompjutacionih ciljeva istovremeno. Oni određuju balans, proporciju u kojoj u svoj kompozitni sud o stanjima okoline kognitivni sistem unosi efekat sopstvenih verovanja (tj. prethodno reprezentovanih informacija) i neposredno datih stanja sredine, suštinski rešavajući *pitanje poverenja*: u kojoj meri prihvatamo ono što su informacije iz aktualne situacije, a u kojoj meri se oslanjamo na ono što o sličnim situacijama znamo iz prethodnog iskustva?

Vidimo da mešoviti statistički modeli predstavljaju direktno otelotvorenje ideje o paralelnom tretmanu više različitih kompjutacionih ciljeva. U visoko varijabilnoj, dinamičkoj sredini u kojoj trpi adaptivne pritiske, kognitivni sistem ne može uvek da se osloni na rešavanje tek jednog kompjutacionog problema da bi se optimalno prilagodio. U problemu kauzalnog učenja, dok ne donese sud o tome da li je kovariranje koje registruje zaista i kauzalni odnos ili ne, kognitivni sistem je potpuno opravdan u primeni *mešovite strategije* kojom u određenoj proporciji tretira kovarijaciju kao puku kovarijaciju i u određenoj proporciji kao pravi kauzalni odnos. U problemu odlučivanja, kognitivni sistem često ne može da zna u kojoj meri su aktualna stanja sredine zaista korespondentna objektivnim verovatnoćama na osnovu kojih će se aktualizovati za njega relevantni ishodi, te je opravdan u primeni mešovite strategije kojom svoja prethodna verovanja o relevantim aspektima sredine u određenoj proporciji koristi za korekciju percepcije tih stanja. Baš kao što u eksperimentalnoj analizi ponašanja nauka sreće problem postojanja neopservabilnih parametara koje kognitivni sistem „skriva“ od objektivnih posmatranja, kognitivni sistem se u svom ekološkom okruženju susreće se drugim kognitivnim sistemima i prirodnim procesima koji od njega „skrivaju“ brojne relevantne parametre. Disciplina matematičke statistike je razvijena upravo da bi nam omogućila tretman

situacija u kojima su naše opservacije fundamentalno ograničene tako da su za nas saznatljive samo *ocene* relevantnih parametara u našem okruženju. Razvoj njene nešto komplikovanije rođake u formalnoj strukturi matematičkih disciplina koje tretiraju rizik i neizvesnost, *teorije igara*, doneo je analitičke koncepte - poput koncepta *mešovite strategije* - koji će nam sada omogućiti da teorijski egzaktno diskutujemo problem paralelnog tretmana više adaptivnih ciljeva i ukažemo na neke fundamentalne posledice po shvatanje racionalnosti saznanja koje će uslediti.

*Pojam kognitivne strategije.* Napraviti „statističku mešavinu" dva ili više „osnovna" modela, od kojih svaki, po pretpostavci, rešava određeni problem adaptacije, da bi se tako rešavao problem paralelnog tretmana više kognitivnih ciljeva, nije motivisano pukom potrebom da se iskoristi maksimum informacija (varijanse) koji svaki od „osnovnih" modela može da nosi u rešenju tog problema. Mešoviti statistički modeli kognitivnih funkcija kakve ćemo sada predstaviti - i pokazati da njihovo prihvatanje predstavlja ujedno i prihvatanje nove teorijske paradigme u okviru KKP - motivisani su konceptom *mešovite strategije* teorije igara. Iako je nemoguće ovde pružiti kompletan formalni tretman ovog pojma, neophodno je da ga bar uvedemo dovoljno jasno kako bismo pokazali njegovu primenu u analizi racionalnosti saznanja. Poslužićemo se čuvenim primerom koji se još uvek često koristi kada je u udžbenicima i teorijskim radovima potrebno motivisati uvođenje koncepta mešovite strategije. Sledeća ilustracija se prvi put pojavljuje u „*Theory of Games and Economic Behavior*" fon Nojmana i Morgenšterna iz 1944. godine (von Neumann & Morgnestern, 1944) a odnosi se na analizu situacije opisane u jednoj od avantura Šerloka Holmsa iz pera Artura Konana Dojla, „*The Adventure of the Final Problem*" iz 1893; postavka problema je sledeća:

> 1. *Šerlok Holmes pokušava da pobegne od svog arhineprijatelja profesora Morijartija, u društvu dobrog doktora Votsona; dogovara se sa Votsonom da se nađu na železničkoj stanici Viktorija u Londonu i krenu vozom do Dovera odakle će preći Lamanš i skloniti se od Morijartijeve hajke na Kontinentu;*
>
> 2. *Na samom polasku voza sa Viktorija stanice, Holms ugleda Morijartija, koji je očigledno prozreo Votsonove i njegove korake, kako se probija kroz staničnu gužvu;*
>
> 3. *Znajući da je Morijartiju na raspolaganju specijalni voz kojim će do Dovera stići pre njih, i predvidajući da će Morijarti sigurno ubiti Holmsa uhvati li ga tamo, Holms i Votson se pitaju da li da voz napuste na međustanici u Kanterberiju;*
>
> 4. *Holms objašnjava Votsonu da se o Morijartiju, ingenzionom čoveku koga je sopstvena priroda odvela u vode najmračnijih zločina, mora razmišljati kao*

*o protivniku intelektualnih sposobnosti ekvivalentnih njegovim, i da će u suštini Morijarti izvesno rezonovati onako kako bi Holms rezonovao sam da je na njegovom mestu; ovo vodi u infinitnu regresiju pitanja o tome da li treba ići vozom do Dovera, predviđajući da će Morijarti predvideti mogućnost pokušaja bekstva u Kanterberiju, ali predvidevši tu mogućnost, predviđa da će Holms pokušati dodatno da ga zavara ne koristeći je i nastaviti do Dovera, gde će sačekati i ubiti jer je predvideo taj korak, ali znajući to, Holms...*

Rešenja ovakvih problema, očigledno, ne nastupaju rezonovanjem poput onog opisanog pod tačkom 4. u postavci problema. Šta Holms - ili, ekvivalentno, šta Morijarti - treba da učini u ovako neizvesnoj situaciji? Fon Nojman i Morgenštern postavljaju problem kao problem teorije igara i ilustruju ga sledećom strateškom formom igre između Šerloka Holmsa i profesora Morijartija:

Tabela 14. Strateška forma igre između Holmsa i Morijartija (prema von Neumman & Morgenstern, 1944).

| | | *Holms* | |
|---|---|---|---|
| | | KANTERBERI | DOVER |
| *Morijarti* | KANTERBERI | (100,-100) | (-50,50) |
| | DOVER | (0,0) | (100,-100) |

Ukoliko se susretnu na istoj stanici, Holms će sigurno poginuti od ruke Morijartija; ukoliko Holms pobegne za Dover, ima šanse da se dočepa inostranstva i preživi. Fon Nojman i Morgenštern, za potrebe primera, određuju isplate igračima u četiri moguća slučaja koja prikazuje tabela 14: (a) obojica su u Kanterberiju, Morijarti hvata Holmsa: visok pozitivni ishod za Morijartija i visok negativan ishod za Holmsa, (b) Morijarti silazi u Doveru a Holms u Kanterberiju, što predstavlja nerešen rezultat, (c) Holms silazi u Dover a Morijarti pokušava da ga nađe u Kanterberiju, što je trenutni poraz Morijartija ali ne toliko negativan koliko su to po Holmsa ishodi u kojima ga Morijarti hvata, i konačno, (d) situacija u kojoj se sreću u Doveru, pri čemu Morijarti ubija Holmsa i tako ponovo imamo visok pozitivan ishod za Morijartija i visok negativan ishod za Holmsa. U kompletnoj formalnoj analizi ovakvih problema odgovarajući ishodi bi bili, naravno, zamenjeni odgovarajućim korisnostima koje karakterišu igrače; mi ćemo, za potrebe primera, nastaviti da radimo sa ishodima kakvi su dati u tabeli.

Prvo, treba primetiti da problem u postavci koju koriste fon Nojman i Morgenštern *nema rešenje koje bi predstavljalo uzajamno najbolju strategiju dva igrača.* Naime, ukoliko Morijarti igra „Kanterberi“, za Holmsa je najbolji potez „Dover“, ali ukoliko Holms odluči da igra „Dover“, najbolji potez za Morijartija nije „Kanterberi“ već takođe „Dover“; ukoliko Morijarti pak igra „Dover“, najbolji Holmsov potez je „Kanterberi“, ali ako Holms igra „Kanterberi“ najbolji Morijartijev potez nije „Dover“ već takođe „Kanterberi“. Za ovakve igre se kaže da nemaju *Nešovu ravnotežu u čistim strategijama*[90] (Stojanović, 2005, Binmore, 2007); upravo činjenica da ne postoji uzajamno najbolja strategija dva igrača je osnova beskonačne regresije predikcija o uzajamnom ponašanju opisana u postavci problema. Za bilo koju igru se lako ustanovljava da li ima Nešovu ravnotežu u čistim strategijama ili ne. Ako fiksiramo strategiju jednog igrača, pitamo se koji je najbolji odgovor koji na nju može da pruži drugi igrač. Pošto ustanovimo najbolju strategiju drugog igrača na fiksiranu strategiju prvog igrača, pitamo se da li je istovremeno i najbolji mogući odgovor prvog igrača na tako odabranu strategiju drugog ona strategija koju smo u početku fiksirali. Ako je odgovor „da“, našli smo par strategija koji predstavlja Nešovu ravnoteža date igre; ako nismo, nastavljamo fiksirajući jednu po jednu raspoloživu strategiju prvog igrača. Međutim, moguće je da ne pronađemo nijedan par strategija koji zadovoljava uslov da su one uzajamno najbolji odgovori jedna na drugu; tada kažemo da igra nema Nešovu ravnotežu u čistim strategijama. Onda se suočavamo sa problemima koji zahtevaju rešenje sasvim drugačije prirode, i to rešenje se nalazi u mešovitim strategijama. Pitamo se: na koji način pristupamo analizi igara u situacijama kada se rešenja ne nalaze u primeni čistih strategija?

Očigledno je da ono što igrač mora da učini u odnosu na drugog igrača jeste to da svoje ponašanje *učini što više nepredvidljivim*; on ne sme da dopusti da onaj drugi „pročita“ njegov potez. Zato, igrač ima na raspolaganju *randomizaciju svog ponašanja*: unošenje slučajnog, probabilističkog faktora u svoje odluke, čime protivnika dovodi u situaciju u kojoj ovaj ima samo ograničenu mogućnost da predviđa odluke svog protivnika. Pretpostavimo da Holms želi da učini svoje odluke nezavisnim od odluka Morijartija na sledeći način: on će odigrati „Dover“ sa verovatnoćom $p$ ili će odigrati „Kanterberi“ sa verovatnoćom *1-p*; onda, ukoliko Morijarti odigra „Kanterberi“, Holms ima očekivani dobitak od $p \cdot 50 + (1-p) \cdot (-100)$, a ukoliko Morijarti odigra „Dover“, $p \cdot (-100) + (1-p) \cdot (0)$. U odnosu na moguće poteze Morijartija, Holmsu će biti svejedno šta da odigra ukoliko se

izjednače vrednosti igre za njega u odnosu na to da li Morijarti igra jednu ili drugu strategiju koja mu je na raspolaganju, dakle, za Holmsa: $p \cdot 100 + (1-p) \cdot (-100) = p \cdot (-100) + (1-p) \cdot (50)$, iz čega direktno sledi da Holms treba da igra „DOVER“ sa verovatnoćom od $p = \frac{2}{5}$ (i „KANTERBERI“ sa verovatnoćom $1 - p = \frac{3}{5}$). Ako Holms napravi ovakvu raspodelu verovatnoća nad svojim potezima odn. strategijama, njemu postaje svejedno šta će odigrati Morijarti; za Morijartija, s druge strane, može istom logikom da se pokaže da treba da igra „DOVER“ sa verovatnoćom $q = \frac{3}{5}$ i „KANTERBERI“ sa verovatnoćom $1 - q = \frac{2}{5}$. Ova rešenja, primenjujući logiku koju smo ovde poštovali da bismo do njih došli, predstavljaju *fon Nojmanova rešenja u mešovitim strategijama* za igru Holmsa i Morijartija; lako se pokazuje da ona takođe predstavljaju i *Nešovu ravnotežu u mešovitim strategijama* za ovu igru, što znači da predstavljaju uzajamno najbolje strategije obojice igrača. Podsetimo se, takve uzajamno najbolje strategije nisu postojale ako su igračima na izboru bile samo čiste strategije, ond. ako oni nisu imali mogućnost da randomizuju svoje ponašanje.

Kako interpretiramo mešovite strategije? Evo egzaktne interpretacije onoga što bi Holms, randomizujući svoje ponašanje u mešovitoj strategiji koja predstavlja optimalan odgovor Morijartiju u ovoj igri, trebalo da učini: Holms bi trebalo da ima ne-fer novčić koji izlazi „*glava*“ $\frac{2}{5}$ vremena i „*pismo*“ $\frac{3}{5}$ vremena - ili neku sličnu „slučajnu jedinicu“ kojom bi mogao da generiše ovakvu distribuciju verovatnoće. Holms bi onda trebalo da izvrši jedan statistički eksperiment bacanja takvog novčića, i ako novčić izađe kao „*glava*“ da siđe u Doveru, a ako izađe kao „*pismo*“ - u Kanterberiju. Slično bi trebalo da učini profesor Morijarti kako bi svoje odluke optimizovao u odnosu na moguće odluke koje donosi Holms.

Mešovite strategije su veoma česte u svakodnevnom kompetitivnom i kooperativnom ponašanju živih organizama. Kod čoveka, neprestano ih srećemo u najrazličitijm formama takmičenja: od ponašanja u sportu i igri, socijalnih procesa u kojima je kompeticija samo ritualne prirode (poput zavođenja), preko biznisa, do politike i vojne strategije. Devojka koja ostavlja nejasnu poruku udvaraču o tome hoće li mu se pridružiti u izlasku naredni dan ponaša se kao da igra mešovitu strategiju u kojoj udvaraču ostavlja da donese subjektivni sud o verovatnoćama njenih akcija (tj. njenih čistih strategija) i korisnosti koju ona pridaje svakoj od njih. Teniseri i drugi sportisti koriste tzv. „varke telom“ kojima pokušavaju da svoje protivnike dovedu u situaciju u kojoj oni samo sa određenom verovatnoćom mogu da predvide da li će akcija koja će uslediti biti kongruentna sa trenutnim položajem tela, brzinom i pravcem kretanja i drugim motornim parametrima ili

ne; jasno je da i protivnicima protiv kojih je takva strategija usmerena ne ostaje ništa drugo do da svoje odgovore planiraju kao mešovite strategije. Međutim, u svim ovim slučajevima, koji predstavljaju tipične situacije u kojima se primenjuju analize teorije igara, govorimo o sistemu koji koristi mešovite strategije u takmičenju ili borbi protiv inteligentnog protivnika. Šerlok Holms smatra Morijartija takvim protivnikom: tek ako pretpostavi da je nivo njegove intelektualne analize ponašanja isti kao njegov, on je opravdan u primeni mešovite strategije - jer samo ona, pod takvim uslovima, predstavlja optimalan odgovor takvom protivniku, i ujedno odgovor koji protivnika forsira da i on igra mešovitu strategiju. Međutim, kognitivni sistem je, ako izuzmemo njegovu ulogu u socijalnim situacijama poput onih iz prethodnih primera, suočen sa jednim „protivnikom“ koji se ne može nazvati inteligentnim: on analizira distribucije informacija u sredini koja ga okružuje i na osnovu njihovih karakteristika treba da proizvede - da izračuna - optimalne bihejvioralne odgovore. Pretpostavljamo da su te distribucije informacija u sredini bar u nekoj meri stabilne - jer ako nisu, priroda problema koji diskutujemo u celoj ovoj raspravi se fundamentalno menja. Na koji način bi, onda, primena mešovitih strategija uopšte mogla da pomogne kognitivnim sistemima u rešenju problema adaptacije?

Odgovor na prethodno pitanje nalazi se upravo u našem prethodnom zaključku da se kognitivni sistem često nalazi u situaciji u kojoj iz istih environmentalnih podataka, odn. istih distribucija informacija koje su mu dostupne, može da izračuna više različitih parametara, od kojih svaki može da ima interpretaciju rešavanja određenog kognitivnog problema, tj. ispunjenja određenog kompjutacionog cilja. Vratimo se Hjumovom problemu kauzalne indukcije još jednom. Kognitivni sistem, pretpostavimo, na raspolaganju ima strategiju (tj. algoritme, kognitivne procese, činove - koji god termin izabrali) koja izračunava stepen kovarijacije između dve varijable, isto kao što ima strategiju koja izračunava intenzitet kauzalnog odnosa pod pretpostavkom da su izvori kovarijacije zaista i kauzalno povezani. Ali, on ovo drugo ne zna, i upravo ima za cilj to da ustanovi da li je određena kovarijacija i kauzalni odnos, ili tek puka kovarijacija. U takvoj situaciji, kognitivni sistem bi racionalno, normativno trebalo da primeni mešovitu strategiju, tj. da pokuša da odredi - na osnovu ma čega drugog što može da zna - sa kojom verovatnoćom je osmotrena kovarijacija zaista znak kauzalnog odnosa (te da sa tom verovatnoćom u budućnosti ocenjuje intenzitet kauzalne veze, i ponaša se u skladu sa tom ocenom), a sa kojom verovatnoćom je osmotrena kovarijacija ipak samo kovarijacija (te da

sa tom verovatnoćom u budućnosti ocenjuje intenzitet kovarijacije, i nastavi da se ponaša u skladu sa tom ocenom). Kao što smo videli, potpuna paralela ovom načinu mišljenja nalazi se i u analizi problema odlučivanja u terminima teorija poput Viskuzijeve teorije perspektivne reference ili naše teorije poverenja. Sa kojim verovatnoćama kognitivni sistem treba da pristupi rešavanju ovakvih problema, tj. kako da raspodeli verovatnoće nad svojim „čistim strategijama" - koje u sadašnjoj raspravi, jasno, predstavljaju „čiste" *formalne kognitivne modele*? Ako on nema nikakve druge informacije na osnovu kojih bi doneo odluku o takvoj raspodeli verovatnoće, racionalno je da se ponaša bejzijanski, polazeći od neinformativne, uniformne distribucije (odn. 50:50 u slučaju kombinovanja samo dva „čista" modela), i onda uči kroz fidbek dok ne stabilizuje raspodelu verovatnoća koja odgovara sredini kojoj se adaptira. Moguća su i druga rešenja. Tako, teorija poverenja definiše precizno koje prethodne informacije bi kognitivni sistem trebalo da već poznaje, i kako da ih kombinuje u procesu formiranja verovanja kako bi izračunao (*i*) verovatnoću sa kojom koristi komponentu očekivane korisnosti na osnovu datih verovatnoća i (*ii*) verovatnoću sa kojom koristi komponentu očekivane korisnosti na osnovu *a priori* verovatnoća. Viskuzijeva teorija, videli smo, upravo parametre koji određuju ove verovatnoće ostavlja kao slobodne parametre.

Pažljiv čitalac je do sada već primetio da mešoviti statistički modeli, kao što su hibridni pseudonormativni model kauzalnog učenja koji diskutuju Perales i Šenks, ili model Viskuzijeve teorije i teorije poverenja, *nisu direktna implementacija* koncepta mešovite strategije teorije igara. U odigravanju mešovite strategije, igrač randomizuje svoje ponašanje tako što na osnovu određene raspodele verovatnoća *bira jednu čistu strategiju* koju ima na raspolaganju i odigrava je; Holms, tako, silazi ili u Doveru, ili u Kanterberiju. Svi mešoviti statistički modeli kognitivnih funkcija koje mi diskutujemo rade nešto drugo: oni verovatnoćama ponderišu svoje modele-komponente (svoje „čiste strategije") i na osnovu tako dobijene linearne kombinacije izračunavaju *jednu vrednost* koju koriste kao ocenu relevantnog parametra okruženja. To je veoma različito od mešovite strategije čija primena izgleda kao što je opisano: slučajnom događaju („izvlačenju" događaja sa određene distribucije) sledi primena samo jedne od mogućih akcija. Možemo da ponudimo teorijsko rešenje koje ispravlja ovu nekonzistentnost. Naime, moguće su dve interpretacije mešovitih statističkih modela koji kao komponente sadrže „čiste" formalne kognitivne modele odgovarajućih funkcija. Pre svega, definišemo *kognitivnu strategiju* kao niz formalnih kognitivnih modela $C_1$, $C_2$,.. $C_n$, i njemu

pridružen vektor verovatnoća $p_1$, $p_2$,.. $p_n$: ($p_1$:$C_1$, $p_2$:$C_2$,.. $p_n$:$C_n$). Sledeće dve interpretacije kognitivnih strategija ilustruju različite moguće načine njihove realizacije u realnom radu kognitivnog sistema:

(I) Kognitivna strategija ($p_1$:$C_1$, $p_2$:$C_2$,.. $p_n$:$C_n$) se realizuje primenom određenog formalnog modela $C_i \in \{C_1, C_2, .. C_n\}$ koji se bira sa odgovarajućom verovatnoćom $p_i$; osmotreno ponašanje je rezultat primene tako izabranog kognitivnog modela i odgovarajuće greške u klasičnom fehnerijanskom modelu merenja. Tako, neka je $R_c$ opservabilna mera (procena na skali, ocena monetarnog ekvivalenta, tj. bilo koja bihejvioralna mera) za koju pretpostavljamo da sledi iz primene modela $C$; mi osmatramo (merimo) $R_c + \varepsilon$. Ako se kognitivna strategija realizuje na upravo opisani način, nazivamo je *pravom mešovitom kognitivnom strategijom.*

(II) Definišemo *funkciju odgovora kognitivne strategije* ($p_1$:$C_1$, $p_2$:$C_2$,.. $p_n$:$C_n$) kao: $R_{pC} = p_1 \cdot R_{C_1} + p_2 \cdot R_{C_2} + .. + p_n \cdot R_{C_n}$, odn. $R_{pc} = \sum_{i=1}^{n} p_i \cdot R_{C_i}$. Tako, funkcija odgovora neke kognitivne strategije jeste linearna kombinacija funkcija odgovora svih njenih komponenti odn. „čistih" formalnih kognitivnih modela koje ona obuhvata. Ova interpretacija je opravdana ukoliko su funkcije odgovora svih komponenti samerljive, tj. ukoliko izračunavanja svih komponenti rezultiraju na istoj skali. Ako se kognitivna strategija realizuje na upravo opisani način, onda mi osmatramo (merimo) funkciju odgovara kognitivne strategije u celini, tj. $R_{pC}$; moguće je da osmatramo „čisto" $R_{pC}$, kao što je moguće da osmatramo $R_{pC} + \varepsilon$ - što bi bila pretpostavka fehnerijanskog modela. Međutim, i ovo je veoma značajno primetiti, pod ovom interpretacijom, *primena fehnerijanskog modela nije nužna*: ono što inače tretiramo kao grešku merenja, *pretpostavljajući da jedan „čist" formalan kognitivni model proizvodi ponašanje koje merimo*, pod ovom interpretacijom može da bude tek varijansa koju generiše kombinacija svih drugih formalnih kognitivnih modela u istoj kognitivnoj strategiji. Nazovimo kognitivne strategije koje se realizuju na ovaj način *integrisanim kognitivnim strategijama.*

Posle ovog teorijskog razjašnjenja odnosa koncepta mešovite strategije teorije igara i koncepta kognitivne strategije preostaje još jedan korak do zasnivanja jedne nove teorijske paradigme u kompjutacionoj kognitivnoj psihologiji, paradigme za koju verujemo da nam otvara vrata ka bitnim zaključcima o problemu racionalnosti saznanja. Pretpostavljamo, sasvim opravdano, da funkcija odgovora svakog formalnog kognitivnog modela $C$ - modela koji predstavlja *čistu strategiju*, odn. određeni algoritam koji kognitivni sistem koristi u adaptaciji određenim strukturama

informacija u okruženju - predstavlja osnovni parametar koji karakteriše adaptivni bihejvioralni odgovor na određeno stanje sredine. Ukoliko se kognitivni sistem nalazi u situaciji u kojoj iste distribucije informacija mogu da budu interpretirane kao znaci različitih, latentnih problema, a realizuje se samo jedan od tih problema - nazovimo ga „*pravim problemom*“ po analogiji sa „pravom vrednošću“ parametra u statistici - jasno je da će funkcija odgovora tog modela $C$, ako je upravo on primenjen, biti korisna tek u određenoj meri. Ako je formalni kognitivni model $C$ upravo model koji odgovara latentnom problemu, ta korisnost je maksimalna, ukoliko nije, ona očigledno ima neku vrednost manju od maksimalne. U odnosu na pitanje kako kognitivni sistem treba da raspodeli verovatnoće u svojim kognitivnim strategijama, jasno je da on može da uči upravo kroz fidbek o *korisnosti* koji svaki od njegovih potencijalnih odgovora $R_C$ može da ima. Takođe, kognitivni sistem može da uči kroz fidbek o $R_{pC}$ - odgovoru integrisane kognitivne strategije - u kom slučaju ga fidbek indirektno informiše o korisnosti celokupne raspodele verovatnoća koju koristi u toj kognitivnoj strategiji (slučaj (II)). Dakle, jasno je da postoji neka funkcija korisnosti $U$ koja preslikava ili $(i)$ vrednosti svake funkcije odgovora „čistih“ kognitivnih modela u odgovarajuće korisnosti za kognitivni sistem, $U : R_C \rightarrow \mathbb{R}$, ili $(ii)$ sve različite raspodele verovatnoća za neku integrisanu kognitivnu strategiju u odgovarajuće korisnosti, indirektno kroz vrednost funkcije odgovora integrisane kognitivne strategije $U : R_{pC} \rightarrow \mathbb{R}$. Bilo koja od ove dve funkcije, u zavisnosti od toga kako se realizuje određena kognitivna strategija (kao prava ili kao integrisana), obezbeđuju kognitivnom sistemu da kroz interakcije sa sredinom odredi raspodelu verovatnoća za ma koju kognitivnu strategiju tj. ma koju kombinaciju formalnih kognitivnih modela koje će koristiti u svojoj adaptaciji sredini. Drugim rečima, čini se da nema nikakve prepreke pretpostavci da kognitivni sistem ima funkcije korisnosti nad svojim „akcijama“, tj. nad formalnim kognitivnim modelima, algoritmima, kognitivnim procesima - kako god odlučili da zovemo njegove kompjutacione adaptivne sposobnosti. Posmatran na ovaj način, kognitivni sistem postaje *mešoviti strateg* teorije igara, igrač sa otvorenom mogućnošću da na osnovu fidbeka o sopstvenim akcijama, planiranim na osnovu rezultata sopstvenih kognitivnih izračunavanja, uči i prilagođava se *dok ne otkrije distribuciju verovatnoća nad tim akcijama koja će ga održavati u adaptivnoj ravnoteži sa okolinom.* Ovo je suštinska tvrdnja i osnovna hipoteza teorijske paradigme koju smo upravo uveli u diskurs KKP: $(i)$ *kao posledica činjenice da isti environmentalni podaci* (iste distribucije informacija) *mogu da ukazuju na različite kompjutacione adaptivne*

*probleme sledi* (*ii*) *da kognitivne funkcije treba posmatrati kao mešovite strategije;* (*iii*) *u skladu sa prethodnim, kognitivni sistem je mešoviti strateg koji poznaje ili otkriva distribucije verovatnoća nad njemu dostupnim rešenjima adaptivnih problema i primenjuje te distribucije tokom svoje adaptacije takvoj sredini realizacijom odgovarajućih kognitivnih strategija.*

U odnosu na prirodu kognitivnih funkcija i kognitivnog sistema koja se otkriva posle predložene promene paradigme, kao i zbog direktne analogije sa konceptima teorije igara, smatramo da termin *strategijska teorija kognitivnih funkcija* odgovara strukturi teorijskog mišljenja o kognitivnom sistemu koje smo demonstrirali u prethodnim redovima. Teorijske posledice za debatu o racionalnosti koje slede iz strategijske teorije kognitivnih funkcija izvodimo odmah.

*Racionalnost saznanja i kognitivne strategije.* Prva, sasvim očigledna posledica po debatu o racionalnosti koju smo detaljno predstavili i diskutovali u III delu ove rasprave, već je razorna za diskurs debate u celini. Naime, iz strategijskog viđenja kognitivnih funkcija koje smo predložili, i za koje smo pokazali da ima i empirijsku snagu i teorijsku osnovu, sledi da *analiza pojedinačnih formalnih kognitivnih modela ne može da kaže ništa o njihovoj racionalnosti*, bez obzira na normativne osnove na kojima oni počivaju. Drugi način da se formuliše ova tvrdnja je sledeći: *osmotreno ograničeno racionalno ponašanje u nekom domenu debate o racionalnosti može da bude direktna posledica primene (mešovite) kognitivne strategije, što znači da je moguće da se ograničeno racionalno ponašanje javlja kao posledica jedne „racionalnosti višeg reda“*. Zašto kažemo „racionalnosti višeg reda“? Naime, primena koncepta mešovite strategije u teoriji igara predstavlja savršeno racionalan čin za svakog igrača koji se suočava sa specifičnim problemom u kome je primena upravo mešovitih strategija normativna. Ukoliko mešovita strategija obuhvata komponente od kojih svaka predstavlja formalni kognitivni model čija bi racionalnost mogla da postane predmet diskusije, onda se diskusija racionalnosti same mešovite strategije postavlja na nivou *iznad* analize tih pojedinačnih modela. Primer: ukoliko je model probabilističkog kontrasta normativno rešenje za problem ocene kovarijacije, a model kauzalne moći za problem ocene intenziteta kauzalnog odnosa, primena mešovite strategije koja koristi ova dva modela kao svoje komponente je (izvesno) normativna za Hjumov problem kauzalne indukcije, odn. pokušaj ustanovljavanja da li je neki izvor kovarijacije znak pravog kauzalnog odnosa ili ne. Ukoliko analiziramo izolovano model kauzalne moći u odnosu na bihejvioralne podatke koji potiču iz kognitivnog sistema koji koristi opisanu mešovitu strategiju, naš zaključak će biti

da model nije u potpunosti empirijski adekvatan; isti zaključak ćemo doneti za model probabilističkog kontrasta. U nekim studijama, sa variranjem eksperimentalnih nacrta i karakteristika uzorka ispitanika, pokazivaće se da je prvi model empirijski adekvatniji od drugog, dok će u drugim studijama drugi model pokazivati bolje empirijske performanse. Naš zaključak, ako se ograničavamo na izbor jednog od ta dva modela, nužno će biti pogrešan dok ne promenimo paradigmu u okviru koje razumemo šta kognitivni sistem uopšte pokušava da učini, *jer pojedinačan model uopšte ne može da bude empirijski adekvatan model ponašanja kognitivnog sistema koji koristi mešovite strategije kao što to predviđa strategijska teorija kognitivnih funkcija*. Debata o racionalnosti, u ovom slučaju, uopšte ne može da se završi nedvosmislenim zaključkom pre promene paradigme na opisani način; u drugim oblastima pored ove koje koristimo kao ilustraciju važiće isto. Takođe, zaključak važi za ma koji broj formalnih kognitivnih modela koje analiza obuhvata. Uopšte, postaviti pitanje o tome da li je pojedinačan formalni kognitivni model racionalan ili nije u odnosu na rešenje nekog adaptivnog problema nema nikakvog smisla ukoliko se analiza sprovodi nad kognitivnim sistemom koji samo sa određenom verovatnoćom zna koji mogući adaptivni problem rešava; sve što nam preostaje jeste ocena te verovatnoće u odgovarajućem mešovitom statističkom modelu - ukoliko uopšte možemo da formulišemo takav model. Mi na ovom mestu želimo da generalizujemo zaključak do kog je došla studija Harisona i Rutstrumove (Harrison & Rutström, 2008) uzimajući u obzir statističku mešavinu teorije izgleda i teorije očekivane korisnosti u odlučivanju. Naime, u svim situacijama u kojima jedinstvena stanja sredine mogu da označavaju prisustvo različitih problema adaptacije, *ne postoji reprezentativan subjekat određene teorije tj. određenog modela neke kognitivne funkcije*, ukoliko kognitivni sistem uopšte ima sposobnost prilagođavanja takvim situacijama.

Zaključak koji smo upravo doneli predstavlja direktnu posledicu konsekventnog sprovođenja logike same racionalne analize. Osnova racionalne analize jeste tvrdnja da kognitivni sistemi optimizuju ponašanje organizama u procesu njihove adaptacije sredini. Mi smo ovoj osnovnoj problemskoj situaciji, kako je predstavlja racionalna analiza, dodali zapažanje o tome da postoje adaptivni problemi koji se kognitivnom sistemu predstavljaju na istovetan način, odn. kroz iste distribucije informacija. Sledeći korak je bio pokušaj da se racionalna analiza konsekventno sprovede za takve slučajeve. Optimizacija ponašanja u takvoj situaciji, ukoliko ona ima ikakve veze sa racionalnim pristupom kakav poznajemo iz matematike teorije odlučivanja i teorije

igara razvijene tokom XX veka, izvesno podrazumeva primenu mešovitih strategija kako smo je mi diskutovali. Dakle, ako je sistem racionalan, u (problematičnom) značenju koje racionalnosti pridaje racionalna analiza (up. diskusiju odnosa pojmova racionalnosti i adaptacije iz III dela), on može - on *mora* - da proizvede ponašanje koje je na nivou eksperimentalne analize ograničeno racionalno kada se posmatra kao posledica primene nekog jedinstvenog formalnog kognitivnog modela. Moguće je, dakle, da je cela debata o racionalnosti u domenu viših kognitivnih procesa posledica *implicitne pretpostavke da postoji jedan pravi model određene kognitivne funkcije koji kognitivni sistem isključivo primenjuje suočen sa određenom klasom problema.* U suštini, mi smo pokazali da ta implicitna pretpostavka ne mora da bude tačna. U Hjumovom problemu kauzalne indukcije, ona sigurno nije tačna. U problemu odlučivanja u uslovima rizika, kada se on postavi na način na koji ga postavljaju Viskuzijeva teorija perspektivne reference i teorija poverenja, ta pretpostavka takođe nije tačna. Naša analiza ne implicira da ne postoje adaptivni problemi za koje kognitivni sistemi nemaju jedno jedinstveno rešenje, naprotiv: naša analiza implicira samo da postoje situacije u kojima kognitivni sistem ne zna tačno koji od problema koje zna da reši *treba* da rešava, što ga forsira da koristi mešovite strategije. Međutim, upotreba mešovitih strategija netrivijalno komplikuje eksperimentalnu analizu ponašanja: naučnik suočen sa analizom takvog kognitivnog sistema sada mora da zna (a) koji formalni kognitivni modeli su na raspolaganju kognitivnom sistemu, i (b) koji skup problema adaptacije implicira određeno jedinstveno stanje sredine. Tek ako poznaje sve to, on može da postavi odgovarajući mešovit, hibridan statistički model i fituje eksperimentalne opservacije do kojih je došao. Mi smo u V delu rada pokazali kako takav postupak može da bude uspešan u proučavanju odlučivanja u uslovima rizika.

Zaključak koji diskutujemo predstavlja našu centralnu, najvažniju tvrdnju o naučnom statusu koncepta racionalnosti saznanja: *na način na koji je ovaj koncept do sada tretiran, on uopšte nema smisla kao naučni koncept.* Određeni kognitivni sistem će, ako poznaje pravu distribuciju javljanja latentnih problema - različitih problema koji mogu da se kriju iza istih environmentalnih stanja - koristiti upravo tu distribuciju verovatnoće u formulisanju svojih kognitivnih strategija. Sistem koji to čini raspolaže mešovitom strategijom koja je racionalna u celini, i čija racionalnost u potpunosti relativizuje racionalnost bilo koje komponente („čiste strategije“) koju obuhvata. Rasprava o tome da li je neki formalni kognitivni model racionalan ili ne u odnosu na određeni problem tako u potpunosti pripada normativnim disciplinama:

logici, oblastima filozofije koje se bave formalizacijom sistema verovanja, teoriji verovatnoće i matematičkoj statistici; on, u pravom smislu te reči, uopšte ne pripada jednoj empirijskoj nauci kao što je kognitivna psihologija.

Kako se odrediti prema tvrdnji da je primena određene kognitivne strategije, ako kognitivni sistem koji je koristi zaista zna pravu distribuciju verovatnoće javljanja problema koje njene komponente rešavaju, *u celini racionalan čin*? Da li je ta tvrdnja dovoljna da se donese zaključak o takvom kognitivnom sistemu kao racionalnom? Verujemo da odgovor na to pitanje nedvosmisleno glasi: *ne*. Ako bismo tako nešto smatrali odlikom racionalnosti, onda bi se celokupna racionalnost saznanja svela na pitanje otkrivanja prave distribucije verovatnoća, što je onda ništa drugo do reći da racionalnost podrazumeva rešavanje određenog optimizacionog problema. Reći da kognitivni sistem pokušava da pruži optimalan odgovor sredini kognitivnim funkcijama kojima raspolaže tako što pokušava da otkrije verovatnoće relevantnih događaja, i reći da je on zbog toga kognitivno racionalan, predstavljalo bi neprihvatljivu trivijalizaciju pojma racionalnosti. S druge strane, ako se kognitivni sistemi uopšte prilagođavaju svojim okruženjima, što je osnovna pretpostavka racionalne analize u savremenoj psihologiji uopšte, nije jasno *šta bi drugo oni mogli uopšte da rade do da svoje akcije dovode u saglasnost sa verovatnoćama javljanja određenih problema* koje znaju da reše u svojoj sredini. Ako bismo izjednačili pojam racionalnost saznanja sa mogućnošću formiranja mešovitih strategija kao forme dinamičkog odgovora sredini, izjednačili bismo pojam racionalnosti sa pojmovima adaptacije i optimizacije. Racionalnost saznanja ima smisla ukoliko se analizira u odnosu na normativne osnove za rešavanje određenog problema; samo taj izvor njenog značenja može da je razlikuje od pojmova adaptacije i optimizacije, koji su dobro definisani i kojima nije potrebna nikakva dopunska konotacija. Strategijska teorija kognitivnih funkcija nam pokazuje kako je moguće da se značaj tih normativnih osnova u potpunosti relativizuje u realnom kognitivnom funkcionisanju. Zbog toga, racionalnost saznanja nema smisla *kao naučni koncept*; diskusije o tome šta je normativno rešenje nekog određenog, interesantnog logičkog, matematičkog ili epistemološkog problema svakako imaju smisla, i ne gube na značaju posle ovog zaključka.

Strategijsko viđenje kognitivnog sistema nam je pružilo mogućnost da artikulišemo argument koji nas je odveo ka ovom, neočekivanom, zaključku o prirodi našeg naučnog istraživanja kognitivnih funkcija i naučnom statusu koncepta racionalnosti saznanja. Promena paradigme koju predlažemo otvara vrata za još

neke interesantne i značajne zaključke kojima ćemo posvetiti pažnju u narednim redovima.

*Stabilnost kognitivnih strategija i neka bitna ograničenja u proučavanju kognitivnih funkcija.* Strategijsko viđenje kognitivnih funkcija u toku prethodne analize nije tretiralo kognitivne strategije kao rešenje za *pravi* problem teorije igara - problem koji, u velikom broju situacija, uopšte može da reši samo primena mešovitih strategija. Podsećamo, mi smo pretpostavili da kognitivni sistem igra protiv prirode: priroda „zna“ pravu distribuciju latentnih problema, problema koje rešavaju „čisti“ formalni kognitivni modeli, i koji se kriju iza istih struktura informacija u okruženju. U sklopu ove pretpostavke, mi smo (implicitno) pretpostavili i to da je distribucija verovatnoće javljanja različitih latentnih problema stabilna, tj. da se ona ne menja sa vremenom. U tom slučaju, kognitivni sistem je uvek opravdan u upotrebi onoga što mi nazivamo (mešovitom) kognitivnom strategijom, a njegov osnovni cilj je da otkrije tj. nauči verovatnoće sa kojima se javljaju različiti latentni problemi u njegovom okruženju. Jednom kada je to učenje završeno, tj. kada i kognitivni sistem zna (odn. aproksimira) pravu distribuciju latentnih problema u prirodi, problem adaptacije je rešen: različiti kognitivni modeli koje kognitivni sistem koristi u rešavanju različitih problema nalaze se u ravnoteži prema distribuciji latentnih problema - ravnoteži koja je obezbeđena upravo kroz naučenu distribuciju verovatnoće javljanja tih problema u nekoj kognitivnoj strategiji ($p_1$:$C_1$, $p_2$:$C_2$,.. $p_n$:$C_n$). Ako je naš zaključak o strategijskoj prirodi kognitivnog sistema tačan, već i ovakav ishod, pokazali smo, ozbiljno dovodi u pitanje smisao pitanja o racionalnosti ma kog formalnog kognitivnog modela ako se on posmatra van određene kognitivne strategije.

Statistička teorija odlučivanja, uopšte, može da se posmatra kao pojednostavljenje komplikovanog sveta teorije igara: jedan igrač pokušava da „očita“ parametre koje priroda krije od njega. Celokupnu matematičku statistiku je tako moguće posmatrati kao specijalan slučaj teorije igara. Sada je vreme da postavimo sledeće pitanje: šta ako je problem adaptacije koji kognitivni sistem rešava *zaista pravi problem strategijskih interakcija*, tj. pravi problem teorije igara? Umesto prirode, sada pretpostavljamo *inteligentnog protivnika* kognitivnom sistemu: protivnika sa kojim kognitivni sistem mora da podeli određene, adaptivno relevantne resurse u okruženju, i koji pokušava da optimizuje svoj bihejvioralni odgovor podjednako kao što i posmatrani kognitivni sistem to pokušava da učini. Sledeći primer će nam pomoći da raspravu povedemo u upravo određenom smeru.

Praksa pokazuje da igre sa sledećom formalnom strukturom često igraju u hodnicima Filozofskog fakulteta Univerziteta u Beogradu; na II spratu, gde je smešteno Odeljenje za psihologiju, čak su veoma učestale i omiljene među studentima. Posmatrajmo igru sa dva igrača između studenta i studentkinje psihologije. Pretpostavimo da se između njih odigrava socijalni ritual kroz koji se student udvara studentkinji, i da tokom tog rituala on prati određene njene reakcije u pokušaju da sazna sve što mu je potrebno da bi planirao svoje poteze. Prepostavimo da je, nezavisno od svojih društvenih interakcija sa studentkinjom čije srce pokušava da osvoji, student bio u prilici da upozna kovarijaciju između sledeće dve varijable: (a) osmehivanja devojaka u prisustvu određenih mladića, i (b) uspeha tih mladića u osvajanju srca tih devojaka. Student, međutim, uviđa dublji problem od pitanja o tome da li su varijable (a) i (b) povezane kauzalno ili ne. Naime, devojke se, u prisustvu mladića, mogu smešiti iz razloga čiste pristojnosti, ili njihov osmeh može da bude tek jedna od uobičajenih ekspresija njihove dražesne ličnosti; one, ako žele da im određeni mladići prilaze, mogu da formulišu sasvim drugačije znake kojima bi ih obaveštavale o tome, možda smatrajući da su osmesi suviše trivijalni da bi preneli tako važnu informaciju. Pretpostavimo da student na osnovu prisustva ili odsustva studentkinjinog osmeha poduzima određene socijalne akcije, i kroz fidbek o njihovim posledicama shvata da li mu studentkinja stavlja do znanja da joj se on dopada ili ne. Dakle, on može da opaža dve binarne varijable: prisustvo odn. odsustvo osmeha, i prisustvo odn. odsustvo znakova dopadanja. Interesuje nas vrednost osmeha za njegov romantični poduhvat: da li on treba da tretira ove dve varijable kao kauzalno povezane, ili pak kao da se one nalaze u pukoj kovarijaciji? Očigledno, pred njim se nalazi instanca Hjumovog problema.

Ono što pokazujemo sada jeste da pristup ovom kognitivnom problemu fundamentalno zavisi od funkcija korisnosti studenta i studentkinje koji učestvuju u opisanom socijalnom ritualu - što će se pokazati kao konačni argument protiv ma kakve mogućnosti donošenja konačnog suda o racionalnosti neke kognitivne funkcije uopšte. Iz strategijske teorije kognitivnih funkcija sledi da je ma kakvo rešenje Hjumovog problema kauzalne indukcije vezano za primenu kognitivne strategije koja kao svoje komponente obuhvata sud o kovarijaciji i sud o kauzalnosti. Podsetimo se da Hjumov problem ne može da se postavi kao problem ocene kauzalne snage pod pretpostavkom da odnos *možda* jeste kauzalan - jer ma kakav nivo intenziteta kauzalne moći bio izračunat pod takvom hipotezom, to još uvek ne znači da odnos nije puka kovarijacija. Jedini način da se pristupi problemu je fundamentalan i

uzima u obzir upravo njegov najteži aspekt, a to je da je odgovor o tome da li je odnos kauzalan ili nije nešto što pripada noumenalnom; dakle, taj problem je moguće rešiti samo pravljenjem dve paralelne pretpostavke, jedne o tome da je odnos kauzalan i da ga treba oceniti kao takvog, i druge da je on puka kovarijacija iz čega sledi da ga treba i oceniti kao kovarijaciju. Strategijska teorija tvrdi da je rešenje problema upravo u primeni mešovite strategije poput one koju predstavlja tzv. hibridni psuedonormativni model koji diskutuju Perales i Šenks. Međutim, mi moramo da prepostavimo i nešto o *ishodima socijalnih interakcija* koje mogu da se jave kao posledice sudova koje učesnici donose jedni o drugima.

Posmatrajmo sledeću tabelu:

Tabela 15. Strateška forma igre između studenta i studentkinje.

| | | *Student* | |
|---|---|---|---|
| | | $C \to E$ | $C \upuparrows E$ |
| *Studentkinja* | $C \to E$ | (2,3) | (3,1) |
| | $C \upuparrows E$ | (4,2) | (2,3) |

U tabeli 15, oznaka $C \to E$ među strategijama kojima raspolaže student znači da on bira da osmehe studentkinje interpretira kao znake pravog dopadanja i evaluira odnos između ta dva fenomena kao kauzalan; oznaka $C \upuparrows E$ među njegovim strategijama označava da on bira da posmatra vezu između osmehivanja i pravog dopadanja kao puku kovarijaciju. Na strani studentkinje, strategije imaju drugačiju interpretaciju: oznaka $C \to E$ među njenim strategijama znači da zaista postoji kauzalna veza između toga da joj se student dopada i činjenice da mu se osmehuje, dok strategija $C \upuparrows E$ kod nje označava da ta dva fenomena tek kovariraju (npr. student joj se zaista dopada, ali osmeh je tu iz pristojnosti i ona ne očekuje da ga on interpretira kao poziv na hrabrije poteze). U ćelijama tabele nalaze se ishodi odigravanja svakog para strategija studenta i studentkinje za svakog od njih dvoje: prvi broj u svakom uređenom paru je ishod za studentkinju, drugi za studenta. Igra prikazana u tabeli 15. nije igra nulte sume: u svim mogućim kombinacijama ishoda, i student i studentkinja profitiraju u određenoj meri; možemo da pretpostavimo da se zaista sviđaju jedno drugom i da postoji izvesna korisnost za njih koja je posledica činjenice da uopšte provode vreme zajedno. Ako oboje igraju $C \to E$,

korisnost za studenta je 3, za studentkinju 2; ako oboje igraju $C \upuparrows E$, korisnost za studenta je opet 3, za studenktinju opet 2. Ako student igra $C \rightarrow E$, a studentkinja $C \upuparrows E$, korisnost za njega je 2, a za nju 4, dok u obrnutom slučaju on osvaja 1, a ona 3. Podsetimo se koncepta Nešove ravnoteže: (mešovite) strategije će se nalaziti u ravnoteži ako nijedan igrač ne može da osvoji više nego što osvaja pod pretpostavkom da drugi igrači ne promene strategije kojih se drže. U čistim strategijama, igra data tabelom 15. nema Nešovu ravnotežu, što se pokazuje veoma lako pregledom korisnosti koju osvajaju student i studentkinja. Nijedan par čistih strategija ne predstavlja uzajamno najbolje odgovore. Igra, međutim, ima Nešovu ravnotežu u mešovitim strategijama: student mora da odredi verovatnoću $p$ sa kojom će igrati $C \rightarrow E$ (što znači da će $C \upuparrows E$ igrati sa verovatnoćom *1-p*), a studentkinja verovatnoću $q$ sa kojom će igrati $C \rightarrow E$ (što znači da će $C \upuparrows E$ igrati sa verovatnoćom *1-q*), tako da ako se oboje drže tih mešovitih strategija nijedno od njih neće odstupiti - jer neće moći da poboljša svoju poziciju ukoliko ono drugo nastavi da se drži svoje mešovite strategije (što je definicija Nešove ravnoteže). Podsetimo se: određivanje mešovite strategije koja uključuje $C \rightarrow E$ i $C \upuparrows E$ kao svoje komponente *za studenta znači primenu kognitivne strategije* u kojoj su ocene intenziteta kauzalnog odnosa i ocena kovarijacije „čisti“ kognitivni modeli. Dakle, rešavanje ove igre - određivanje ravnotežnog para strategija koji je održava u Nešovoj ravnoteži - za jednog od igrača (studenta) znači i određivanje načina na koji on u kontekstu igre treba da rešava Hjumov problem kauzalne indukcije. Da se vratimo liniji argumenta koji razvijamo: uskoro ćemo pokazati - dok je čitalac donekle upućen u teoriju igara verovatno već prozreo naš argument - da način na koji student treba da pristupi rešavanju problema kauzalne indukcije zavisi od funkcije korisnosti koja ga odlikuje.

Postupak određivanja Nešove ravnoteže za igru sa dva igrača i po dve strategije je jednostavan. Treba izračunati verovatnoću $p$ sa kojom student igra $C \rightarrow E$ tako da je studentkinji svejedno koju od dve strategije koje su njoj na raspolaganju bira, i verovatnoću $q$ sa kojom studentkinja igra $C \rightarrow E$ tako da je studentu svejedno koju od dve svoje strategije bira; za dati par vrednosti ($p$,$q$) igra će biti u Nešovoj ravnoteži. Ako student igra $C \rightarrow E$ sa verovatnoćom $p$ i $C \upuparrows E$ sa verovatnoćom 1-$p$, studentkinja ima očekivanu korisnost od 2p+3(1-p) ako igra $C \rightarrow E$, i očekivanu korisnost od 4p+2(1-p) ako odigra $C \upuparrows E$. Izjednačavajući dva izraza očekivane korisnosti (izražavajući, dakle, situaciju u kojoj je studentkinji svejedno koju od dve svoje strategije će odigrati) nalazimo da je $p = \frac{1}{3}$. Istim postupkom rešavamo

vrednost $q$ sa kojom studentkinja treba da igra $C \rightarrow E$ i nalazimo vrednost $q = \frac{1}{3}$; igra ima Nešovu ravnotežu $(p,q)=(\frac{1}{3},\frac{1}{3})$. Sa ovakvim rešenjem, sada znamo da student treba da rešava problem kauzalne indukcije kognitivnom strategijom koja kombinuje ocenu intenziteta kauzalnog odnosa i ocenu kovarijacije u odnosu $\frac{1}{3} : \frac{2}{3}$, jer je opravdan u očekivanju da će studentkinja sa verovatnoćom $q = \frac{1}{3}$ zaista deliti osmehe kao kauzalne posledice toga što joj se on dopada, dok će sa verovatnoćom od $1 - q = \frac{2}{3}$ ti osmesi biti u pukoj kovarijaciji sa činjenicom o dopadanju.

Prvi deo argumenta koji razvijamo je izveden: proces formulisanja određene kognitivne strategije, odn. nalaženja distribucije verovatnoća sa kojom se (a) bira i koristi određeni „čist“ kognitivni model (u pravim mešovitim kognitivnim strategijama), ili (b) na osnovu koje se razvija linearna kombinacija čistih modela (u integrisanim kognitivnim strategijama), *može da bude proces determinisan rešenjem problema strategijskih interakcija* ako kognitivni sistem „za protivnika“ nema prirodu, već složenijeg, inteligentnog takmaca - poput drugog ljudskog bića. Smatramo da je moguće govoriti o određenoj hijerarhiji u kompleksnosti formulisanja kognitivnih strategija, koja sadrži jednostavnije, specijalne slučajeve otkrivanja pravih distribucija latentnih problema („igra protiv prirode“) na nižim, i kompleksnije slučajeve, kao što je upravo diskutovan problem u kome prave distribucije zavise od strategijskih interakcija, na višim nivoima hijerarhije. Jednostavniji fizički procesi koje odlikuju stabilne distribucije verovatnoća mogu biti izvor latentnih problema niže u ovoj hijerarhijskoj organizaciji, ali veću proporciju biosocijalnih procesa u našem okruženju svakako nalazimo na višim nivoima te hijerarhije, gde se prave distribucije koje naše kognitivne strategije moraju da otkriju javljaju u funkciji naše interakcije sa sistemima prema kojima usmeravamo naše kognitivne kapacitete. Nažalost, situacija naučne analize kognitivnog sistema koji rešava probleme u takvom kontekstu se progresivno usložnjava.

Posmatrajmo sledeću tabelu koja predstavlja stratešku formu iste igre između studenta i studentkinje. U tabeli 16. uneli smo minimalnu promenu u odnosu na tabelu 15.

Tabela 16. Strateška forma igre između studenta i studentkinje sa minimalnom promenom u funkciji korisnosti studenta u odnosu na igru iz tabele 15.

| | | *Student* | |
|---|---|---|---|
| | | $C \to E$ | $C \Uparrow E$ |
| *Studentkinja* | $C \to E$ | (2,5) | (3,1) |
| | $C \Uparrow E$ | (4,2) | (2,3) |

Jedina razlika u odnosu na igru kako je prethodno formulisana je ta što prema tabeli 16. student osvaja 5 jedinica korisnosti ako i on i studentkinja odigraju $C \to E$ (gornje levo polje tabele), umesto 3 koliko je osvajao u toj situaciji u igri koju opisuje tabela 15. Promenjena je minimalno funkcija korisnosti koja odlikuje studenta. Ako sada pokušamo da nađemo ravnotežno rešenje igre, uvidećemo da se verovatnoća $p$ sa kojom student treba da igra $C \to E$ ne menja i iznosi $p = \frac{1}{3}$ kao i u prethodnoj igri. Međutim, verovatnoća sa kojom studentkinja treba da igra $C \to E$ se menja i sada iznosi $q = \frac{1}{5}$, tako da igra ima Nešovu ravnotežu za par $(p,q)=(\frac{1}{3},\frac{1}{5})$. Zaključak je da je student sada opravdan da pristupi rešavanju problema kauzalne indukcije tako što će doprinos ocene intenziteta kauzalnog odnosa i ocene kovarijacije držati u proporciji $\frac{1}{5} : \frac{4}{5}$, što je bitno drugačija kognitivna strategija od one koju je prethodno trebalo da igra, sa odnosom $\frac{1}{3} : \frac{2}{3}$!

Mala promena funkcije korisnosti jednog od igrača dovodi do promene distribucije verovatnoća koju on treba da koristi u (mešovitoj) kognitivnoj strategiji $(p_1{:}C_1,\ p_2{:}C_2,..\ \ p_n{:}C_n)$. U prethodnoj sekciji postavili smo pitanje o tome da li upotreba mešovitih strategija u rešavanju kognitivnih problema adaptacije znači da su kognitivni sistemi koji to čine racionalni u celini - na „višem nivou analize“ od nivoa na kome se analizira racionalnost pojedinačnih kognitivnih modela. Rezultat do koga smo sada došli pokazuje da u složenijim problemima adaptacije nema ni govora o oceni racionalnosti kognitivnog sistema na takvom „višem nivou analize“: distribucija verovatnoće koju kognitivni sistem treba da koristi u formulisanju adekvatne kognitivne strategije, kao što vidimo, zavisi od samih strategijskih interakcija u koje on ulazi - i od takvih parametara kao što su oni koji određuju njegovu funkciju korisnosti. Kako izgleda praksa eksperimentalne analize ponašanja kognitivnog sistema sa ovakvom osobinom? U funkciji parametara koji se odnose na stav tog kognitivnog sistema prema strategijskim interakcijama

u koje ulazi, npr. parametara koji određuju njegovu funkciju korisnosti, mi u eksperimentalnim zadacima kroz koje prikupljamo podatke za ocenu modela ocene kauzalne moći ili ocene kovarijacije dolazimo do zaključaka o individualnim razlikama - javljaju se različite distribucije verovatnoća u mešovitim modelima tj. kognitivnim strategijama - koje nemaju nikakve veze sa samim procesom rešavanja problema kauzalne indukcije. Još jednom, vidimo da će svaki naš pokušaj da raspravljamo racionalnost saznanja biće neuspešan, tj. vodiće nas u protivrečnosti koje će proizvoditi kontradiktorne empirijske opservacije od kojih će neke svedočiti *pro*, a neke *contra* racionalnosti određene kognitivne funkcije u odnosu na određeni normativni kriterijum - dok će se pravi izvor problema, kao što smo pokazali, nalaziti daleko van fokusa rasprave.

Podsetimo se na trenutak samog početka naše rasprave, kada smo se jasno ogradili od pitanja *racionalnosti ciljeva* i postavili pitanje o racionalnosti saznajnog procesa kao takvog. Ako analiziramo kognitivni sistem koji zna koji cilj među potencijalnim latentnim ciljevima treba da rešava - na primer, zna da su verovatnoće na lozovima u odlučivanju objektivne i nezavisne od njegovih prethodnih iskustava, ili zna da suočen sa strukturom Hjumovog problema ima posla sa pukom kovarijacijom - mi možemo da analiziramo racionalnost kognitivne funkcije kojom on pokuđava da reši taj određeni kompjutacioni cilj. Pitanje je, naravno, kada možemo da tvrdimo da kognitivni sistem zaista zna koji od potencijalnih kompjutacionih ciljeva treba da zadovolji, ali recimo da u nekoj hipotetičkoj, idealnoj situaciji to izvesno zna; ponovimo, onda je analiza racionalnosti neke kognitivne funkcije moguća u odnosu na normativne standarde. S druge strane, pokazali smo da rešavanje problema postojanja više latentnih kompjutacionih ciljeva, koje je, verujemo, moguće samo ako se tom problemu pristupi kao problemu teorije igara, na način koji smo upravo demonstrirali u razvoju strategijske teorije kognitivnih funkcija, ne ostavlja mnogo prostora za zaključke o racionalnosti kognitivnih funkcija u odnosu na neke jedinstvene normativne standarde. Zaključak koji sledi je da je *racionalnost, uopšte, moguće diskutovati samo ukoliko se ne gubi iz vida to u odnosu na koji cilj se određeni kognitivni čin analizira kao racionalan: govoriti o racionalnosti saznanja po sebi nema smisla.*

Vratimo se na trenutak rezultatu prethodne analize strategijske interakcije u našem izmišljenom primeru igre studenta i studentkinje. Značajan rezultat, koji predstavlja činjenica da promena funkcije korisnosti jednog od igrača u strategijskim interakcijama menja formu koju treba da uzmu njegovi kognitivni procesi, posledica

je toga što se oni primenjuju upravo u proceni situacije strategijske interakcije, a ne u proceni situacije u kojoj se podrazumeva neka stabilna distribucija verovatnoća koju diktira prirodno okruženje. Ovaj uvid omogućen je tek strategijskom analizom kognitivnih funkcija, odn. tek analizom u kojoj određena strategija u teoriji igara više nije čin koji ima posledice tek u samoj strateškoj interakciji, već ima posledice po način na koji igrač opaža i interpretira parametre situacije koji su za njega značajni. Bez namere da ovde dalje razvijamo posledice strategijske teorije kognitivnih funkcija, čini se da ona otvara mogućnost za još interesantnih uvida poput onog koji smo diskutovali u našem jednostavnom primeru. Interesantno je da tokom istorije kognitivne psihologije, koliko god ona u debati o racionalnosti bila isprepletana sa fundamentalnim raspravama u oblasti odlučivanja i teorije igara, do sada nije postojao predlog da se same kognitivne funkcije posmatraju kao strategije, a kognitivni sistem koji njima raspolaže kao sistem koji uzima strategijski odnos prema visoko varijabilnoj, dinamičkoj sredini koja ga okružuje - od fizičkih preko bioloških do najsloženijih socijalnih, simboličkih procesa. Kao što smo već predložili, čini nam se da je moguće uspostavljanje jedne hijerarhije rastuće nestabilnosti kognitivnih funkcija, kako se krećemo od skupova latentnih problema čiju pojavu kontrolišu stabilne distribucije verovatnoća, ka problemima u kojima su te distribucije verovatnoća posledice strategijskih interakcija. Na što višem nivou se nalazimo u toj hijerarhiji, naše analize postaju sve manje i manje u stanju da nam pruže makar i elementarne predikcije ponašanja u standardnim eksperimentalnim paradigmama kognitivne psihologije, da ne govorimo o zaključcima poput onih koji se odnose na tako komplikovane osobine poput „racionalnosti“ kognitivnih funkcija.

*Još neke posledice strategijske analize kognitivnih funkcija i njena ograničenja.* Strategijsku teoriju kognitivnih funkcija koju smo skicirali kroz naše prethodne analize nećemo ovde u potpunosti razvijati. Pre svega, uz malo više ulaganja u formalni razvoj, nju je lako izraziti na način koji bi zadovoljio tehničke, formalne standarde; u odnosu na našu raspravu o racionalnosti saznanja, međutim, to bi značilo samo još jednu digresiju u oblast primene matematike u kognitivnoj psihologiji. Sledeći pretpostavke koje smo eksplicirali u prethodnim redovima, svako sa minimalnim razumevanjem matematičkih pretpostavki teorije igara lako može da sistematizuje ovo shvatanje i u potpunosti ga formalizuje - ako to smatra neophodnim.

Postoji više interesantnih načina da se u okviru strategijskog viđenja kognitivnih funkcija demonstrira problematičan naučni status pojma racionalnosti saznanja.

Jedan od njih, koji nismo razvili u prethodnim redovima, polazi od pretpostavke da kognitivni sistemi planiraju adaptivne bihejvioralne odgovore u sredini u kojoj su okruženi drugim sličnim sistemima i da se svi jedni prema drugima odnose isključivo kompetitivno. Ova pretpostavka, u čistom obliku koji sada razmatramo, izvesno nije odlika realnog ljudskog okruženja u kome se prožimaju kooperativnost i kompetitivnost. Ipak, samo za trenutak želimo da se fokusiramo isključivo na kompetitivne situacije. Posmatrajmo formu jedne kognitivne strategije kao što je ona koju predstavlja jednačina (93) Viskuzijeve teorije perspektivne reference. Ono što je karakteristično za ovakav model jeste činjenica da njegove parametre nije moguće oceniti simultano. U slučaju Viskuzijeve teorije, radi se o parametrima koji određuju stepen u kome se u oceni rizičnih lozova oslanjamo na naša prethodna verovanja i stepen u kome se u toj oceni oslanjamo na aktualne informacije. Teorija poverenja, videli smo, rešava ovaj problem tako što praktično fiksira vrednost jednog od ova dva parametra. Međutim, nema nikakve prepreke za hipotezu da rešenje koje predlaže teorija poverenja ne funkcioniše u svim zamislivim situacijama u kojima je ocena rizičnih lozova adaptivno relavantan čin. Argument je jednostavan: ukoliko se kognitivni sistem koji analiziramo adaptira u sredini u kojoj je okružen drugim sličnim sistemima sa kojima može samo da podeli raspoložive resurse, *u najdubljem je interesu takvog kognitivnog sistema da nikome ne omogući ocenu parametara modela koje on koristi za adaptaciju toj sredini, jer na taj način omogućava predikciju svog ponašanja i time umanjuje svoje šanse u borbi za raspoložive resurse.* Diskusiju ove hipoteze započeli smo u sekciji 8. III dela naše rasprave. Viskuzijeva teorija, koja bi pod standardnom metodološkom paradigmom kognitivne psihologije i eksperimentalne psihologije trpela kritiku kao model čije parametre nije moguće simultano oceniti - govorilo bi se, dakle, o *kvalitetu modela* - u takvoj situaciji predstavlja paradigmatično rešenje kojem bi kognitivni sistemi *trebalo* da pribegnu kako bi zaštitili sebe od bilo čijeg pokušaja da vrši predikciju njihovog ponašanja. Pošto je predikcija ponašanja jedan od osnovnih ciljeva naučne analize u kognitivnoj psihologiji uopšte, *jasno je da priroda kognitivnog sistema može sasvim opravdano da predstavlja najveću prepreku njegovoj naučnoj analizi uopšte.* Kao što nam nije poznato da je pre ovog rada uopšte diskutovana mogućnost da su kognitivne funkcije strateške prirode, te da se kognitivni sistemi ponašaju kao da koriste postulate teorije igara u kognitivnim adaptacijama sredini, nije nam poznato da je neko izdvojio ovaj - posle promene perspektive, toliko očigledan - argument o prirodi kognitivnog sistema. Sistem koji, na primer, planirano varira distribuciju verovatnoća koja

formuliše njegove kognitivne strategije sa ciljem da onemogući predikciju sopstvenog ponašanja, u odnosu na normativnu analizu ne može da se karakteriše kao racionalan jer je njegovo opservabilno ponašanje ograničeno racionalno (pošto ne koristi distribuciju verovatnoća koja se poklapa sa pravom distribucijom u okolini), dok je na latentnom nivou analize - za onog koji poznaje njegov „unutrašnji rezon“ da zavara protivnike - on savršeno racionalan. Da bi se pokazalo da ovakva strategija može da bude racionalna i uprkos tome proizvodi ponašanje koje je *prima facie* ograničeno racionalno, dovoljno je pretpostaviti da sistem poznaje određenu cenu $C$ koju u jedinicama korisnosti plaća proporcionalno mogućnostima drugih sličnih sistema da izvrše predikciju njegovog ponašanja; direktno sledi da je optimalna primena strategije u kojoj sistem odstupa od optimalnog odgovora „na prvu loptu“ kako bi onemogućio konkurenciji da predviđa njegovo ponašanje. Eto još jednog argumenta u prilog tezi da je racionalnost saznanja nemoguće ustanoviti - ukoliko taj pojam uopšte ima smisla u naučnoj analizi kognitivnog sistema.

Vratimo se još jednom Bernštajnovom problemu kakav prepoznaju napori da se reši problem koordinacije u oblasti senzomotornih procesa. Sada je potpuno jasno da se i u analizi ovog naučnog problema suočavamo sa formalno *istom* situacijom koju smo upravo diskutovali. Dok je višak broja stepeni slobode na strani kontrolnih procesa centralnog nervnog sistema prepreka za naučnu analizu problema koordinacije - problema egzaktne predikcije motornog plana koji će realizovati određeni pokret - za sistem koji rešava problem koordinacije *višak stepeni slobode predstavlja čistu adaptivnu prednost* u odnosu na potencijalno kompetitivno okruženje. Svako ko se ikada igrao sa mačkom i saterao je kroz tu igru u ćošak u kome je ona zauzela karakterističan stav - kao da su zategnuti svi mišići njenog tela, ali tako da se istovremeno nalazi u položaju koji nije prirodan startni položaj za ma koji pokret koji bi ona mogla da izvede, iščekujući da li će taj koji se igra sa njom da se pomeri i minimalno u neku stranu - da bi zatim na neverovatan, naizgled potpuno nepredvidljiv način uspela da se izvuče iz zatečene situacije, može da razmisli da li je baš u takvoj situaciji prisustvovao *adaptivnoj upotrebi* mogućnosti da mu se postavi upravo Bernštajnov problem tokom igre. Pitanje je za koga je Bernštajnov problem - problem. U naučnoj analizi kognitivnog sistema, to je problem za analitičara; ne za organizam koji raspolaže viškom stepeni slobode u kontroli složenih sistema poput senzomotornog. Ako je taj kognitivni sistem koji kontroliše senzomotorni sistem ujedno i mešoviti strateg teorije igara, višak stepeni slobode u sistemu kojim upravlja zapravo nikog ne bi trebalo da iznenađuje. To je upravo ono što obezbeđuje

nepredvidljivost, odn. *onu neizvesnost koju fundamentalno nije moguće redukovati u predikciji ponašanja složenih kognitivnih sistema koji igraju mešovite strategije u svom okruženju* (up. Glimcher, 2004, za neke veoma ilustrativne analize na liniji ovakvih argumenata).

Konačno, neophodno je postaviti jasna ograničenja analizama koje smo predstavili u okviru strategijskog viđenja kognitivnih funkcija. Pre svega, mnogi kognitivni problemi izvesno ne predstavljaju tip problema u kome iste environmentalne distribucije informacija formulišu više latentnih ciljeva adaptacije, ostavljajući kognitivni sistem u situaciji u kojoj mora da procenjuje kada koji problem zapravo rešava. Na samom početku naše rasprave, podsetimo se, ogradili smo se od problema u oblasti percepcije i problema u oblasti psiholingvistike, fokusirajući se na oblast (preostalih) viših i simboličkih kognitivnih funkcija. Za veliki broj funkcija koje vrši perceptivni aparat čoveka (i sličnih organizama) mi verujemo da problemi koje smo mi diskutovali jednostavno nisu relevantni. Da percepciju odlikuje problem postojanja latentnih kompjutacionih ciljeva, ona verovatno uopšte ne bi predstavljala evoluciono prihvatljivo rešenje za adaptaciju čoveka i drugih organizama sa sličnim perceptivnim sistemom. Neke vizuelne iluzije, na primer, poput onih u kojima jedinstvena vizuelna stimulacija pruža mogućnost formulisanja alternativnih percepta (up. Leopold & Logothetis, 1999) u formi dvosmislenih slika, izvesno mogu da se analiziraju o okviru strategijskog viđenja kognitivnih funkcija koje smo mi predstavili. Ipak, postavlja se pitanje u kojoj meri su takvi fenomeni česta, tipična karakteristika našeg okruženja, što direktno otvara pitanje njihove relevantnosti za analizu naše adaptacije - iako ova primedba svakako ne zadovoljava uvek opravdanu naučnu znatiželju. Analiza nekih kognitivnih funkcija koje realizuju ljudski jezik, sa druge strane, izvesno ne postavlja tako teške probleme kao one koje smo analizirali. Interesantno je da je u debati o racionalnosti malo reči o sledećoj činjenici: postoji domen kognitivnih funkcija koji operiše sa *izuzetno složenim strukturama informacija*, a koji se već i pri površnim analizama ukazuje kao optimalan i normativno opravdan skoro bez izuzetka: to je, naravno, domen razumevanja i analize sintaksičkih struktura, kao i njihove upotrebe u produkciji jezičkih iskaza. Veoma složene sintaksičke strukture smo u stanju da razumemo zahvaljujući kognitivnim procesima koji se realizuju na skali od nekoliko desetina do nekoliko stotina milisekundi; ne čini se da nam je mnogo više vremena potrebno da upotrebimo te kognitivne procese u jezičkoj produkciji. S druge strane, gramatika predstavlja rigidan sistem uređen

pravilima koja ne trpe ni najmanja odstupanja. Kada analiziramo gramatičke strukture matematički, formalno, suočavamo se sa nekim od najsloženijih problema formalne analize uopšte. Međutim, kognitivni sistem ne pokazuje nikakve probleme u obradi informacija, i skoro nikakva odstupanja od normativnog okvira - jer sintakse upravo definišu određene normativne okvire - tokom kognitivne analize tako složene formalne strukture kakva je gramatika prirodnih jezika. Ako postoji markatan primer one racionalnosti i optimalnosti o kojoj govori racionalna analiza, to je skup psiholingvističkih funkcija koje realizuju sintaksičke analize i upotrebu sintaksičkih regularnosti u govornoj produkciji. Ne bi bilo tačno reći da strategijska analiza kognitivnih funkcija u ovom slučaju ne bi našla svoju primenu. Produktivnost, kao suštinska osobina jezika, podrazumeva mogućnost primene rekurzivnih statističkih struktura da bi se proizveo potencijalno beskonačan skup rešenja za određeni problem komunikacije (Hauser, Chomsky & Fitch, 2002). Ali, pretpostavljajući da *sadržaj koji treba komunicirati ne varira*, mi znamo da će sva potencijalna sintaksička rešenja problema komunikacije tog sadržaja rešavati *jedan isti* kognitivni problem, i dalje, govoreći o sintaksi jednog istog jezika, mi znamo da bi sva ta potencijalna rešenja rešila taj problem *podjednako dobro sa normativnog stanovišta*. Strategijska analiza psiholingvističkih, sintaksičkih funkcija kognitivnog sistema bi u ovom slučaju eventualno mogla da otkrije koje su distribucije verovatnoća nad potencijalnim sintaksičkim rešenjima istog problema preferentne - ali onda preostaje veoma težak problem objasniti zašto bi određene sintaksičke strukture za rešenje nekog problema komunikacije bile preferentne u odnosu na članove sopstvene klase ekvivalencije[91].

Prethodne analize, dakle, jasno treba ograničiti na sve one slučajeve u kojima iste environmentalne distribucije informacija mogu da predstavljaju znake različitih, latentnih problema adaptacije, koji onda zahtevaju zadovoljenje različitih kompjutacionih ciljeva na strani kognitivnog sistema. Mi smo se u ovoj raspravi usresredili na samo dva takva problema: problem kauzalne indukcije i problem odlučivanja u uslovima rizika. Verujemo da spisak svakako nije iscrpljen; broj situacija u kojima dostupne informacije ne determinišu jednoznačno koji kompjutacioni cilj kognitivni sistem treba da zadovolji u funkciji adaptacije, naprotiv, čini se ogromnim, ocrtavajući onu fundamentalnu neodređenost, onu neizvesnost naših interakcija sa složenim fizičkim i biosocijalnim okruženjem koju ni u principu ne možemo da redukujemo.

## 13.3 Argument III: Analiza racionalnosti elementarnih simboličkih funkcija

Naredna analiza će nas odvesti ka još jednom, završnom argumentu koji pokazuje da koncept racionalnosti saznanja nema upotrebnu vrednost u naučnom diskursu. Argument koji formulišemo odnosi se na simboličke funkcije ljudskog kognitivnog sistema. Specifično, argument se odnosi na kognitivnu funkciju kategorizacije, odn. probleme interpretacije i kreiranja karakteristika koje omogućavaju konceptualnu organizaciju znanja - onako kako smo te probleme diskutovali u sekciji 7.6. Diskusija je organizovana na sledeći način. Prvo pokazujemo da fleksibilnost u interpretaciji značenja simbola kojima kognitivni sistem može da raspolaže implicira isto ograničenje koje smo demonstrirali u sekciji 6.2 o falsifikabilnosti kompjutacionističkog programa uopšte. Slično problemu pijaniste i/ili političara Paderevskog (up. sekciju 6.2), koji je sa namerom formulisan tako da obezbedi diskusiju problema odnosa reference i smisla u semantici, mogućnost *reinterpretacije značenja određenog simbola* - inherentna samoj ljudskoj simboličkoj funkciji - vodi pravo u metodološku situaciju u kojoj bitna pitanja poput onog o broju mentalnih reprezentacija koje se odnose na isti termin nisu odlučiva. Neko ko ne može da pokaže ni sa kojim brojem entiteta određena kognitivna funkcija, poput simboličke, barata, verujemo da ne može ni da diskutuje složene osobine kao što je racionalnost takvih funkcija. Ipak, to je tek manje bitno ograničenje u diskusiji o racionalnosti simboličke funkcije od onog koje naredno predstavljamo. Elementarna intuicija o upotrebi simbola u ljudskoj komunikaciji i mišljenju, kao i već pominjani zbunjujući eksperimentalni nalazi (up. studiju Visnievskog i Medina, 1994, sekcija 7.6) vezani za problem interpretacije karakteristika, ukazuju na mogućnost praktično beskonačne reinterpretacije *značenja jednog istog znaka*. Oslanjajući se na konvencionalnost uređenja odnosa između znakova i denotata, i formulišući ovaj odnos kao instancu problema teorije igara - tačno sledeći čuvenu Luisovu (Lewis, 1969) analizu konvencija - analiziramo kognitivnu funkciju interpretacije u okviru strategijske teorije kognitivnih funkcija i pokazujemo da nema nikakvog smisla postaviti pitanje o tome da li je i kada *određena interpretacija nekog simbola racionalna ili nije*.

Kada je 1957. godine u knjizi „*Sintaksičke strukture*“ upotrebio rečenicu „*Bezbojne zelene ideje spavaju besno*“ („*Colorless green ideas sleep furiously*“, u engleskom originalu, Chomsky, 1957), Noam Čomski je to učinio kako bi predstavio primer rečenice čija je sintaksička kompozicija besprekorna i koja je istovremeno

potpuno besmislena. Bezbojne zelene ideje, svakako, ne bi mogle da spavaju besno: postoji čitav niz restrikcija koje se odnose na kombinovanje - *slaganje* - karakteristika denotata na koje upućuju elementi ove rečenice, zbog kojih smatramo da takvo *stanje stvari* ne može biti realizovano. Međutim, mi smo već naučili da je rasprava o karakteristikama koncepata i objekata na koje se oni odnose veoma klizava. Uz nešto mašte i narativne veštine, lako bismo konstruisali prihvatljiv alternativni diskurs u kome bi kultna rečenica Čomskog dobila prihvatljivu interpretaciju. Svakako, takav diskurs bi nas odveo na tragove literature poput Kerolove, ili poezije poput nadrealističke. Ali to je za našu diskusiju manje bitno: ono što je suštinski bitno jeste činjenica da prigodnom manipulacijom konteksta u kome se određeni jezički izraz nalazi mi *uvek* možemo da kreiramo alternativni - makar i samo simbolički - svet u kome taj jezički izraz nalazi interpretaciju koju u svom „stabilnom stanju" - stanju u kome denotira svet svakodnevnice - nema. Da to nije tako, instrukcije date dvema grupama ispitanika u studiji Visnievskog i Medina (up. sekciju 7.2) ne bi ni u principu mogle da indukuju različite interpretacije fizički istih karakteristika na fizički istim crtežima. Da to nije tako, ljudsko mišljenje nikada ne bi moglo da rešava probleme koji zahtevaju reinterpretaciju značenja simbola ili objekata kojima smo svakodnevno okruženi, ostajući zauvek zarobljeno u stanju funkcionalne fiksiranosti (Davidson, 2003). Da to nije tako, proza Luisa Kerola, poezija Artura Remboa, kubizam Braka i Pikasa ili Dišanovi *ready-mades* nikada ne bi imali nikakvog smisla, a mi dobro znamo da živimo u civilizaciji u kojoj svi ovi artefakti i te kako imaju smisla[92]. U odnosu na standarne modele konceptualne organizacije i kategorizacije u kognitivnoj psihologiji, rekli smo, nalazi Visnievskog i Medina predstavljaju potpuni eksces. Onda, u standardnim modelima konceptualne organizacije i kategorizacije, mnoga bitna rešenja do kojih procesi ljudskog mišljenja mogu da dođu, i koja kroz proces komunikacije mogu da postanu deo zajedničkog znanja, a samim time i semantičkih memorija članova određene zajednice, ostaju zagonetna. U odnosu na elementarnu intuiciju o procesima ljudskog mišljenja, intuiciju koja nam govori da je proces interpretacije nečega *kao nečega*, odn. proces semioze - tvorenja znakovne funkcije uopšte (Eco, 1997/2000) - fundamentalno zavistan od mogućnosti da se aktualni kontekst izmešta u alternativne kontekste, takvi eksperimentalni nalazi ne deluju problematično. Ono što izostaje je, očigledno, dobra teorija o procesu interpretacije, ili makar preciznija analiza razloga zbog kojih standardni modeli konceptualne organizacije i kategorizacije ne mogu da uklope nalaze koji se odnose na fenomene interpretacije.

Kao što smo nagovestili, diskusija ovakvih pitanja nužno je ograničena nekim za sada nepremostivnim metodološkim granicama. Interpretacija određene karakteristike nekog koncepta ili objekta uvek podrazumeva uspostavljanje *nove denotacije*. Jedna fizička struktura stimulacije - određeni crtež u eksperimentu Visnievskog i Medina - denotira dete koje pleše u jednoj, i dete koje se penje uz nešto u igralištu u drugoj grupi njihovih ispitanika. Jednostavno možemo da pokažemo da nas svaka promena denotacije određenog znaka vodi pravo u metodološke probleme tipa broja mentalnih reprezentacija Paderevskog koji smo diskutovali u sekciji 6.2. Uzmimo za primer koncept PRAVE u matematici pre i posle oktića neeuklidskih geometrija u XIX veku. Koncept PRAVE je svakako obuhvatao čitav niz apstraknih osobina pre nego što su otkrivene neeuklidske geometrije: beskonačnost je jedna od njih, dok se druge - ne manje apstraktne - odnose na specifičnu *sintaksu* geometrijske teorije, sintaksu koja npr. nije dozvoljavala da kroz tačku van prave prođe više njoj paralelenih prava od jedne. Pošto je otkrivena mogućnost neeuklidskih geometrija, poput hiperbolične ili eliptične geometrije, pojam PRAVE - poput i svih ostalih geometrijskih pojmova - morao je da doživi temeljnu reviziju, i počne da denotira jedan simbolički objekat sa skupom apstraktnih karakteristika koji nije isti kao skup karakteristika koji je odlikovao PRAVE pre otkrića novih geometrijskih sistema. Koncept PARALELNOSTI je odjednom morao da dobije novo značenje i počne da denotira čitavu novu klasu fenomena. Uzmimo za primer analizu kognitivnog funkcionisanja studenta koji se, posle upoznavanja sa euklidskom geometrijom sreće sa neeuklidskim geometrijama i uspešno ih savladava: sa koliko koncepata - *mentalnih reprezentacija* - PRAVE on barata? Da li on razvija koncept $\text{PRAVA}_2$, koji figurira samo u diskursu rasprave neeuklidskih geometrija, nasuprot konceptu $\text{PRAVE}_1$ koji ostaje rezervisan za „obične“ prave euklidske geometrije? Da li on nastavlja da koristi jednu istu mentalnu reprezentaciju pojma PRAVA koja se proširuje novim karakteristikama praćenim uslovima njihove upotrebe tj. specifikacije diskursa rasprave u kojima jedne, a ne neke druge karakteristike, određuju identifikacione procedure koje ustanovljavaju da li nešto jeste PRAVA ili nije? Podsetimo se diskusije iz sekcije 6.2: *ne postoji bihejvioralni test koji pod ma kojim uslovima može da nam da odgovor na pitanje o tome koliko mentalnih reprezentacija koristi kognitivni sistem u ovakvim situacijama.* Matematičari će, uobičajeno, u diskursu svoje discipline govoriti jednostavno o pravama; mi ćemo, na osnovu konteksta rasprave (euklidske ili neeuklidskih geometrija) znati *na šta oni misle kada govore*, ali nikada nećemo znati *kako misle to o čemu govore*, jer nam

bihejvioralni testovi (tj. opservabilno verbalno ponašanje) ne ostavljaju mogućnost da prebrojimo njihove mentalne reprezentacije. Već ovo ograničenje onemogućava mnoga interesantna pitanja o racionalnosti kognitivnih funkcija koja bi mogla da se postave i diskutuju, kao što je na primer pitanje ekonomičnosti reprezentacionog sistema.

Primer koji smo upravo diskutovali nije izabran slučajno niti je naš izbor vođen plemenitim naučnim poreklom geometrijskih pojmova. Primer koji koristi upravo apstraktne geometrijske pojmove - za koje svi dobro znamo *da jesu poučivi* - demonstrira kako je sasvim svejedno da li u diskusiji procesa interpretacije karakteristika tj. značenja pojmova govorimo o pojmovima koji posreduju između znakova i fizičkih objekata ili između znakova i *čisto simboličkih objekata.* Argument koji bi mogao da se suprostavi našoj inicijalnoj tvrdnji o tome da je u principu uvek moguće konstruisati diskurs u okviru kojeg dati simboli dobijaju alternativne interpretacije, tvrdeći da se u te svrhe nekad koriste mogući svetovi koji su simbolički i tek zamislivi, a ne realni kao svet koji odlikuju fizička ograničenja nad osobinama objekata, gubi kritičku snagu zato što se čisto simbolički objekti poput diskutovanih odnose upravo na taj empirijski svet fizikalnih nauka za koji verujemo da nas okružuje. „Obične“ prave euklidske geometrije su savršeno opisivale kretanje pod inercijom u „običnoj“ Njutnovoj mehanici, baš kao što se pokazalo da „neobičan“ svet Ajnštajnove relativističke mehanike zahteva neeuklidski geometrijski opis da bi se opisalo isto to kretanje. Promena u interpretaciji značenja pojmova koja počiva na izmeštanju aktualnog konteksta njihove upotrebe u neki alternativni kontekst, dakle, ne vodi ujedno i ka njihovom odstranjivanju iz diskursa u kome vlada mogućnost opovrgavanja (ili verifikacije) makar i u principu u neki diskurs koji je blizak tek zečevima iz Kerolove „*Alise u zemlji čuda*“; svaki alternativni svet u kome se interpretira značenje nekog simbola može podjednako da bude predmet ograničenja koja ne dozvoljavaju da *svi* odnosi unutar datog diskursa budu arbitrarni.

Studija Visnievskog i Medina iz 1994. godine može da se posmatra kao potpuno nova eksperimentalna paradigma u proučavanju učenja koncepata. U ovoj studiji, čije smo najbitnije rezultate diskutovali u sekciji 7.2, autori nisu propustili priliku da ukažu na predominantnost jednog teorijskog modela učenja koncepata koji su nazvali *standardnim modelom.* Standardni model, prema njima, odlikuju sledeće osobine: (*i*) *ajtemi na osnovu kojih se uči nova kategorija* (tzv. trening ajtemi u učenju kategorija) *su opisani kroz predefinisan i nedvosmislen prostor karakterisika*, (*ii*) *učenje obuhvata selekciju karakteristika iz ovog prostora*, (*iii*)

*proces selekcije karakteristika se oslanja na statističke osobine tih karakteristika* (tj. na raspodelu karakteristika u skupu trening ajtema) i (*iv*) *klasifikacija se bazira na poklapanju karakteristika koje obuhvata sintaksičke identitete*[93] (Wisniewski & Medin, 1994). Ako uporedimo stimuluse (dečji crteži iz Haris-Gudinaf testa) koje Visnievski i Medin koriste u svojoj studiji sa opisom stimulusa koje Šins, Goldstoun i Tibo preporučuju za studije učenja kategorija i kategorizacije u fleksibilnim prostorima karakteristika (up. sekciju 7.2), vidimo da je ova studija učenja kategorija paradigmatičan primer nove eksperimentalne paradigme koju preporučuje potonji teorijski rad (up. Tabelu 1. na str 8. u Schyns, Goldstone & Thibaut, 1998). Uzimajući ovu eksperimentalnu studiju i teorijski rad Šinsa, Goldstouna i Tiboa kao polazne tačke u određenju nove paradigme proučavanja kategorizacije u prostorima *interpretabilnih karakteristika* - gde karakteristike nisu unapred jasno definisane i gde tokom učenja i kategorizacije može da dođe do redefinisanja koncepata u smislu karakteristika koje ih određuju - nalazimo da se standardni model, čiju kritiku ona predstavlja, u potpunosti poklapa sa pretpostavkama *komponencijalnih semantičkih teorija* u formalnoj semantici. Prema komponencijalnim semantičkim teorijama, značenje određene rečenice je funkcija značenja njenih konstituenata. Paralela je jasna: ako su koncepti predstavljeni u simboličkom kodu u semantičkoj memoriji, onda oni predstavljaju određenu vrstu propozicionog znanja. Za tu vrstu propozicionog znanja onda važi da je značenje celokupnog koncepta (totalne propozicije koja ga opisuje u pogodnom „jeziku uma“ tj. simboličkom kodu) determinisano značenjem njegovih konstituenata, gde karakteristike u standardnom modelu učenja kategorija predstavljaju elementarne konstituente koncepata, dok njihove kompozicije (poput disjunkcija i konjukcija karakteristika) grade složenije konstituentne izraze u konceptualnim reprezentacijama.

Ako procesi interpretacije karakteristika u učenju kroz fleksibilne prostore karakteristika falsifikuju standardni model, možemo da se zapitamo da li postoji neka alternativa formulaciji problema koja bi nam pomogla da razumemo učenje koncepata koje obuhvata interpretaciju karakteristika. Mi verujemo da se ta alternativa nalazi u tradiciji analize *jezičkih konvencija* kroz koncepte teorije igara koja polazi od čuvenih analiza filozofa Dejvida Luisa (Lewis, 1969, 1975, 1979, Nolan, 2005) i predstavlja suštinski deo savremene neograjsijanske pozicije u *intencionalnoj semantici* (engl. *Intention-Based Semantics*, skr. IBS, Schiffer, 1982) koja se tradicionalno suprostavlja kompozicionalnim semantičkim teorijama. Analiza simboličke funkcije u terminima teorije igara odlično korespondira sa

strategijskom interpretacijom kognitivnih funkcija koju smo uveli u prethodnim diskusijama. Pogledajmo sada na koji način nas ta analiza vodi ka zaključku o izlišnosti primene koncepta racionalnosti saznanja u domenu elementarnih ljudskih simboličkih funkcija poput interpretacije karakteristika.

Prvo, fenomen interpretacije karakteristika ne dozvoljava da formalna postavka problema kategorizacije (ili učenja kategorija) ostane ista kakva jeste u tradicionalnim studijama ovog procesa. Tradicionalna formulacija bi bila da je neki koncept $X$ određen skupom karakteristika F = $\{f_1,..\ \mathrm{f}_n\}$, te da je odluka o njegovoj pripadnosti kategoriji $\mathrm{CAT}_1$ ili $\mathrm{CAT}_2$ neka funkcija tih karakteristika; dodavanjem pondera važnosti na karakteristike dobijamo klasičan model prototipa u učenju kategorija, dok dodavanjem relacionih struktura koje povezuju kategorije razvijamo neki od modela iz klase „teorije teorija“ (poput generativne teorije klasifikacije). Fenomen interpretacije karakteristika, međutim, nalaže da skup karakteristika F koje odlikuju koncept $X$ shvatimo skup koji *nije unapred definisan* - već upravo obrnuto, skup koji će biti određen tek tokom procesa kategorizacije ili učenja kategorija. Upravo ovo je suština kritike standardnog modela koju pružaju Visnievski i Medin. Pitamo se, onda, da li postoji odgovarajući formalizam kojim možemo da opišemo koncept $X$ ako on može da odgovara *potencijalno ma kojoj deskripciji* kroz neki arbitrarni skup karakteristika F? Susrećemo se sa veoma teškim pitanjem ekstrakcije karakteristika (koje smo već kratko diskutovali u III delu rasprave) odn. pitanjem „dekompozicije“ koncepta u odgovarajuće deskriptivne karakteristike. Formalno, mi možemo bez ikakvih posledica po našu diskusiju interpretacije karakteristika da tvrdimo da je iz koncepta $X$ moguće ekstrahovati *ma koji* skup karakteristika F. Očigledno je da se u ovoj analizi obraćamo težoj formi pitanja o interpretaciji (i kreiranju) karakteristika od forme koju diskutuju Šins, Goldstoun i Tibo. Ako proces interpretacije karakteristika posmatramo na ovaj način, polazeći od potpuno nedefinisanog, potencijalno beskonačnog skupa karakteristika F ma kog koncepta $X$, mi činimo nešto slično onome što je učinio Fodor u teoriji reprezentacije kada je *sliku* kao formu prekonceptualne reprezentacije definisao preko *principa slike* (Fodor, 2008): Ako je P slika nekog X, onda su delovi P slike delova X. Fodorovo shvatanje slike (*ikone*, kako se još koristi ovaj termin za prekonceptualne reprezentacije u formalnoj semantici i semiotici) daje onu formu konceptualne neodređenosti (otud njena *pre*konceptualnost) koja nam je potrebna za shvatanje koncepta koji odlikuje potencijalno ambivalentan i ne uvek definisan skup karakteristika. Možemo reći, analogno Fodorovoj definiciji slike, da

pre kategorizacije koncepta $X$ - tj. pre njegovog shvatanja *kao nečega* - bilo koji skup denotacija „delova“ koncepta X predstavlja njegovu interpretaciju, ostavljajući određenje „delova“ koncepta arbitrarnim onoliko koliko Fodorov princip slike ostavlja arbitrarnim način na koji sliku „sečemo“ upravo da bi naglasio konvencionalnu prirodu određenja „dela slike“ i „dela objekta“. Međutim, Fodor je komponencijalni semantičar, i za njega ne bi bio karakterističan sledeći korak naše analize.

Neka je $X$ koncept nekog subjekta $S$. Neka je $D$ skup referenata koji mogu da predstavljaju interpretacije arbitrarnih karakteristika („delova“, „komponenti“) nekog koncepta $X$. Kažemo da $d \in D$ *interpretira karakteristiku* $f$ koncepta $X$ subjekta $S$ ako $f$ denotira $d$ za $S$: $I(S, f) = d$. Neka je $\delta$ neki drugi subjekt, i generalno, neka je $\Delta$ skup svih drugih subjekata koji stupaju u simboličke interakcije sa subjektom $S$. Uvodimo sledeći princip: ako postoji subjekat $\delta$ (odn. subjekti $\Delta$) za koga je $I(\delta, f) = d'$ (odn. $I(\Delta, f) = d'$ u opštijem slučaju), i $d \neq d'$, onda za subjekta S važi $I(S, f) = d'$ *ukoliko on ima interesa da uđe i ulazi u koordinaciju sa subjektom* $\delta$ (odn. subjektima $\Delta$). Opisani proces u kome interpetacija $I(S, f) = d$ prelazi u $I(S, f) = d'$ nazivamo *procesom reinterpretacije karakteristike*; on očigledno počiva na značenju termina koordinacije između subjekata na koji smo se pozvali. Termin koordinacija koristimo na način koji je uveo Luis u prvim analizama konvencija (Lewis, 1969) i koji je kasnije (u nepromenjenom obliku) primenjen na analizu *jezičkih konvencija* (Lewis, 1975). Luis probleme koordinacije između određenih aktera vidi kao strategijske interakcije koje opisuje teorija igara, ali sa tom osobinom da one moraju da imaju *više ravnotežnih stanja*. Ako se podsetimo naše strategijske analize kognitivnih funkcija u prethodnoj sekciji (13.2), strategijske interakcije - igre - između dva ili više aktera nalaze ravnotežna stanja kao vrste karakterističnih rešenja koja zadovoljavaju određene uslove, poput Nešove ravnoteže u kojoj nijedan igrač nema interes da promeni ravnotežnu strategiju pod pretpostavkom da je neće promeniti ni drugi. Činjenicu da strategijske interakcije mogu da imaju više ravnotežnih stanja (npr. više Nešovih ravnoteža) do sada nismo koristili; ova činjenica uvodi klasu notorno teških problema selekcije odgovarajućeg ekvilibrijuma u teoriji igara. Zašto je za Luisovu analizu konvencija bilo neophodno analizirati igre sa više ravnotežnih stanja? Upravo da bi se definisala ključna karakteristika *konvencionalnosti* koja se odnosi na upotrebu simboličkih sistema u interakcijama. Ako problem izbora simboličkog sistema koji koristimo predstavlja problem koordinacije koji može da se operacionalizuje kao određena igra, konvencionalnost kao suštinska odlika simboličkih sistema nalaže to

da je skup aktera koji koristi određenu konvenciju podjednako mogao da koristi *neku drugu* konvenciju da bi rešavao problem koordinacije u kome se nalazi. Luisova analiza je formalne prirode i toliko puta ponovljena u literaturi o konvencijama i filozofiji jezika da nema potrebe da je eksplicitno uvodimo u našu diskusiju. Pod koordinacijom u gornjoj definiciji procesa reinterpretacije podrazumevamo upravo neko *konvencionalno* rešenje određenog problema koordinacije kako ga sada razumemo[94].

Teorijski aparat kojim sada raspolažemo omogućava nam da problem interpetacije karakteristika u učenju kategorija i kategorizaciji više ne razmatramo isključivo kroz prizmu komponencijalne semantike, već da *analizu premestimo na teren pragmatike.* Sada možemo da kažemo: u funkciji eksperimentalne manipulacije koja se koristi u paradigmatičnim eksperimentima poput onih u studiji Visnievskog i Medina, *subjekti ulaze u koordinaciju sa verovanjima koje sadrži eksperimentalna manipulacija* (odn. nivoi odgovarajuće nezavisne varijable: „*ovo su crteži gradske dece*“, „*ovo su crteži kreativne dece*“, itd). Da bi rešili taj problem koordinacije u koji su postavljeni, oni adaptiraju svoje dekompozicije koncepata (ili objekata) odn. skupove karakteristika F u skupove koji sadrže karakteristike *interpretabilne u koordinaciji sa verovanjima koje sadrži kontekst* indukovan eksperimentalnom manipulacijom. Podsetimo se: pre kategorizacije koncepta $X$, njegovog interpretiranja *kao nečega*, njega shvatamo kroz analogiju sa Fodorovim prekonceptualnim, ikoničkim reprezentacijama: svaka dekompozicija u karakteristike F koja se odnosi na koncept $X$ je podjednako dobra. Uvođenjem problema koordinacije uvodi se skup ograničenja na ekstrakciju karakteristika F iz koncepta $X$. Naš predlog, suštinski, počiva na vrsti „*uvođenja značenja u koncept spolja*“: kontekst indukovan eksperimentalnom manipulacijom razvija određene „linije otpora u kontinuumu“ (Eco, 1997/2000) u prekonceptualnoj reprezentaciji $X$, tako da se sada po tim linijama naziru dekompozicije koncepta $X$ u karakteristike F jer one mogu da se izvedu po *analogiji* sa dekompozicijama u skupove karakteristika koji su kanonički za eksperimentom definisan kontekst. Zbog potrebe da se moguće dekompozicije prenesu iz jednog semantičkog domena (kontekst koji je indukovan eksperimentalnom manipulacijom) u drugi („prekonceptualni koncept“ $X$ čija se kategorizacija razmatra), jasno je da procesi učenja kategorija i kategorizacija u paradigmi koja dopušta interpretaciju karakteristika moraju biti procesi *strukturalnog mapiranja* (Gentner, 1983); ako je to tako, procesi učenja kategorije i kategorizacije u paradigmi interpretacije karakteristika se suštinski

izjednačavaju sa simboličkim procesima koji realizuju analoško mišljenje i slične simboličke funkcije poput razumevanja i upotrebe metafora.

Posle analize predstavljene u prethodnoj sekciji (13.2), sledeći korak je sasvim sigurno očigledan: ako je problem učenja kategorija i kategorizacije u paradigmi interpretacije karakteristika moguće postaviti kao Luisov problem koordinacije, *to implicira da kognitivne funkcije kategorizacije u kontekstima gde dolazi do interpretacije i kreiranja karakteristika uzimaju strategijsku formu.* Tačno kako pokazuju Visnievski i Medin, promene konteksta koje indukuju eksperimentalne manipulacije u ovakvim eksperimentima vode ka ponovoljenim *reinterpretacijama fizički istih karakteristika vizuelnih stimulusa*; ne vidimo drugi formalizam koji bi mogao da odgovara konceptualnim reprezentacijama koje pokazuju ovakav nivo fleksibilnosti do strategijskog, koji u ovom slučaju predstavlja distribuciju verovatnoće nad potencijalnim interpretacijama karakteristika koncepta $X$.

Sledeća implikacija nas onda direktno vodi, baš kao u analizama u sekciji 13.2, u zaključak o izlišnosti ma kakvog pokušaja analize racionalnosti simboličkih funkcija poput interpretacije karakteristika. Ukoliko kognitivne funkcije kategorizacije uzimaju strategijsku formu, pripremajući tako konceptualni sistem na fleksibilne odgovore u funkciji promene konteksta, onda je distribucija verovatnoće koja odlikuje relevantne kognitivne strategije proizvod strategijskih interakcija kognitivnog sistema sa svojom okolinom, iz čega - videli smo u sekciji 13.2 na primeru analize Hjumovog problema kauzalne indukcije - sledi da će forma tih kognitivnih strategija zavisiti od funkcija korisnosti koje odlikuju aktere u strategijskim interakcijama. Čitalac koji u ovom argumentu vidi samo suviše složen način da se kaže da će se ljudski konceptualni sistem prilagođavati onom semantičkom kontekstu u kome se nađe izlaže se opasnosti da propusti suštinu argumenta. Iz naših analiza sledi da će osobine kognitivne funkcije kategorizacije biti posledica funkcija korisnosti sa kojima će taj sistem ući u strategijske interakcije u datom semantičkom kontekstu: dakle, osobine kognitivnih funkcija zavise od osobina mentalnih stanja koje suštinski *nisu kognitivne prirode.* Funkcije korisnosti odlikuju bilo kog kognitivnog aktera, ali govore samo o tome kako on subjektivno procenjuje objektivne vrednosti koje su predmet strategijskih interakcija sa okolinom. Konačno, još jednom: ne postoji nikakva *racionalnost saznanja.* Racionalnost je koncept koji je moguće razmatrati samo u odnosu na cilj koji neki akter - kakvim god kognitivnim sistemom on raspolagao - može sebi da postavi. Promenite taj cilj, promenićete i formu kognitivnih funkcija koje će taj akter pokušati da iskoristi da bi ispunio novi

cilj.

Dopustićemo sebi još jedno zapažanje o prirodi elementarnih simboličkih funkcija koje nas upućuje na dodatno razmišljanje o problemu racionalnosti saznanja. Konvencionalnost znakova u ma kom simboličkom sistemu smatra se osnovnom osobinom simboličkih sistema uopšte. Simboli koji se koriste u ma kom simboličkom kodu ne nose nužno nikakvu sličnost sa osobinama svojih referenata. Zbog toga znamo da su oni proizvod konvencije. Umberto Eko u semiotičkoj studiji „*Kant i kljunar*“ (Eco, 1997/2000) primećuje kako je onda modalitet tumačenja nekog stimulusa *kao simbola* od suštinskog značaja za naše tumačenje simboličkog koda uopšte. Eko predlaže da razlikujemo modalitete $\alpha$ i $\beta$: bilo šta posmatrano u modalitetu $\alpha$ ne predstavlja izraz simbola (tj. predstavlja stimulus na koji očekujemo reakcije karakteristične za njega samog), dok bilo šta posmatrano u modalitetu $\beta$ predstavlja izraz simbola (i tako predstavlja stimulus na koji očekujemo reakcije karakteristične za *nešto drugo* do njega samog - naime, na ono umesto čega on kao simbol stoji). Na slici 43. prikazan je sklop tri stilizovana pokreta četkicom umočenom u crno mastilo: da li možete biti sigurni da autor nije na umu imao univerzalni kvantifikator predikatskog računa? Zanimljiv princip koji, čini nam se, daje prigodnu *naučnu deskripciju* simbola, je onda sledeći: *simbol ne može biti simbol ako ga nije moguće interpretirati kao da nije simbol.* Iz ovog principa direktno sledi konvencionalna priroda simbola. Naše prethodne analize u okviru strategijske teorije kognitivnih funkcija nas onda vode ka jednoj veoma jednostavnoj formalizaciji konvencionalnosti simbola. Koristićemo oznaku $d$ kao skraćenicu za relaciju denotacije. Onda, $d(x) = x$ predstavlja denotaciju $x$ koji nije simbol (što ne znači da $x$ ne može biti indeks, ikona ili nešto drugo poreklom iz teorijskog rečnika semiotike, ali znači da je sada $d$ denotacija samo u uslovnom smislu te reči), dok $d(x) = \neg x$ predstavlja denotaciju $x$ koji jeste simbol. Već smo razumeli da za svako $x$ koje može da se interpretira kao simbol mora da postoji sumnja da ne može da se interpretira kao simbol jer u suprotnom $x$ gubi svoju konvencionalnu prirodu i prestaje da bude simbol. Dakle, jedini formalizam koji odgovara ovakvim osobinama $x$ ako $x$ jeste simbol je onaj u kome kognitivni sistem pod nekom konvencijom koju poštuje njegov korisnik tretira interpretaciju $x$ kao mešovitu strategiju: ($p$: $d(x) = \neg x$, *1-p*:$d(x) = x$), sa $p > 0$ i $p < 1$; ograničenja na verovatnoću $p$ su, očigledno, ključna. Ovaj rekurzivni princip kojim se određuje priroda simbola kao suštinski konvencionalnog elementa procesa tvorenja značenja (semioze) možemo da nazovemo *principom fundamentalne neodređenosti simbola.*

Slika 43. *Da li je ovo univerzalni kvantifikator predikatskog računa?*

Za ustaljene simboličke sisteme poput prirodnog jezika koji koristimo u referentnoj sociolingvističkoj zajednici mi možemo da pretpostavimo da $p \to 1$ za svaki simbol; kada, i pod kojim uslovima, će kognitivni sistem rešiti da promeni distribuciju verovatnoće koja odlikuje interpretaciju određenog simbola, koji u opštem slučaju uošte ne mora da odlikuje stabilnost denotata koja odlikuje simbole prirodnog jezika, pitanje je strategijskih interakcija u koje taj kognitivni sistem ulazi. *Zato je racionalnost saznanja koncept koji nema smisla u naučnoj analizi simboličkih funkcija*: naši kognitivni procesi slediće naše interese i kompromise koje ćemo morati da prihvatimo u potrazi za optimalnim rešenjima strateških interakcija, ne obrnuto. Oni će nam pružiti svu moć koju je prirodna evolucija uspela da nam podari u rešavanju takvih problema, ali u redu izvođenja psihološke analize njihove osobine ne mogu da prethode osobinama problema koje oni treba da rešavaju i ciljeva koje treba da zadovolje.

# Deo VII

# Zaključci o racionalnosti saznanja

U izgradnji pokušaja da se analizira problem racionalnosti saznanja u ovoj raspravi, prihvatili smo zatečeni pluralizam hipoteza i metodologija, mnoge neuredne argumente, inkonzistentne eksperimentalne rezultate i analizirali teorijske pozicije koje često nisu do kraja eksplicirane. Sve zatečeno, zatečeno je u obimnoj naučnoj periodici posvećenoj raznim problemima koji konstituišu savremenu debatu o racionalnosti u kognitivnoj psihologiji. Ono što mi nismo želeli da dopustimo, to je da pitanje ozbiljne naučne rasprave postane pitanje *žanra* u ma kom trenutku: kad god smo bili u situaciji da upoznajemo teorije, modele, argumente i metode koje eksperimentalisti ili teoretičari u diskursu debate o racionalnosti najčešće ne sučeljavaju, mi smo se trudili da izgradimo ili proširimo konceptualni aparat tako da ih učinimo uporedivim. Dugi niz godina već kognitivna psihologija ne obraća pažnju na ozbiljan problem koji nastaje zahvaljujući nekoj vrsti žanrovskog izbora metodologija i pristupa matematičkom modeliranju: laboratorije i grupe istraživača koje jednom napadnu određeni problem, konzistentno koriste iste ili slične eksperimentalne metode i iste ili slične pristupe modeliranju. Male promene

eksperimentalnih paradigmi su onda u stanju da obesmisle velike napore u teorijskom objašnjenju zahvaljući tome što nameću ili inspirišu sasvim drugačiji pristup modeliranju, na primer. Činjenica da se tako nešto događa u eksperimentalnoj i teorijskoj raspravi o racionalnoj prirodi procesa suđenja (up. 7.4), koji je istovremeno toliko fundamentalan i toliko elementaran, u najmanju ruku pruža razloge za zabrinutost.

Naša rasprava je tako bila motivisana i zapažanjem da se nešto *neuobičajeno* događa u novijoj istoriji kognitivne psihologije: naime, od naučne revolucije XVII veka do danas, ako uzmemo u obzir samo nauke koje se baziraju na eksperimentalnom metodu i matematičkom modelu, nismo upoznali nijednu u kojoj su se javljali zaključci inkonzistentni u onoj meri u kojoj se javljaju u kognitivnoj psihologiji. Prirodno, morali smo sebi da postavimo zadatak da raspravu privedemo većoj teorijskoj, konceptualnoj disciplini. Zato ne treba zaboraviti da zaključci koje navodimo u sledećim redovima važe samo uslovno. Oni, naime, važe za istraživanja u naučnoj disciplini u kojoj je čvrsto doneta odluka o standardizaciji eksperimentalnih postupaka, usvojen bar minimalan koncenzus o elementarnim koracima u matematičkom modeliranju empirijskih fenomena, i tako mogućnost *jasne i nedvosmislene naučne komunikacije* postavljena na čvrste osnove. Nismo želeli da dopustimo da se naša rasprava završi bez pominjanja ovog važnog problema koji prožima celu debatu o racionalnosti, celokupnu kognitivnu psihologiju (Milovanović, 2010), i koji je vidljiv na svakoj stranici ovog teksta na kojoj se raspravlja o inkonzistentnim eksperimentalnim rezultatima u pristupu *istim* problemima - a takvih rezultata, videli smo, ima pregršt.

Konačno, suštinska odlika naše rasprave je u tome što *nismo dopustili da se teorijski diskurs vodi bez neprekidne reference na diskurs eksperimentalnih posmatranja - i vice versa.* Teorijske rasprave u KKP na koje je disciplina naviknuta skoro po pravilu propuštaju da se blisko drže empirijskih opservacija. To omogućava izvođenje arbitrarnog broja zaključaka, izvođenje koje se ogleda u trenutnom haosu na teorijskoj sceni kognitivnih nauka, koji nije ništa uredniji od onog koji zatičemo u časopisima koji mahom objavljuju eksperimentalne radove. O izostanku teorijskih napora na drugoj strani, verujemo, posle diskusija u III delu ove teze, izlišno je dalje govoriti.

Zaključci koje navodimo korespondiraju opštim ciljevima naše rasprave kako su određeni još u I delu. Mi verujemo, dakle, posle svega do sada iznetog, da se pod uslovima preciznog postavljanja problema i nedvosmislene primene istih

eksperimentalnih metoda i principa modeliranja kognitivnih funkcija, na problem racionalnosti saznanja odnosi sledeće:

A. Racionalnost saznanja, pre svega, nije bila precizno određen pojam u teorijskoj konstrukciji kompjutacione kognitivne psihologije. Univerzalno primenljiva formulacija koju smo mi koristili, oličena u formalizmu $S(\psi|G,E)\rightarrow B$, koja je potpuno transparentna u metodologiji racionalne analize, interpretirana je kao tvrdnja o tome da kognitivni sistem pokušava da učini sve što može da bi optimizovao svoje ponašanje u odnosu na ciljeve i sredinu. Razlikovanje RACIONALNOSTI$_1$ od RACIONALNOSTI$_2$, koje smo uveli još u I delu ove rasprave, omogućilo nam je uvid u činjenicu da je pitanje koje se diskutuje u debati o racionalnosti zapravo pitanje o tome da li forma kognitivnih funkcija odgovara izboru ciljeva koje kognitivni sistem bira u procesu adaptacije. Naše analize iz VI dela rada, međutim, pokazuju da promene u formi kognitivnih funkcija - u strategijskoj interpretaciji - moraju da prate promene u izboru ciljeva. U tom smislu reči, *pretpostavka o racionalnosti saznanja je zaista nešto što se nalazi u najdubljim teorijskim osnovama KKP, ali je u pitanju tautološka, neinformativna tvrdnja*, tvrdnja koja ni na koji način ne može da uveća naše znanje o kognitivnim sistemima.

Na najdubljem teorijskom nivou, još u IV delu naše rasprave smo to pokazali, zanemareno je više temeljnih problema u logičkom razvoju teorije kompjutacione kognitivne psihologije. Ovo se odnosi na teorijsku konstrukciju ove nauke bez obzira na to koja se paradigma unutar nje „specijalizuje“: simbolicistička, konekcionističko/dinamička ili konstruktivističko/enaktivistička. Istorija nauke i filozofije nas uči da se, kao po pravilu, takve situacije jednom vrate po svoje; tačka u kojoj će kompjutaciona kognitivna psihologija morati da ispita mogućnost svoje sinteze sa domenom fenomenološke rasprave čini nam se posebno osetljivom po tom pitanju, pošto su problemi vezani za izvesne „*idealizacije*“ saznajnog i delatnog subjekta koji izgrađuje formalno-simbolički univerzum diskursa KKP poreklom upravo fenomenološke prirode.

Naši zaključci u VI delu rasprave pokazali su kako se „raspada“ postupak racionalne analize na neposrednim primerima u kompjutacionoj kognitivnoj psihologiji, dakle daleko od dubokih, teorijskih pitanja utemeljenja ove nauke, pošto se već prihvate sve pretpostavke na kojima je izgrađen njen diskurs. Postupak racionalne analize, videli smo, nikad ne vodi ka nedvosmislenom zaključku o racionalnosti kognitivnih funkcija koju pokušava da ustanovi; često je moguće

pokazati da se do ambivalencije u sudu o racionalnosti u nekom bihejvioralnom domenu dolazi tako što se pretpostavke same paradigme racionalne analize izvedu do svojih krajnjih konsekvenci.

B. *U okviru kognitivne psihologije, sa teorijskim i metodološkim aparatom kakav danas poznajemo, moguće je postojanje više različitih, egzaktnih, modela ili teorija koje objašnjavaju iste bihejvioralne opservacije.* Ovaj zaključak, potpuno kontraintuitivan, sledi iz zaključka o razmenljivosti eksplanatornih struktura dispozicionih i reprezentacionih teorija koji smo izveli u VI delu rasprave. U ovom slučaju, RACIONALNOST$_2$, koja se navodno ispituje empirijski za neku kognitivnu funkciju i koja može biti ustanovljena ili ne, uzima onu formu koju joj diktira prethodni izbor strukture naučne teorije, koja određuje verovanje u neku RACIONALNOST$_1$, verovanje koje prethodni empirijskim koracima. Kako se menja naše verovanje o odgovarajućoj eksplanatornoj strukturi - ograničeno racionalnoj (dispozicionoj) ili racionalnoj (reprezentacionoj) - menja se i interpretacija „empirijske" RACIONALNOSTI$_2$. Naš prvi predlog za izlaz iz ove neuobičajene situacije jeste da se kognitivnom teorijom određenog problema smatra *klasa ekvivalencije svih modela koji uopšte pokazuju dobru korespondenciju sa eksperimentalnim podacima.* Drugo rešenje leži u idejama *strategijske teorije kognitivnih funkcija* i konceptu *kognitivne strategije* koje smo predstavili u VI delu ove rasprave. Ako se takav teorijski pristup pokaže plodnim, međutim, shvatanje teorije u kompjutacionoj kognitivnoj psihologiji će morati radikalno da se promeni, a debata o racionalnosti - bar na način na koji je do sada vođena - napusti. Pod strategijskom interpretacijom kognitivnih funkcija i procesa, nijedno normativno opravdanje za određenu kognitivnu funkciju nije dovoljno da se u procesu integracije empirijskih opservacija i teorije donese zaključak o racionalnosti, pošto kognitivne funkcije zapravo predstavljaju ansambl različitih formalnih modela koji sa određenim verovatnoćama doprinose rešenju mogućih *različitih problema adaptacije* u *istim eksperimentalnim situacijama.*

C. *Teorijske pretpostavke kompjutacione kognitivne psihologije koje određuju njenu racionalnu strukturu kao naučne teorije* (RACIONALNOST$_1$) *su posledica krupne promene paradigme koja je u epohu simboličke mašine, u zeitgeist kompjutacionizma, uvela sve prirodne nauke. U tom smislu, kompjutaciona kognitivna psihologija je reprezent, tipičan primerak, jedna od centralnih nauka savremenice, ona nikako nije ona „stara naučna psihologija" koja je posmatrala model fizikalnih nauka kao ideal.* Poreklo pojma racionalnosti koje ova nauka

koristi nalazi se u najmanje tri povezane linije istorijskog razvoja nauke od XVII veka do danas: (a) u matematičkim raspravama koje vode razvoju teorije verovatnoće, (b) u filozofskim raspravama o mogućnosti mehanizacije umnih funkcija koja kulminira u filozofiji matematike prve polovine XX veka, i (c) u naučnim raspravama kroz koje model termodinamičke mašine počinje da smenjuje terminološki i konceptualni aparat klasične mehanike, kada u XIX veku počinje razvoj onog mišljenja koje će u drugoj polovini XX veka povezati pojmove racionalnog, adaptacije i optimizacije u jedinstvenu paradigmu racionalne analize. Dela poput danas klasičnog Šredingerovog eseja „*Šta je život*?“ (Schrödinger, 1944) svedoče o tome da savremene kognitivne nauke i evoluciona biologija dele zajednički istorijski uticaj smene paradigmi u fizici XIX veka.

D. *Opšti zaključak naše rasprave je da racionalnost saznanja nije naučni pojam i, u najstrožem smislu, ne treba uopšte da ima tretman unutar diskursa kompjutacione kognitivne psihologije.* Ovo ni na koji način, ponovimo to još jednom, ne implicira da je analiza normativnih okvira za rešenje određenih logičkih ili semantičkih problema, ili problema teorije verovatnoće, postupak koji nije naučne prirode. Ono što naš zaključak implicira jeste da takva analiza nije postupak koji može da ima *značenje unutar diskursa ma kakve empirijske nauke o kognitivnim funkcijama.*

*Debata o racionalnosti je pogrešno motivisana i nikada neće imati definitivan zaključak, što predstavlja najbolju preporuku da se debata prekine, bar u onim terminima u kojima se još uvek vodi.* Sva tri teorijska argumenta koja smo predstavili u delu VI ove rasprave upućuju na to da koncept racionalnosti saznanja nema nikakvu *informativnu vrednost.* Kada se ispune određeni uslovi pod kojima kognitivne funkcije optimalno slede ono što su definisani ciljevi kognitivnog sistema u nekom kontekstu, eventualno možemo da konstatujemo nekakvu kognitivnu racionalnost ali pojmovno potpuno izjedačenu sa pojmom optimizacije; mi smo izneli argumente zbog kojih smatramo da je ovakvo shvatanje pojma racionalnosti saznanja trivijalno. Kognitivne funkcije i procesi, prema shvatanju koje smo ovde razvili, nisu ni racionalni, ni iracionalni: to su prirodni procesi čijoj studiji možemo da pristupimo deskriptivno, kao studiji ma kojih drugih procesa u prirodi. Ta studija će biti ograničena više (veoma teškim) problemom merenja neopservabilnih konstrukata nego teorijskim okvirom koji bi nas obavezivao da kognitivne funkcije diskutujemo kao više ili manje racionalne. Naš drugi i treći argument u VI delu rasprave u suštini pokazuju da *sam proces formiranja verovanja*, kao deo celokupnog procesa $S(\psi|G,E)\rightarrow B$ koji generiše neko ponašanje u relevantnoj sredini, uzima onu

formu koju mu diktira uređenje strukture ciljeva kognitivnog aktera $S$. U oblasti viših kognitivnih funkcija, i sasvim izvesno u oblasti simboličkih funkcija koje fundamentalno počivaju na konvencionalnosti simbola uopšte, ne postoje nikakvi procesi koji bi vodili formiranju nekih verovanja $\psi$ koji bi bili nezavisni od izbora određenih ciljeva $G$. Naš prvi argument u VI delu rasprave na nešto drugačiji način pokazuje isto to: razlika između teorija ograničene racionalnosti i teorija iz tradicije racionalne analize, ako one uspevaju da objasne isti skup bihejvioralnih opservacija, počiva na načinu na koji se formulišu funkcije formiranja verovanja u njima; dva deskriptivna jezika koje smo prepoznali su potpuno razmenljivi, još jednom pokazujući da nešto poput kognitivnih funkcija uopšte ne može da se karakteriše predikatom „racionalnosti“ ili njenog odsustva. Eksplanatorna moć ovog koncepta je nikakva; šta više, mi tvrdimo da količina informacija koje on nosi iznosi nula. Ponovimo još jednom, to ne znači da je nemoguće oceniti u celini neko ponašanje kao racionalno ili ne, ili racionalno u većoj ili manjoj meri; to znači tačno to da je takvu ocenu moguće doneti samo sa stanovišta analize ciljeva, ali ne i sa stanovišta analize prirode samih kognitivnih funkcija koje su odgovorne za to ponašanje. Ali, rasprava o racionalnosti ciljeva je nužno normativna rasprava, i kao takva pripada normativnim disciplinama. Kognitivne funkcije su proizvod procesa evolucije, i poslužiće u adaptaciji svetu ma kakve ciljeve sebi subjekt postavio, ali će ukupna racionalnost tog subjekta moći da se oceni samo sa stanovišta izbora tih ciljeva, i nikako na osnovu prirode kognitivnih procesa koji mu omogućavaju da ih ispuni.

E. Na margini jednog Gedelovog neobjavljenog rukupisa nalazi se komentar: „*Mind: an ego using reason*” (Cassou-Nogues, 2005). Ova Gedelova jezgrovita rečenica možda predstavlja minimalani zapis suštine ideje o analizi kognitivnih funkcija koju zastupamo; možda kompjutaciona kognitivna psihologija treba da se podseti Kantove distinkcije između „*uma*“ i „*razuma*“ i povede diskusiju o svojoj reintegraciji u psihologiju uopšte, gde bi se - poput u Gibsonovom romanu „*Neuromancer*“ - odigrao susret između prethodno rasutih veštačke inteligencije i njoj komplementarne ličnosti. Međutim, pod ovakvim scenarijom, psihologija će se suočiti sa pitanjem racionalnosti izbora ciljeva, neminovno povezanim sa izborom forme kognitivnih funkcija (koje se pogrešno analiziraju kao „racionalne“ ili „ograničeno racionalne“), što znači da bi se ona ponovo suočila sa bolnim pitanjem svoje nepotpune separacije od filozofije.

Njena druga moguća budućnost je prerastanje u opštu nauku o kognitivnim

procesima - u kognitivnu nauku, jednostavno - koja bi više predstavljala neku vrstu matematičke epistemologije do oblast psihologije. Takvu naturalizovanu, matematičku epistemologiju interesovalo bi pitanje nastanka, razvoja i strukture kognitivnih sistema među drugim prirodnim sistemima *uopšte*, bez obzira na to da li oni nastaju u kontekstu nekog subjekta voljne radnje - nekakve ličnosti - čije određene ciljeve treba da zadovolje ili ne. Takva kognitivna nauka nalazila bi se u potpunom kontinuitetu sa fizikalnim i biološkim naukama, i potpuno lišena potrebe da raspravlja ma kakvo pitanje racionalnosti; ono bi za nju zauvek moglo da ostane pitanje filozofije nauke. Razvoj takve naturalističke discipline i te kako predstavlja izazov.

Tokom naše rasprave upoznali smo detaljno teorijsku konstrukciju kompjutacione kognitivne psihologije. Prema našem, već iznetom, mišljenju, najozbiljniji problem ove nauke je problem merenja, koji se u njenom slučaju pokazuje izuzetno složenim - daleko složenijim od odgovarajuće problematike u fizikalnim naukama ili čak biologiji. Međutim, postoji način shvatanja problema kompjutacione kognitivne psihologije koji otkriva da možda to što ona naziva problemom merenja zapravo *nije* njen pravi problem. Više puta smo povlačili analogiju između problema merenja mnoštva neopservabilnih varijabli, koji karakteriše kognitivnu psihologiju ograničenu bihejvioralnom metodologijom, i Bernštajnovog problema senzomotorne koordinacije. Sistem koji odlikuje veliki broj redundantnih stepena slobode je sistem koje je prilagođen tome da koristi svoje resurse strategijski, u onom smislu u kome se to u teoriji igara opisuje kao upotreba mešovitih strategija. Ako pretpostavimo da kognitivni sistem pristupa rešavanju kognitivnih problema na taj način, mi možemo da „grešku merenja“, koja se javlja kao ostatak informacije iz sistema koja nije mogla biti „uprosečena“ niti „randomizovana“ potezima koje preporučuje Ronald Fišer, sada prepoznamo kao nedostatak našeg teorijskog, matematičkog modela *koji propušta da opiše sve moguće načine* na koje bi sistem mogao da reši problemsku situaciju u kojoj ga mi, po pretpostavci, posmatramo. Ponovimo: Bernštajnov problem je problem za analitičara sistema sa velikim brojem redundantnih stepeni slobode, ne za sistem koji kontroliše te stepene slobode - ako je taj sistem u stanju da pristupi rešavanju problema adaptacije kao mešoviti strateg teorije igara. Ne može se greškom merenja nazivati informacija rasuta od strane kognitivnog sistema koji čini sve što može da umanji moć predikcije sopstvenog ponašanja od strane drugih kognitivnih sistema, ili koji pokušava da uporedo zadovolji više potencijalnih ciljeva ne znajući koji od njih će se realizovati u adaptivno relevantnom okruženju. Kao što

primećuje Džon Mejnerd Smit, onda je sasvim nebitno da li taj sistem posmatramo kao da je inherentno u stanju da proizvede slučajno ponašanje ili ne - on se za sve praktične potrebe naših analiza *ponaša kao da to čini* (Maynard Smith, 1982). Naš poslednji zaključak u ovoj raspravi nastavlja se na ovu primedbu velikog britanskog biologa i glasi: sve rečeno sugeriše *da priroda naučne analize takvog sistema mora da se promeni*, kako bi se razbila slika o prisustvu velike greške merenja u bihejvioralnim naukama i nestabilnosti ponašanja koje se proučava - slika koja se, posle promene paradigme koju smo predložili u ovoj raspravi, pokazuje kao pogrešna.

# Beleške

[1]Interesantno je da se nešto slično, po svemu sudeći, dogodilo Johanesu Kepleru tokom inferencija koje su vodile postavljanju njegovih znamenitih zakona, prethodnice Njutnove velike mehaničke sinteze, u osvit doba koje će obeležiti razvoj naučnog racionalizma. Prema Arturu Kestleru, Keplerovi proračuni vezani za analizu putanje Marsa zauzimali su nekih devet stotina stranica (!) sitnog rukopisa. Mogućnost greške u ručnim proračunima je, tako, sigurno bila visoka. Sam početak Keplerovih proračuna koji su vodili ka ideji njegovog prvog zakona obeležen je pogrešnim prepisivanjem tri numeričke vrednosti opservacija iz Braheovog atlasa; posle ogromnog numeričkog napora, pri samom kraju te inferencije, Kepler pravi nekoliko banalnih aritmetičkih grešaka koje skoro u potpunosti „potiru“ greške koje slede iz početnih pogrešnih podataka. Računica, pošto je naknadno proverena, sa ispravljenim vrednostima i bez aritmetičkih grešaka, poklapa se veoma blisko sa Keplerovim originalnim rezultatom. Nešto drugačija, ali po strukturi slična greška, dogodila se u njegovom izvođenju drugog zakona (Koestler, 1959).

[2]Uz ovu ogromnu većinu se naravno ne svrstavaju pristalice savremenih pokušaja naturalizacije fenomenologije (Petitot, Varela, Pachoud & Roy, 1999), čiji su napori toliko vredni poštovanja, ali čije glasove u ovom trenutku ne vidimo na koji način da uključimo u našu diskusiju racionalnosti saznanja.

[3]Pregled problema racionalnog izbora u psihologiji odlučivanja dala je na srpskom jeziku Dubravka Pavličić u časopisu „*Psihologija*“ (1997).

[4]Bernulijev rad, predstavljen originalno 1731. godine, sadržao je kao prilog pismo švajcarskog matematičara Gabrijela Kramera iz 1728, koje dokazuje da je Kramer došao do hipoteze o očekivanoj korisnosti nezavisno od Bernulija (Basset, 1984).

[5]Latinski izvornik je rad „*Specimen Theoriae Novae de Mensura Sortis*“, *Comentarii Academiae Scientiarum Imperialis Petroolitanae*, *Tomus V*, 1738, str. 175-192; referenca koju mi koristimo je prevod doktora Luisa Somera sa Američkog univerziteta u Vašingtonu, objavljena u časopisu „*Econometrica*“, Vol. 22, Br. 1, (Jan., 1954), str. 23-36.

[6]Naša iskustva u svakodnevnim raspravama sa kolegama i prijateljima o problemu Sv. Petrovgrada pokazuju da, raspravljujući vrednosti u evrima, ljudi nisu spremni da ponude više od dvadesetak evra za učešće u ovakvoj igri; dominiraju odgovori u malim iznosima od oko pet evra.

[7]Funkcija korisnosti koju prikazujemo na Slici 1a. je stepena funkcija. Bernulijeva originalna formulacija je podrazumevala logaritamsku funkciju korisnosti. Stepene funkcije se danas uobičajeno koriste za opis empirijskih funkcija korisnosti; kasnije ćemo videti da su tokom istorije diskutovane razne forme ovih funkcija sa različitim osobinama. Ključna osobina ove funkcije, posle Bernulijeve analize, jeste konkavnost; stepena funkcija, sa eksponentom između 0 i 1, kao i logaritamska funkcija, konkavne su na celom domenu vrednosti. Interesantno je da je upravo Bernulijev rad iz prve polovine XVIII veka predstavljao deo inspiracije za Fehnerov razvoj psihofizičke funkcije u XIX veku (Masin, Zudini, & Antonelli, 2009, Dzhafarov & Colonius, 2011).

[8]Pažljiv čitalac je sada mogao da primeti da koristimo oznaku $U(\cdot)$ za funkciju korisnosti umesto prvouvedene oznake $u(\cdot)$. Nije u pitanju štamparska greška: ova finesa u notaciji je bitna, a objasnićemo je u trenutku kada budemo mogli da motivišemo razliku između Bernulijeve funkcije korisnosti i tzv. fon Nojman-Morgnešternove funkcije korisnosti za rizične lozove.

[9]Bernulijevo izvođenje počiva na eksplicitnom izlaganju nekoliko principa na koje se oslanja, ali u slučaju njegovog rada, kao što ćemo uskoro videti, još je rano govoriti o aksiomatskoj analizi odlučivanja u uslovima rizika.

[10]Sistem aksioma koji ovde predstavljamo se razlikuje od originalne formulacije fon Nojmana i Morgnešterna koja se danas smatra donekle anahronom i manje elegantnom od ovde izložene. Aksiomatizacije teorije očekivane korisnosti inače variraju u tehničkim detaljima ali uvek vode u iste zaključke; za jednu varijaciju veoma sličnu ovde iznetom sistemu upućujemo čitaoca na referencu Fishburn, 1988. Aksiomi A1 i A2 se često izlažu kao jedan aksiom, tako da se sreću aksiomatike koje sadrže samo tri aksioma. Takođe, za proširenje teorije na kontinuiran slučaj, neophodno je uvođenje nekih aksioma od čisto tehničkog značaja. Danas se najčešće diskutuje Sevidžova aksiomatizacija tzv. *subjektivne očekivane korisnosti* (engl. Subjective Expected Utility, SEU, Savage, 1954/1972), ali je ona tehnički suviše zahtevna za potrebe ove uvodne diskusije.

[11]Često navođen „čuveni" uzorak Aleovih ispitanika, među kojima su se nalazili učesnici konferencija u Luvenu (1951) i Parizu (1952), skoro svi slavni ekonomisti i matematičari, uključujući nekoliko nobelovaca među njima, ipak je samo mit. Ovo se često koristi da bi se retorički naglasila snaga Aleovih empirijskih nalaza. Skorije istorijske analize (Jallais & Pradier, 2005) pokazuju da su, od Aleovih ispitanika, Kenet Erou, Bomol, Fridman, Maršak i Sevidž uspešno (normativno racionalno) rešili anketu koju je Ale pripremio izgleda *tek posle* konferencije u Parizu 1952, Oskar Morgneštern i Pol Samjuelson uopšte nisu odgovorili na nju, dok su se Melinvud, de Fineti (!) i Šekl „uhvatili" u Aleove eksperimentalne zamke - ali ne one koje se odnose na čuveni paradoks, već neke druge. Izgleda da je jedini proslavljeni naučnik koga je Ale uspeo da „prevari" paradoksom bio Leonard Džimi Sevidž (!!), autor teorije subjektivne očekivane korisnosti, kome je Ale postavio dvadesetak pitanja na pauzi za ručak konferencije iz 1951. i ustanovio da niti jedan njegov odgovor nije bio u skladu sa aksiomatikom racionalnog izbora.

[12]Up. Harperovo istraživanje, prema Glimcher, 2004, str. 294; ponašanje *pataka* u ovom istraživanju zadovoljava neke normativne kriterijume teorije igara koji su daleko komplikovaniji od ma kog problema teorije individualnog racionalnog izbora.

[13]Terminologija teorije izgleda dovodi do zbrke koja nije neophodna: ono što smo u dosadašnjem izlaganju zvali funkcijom korisnosti, Kaneman i Tverski nazivaju funkcijom vrednosti.

[14]U kumulativnoj teoriji izgleda, ponderisanje verovatnoće za dobitke se ne poklapa nužno sa ponderisanjem verovatnoća za gubitke, upravo kao što ni funkcija vrednosti za dobitke ne mora (i u opštem slučaju nije) ista kao i funkcija vrednosti za gubitke.

[15]Pod *kognitivnim konstruktivizmom* podrazumevamo čitav spektar teorijskih polazišta u kognitivnoj psihologiji za koje je zajednički stav da um nije sredstvo pasivne recepcije sveta koji postoji objektivno i nezavisno od subjekta, već da *subjekat svojom aktivnošću nameće značenja i red informacijama* iz okoline tako konstituišući ono što je svet *za njega*.

[16]Iako je esej „*Istina i verovatnoća*" jedini Remzijev naučni rad koji je direktno relevantan za naučnu psihologiju, on otkriva da je Remzi bio odlično informisan o diskusijama merenja u

psihologiji aktualnim u njegovom vreme, npr. diskutujući problem merenja subjektivnih verovanja u terminima poput „jedva primetnih razlika“.

[17]Dok ovaj stav deluje sasvim prirodno nama koji nasleđujemo poziciju razvijene kognitivne psihologije, on je u potpunom neskladu sa *teorijom bihejviorizma*, koja potiče iz istog vremena kada je pisan Remzijev esej „*Istina i verovatnoća*“. Razvoj ovakve bihejvioralne metodologije, koja sebi za osnovni cilj postavlja merenje neopservabilnih konstrukata na osnovu opservabilnog ponašanja, nije u skladu sa teorijskom pozicijom bihejviorizma u kojoj je govor o neopservabilnim konstruktima u nauci zabranjen. Interesanto je, iz istorijske perspektive, postaviti pitanje o tome kako jedan ovakav impuls - od Remzijevog eseja iz 1926. godine do vNM aksiomatike iz 1947 - nije došao u dodir sa mejnstrimom psihološke misli.

[18]Pored mnogih drugih opisa koji se sreću u literaturi: *psihologija zdravog razuma* (engl. *Common-Sense Psychology*, Stich & Nichols, 2003), *intencionalna psihologija* (engl. *Intentional Psychology*, Fodor, 1987) i sl.

[19]Tačno značenje engleske reči *token*, koja se u izvornom obliku koristi u srpskom, teško je jednostavno objasniti. Token je osnovni objekat nekog sistema koji nosi neku vrednost: recimo, slovo *a* je token nekog jezika. U lingvistici i kompjuterskim naukama, token se shvata kao realizacija određenog *tipa*. Na primer, u sledećem izrazu: „*Ana je budna*“, postoje tri tokena '*a*' istog tipa „a“ (slova „a“ u srpskom jeziku). U šahu, ako ga posmatramo kao formalni sistem tj. formalnu, matematičku igru sa striktnim pravilima, svaka figura je token. Postoji osam tokena istog tipa „*pešak*“ na strani svakog igrača, jedan token tipa „*kraljica*“ na strani svakog igrača, itd. U šahu, pravila igre su pravila manipulacije tokenima: šah je jedan formalni sistem.

[20]Treba da bude jasno da su diskusije u filozofiji uma o kojima je reč izuzetno značajne i plodne, ali ovde je reč o tome da je za analizu racionalnosti u okviru KKP dovoljno i ovakvo široko, fundamentalno i apstraktno predstavljanje FP. Pokazećemo da plodna analiza koncepta racionalnosti može da se izvede i u odnosu prema samo ovim fundamentalnim osobinama standardne paradigme.

[21]Englesku reč *emergentism*, kao i glagol *to emerge*, teško je prevesti na srpski jezik u pokušaju da se zadrži njegovo izvorno značenje. Latinski glagol *ēmergere* - izroniti, pojaviti se - ne predstavlja dobar putokaz, iako *to emerge* na engleskom može da upućuje i na radnju izranjanja iz vode ili podizanja nečega iz zemlje. Emergentni proces bismo mogli da opišemo jezikom geštalt psihologije kao proces u kome se figura izdvaja u odnosu na pozadinu, proces kroz koji se prethodno nevidljiva struktura pomalja kao opservabilna celina iz prethodno nediferenciranog perceptivnog polja. Specifično značenje koje ovaj koncept ima u kognitivnim naukama vezano je upravo za opisani proces razvoja određene globalne, makroskopske odlike sistema koja se prepoznaje kao semantički interpretabilna celina u masi interakcija među elementima sistema na nižem, molekularnijem nivou opisa.

[22]Neki autori smatraju da je moguće govoriti o kontinuumu emergentnih karakteristika koji polazi od slabe emergentivnosti, u kom slučaju je globalno ponašanje sistema moguće objasniti polazeći od opisa nad lokalnim elementima posle posmatranja ponašanja, ka jakoj emergentivnosti u kojoj je ekstremno teško dovesti globalne odlike u vezu sa opisom na mikro-nivou i konačno maksimalne emergentivnosti koja predstavlja slučaj kada svođenje uopšte nije moguće (Assad & Packard, 1992/2008).

[23]Imena pravila u Volframovoj notaciji su brojevi u dekadnom sistemu koji nastaju prevođenjem iz binarnog sistema niza jedinica i nula koji opisuje stanje aktivacije ćelije u celularnom automatu u odnosu na prethodno stanje te ćelije i njenih suseda. Kada se konfiguracije početnih stanja poređaju po konvenciji koja se uvek poštuje, dobija se jedinstven sistem za imenovanje pravila. Ista konvencija koju poštuje Volfram u „*A New Kind of Science*“ (Wolfram, 2002) poštovana je u imenovanju pravila i ovde (Slika 8).

[24]Ovakvi matematički dokazi su veoma složeni i zasnovani na postupku kojim se pokazuje da kompjutacioni sistem za koji se tvrdi univerzalnost može da simulira rad drugog kompjutacionog sistema za koji je univerzalnost već dokazana. U dokazu univerzalnosti za automat sa pravilom 110, Kuk je morao da analizira sve stabilne strukture koje on razvija i njihove interakcije, da bi ustanovio koje od njih mogu da se iskoriste da bi se simulirao sistem tzv. cikličnog tagovanja, za koji je od ranije poznato da je Tjuring-ekvivalentan (Cook, 2004).

[25]Naš prevod engleskog *to enact* kao *ustanovljavanje* je najpribližniji izvornoj konotaciji koju enaktivistički autori koriste. U engleskom jeziku koristimo glagol *to enact* da označimo odluku o donošenju nekog zakonskog akta. S druge strane, *to enact* ima i konotaciju *odigrati*, kao za odigravanje nekog pozorišnog komada. U tom smislu, kognitivni sistem u interakciji sa okolinom uzima određenu ulogu dok istovremeno počinje da tu okolinu perceptivno i kategorijalno opaža kao zakonitu, regularnu. Regularnosti okoline koje će neki sistem prepoznati zavise i od strukture okoline i od strukture kognitivnog sistema koji se strukturalno povezuju. Konotacija vezana za donošenje nekog pravnog akta nam omogućava da semantički naglasimo to da svet, prema enaktivistima, nije nešto unapred definisano i dato kognitivnom sistemu, već nešto što on ustanovljava kroz svoje interakcije sa okolinom koja *a priori* nudi ogroman broj interpretacija.

[26]Čitaoce knjige „*The embodied mind: Cognitive science and human experience*“ Varele, Tompsona i Rošove lako će zbuniti njihova notacija pravila za elementarne celularne automate koja odstupa od danas kanoničke notacije koju koristi Volfram u „*A New Kind of Science*“ (Wolfram, 2002). Wolframovo pravilo 133 (binarno: 10000101) enaktivisti koriste u ovoj knjizi u notaciji 01101110 (Slika 8.4, Varela, Thompson & Rosch, 1991).

[27]Autori u savremenoj filozofiji nauke primećuju da veza između Marovog algoritamskog nivoa 2 i kompjutacionog nivoa 3 nije veza karakteristična za objašnjenje putem mehanizma (Wright & Bechtel, 2007). U objašnjenju putem mehanizma, pojave na kompjutacionom nivou 3 kognitivne analize morale bi da budu dekomponovane u strukture i operacije na algoritamskom nivou 2, dok se u Marovoj formulaciji ova dva nivoa razlikuju po tome što pružaju različit *opis* fenomena. Ipak, algoritamski nivo je jasno formulisan kao nivo na kome se očekuje otkriće kognitivnih mehanizama; filozofija nauke onda ukazuje na to da bi algoritamski nivo 2 i implementacioni (neurofiziološki) nivo 1 mogli da budu jedna ista stvar (Wright & Bechtel, 2007). Karakteristika KKP je, međutim, razvoj procesnih teorija na nivou 2 čak i kada ne postoje neurofiziološka ograničenja sa implementacionog nivoa.

[28]Rudolf Karnap povlači preciznu granicu između *dispozicionih* i *teorijskih* pojmova koja će kasnije biti od suštinskog značaja za neke naše argumente. Uviđajući (ne bez potonje kritike u istoriji filozofije nauke) da je jezik svake naučne teorije moguće podeliti na opservacijski jezik, $L_0$, sa vokabularom $V_0$, i teorijski jezik, $L_T$, sa vokabularom $V_T$, Karnap predlaže da su dispozicioni pojmovi deo proširenog opservacijskog jezika $L'_0$: to su svi pojmovi koji nastaju kada se pokaže

da pod uslovima S, na nekom predmetu posmatranja osmatramo fenomen R; onda kažemo da taj predmet karakteriše dispozicioni pojam $D_{SR}$(Carnap, 1956/1985). Vidimo da je Karnap dispozicione pojmove razumeo kao bliže opservacionom jeziku nego teorijskom jeziku, iako su npr. većina teorijskih koncepata psihologije upravo dispozicioni pojmovi.

[29]Za razliku od neuropsihologije, koja raspolaže metodama poput odslikavanja funkcionalnom magnetnom rezonancom (fMRI), magnetne encefalografije (MEG) ili transkranijalne magnetne stimulacije (TMS); ipak, ni neuropsihološki pristup ne može da izgradi teoriju o kognitivnim procesima bez referenci na opservabilna ponašanja.

[30]Ovde bi trebalo dodati „*...i među drugim ljudima*" - kognitivni sistem čoveka u svom ekološkom okruženju nalazi i druge kognitivne sisteme na čije se kapacitete rutinski oslanja u rešavanju problema adaptacije. Ipak, diskusija socijalnog aspekta kognicije bi nas na ovom mestu odvela predaleko od problema kojima je diskusija posvećena.

[31]Markus ne koristi originalno ovaj argument u debati o racionalnosti, već u drugoj, takođe veoma razgranatoj teorijskoj debati u kognitivnim naukama koja se odnosi na pitanje *modularnosti kognitivnih funkcija* (up. Marcus, 2006, Fodor, 1983, 2000).

[32]Sume u deliocima termina za verodostojnost i *a priori* verovatnoće očigledno samo vrše neophodnu normalizuju da bi se *a posteriori* verovatnoće dovele u formu pravih funkcija gustina verovatnoće. Bez ikakvog gubitka informacije termini u deliocima mogu da se uklone, i u literaturi često srećemo formu Bejzove teoreme koja ih ne koristi: $P(H|A) \propto P(A|H) \cdot P(H)$.

[33]Engleski termin je „*degenerate distribution*", mi ga prevodimo kao *svedeni loz*, u nadi da je taj prevod više u duhu našeg jezika.

[34]Naša ilustracija koncepta komonotonosti odnosi se na slučaj rizika u kome su verovatnoće objektivno poznate; u slučaju neizvesnosti ona ostaje ista, dok se u formalizaciji koristi tek nešto drugačija notacija.

[35]Uz jedan jednostavan dopunski uslov *konzistencije gubitaka i dobitaka* koji ovde ne diskutujemo (up. Wakker & Tversky, 1993).

[36]Lavice, inače, češće love u grupi nego same, iako se čini da je njihov uspeh u lovu podjednak u te dve situacije, čineći problem objašnjenja prosocijalnog ponašanja lavica u lovu jednim od najinteresantnijih problema bihejvioralne ekologije (Mangel & Clark, 1988).

[37]Čitalac koji je upućen u problematiku probabilističke kauzalnosti razumeće zašto biramo da diskutujemo ovu oblast kroz razmatranje isključivo binarnih varijabli; verujemo da će se čitalac koji nema ekspertsko znanje u oblasti probabilističke kauzalnosti sam uveriti koliko je duboka problematika vezana za ovaj „pojednostavljen" svet binarnih varijabli. Svi suštinski problemi probabilističke kauzalnosti i kauzalne indukcije mogu da se demonstriraju na primeru binarnih varijabli. U tom smislu, smatramo da nije neophodno dalje komplikovati ovu formalno veoma složenu oblast uvođenjem bilo diskretnih uzroka i posledica koji imaju više nivoa, ili uzroka i posledica koji su kontinuirane prirode.

[38]Periodika je puna opomena o tome koliko ovaj termin daje pogrešnu konotaciju razumevanju kauzalnih mreža. Parametri ovakvih modela podjednako mogu da se ocenjuju bejzijanskim i klasičnim metodama, a upotreba termina „bejzijanske mreže" se bazira na labavoj analogiji prema kojoj analiza probabilističke kauzalnosti počiva na prethodnom poznavanju relevantnih verovatnoća.

[39]Očigledno svaku probabilističku kauzalnu mrežu posmatramo kao da je potpuno umrežena u probabilistički kauzalni *Univerzum*, tako da „prvi uzroci“ u svakoj kauzalnoj mreži, tj. oni koji nemaju druge uzroke koji bi objašnjavali verovatnoću njihove pojave, moraju da imaju egzogeno date verovatnoće pojavljivanja.

[40]Odnos između probabilističkog kontrasta, $\Delta P$, i kauzalne moći, $p_c$, je fundamentalan. Za njegovu eksplikaciju potrebno je razmatrati tzv. *parametrizacije kauzalnih mreža*, pri čemu se jednostavno pokazuje da i jedna i druga mera predstavljaju normativnu ocenu intenziteta kauzalnog odnosa: $\Delta P$ pod tzv. *linearnom parametrizacijom*, a $p_c$ pod tzv. *Noisy-Or za generativne* i *Noisy-And-Not za preventivne uzroke* (Griffiths & Tenenbaum, 2005).

[41]U nekim studijama je ovo opterećenje memorijskih kapaciteta tokom kauzalnog učenja poželjan efekat u odnosu na ciljeve studije, npr. Rehder & Milovanović, 2007.

[42]Ovo predstavlja dodatni razlog zašto ne treba poistovećivati $\Delta P$ sa RW modelom, uprkos tome što prvi predstavlja asimptotsku vrednost drugog u širokoj klasi eksperimentalnih nacrta.

[43]Eksperimenti 3a i 3b studije Peralesa i Šenksa iz 2008. zapravo obezbeđuju podršku za teoriju kauzalne moći Patriše Čeng; Perales i Šenks pokazuju da evidencija koju ti eksperimenti pružaju nije dovoljna tek na nivou analize odgovora pojedinačnih ispitanika (up. Perales & Shanks, 2008). Ne možemo da se uzdržimo od komentara da analiza odgovora pojedinačnih ispitanika skoro po pravilu izostaje u studijama kauzalnog učenja.

[44]Optimalne vrednosti parametra su određene postupkom *grid-pretrage* prostora parametara, up. Perales & Shanks, 2007, str. 587-586.

[45]Racionalnu bejzijansku teoriju Andersona ovde ne razmatramo pošto su sve savremene bejzijanske teorije kauzalnog učenja formulisane u jedinstvenom teorijskom okviru kauzalnih modela koji u vreme razvoja Andersonovog modela još nije bio široko prihvaćen (Anderson & Sheu, 1995).

[46]Dobijeni $R^2$ je .66 za model kauzalne podrške, ali studija Peralesa i Šenksa ostavlja neke nejasnoće oko formalnog tretmana ovog modela. Autori tvrde, dosta neuobičajeno, da je bilo neophodno „uprosečiti“ predikcije modela kauzalne podrške za generativne i preventivne uzroke da bi se evaluirala njegova predikcija, što tim rečima formulisano nema puno smisla. Lu i saradnici su u studiji iz 2008. godine izvršili reanalizu prediktivne moći modela kauzalne podrške i podataka koje koriste Perales i Šenks i dobili daleko višu vrednost $R^2$ (vrednost koja je porediva sa vrednošću najprediktivnijeg modela koji je ta studija otkrila; up. Perales & Shanks, 2007, Lu et al, 2008).

[47]Luova i saradnici koriste drugačiju terminologiju: prema njima, takve apriori verovatnoće su *SS*, što je skraćen od engl. „*Sparse and Strong causes*“ (up. Lu et al, 2008).

[48]Namerno izbegavamo upotrebu određenja „normativni modeli“ na ovom mestu.

[49]Čitaoca upućujemo na studiju Raaijmakers, 2003, za pregled i modeliranje više klasičnih studija efekta raspodeljenog vežbanja, i Glenberg, 1979, gde eksperiment 2 (Slika 4. u originalnom radu) ilustruje veoma jasan efekat interakcije raspodeljenog vežbanja sa odlaganjem testa slobodne reprodukcije.

[50]U radu iz 1997. godine Skuler i Anderson razvijaju vezu između faktora $Q$ i $H$ u racionalnom modelu (Schooler & Anderson, 1997)

[51]Anderson i Skuler tvrde da je takvu predikciju dao još Herbert Sajmon u razvoju funkcije distribucije verovatnoće koja danas nosi ime Jul-Sajmonova distribucija (Simon, 1955b).

[52]Zahvaljujem se Milanu M. Ćirkoviću na ovoj sugestiji o broju parametera standardnog i klasičnog kosmološkog modela.

[53]Misli se na spejs-šatl Čelendžer američke svemirske agencije NASA, u čijoj je tragičnoj eksploziji pri poletanju 28. januara 1986. godine život izgubilo svih sedam članova posade.

[54]Preko ovog argumenta, ako su ga uopšte razmatrali, Anderson i Milson olako prelaze, pretpostavljajući da je distribucija kontekstualnih znakova nezavisna od distribucije prethodne upotrebe za dati memorijski trag $A$; u fusnoti 2. u zajedničkom radu iz 1989. godine navode kako ta nezavisnost možda nije ograničenje koje karakteriše ljudsko pamćenje i da bi bilo interesantno proveriti kako odustajanje od te pretpostavke utiče na rezultate njihovog modela (up. Anderson & Milson, 1989, str. 704).

[55]Termin *inklinacija* koristimo kao prevod na srpski jezik originalnog engleskog termina „*bias*“, pa istraživački program „*Heuristics and biases*“ prevodimo kao istraživački program *heuristika i inklinacija*. Na ovom mestu je neophodna napomena o tome da je praktično nemoguće prevesti englesku reč „*bias*“ na srpski jezik tako da se očuva kompletna konotacija koju ona nosi u kontekstu debate o racionalnosti. Mnogi autori u objašnjenju značenja termina „*bias*“ u debati o racionalnosti povlače analogiju sa konceptom *nepristrasne ocene* u statističkoj teoriji. U statistici, statistik koji sistematski odstupa od prave ocene parametra u populaciji, *nije pristrasan*. Kvalitetne statističke ocene, poput ocene maksimalne verodostojnosti, jesu nepristasne. Dakle, ukoliko kognitivni sistem nije „intuitivni statističar“, kako tvrde pristalice racionalnih paradigmi, već su njegove ocene „pristrasne“, on pokazuje određeni „*bias*“. Međutim, koristiti termin pristrasnosti kao prevod za engleski „bias“ menja konotaciju izvornog termina jer u duhu srpskog jezika nosi previše socijalnu konotaciju (npr. „pristrasnosti u oceni nečijeg rada“). Drugo, u samoj debati o racionalnosti termin „*bias*“ ima konotaciju koja je šira od one omogućene analogijom sa statističkom ocenom parametara. Termin „greška“ na srpskom, opet, nije pogodan, jer nisu sve greške koje mogu da nastanu u radu kognitivnog sistema takve da bi bile svrstane u klasu „*biases*“. Konačno, opredelili smo se sa izraz *inklinacije,* u nameri da se oslonimo na konotacije izraza poput „...toga da neka osoba pokazuje *inklinaciju* da misli na taj i taj način...“ ili „... da je osoba X *inklinirana* da reaguje na takvo ponašanje uvek na isti način...“ i sl. Termin „*inklinacija*“ u svojoj konotaciji nosi to da je u pitanju karakteristika koja je *inherentna* određenoj osobi, određenom načinu mišljenja ili određenim postupcima, što pokriva bitan deo značenja termina „bias“ u debati o racionalnosti. Pored toga, ovaj termin može da se upotrebi tako da označava odstupanje od normativnog rešenja, kao u sledećem primeru: „U suštini, svaki put kada se takvo mišljenje suoči sa novim tipom problema, one ne uspeva da ga reši, jer inklinira ka pokušaju da ga reši kao da je u pitanju problem iz već poznate klase“.

[56]Toliko o efikasnosti nastave statistike na postdiplomskim studijama psihologije.

[57]Heuristika referente tačke i podešavanja je naš prevod na srpski originalnog engleskog termina „*anchoring and adjustment heuristic*“.

[58]Poreklo logaritma u odnosu verodostojnosti je u činjenici da logaritamski oblik Bejzove teoreme, kada se ona izrazi kao odnos *a posteriori* verovatnoća da je pravi model koji generiše podatke $D$ (engl. *bayesian odds*) jedan od dva alternativna ($h$ ili $\bar{h}$), postaje aditivan, pa postaje očigledno da član (39) aditivno doprinosi evidenciji da je tačan model $h$ preko njegove verovatnoće *a prirori*. Jednačina (39) se tek neznatno komplikuje kada je potrebno uključiti više od dva modela.

Ne treba propustiti da model kauzalne podrške, istih autora, diskutovan u sekciji 7.2 o kauzalnom učenju, uzima upravo ovu formu.

[59]Tek smo donekle parafrazirali pravilo kako je tačno navedeno u širim eksperimentalnim instrukcijama Čengove i Holiuka (eksperiment 1, Cheng & Holyoak, 1985)

[60]Preciznije: kao ograničenja na moguće, dozvoljene distribucije verovatnoća pod pravilom određene strukture.

[61]Čitaoca može da zbuni činjenica da hipoteza koja se testira u zadatku selekcije, *Modus Ponens*, ne potiče iz modela, već je određena strukturom eksperimenta. U optimalnoj selekciji podataka hipoteze i podaci igraju različitu ulogu u odnosu na tipičan bejzijanski model. U bejzijanskom modelu reprezentativnosti Tenenbauma i Grifitsa, kognitivni sistem mora da generiše hipoteze o različitim osobinama novčića čija bacanja generišu osmotrene sekvence GGGPPPG, GPGPGP, GPPPGPG i sl. Tako se u ovom modelu *a priori* distribucija odnosi na hipoteze. U modelu Oaksforda i Čatera, hipoteza je data eksperimentom, a *a priori* distribucija se odnosi na moguće *ishode* eksperimenta okretanja karata u Vosonovom zadatku. Pošto okretanje karata rezultira u skupu $\{p, q, \neg p, \neg q\}$, *a priori* distribucija za ovaj tip bejzijanske inferencije je zajednička distribucija verovatnoće nad $\{p, q\}$.

[62]Počeci savremenog pristupa istraživanju koncepata i kategorizacije su u istraživanju organizacije semantičke memorije. I danas je moguće prepoznati te dve povezane tradicije istraživanja koje se obe odnose na koncepte: istraživanja organizacije semantičke memorije i istraživanja u tradiciji „koncepata i kategorija“ koja su više okrenuta pitanjima učenja koncepata, primene konceptualnog znanja i razvojnoj psihologiji konceptualnog sistema.

[63]Ne strogo; modeli primeraka se lako parametrizuju da uključe parametre zaboravljanja u izvesnoj meri, što je više realistično.

[64]Up. Milovanović, 2000 za pregled ovih teorija na srpskom jeziku

[65]Kao što neki autori s pravom primećuju, oblast prepoznavanja objekata u percepciji i oblast učenja kategorija i kategorizacije u psihologiji viših kognitivnih procesa potpuno se neopravdano posmatraju kao različite discipline. Obe se, naime, odnose na isti problem: na osnovu čega i oslanjajući se na koje procese kogntivni sistem odlučuje da su neki podaci $D$ evidencija o objektu/pojmu klase $C$ (Schyns, Goldstone & Thibaut, 1998.)

[66]Fodor, na primer, izražava nesigurnost da su problemi poput problema okvira uopšte rešivi unutar kompjutacionističke paradigme (Fodor, 2008)

[67]Pored Paretove, i druge klase distribucija su predložene kao kandidati koji opisuju ovaj univerzalni odnos (Kleiber & Kotz, 2003). Ipak, sve takve distribucije odlikuje nelinearni pad verovatnoće raspodele bogatstva sa njegovim porastom.

[68]Interesantno je da je još 1956. godine Herbert Sajmon razmatrao formalni model organizma koji u svom okruženju mora da optimizuje rešavanje nekoliko ciljeva uporedo, iako je ova formulacija sasvim van konteksta trenutne analize u kojoj govorimo o mogućnosti postojanja više uporednih ciljeva u odnosu na formalno istu environmentalnu strukturu podataka (Simon, 1956, str. 133).

[69]U sekciji 9.4 daćemo formalno još preciznije određenje, diskutujući poreklo koncepta optimizacije u savremenim raspravama o racionalnosti kognitivnih funkcija.

[70]Džouns i Lav u teorijskoj diskusiji bejzijanskih modela takođe primećuju da uvek postoji više racionalnih modela istog zadatka, ali pitanje ne postavljaju u kontekstu mogućeg postojanja

više kompjutacionih ciljeva koje kognitivni sistem može da postavi u odnosu na isti zadatak, već kritikuju izostanak testova različitih bejzijanskih modela istih zadataka (Jones & Love, 2011).

[71]Dok je Galton verovao da je inteligencija fundamentalno određena brzinom psihofizičkih psihofizioloških procesa, savremena paradigma je verzuje za brzinu kognitivnog izračunavanja uopšte.

[72]Isti problemi, paradoksalno, snažno privlače savremenu KKP Frojdovoj koncepciji subjektivnosti, iako je reč o dve teorije između kojih bi u diskusiji ljudske racionalnosti moglo da vlada samo stanje teorijskog rata.

[73]Up. priloge radu iz 1979. u „*Econometrica*“ u kome su dati bitni dokazi Dejvida Kranca za teoriju izgleda, i prilog radu o kumulativnoj teoriji izgleda, Tversky & Kahneman, 1992.

[74]Termin *psihoekonomija* se češće koristi kao drugo ime za bihejvioralnu ekonomiju, oblast koja integriše kognitivnu psihologiju i psihološke eksperimente u ekonomske analize; Glimčer u svojoj knjizi „*Neuroekonomija*“ eksplicitno tvrdi da je jezik matematičke bejzijanske teorije verovatnoće, u potpunosti svojstven ekonomskim analizama, jezik koji predstavlja odgovarajuću deskripciju senzomotornog problema koji se nalazi u fokusu neurobiologije (Glimcher, 2004).

[75]Gibsova slobodna energija se definiše izrazom $G(p,T) = U + pV - TS$, gde je $G(p,T)$ - Gibsova slobodna energija na pritisku $p$ i temperaturi $T$, $U$ - ukupna interna energija sistema, $V$ - zapremina, $S$ - entropija sistema.

[76]Ovo tvrđenje predstavlja samo još jedan mogući način da se izrazi II zakon termodinamike.

[77]Ponekad se koristi izraz „logicistička kognitivna psihologija“ da se označi ovaj pristup (up. Oaksford & Chater, 2009), ali mi izbegavamo upotrebu ovog termina koji omogućava da se jedna teorijska paradigma KKP pomeša sa jednim značajnim pravcem u filozofiji matematike prve polovine XX veka.

[78]Za praćenje redova u ovoj sekciji bitno je razumeti sledeću distinkciju koju uvodimo. Pod *konstruktivističkim postupkom* uvođenja matematičkih objekata, sve dok ne govorimo o stritkno intuicionističkim shvatanjima, mislimo na osobinu postupaka uvođenja novih matematičkih objekata širu od značenja *konstruktivizma* kako se ono uobičajeno vezuje za ovaj termin u filozofiji matematike. Pod konstruktivističkim postupkom u širem smislu smatraćemo svaki postupak uvođenja novog matematičkog objekta koji se odvija tako što se pokazuje način na koji se taj objekat uvodi u univerzum diskursa, tako da se on nalazi *u saglasnosti sa svime prethodno demonstriranim u tom univerzumu diskursa.* Jasno je, na primer, da je demonstracija egzistencije kardinalnog broja kontinuuma ($\mathbb{R}$) konstruktivička u ovom smislu, i ako se uvodi pomoću argumenta strukture kontrapozicije, oslanjajući se na zakon isključenja trećeg, koju intuicionisti sa svojim shvatanjem konstruktivizma u užem smislu ne bi prihvatili kao konstruktivistički. Postupci koji su konstruktivistički u ovom širem smislu koji uvodimo su nešto na šta će se oslanjati svi autori u ovde analiziranim pravcima u istoriji filozofije; niti jedan matematički objekat ne može biti uveden u univerzum diskursa ukoliko to uvođenje na neki način ne pokaže njegovu saglasnost sa prethodno demonstriranim “potezima” u istom diskursu; razlike između formalista i intuicionista počinju upravo u shvatanju te vrste “saglasnosti” o kojoj govorimo. Na primer, dok će Hilbert i formalisti smatrati uvođenje $\aleph_1$ savršeno saglasnim sa univerzumom diskursa jer oslanjanje na pravilo isključenja trećeg ne remeti ništa u sintakstičkoj koherenciji formalnog sistema izgradnje matematike, intuicionisti će na osnovu svog konstruktivističkog stava (koji ovde nazivamo

*konstruktivizmom u užem smislu*) insistirati na postupku ukazivanja na $\aleph_1$ *finitinim izračunavanjem odgovarajuće aritmetičke funkcije koja produkuje kardinalne brojeve*. Takav postupak, kao što je poznato, do danas nije ustanovljen, ostavljajući intuicioniste u svetu *potencijalne*, a formaliste u svetu *aktualne beskonačnosti*.

Takođe, treba imati na umu da kada kažemo da je Kantorovo definisanje beskonačnosti kao kardinalnih brojeva postupak zasnovan na finitističkim sredstvima, hoćemo da kažemo sledeće: njegov postupak ne zahteva od samog ljudskog uma beskonačnu iteraciju, niti *a priori* koncept nekog esencijalnog beskonačnog, da bi bio uveden u univerzum diskursa. Postupak koji sprovodi Kantor u konstrukciji $\aleph$-brojeva se oslanja samo na korake u matematičkom diskursu koji su sami konačni. Ovo shvatanje finitizma nije istovetno sa shvatanjem finitizma koje se javlja u mnogim raspravama u filozofiji matematike u XX veku; u tom smislu, prihvatimo to i da finitizam ovde koristimo u upravo opisanom širem smislu, a ako budemo o njemu govorili u užem smislu, to će biti naglašeno.

[79]Osnove teorije poverenja prvi put su predstavljene 14. maja 2009. godine na Tribini Laboratorije za eksperimentalnu psihologiju Filozofskog fakulteta u Beogradu. Ovde predstavljena forma teorije poverenja koristi značajno drugačije teorijske mehanizme u odnosu na ranu verziju diskutovanu tada.

[80]Očekivana vrednost *a posteriori* distribucije *nije* MAP (*maksimum a posteriori*) bejzijanska ocena, što čistuncu može da zasmeta. Upotreba očekivanja *a posteriori* distribucije, međutim, uvodi se da bi se izbegli problemi moguće bimodalnosti Beta distribucije za veoma male vrednosti $\alpha, \beta$. Rešenje koje tako primenjujemo sasvim je često u praksi bejzijanske inferencije. Isto se odnosi na sve buduće upotrebe Dirišleove distribucije u slučaju lozova $(X,P)$ sa većim brojem ishoda.

[81]Egzaktna nova veličina uzorka je *N+100-2*; iz daljih jednačina smo izbacili član *-2* koji se javlja kao posledica beta-binomijalne bejzijanske inferencije. Prvo, kao što ćemo videti taj član ima zanemarljiv efekat na novu veličinu uzorka za sve praktično očekivane vrednosti $N$; drugo, ovo je u skladu sa Dirišle-Multinomijalnom bejzjanskom inferencijom koju ćemo uskoro početi da koristimo kao generalizaciju za lozove koji sadrže više od dva ishoda.

[82]Preciznije: u originalnoj formulaciji teorije izgleda. Kumulativna teorija izgleda (Tversky & Kahneman, 1992) opisuje ponderisanje marginalnih doprinosa kumulativnih i dekumulativnih verovatnoća, kao što je prikazano na slici 4a, te je korespondencija daleko komplikovanija od one sa originalnim predlogom iz 1979.

[83]Poznata još i kao *Lomaksova distribucija*.

[84]Već tokom rada na ovoj doktorskoj tezi razvijeni su alternativni predlozi o mehanizmu formiranja verovanja u teoriji poverenja.

[85]Tverski i Kaneman u oceni parametara kumulativne teorije izgleda koriste vrednosti medijana monetarnih ekvivalenata, ne proseke koje mi analiziramo. Upotreba medijana u radu iz 1992. - koja je kasnije, čini se po inerciji, inspirisala i druge da koriste medijane a ne proseke (up. Gonzales & Wu, 1999), u tom radu ničim nije motivisana (Tversky & Kahneman, 1992). Pri tom, u istom tom radu kojim se uvodi kumulativna teorija izgleda, Tverski i Kaneman ne navode vrednosti medijana monetarnih ekvivalenata za mešovite lozove, što onemogućava ocenu averzije prema gubicima iz njihovih podataka.

[86]Nadamo se da će nam biti oprostena ova kratka digresija, dovoljna da saopšti kako bi bolja

komunikacija teorijskih rezultata i eksperimentalnih nalaza svakako prijala ovoj oblasti.

[87]Rezultate svih analiza za eksperiment 2b, kao i odgovarajuće grafikone, moguće je dobiti od autora direktnim zahtevom na *e-mail* adresu: goran.milovanovic@fmk.edu.rs

[88]U praksi, najčešče se minimizuje negativna logaritamska funkcija verodostojnosti modela, zato što većina softverskih implementacija Nelder-Midove i drugih metoda optimizacije po konvenciji traži minimume, a ne maksimume, objektnih funkcija.

[89]Up. analizu ekvivalencije teorije poverenja i kumulativne teorije izgleda u Prilogu A. Iako je ovu analizu nemoguće izraziti tako da se egzaktno pokažu uslovi pod kojima su dva modela ekvivalentna, jasno je da će ta ekvivalencija važiti samo za neke slučajeve u kojima su vrednosti *a priori* i objektivnih verovatnoća upravo regularno povezane. Ukoliko bismo kumulativnu teoriju izgleda proširili tako da dobijemo više funkcija ponderisanja verovatnoća umesto jedne, možda bismo lakše mogli da pokažemo ekvivalenciju, ali u odnosu na eksperimentalne rezultate u V delu znamo da bi onda ta funkcija verovatno morala da uzme formu funkcije dve promenljive - zbog eksperimentalno demonstrirane zavisnosti ponderisanja verovatnoća od visine samih ishoda na lozovima, a ne samo od ranga njihovih verovatnoća.

[90]Čitaoca upućujemo na odličan, neformalan uvod u koncepte teorije igara na srpskom jeziku, autora Bože Stojanovića, „*Teorija igara: elementi i primena*“, Službeni glasnik i Institut za evropske studije, 2005. Svi koncepti teorije igara koje koristimo u narednim redovima uvedeni su i diskutovani u tom udžbeniku. Pored Stojanovićevog udžbenika, tehnički tek nešto zahtevnija „*Playing for Real: A Text on Game Theory*“, Ken Binmore, Oxford University Press, 2007, predstavlja naš drugi osnovni izvor za sve koncepte teorije igara na koje se oslanjamo u diskusijama koje slede.

[91]Naravno, isključujući trivijalne probleme vezane za ekstra-lingvističke faktore kao što bi bilo opterećenje operativne memorije usled potrebe da se procesiraju predugačke rečenice i sl.

[92]Makar: živimo u kulturi u kojoj ovi semiotički artefakti, simbolički izrazi, imaju smislu o odnosu na normu o smislenosti koju prihvata određena publika. To što je moguće da određeni simbolički sistem ima smisla samo za podskup korisnika opštijeg sistema simbola u kome je izražen nas uopšte ne oslobađa obaveze da ga podjednako tretiramo kao empirijski nalaz čiju strukturu psihološka teorija simboličkih funkcija mora da objasni .

[93]Pod „sintaksičkim identitetom“ se u učenju kategorija podrazumeva poklapanje parova *atribut*:*vrednost* u učenju kategorija (npr. boja:zeleno i boja:crveno ne zadovoljavaju sintaksički identitet da bi dva ajtema bila kategorisana kao ista ako je njihova boja jedina relevantna karakteristika u kategorizaciji).

[94]Originalna filozofska analiza Dejvida Luisa (Lewis, 1969) analizira konvecionalnost rečenica, ne elementarnih entiteta poput singularnih termina; mi primenjujemo Luisovu analizu konvencionalnosti na nivo simbola. Ovo nema nikakvih posledica po strukturu njegove originalne analize konvencionalnosti.

# Bibliografija


Abdellaoui, M. (2009). Rank Dependent Utility. U: Anand, P., Pattanaik, P. & Puppe, C. (ur.), Handbook of Rational and Social Choice. Oxford University Press, Oxford, 2009.

Agrawal, K. (2010). To study the phenomenon of the Moravec's Paradox. arXiv:1012.3148v1 [cs.AI], submitted on 14 Dec 2010, URL: http://arxiv.org/abs/1012.3148

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. Bulletin of the Psychonomic Society, 15, 147-149.

Allan, L. G. (1993). Assessing power PC. Learning& Behavior, 31(2), 192-204.

Allan, L. G. (1993). Human contingency judgments: rule based or associative? Psychological Bulletin, 114(3):435-48.

Anderson, J. R. (1983). The Architecture of Cognition. Cambridge: Harvard University Press.

Anderson, J. R. (1989). A rational analysis of human memory. U: H. L. Roediger, III and F. I. M. Craik (ur.) Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving. Hillsdale, NJ: Erlbaum, 195-210.

Anderson, J. R. (1991a). The place of cognitive architectures in a rational analysis. U: K. Van Len (ur.), Architectures for Intelligence. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991b). Is human cognition adaptive? Behavioral and Brain Sciences, 14, 471-484.

Anderson, J. R. (1991c). The adaptive nature of human categorization. Psychological Review, 98, 409-429.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. American Psychologist, 51, 355-365.

Anderson, J. R. & Matessa, M. (1998). The rational analysis of categorization and the ACT-R architecture. U: M. Oaksford & N. Chater (ur.) Rational models of cognition, pp. 197-217. Oxford: Oxford University Press.

Anderson, J. R. & Milson, R. (1989). Human Memory: An Adaptive Perspective. Psychological Review, 96, 703-719.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, 2, 396-408.

Anderson, J. R. & Schooler, L. J. (2000). The adaptive nature of memory. U: Tulving, E. & Craik, F. I. M. (ur.) Handbook of memory, 557-570. New York: Oxford University Press.

Anderson, J. R. & Sheu, C. F. (1995). Causal inference as perceptual judgments. Memory & Cognition, 23, 510-524.

Ashby, W. A. (1956). An Introduction to Cybernetics. Chapman & Hall, London.

Assad, A. & Packard, H. (1992/2008). Emergent Colonization in an Artificial Ecology. U: Bedau, M. & Humphreys, P. (ur.) Emergence: contemporary readings in philosophy and science. Cambridge, MA: MIT Press.

Baddeley, A. (1997/2004). Ljudsko pamćenje. Zavod za udžbenike i nastavna sredstva, Beograd, 2004.

Banks, W.P., Clark, H.H., Lucy, P., 1975. The Locus of the Semantic Congruity Effect in Comparative Judgments. Journal of Experimental Psychology: Human Perception and Performance 104 (1), 35–47.

Bar-Hillel, Maya (1980). The base-rate fallacy in probability judgments. Acta Psychologica 44: 211–233.

Barsalou, L.W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. U: T.K. Srull & R.S. Wyer (ur.), Advances in social cognition, Volume III: Content and process specificity in the effects of prior experiences (pp. 61-88) Hillsdale, NJ: Lawrence Erlbaum Associates.

Basset, G. W. (1984). The St. Petersburg paradox and bounded utility. History of Political Economy 19:4, 517-523.

Bechtel, W. & Abrahamsen, A. (2005). Explanation: A Mechanistic Alternative. Studies in History and Philosophy of the Biological and Biomedical Sciences , 36, 421-441.

Becker, G. M., De Groot, M.H., Marschak, J. (1964). Measuring utility by a single-response sequential method. Behavioral Science 9(3), 226–32.

Bedau, M. A. & Humphreys, P. (ur.) (2007). Emergence: Contemporary Readings in Philosophy and Science. MIT Press: London.

Bell, D. (1988). Disappointment in Decision Making Under Uncertainty. U: Bell, D., Raiffa, H. & Tversky, A. (ur.). Decision Making: Descriptive, Normative, and Prescriptive Interactions. Cambridge University Press.

Bermudez, J. L. (2005). Philosophy of Psychology: A Contemporary Introduction. Routlege, 2005.

Bernoulli, D. (1954/1738). Exposition of a New Theory on the Measurement of Risk . Econometrica, Vol. 22, No. 1. (Jan., 1954), pp. 23-36. Prevod na engleski sa latinskog od strane Dr. Louise Sommer, The American University, Washington, D.C., izvornik: Specimen Theoriae Novae de Mensura Sortis. Commentarii Academiae Scientiarum Imperialis Petropolitanae, Tomus V, 1738, pp. 175-192.

Beth, E. W. (1955/1987). Semantičko nasleđivanje (istinitosti) i formalna izvedivost. Prevod originalnog rada Beth, E. W., (1955). Semantic entailment and formal derivability. Mededlingen van de Koninklijke Nederlandse Akademie van Wetenschappen, Afdeling Letterkunde, N.R. Vol 18, no 13, 1955, pp 309–42. U zborniku "Novija filozofija matematike", Šikić, Z. (ur.), Nolit, Beograd, 1987.

Bickhard, M. H. (1996). Troubles with Computationalism. U: O'Donohue, W. & Kitchener, R. F. (ur.) The Philosophy of Psychology. (173-183). London: Sage.

Binmore, K. (2007). Playing for Real – A Text on Game Theory. New York: Oxford University Press.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. Psychological Review, 115, 463-501.

Blavatskyy, P. (2006). Stochastic Expected Utility Theory. Journal of Risk and Uncertainty, 34(3), 259-286.

Blavatskyy, P. (2011). Which Decision Theory? Vienna Joint Economics Seminar, Thursday, 24. November 2011. URL:

http://econ.univie.ac.at/uploads/tx_cal/media/Blavatskyy_paper_an.pdf

Boden. M (2006). Mind As Machine: A History of Cognitive Science. Oxford University Press.

Borg, I. and Groenen, P.J.F. (2005). Modern multidimensional scaling. 2nd edition. New York: Springer.

Bott, L., Hoffman, A., & Murphy, G. L. (2007). Blocking in category learning. Journal of Experimental Psychology: General, 136, 685-699.

Bradley, I., al-Nowaihi, A. & Dhami, S. (2008). The Utility Function Under Prospect Theory. Economics Letters, 99, 337-339.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). Making choices without trade-offs: The priority heuristic. Psychological Review, 113(2), 409–432.

Brooks, R. A. (1991). Intelligence without representation. U: Haugeland, J. (ur.). Mind Design II, 395-420.

Brouwerm L.E.J. (1912). Intuitionism and Formalism. U: Philosophy of Mathematics. Selected Readings, Benacerraf, P. & Putnam, H. (ur.), Prentice-Hall, NJ, 1964.

Buehner, M., Cheng, P.W., Clifford, D. (2003.) From Covariation to Causation: A Test of the Assumption of Causal Power. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 1119-1140.

Carbone, E. & Hey, J. (2000). Which Error Story is Best? Journal of Risk and Uncertainty, 20(2), 161-176.

Carnap, R, (1956/85). Metodološka narav teorijskih pojmova. U Sesardić, N. (ur.), Filozofija nauke, Nolit, Beograd, 1985.

Cartwright, N. (1989). Nature' s Capacities and their Measurement. Oxford University Press, 1989.

Cassou-Nogues, P. (2005). Gödel and the 'objective existence' of mathematical objects. History and Philosophy of Logic, 26 (September 2005), 211-228.

Chapman , G. B. Robbins , S. I. (1990). Cue interaction in human contingency judgment. Memory & Cognition, 18, 537-545.

Chateauneuf, A. & Wakker, P. P. (1999). An Axiomatization of Cumulative Prospect Theory for Decision under Risk. Journal of Risk and Uncertainty 18, 137-145.

Chater, N. & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. Cognitive Psychology, 38, 191-258.

Chater, N. & Vitányi, P. (2003). The generalized universal law of generalization. Journal of Mathematical Psychology, 47, 346-369.

Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behaviour. Synthese, 122, 93-131.

Chater, N., Tenenbaum, J. B. & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. Trends in Cognitive Sciences Vol.10 No.7 July 2006.

Cheng, P. & Holyoak, K. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Cheng, P.W. (1997). From covariation to causation: A causal power theory. Psychological Review, 104, 367-405.

Cheng, P.W. (2000). Causality in the mind: Estimating contextual and conjunctive causal power. U: Keil, F. & Wilson, R. (ur.), Cognition and Explanation (pp. 227-253). Cambridge: MIT Press.

Cheng, P.W., & Novick, L.R. (1990). A probabilistic contrast model of causal induction. Journal of Personality and Social Psychology, 58, 545-567.

Cheng, P.W., & Novick, L.R. (1992). Covariation in natural causal induction. Psychological Review, 99, 365-382.

Cheng, P.W., Park, J., Yarlas, A., & Holyoak, K.J. (1996). A causal-power theory of focal sets. U: Shanks, D.R. , Holyoak, K.J. & Medin, D.L. (ur.), The psychology of learning and motivation, vol. 34: Causal learning (pp. 313-355). New York, NY: Academic Press.

Chomsky, N. (1957). Syntactic Structures. The Hague/Paris: Mouton

Chomsky, N. (1959/1967). A Review of B. F. Skinner's Verbal Behavior. Language, 35, No. 1 (1959), 26-58. U Readings in the Psychology of Language, ed. Leon A. Jakobovits and Murray S. Miron (Prentice-Hall, Inc., 1967), pp.142-143.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.

Clark, A. (1998). Embodied, situated, and distributed cognition. U: W. Bechtel & G. Graham (ur.), A companion to cognitive science. (pp. 506-517). Malden, MA: Blackwell.

Clauset, A., Shalizi, C. R. & Newman, M. E. J. (2009). Power-law distributions in empirical data. SIAM Review 51, 661-703 (2009).

Cohen, L. J. (2008/1981). Can Human Irrationality Be Experimentally Demonstrated? Behavioral and Brain Sciences, 4, 317–370. U: Adler, J. E. & Rips, L. J. (ur.), Reasoning. Studies of Human Inference and Its Foundations. Cambridge University Press, New York, 2008.

Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Collins, D. J. & Shanks, D. R. (2006). Conformity to the power PC theory of causal induction depends on the type of probe question.Quarterly journal of experimental psychology, 59(2), 225-32.

Conlisk, J. (1989). Three Variants on the Allais Example. American Economic Review, 79, 392–407.

Cook, M. (2004). Universality in Elementary Cellular Automata. Complex Systems, vol. 15 (1), 2004, pp. 1-40.

Corter, J. E. (1991). Normative Theories of Categorization. Behavioral and Brain Sciences, 14, 471-484.

Cosmides, L. (1985). Deduction or Darwinian Algorithms? An explanation of the "elusive" content effect on the Wason selection task. Doctoral dissertation, Harvard University.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition, 31, 187-276.

Cosmides, L. & Tooby, J. (1987). From evolution to behavior: Evolutionary psychology as the missing link. U: J. Dupre (ur.), The latest on the best: Essays on evolution and optimality. Cambridge, MA: The MIT Press.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition, 58, 1-73.

Couturat, L. (1901/2002). The Logic of Leibniz [Logique

de Leibniz]. Anciene Librairie Germer Bailliere, Paris. URL: http://philosophyfaculty.ucsd.edu/faculty/rutherford/Leibniz/contents.htm

Coveney, P. & Highfield, R. (1995). Frontiers of Complexity. Fawcett Columbine, New York.

Cox, J. C. (2008). Preference Reversals. U: Plott, C. R. & . Smith, V. L. (ur.), Handbook of Experimental Economics Results, Volume 1.New York: Elsevier Press.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. Journal of Mathematical Psychology, 47, 109-121.

Danks, D., Griffiths, T. L., & Tenenbaum, J. B. (2003). Dynamical causal learning. U: S. Becker, S. Thrun, & K. Obermayer (ur.), Advances in neural information processing systems 15(pp. 67-74). Cambridge, MA: MIT Press.

Davidson, J. E. (2003). Insights about Insightful Problem Solving. U: The Psychology of Problem Solving, Davidson, J. E. & Sternberg, R. J. (ur.). Cambridge University Press.

de Finetti, B. (1970/1974). Theory of Probability: A Critical Introductory Treatment. John Wiley & Sons, New York.

de Saussure, F, (1916/1977). Opšta lingvistika. Nolit, Beograd, 1977.

Descartes (1644). Principles of Philosophy. URL: http://www.earlymoderntexts.com/descprin.html

Diecidue, E. & Wakker, P. P. (2001). On the Intuition of Rank-Dependent Utility. Journal of Risk and Uncertainty, 23, 281-298.

Dreyfus, H. L. (1972/1977). Šta računari ne mogu? Nolit, Beograd, 1977.

Dzhafarov, E. N. & Colonius, H. (2011). The Fechnerian idea. American Journal of Psychology, 124, 127-140.

Ebbinghaus, H. (1885/1913). Memory: A Contribution to Experimental Psychology. Prevod Henry A. Ruger & Clara E. Bussenius (1913). Originalno objavljeno u Njujorku: New York, Teachers College, Columbia University. URL: http://psychclassics.yorku.ca/Ebbinghaus/index.htm

Eco, U. (1997/2000). Kant i kljunar. PAIDEIA, Beograd, 2000. Originalno izdanje: Umberto Eco: Kant e l' ornitornico. 1997 R.C.S. Libri S.p.A. I edizione Studi Bompiani ottobre 1997.

Elman, J. (1995). Language as dynamical system. U: Port, R. & van Gelder, T. (ur.), Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge MA: MIT press.

Elman, J. L. (1998). Connectionism, artificial life, and dynamical systems. U: Bechtel, W. & Graham, G. (ur.), A companion to cognitive science. Malden, MA: Blackwell.

Elman, J. L., Bates, E. A., Johnson, M. A., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). Rethinking Innateness. A Connectionist perspective on development. A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England.

Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.

Feldman, J. (2000) Minimization of Boolean complexity in human concept learning. Nature, 407, 630–633.

Fishburn, P. (1986). The Axioms of Subjective Probability. Statistical Science, Vol. 1(3), 335-345.

Fodor, J. (1987). Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, MA: MIT Press.

Fodor, J. (2000). The Mind Doesn't Work That Way; The Scope and Limits of Computational Psychology. MIT Press.

Fodor, J. A. (1975). The Language of Thought, Harvard University Press, 1975.

Fodor, J. A. (1983). The Modularity of Mind:An Essay in Faculty Psychology. The MIT Press.

Fodor, J. A. (1997). Connectionism and the Problem of Systematicity (Continued): Why Smolensky's Solution Still Doesn't Work. Cognition 62:109-19.

Fodor, J. A. (2008). LOT 2: The Language of Thought Revisited, Oxford University Press, 2008.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. U: Pinker, S. & Mehler, J. (ur.), Connections and symbols. (pp. 3-72). Cambridge, MA: Bradford.

Foerster, von H. (2002). Understanding Understanding: Essays on Cybernetics and Cognition Springer-Verlag.

Fox, C. R. & Poldrack, R. A. (2009). Prospect Theory and the Brain. U: Glimcher, P., Camerer, C. F., Fehr, E. & Poldrack, R. A. (ur.), Neuroeconomics. Decision Making and the Brain. Elsevier Inc., 2009.

Frege, G., (1892/1960). On Sense and Reference. U: Translations from the Philosophical Writings of Gottlob Frege, Geach, P. & Black, M. (ur.). Oxford: Basil Blackwell.

Friedman, M. (2009). Carnap on Theoretical Terms: Structuralism without Metaphysics. Predavanje na Theoretical Frameworks and Empirical Underdetermination Workshop (Düsseldorf April 10-12, 2008).

Gachter, S. Johnson, E. J. & Herrmann, A. (2007). Individual-Level Loss Aversion in Riskless and Risky Choices. IZA Discussion Papers 2961, Institute for the Study of Labor (IZA).

Gardenfors, P. (1988). Knowledge in Flux. Modeling the Dynamics of Epistemic States, Cambridge, MA: MIT Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155-170.

Gergen, K. (1999). An Invitation to Social Construction. SAGE Publications.

Gigerenzer, G. (1991). Does the environment have the same structure as Bayes' theorem? Behavioral and Brain Sciences, 14, 495-496.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: beyond „heuristics and biases". U: Stroebe, W. & Hewstone, M. (ur.), European review of social psychology (Vol. 2, pp. 83-115). Chichester, UK: Wiley.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. U: Manktelow, K. I. & Over, D. E. (ur.), Rationality: psychological and philosophical perspectives (pp. 284-313). London: Routledge.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). U: G. Wright, & P. Ayton (ur.), Subjective probability (pp. 129- 161). Chichester, UK: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky (1996). Psychological Review, 103, 592-596.

Gigerenzer, G. (2008). Rationality for mortals: How people cope with uncertainty. New York: Oxford University Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review, 102, 684-704.

Glasersfeld, von E. (1984). An introduction to Radical Constructivism. U: Paul Watzlawick (ur.) The Invented Reality. New York: Norton.

Glasersfeld, von E. (1989). Cognition, Construction of Knowledge, and Teaching Synthese 80 (1): 121-140. Reprodukovano u: Matthews, M. R. (ur.) (1991) History, philosophy, and science teaching. New York: Teachers College Press.

Glasersfeld, von E. (1995). Radical Constructivism. Falmer Press.

Glasersfeld, von E. (1999). How Do We Mean? Constructivist Sketch of Semantics Cybernetics and Human Knowing 6 (1): 9-16.

Glasersfeld, von E. (2001). The Radical Constructivist View of Science Foundation of Science. Specijalno izdanje "The Impact of Radical Constructivism on Science", vol. 6, nos. 1-3: 31-43.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. Memory & Cognition, 7(2), 95-112.

Glimcher, P. (2004). Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics. MIT Press/Bradford Press.

Glimcher, P., Camerer, C. F., Fehr, E. & Poldrack, R. A. (ur.) (2009). Neuroeconomics. Decision Making and the Brain. Elsevier Inc., 2009.

Glymour, C. (2003). Learning, Prediction and Causal Bayes Nets. Trends in Cognitive Sciences 7(1), 43-48.

Glymour, C. & Eberhardt, F. (2011). Hans Reichenbach. The Stanford Encyclopedia of Philosophy (Summer 2011 Edition), Edward N. Zalta (ed.), URL: http://plato.stanford.edu/archives/sum2011/entries/reichenbach/

Glymour, C. N. (2001). The Mind's Arrows: Bayes nets and graphical causal models in psychology. Cambridge, MA: The MIT Press.

Gonzalez, R. & Wu, G. (1999). On the shape of the probability weighting function. Cognitive Psychology 38, 129-166.

Gould, S. J. & Lewontin, R. (1979). The spandrels of San Marco and the Panglossion paradigm: a critique of the adaptationist programme. Proc R Soc Lond B 205 (1161): 581–598.

Griffin, D. & Brenner, L. (2004). Perspectives on Probability Judgment Calibration. U Blackwell Handbook of Judgment and Decision Making, Koehler, D. J. & Harvey, N. (ur.). Blackwell Publishing Ltd.

Griffiths, T. L. (2005). Causes, coincidences, and theories. Neobjavljena doktorska disertacija, Stanford University, Stanford CA.

Griffiths, T. L. and Tenenbaum, J. B. (2007). Two proposals for causal grammars. U: Gopnik, A., & Schulz, L. (ur.), Causal learning: Psychology, philosophy, and computation. Oxford University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. Cognitive Psychology, 51, 354-384.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. Psychological Science, 17, 767-773.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. Psychological Review, 116, 661-716.

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. U: A. Gopnik & L. Schulz (ur.), Causal learning: Psychology, philosophy, and computation (pp. 86-100). Oxford: Oxford University Press.

Hampton, J.A. (1979). Polymorphous concepts in semantic memory. Journal of Verbal Learning and Verbal Behavior, 18, 441-461.

Hampton, J.A. (1997). Psychological representations of concepts. In M.A.Conway (Ed.) Cognitive models of memory. pp. 81-110. Hove: Psychology Press.

Hampton, J.A. (2001). Prototype concepts. Article for the Encyclopedia of Cognitive Science. New York: Nature Publishing Group, Macmillans.

Hampton, J.A. (2006). Concepts as Prototypes. In Ross, B.H. (Ed) The Psychology of Learning and Motivation: Advances in Research and Theory, 46, 79-113.

Harless, D. (1993). Experimental tests of prospective reference theory. Economics Letters, 43(1), 71-76.

Harless, D. W. & Camerer, C. F. (1994). The Predictive Utility of Generalized Expected Utility Theories. Econometrica,vol. 62(6), pages 1251-89

Harrison, G. W. & Rutström, E. (2009). Expected Utility And Prospect Theory: One Wedding and Decent Funeral. Experimental Economics, 12(2), 133-158.

Harvey, I., (2005). Evolution and the Origins of the Rational. U: Zilhão, A. (ur.) Cognition, Evolution, and Rationality. London, Routledge, 2005.

Hattori, M. & Oaksford, M. (2007a). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. Cognitive Science, 31(5), 765-814.

Haugeland, J. (1981). Semantic Engines. U: Haugeland, J. (ur), Mind Design. Cambridge, Mass.: MIT Press.

Haugeland, J. (Ed.) (1997). U: Haugeland, J. (ur), Mind design II: Philosophy, Psychology and Artificial Intelligence. MIT Press, 1997.

Hauser, M., Chomsky, N. & Fitch, W. T. (2002). The Language Faculty: What is it, who has it, and how did it evolve? Science, 298, 1569-1579

Haykin S, 1999, Neural Neworks: a Comprehensive Foundation. Prentice Hall

Hempel, C. (1945). Studies in the Logic of Confirmation. Mind, 54: 1–26, 97–121.

Hempel, C. & Oppenheim, P. (1965/2008). On the Idea of Emergence. U: Bedau, M. & Humphreys, P. (ur.) Emergence: contemporary readings in philosophy and science. Cambridge, MA: MIT Press.

Herre, H. & Schroeder-Heister, P. (1998). Formal languages and systems. Routledge Encyclopedia of Philosophy, London 1998.

Hey, J. D. & Orme, C. (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. Econometrica, 62(6), 1291-1326.

Hick, W. E. (1952). On the rate of gain of information. Quarterly Journal of Experimental Psychology, 4:1, 11-26.

Hilbert, D. (1925). On the Infinite. U: Benacerraf, P. & Putnam, H. (urednici), Philosophy of Mathematics. Selected Readings. Prentice-Hall, NJ, 1964.

Holyoak, K.J. & Cheng, P.W. (2010). Causal learning and inference as a rational process: The new synthesis. Annual Review of Psychology, 62: 23.1-23.29.

Horst, S. (2011). The Computational Theory of Mind. The Stanford Encyclopedia of Philosophy (Spring 2011 Edition), Edward N. Zalta (ed.), URL:http://plato.stanford.edu/archives/spr2011/entries/computational-mind/.

Hsu, F. (2002). Behind Deep Blue: Building the Computer That Defeated The World Chass Champion. Princeton University Press, Princeton and Oxford.

Hume, D. (1739/1983). Rasprava o ljudskoj prirodi. Veselin Masleša, 1983. Sarajevo.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. Psychological Review, 110, 220-264.

Iemhoff, R. (2012). Intuitionism in the Philosophy of Mathematics. The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), URL:http://plato.stanford.edu/archives/win2012/entries/intuitionism/.

Isoni, A., Loomes, G. & Sugden, R. (u štampi). The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect", Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: A Reassessment. American Economic Review.

James, W. (1890/1950). The Principles of Psychology. Vol 1-2, Dover Publications, 1950.

Johnson-Laird, P. N. & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. Cognitive Psychology,1, 134-138.

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1994). Continuous univariate distributions Vol 1. Wiley Series in Probability and Statistics.

Jones, M. & Love, B.C. (2011). Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition. Behavioral and Brain Sciences, 34, 169-231.

Jovanović, G. (1997). Frojd i moderna subjektivnost. Svetovi, Novi Sad.

Juran, M. J. (1975). The Non-Pareto Principle: Mea Culpa. Quality Progress. New York: American Society for Quality Control, May, 1975.

Kahneman, D. (1981) Who shall be the arbiter of our intuitions? Behavioral and Brain Sciences 4:339 – 40.

Kahneman, D. (December 8, 2002). Maps of Bounded Rationality: A Perspective on Intuitive Judgement and Choice (Nobel Prize Lecture). NobelPrize.org. The Nobel Foundation. Retrieved 2009-06-13.

Kahneman, D. & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. U: Gilovich, T., Griffin, D., & Kahneman, D. (ur.). Heuristics of Intuitive Judgment. New York. Cambridge University Press.

Kahneman, D. & Frederick, S. (2005). A Model of Heuristic Judgment. U: Holyoak, K.J. & Morrison, R. G. (ur.), The Cambridge Handbook of Thinking and Reasoning. Cambridge University Press. pp. 267–294.

Kahneman, D. & Tversky, A. (1979). Prospect Theory. An Analysis of Decision under Risk. Econometrica, 47:2, 263-91.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430–454

Kahneman, D., Knetsch, J., & Thaler, R. (1990). Experimental Test of the endowment effect and the Coase Theorem. Journal of Political Economy 98(6), 1325-1348.

Kahneman, D.; Tversky, A. (1973). On the psychology of prediction. Psychological Review 80 (4): 237–251.

Kelly, G. A. (1955/1991). The psychology of personal constructs. Routlege, New York.

Kemp, C. and Tenenbaum, J. B. (2008).The discovery of structural form. Proceedings of the National Academy of Sciences.105(31), 10687-10692

Kleene, S. C. (1952/1987). Izračunljivost, odlučivost i teoremi nepotpunosti. Prevedeno 5. poglavlje Kleenejeve "Matematičke logike" koje nosi naslov "Izračunljivost i odlučivost", u zborniku "Novija filozofija matematike", Šikić, Z. (ur.), Nolit, Beograd, 1987.

Kleiber, C. & Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences. New York: Wiley.

Knight, F.H. (1921). Risk, Uncertainty, and Profit. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Knoblauch, K. & Maloney, L. T. (2008), MLDS: Maximum likelihood difference scaling in R. Journal of Statistical Software, 25(2), 1-26 .

Köbberling, V. & Wakker, P. P. (2005). An Index of Loss Aversion. Journal of Economic Theory 122, 119-131.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. Behavioral and Brain Sciences 1996;19(1):1-53.

Koestler, A. (1959). The Sleepwalkers. London: Penguin Group.

Kolmogorov, A. (1933/1956). Foundations of the Theory of Probability (2nd ed.). New York: Chelsea.

Kostić, A. (2006). Kognitivna psihologija. Zavod za udžbenike, Beograd (2006-2010).

Lee, P. M. (1989/2004). Bayesian Statistics. Third Edition. Hodder Arnold.

Leopold, D. A. & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. Trends in Cognitive Sciences, Vol.3, No. 7, July 1999.

Lewis, D. (1969). Convention. Cambridge: Harvard University Press.

Lewis, D. (1975). Languages and Language. Reprodukovano u: Philosophical Papers, vol. 1. Oxford: Oxford University Press.

Lewis, D. (1979). Scorekeeping in a language game. Journal of Philosophical Logic 8 (1), 339-359.

Link, S.W. (1994). Rediscovering the past: Gustav Fechner and signal detection theory. Psychological Science, 5(6), 335-339.

Lober, K., Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). Psychological Review, 107(1), 195-212.

Loomes, G., Moffat, P. & Sugden, R. (2002). Microeconometric Test of Alternative Stochastic Theories of Risky Choice. Journal of Risk and Uncertainty, 24, 103-130.

Lu, H., Yuille, A., Liljeholm, M., Cheng, P.W., & Holyoak, K.J. (2008). Bayesian generic priors for causal learning. Psychological Review, 115, 955-984.

Lu, H.J., Yuille, A.L, Liljeholm, M., Cheng, P.W., & Holyoak, K .J. (2007). Bayesian models of judgments of causal strength: A comparison. Proceedings of the 29th Annual Conference of the Cognitive Science Society. pp. 1241-1246. August 2007.

Luce, R. D. (1959). Individual Choice Behavior: A Theoretical Analysis. New York: Wiley.

Luce, R. D. (1977). The choice axiom after twenty years. Journal of Mathematical Psychology 15 (3): 215–233.

Luce, R.D. & Tukey, J.W. (1964). Simultaneous conjoint measurement: a new scale type of fundamental measurement. Journal of Mathematical Psychology 1, 1–27.

MacCorquodale, K. & Meehl, P. R. (1948). On a Distinction Between Hypothetical Constructs and Intervening Variables. Psychological Review, 1948, 55, 95-107.

Maloney, L. T. (2002). Statistical decision theory and biological vision. In Heyer, D. & Mausfeld, R. [Eds], Perception and the Physical World: Psychological and Philosophical Issues in Perception. New York: Wiley, pp. 145-189.

Maloney, L. T. & Yang, J. N. (2003), Maximum likelihood difference scaling. Journal of Vision, 3, 573-585.

Mangel, M. & Clark, C.W. (1988). Dynamic Modeling in Behavioral Ecology. Princeton University Press, Princeton, NJ.

Marcus, G. F. (2001). The Algebraic Mind: Integrating Connectionism and Cognitive Science. Cambridge, MA: MIT Press.

Marcus, G. F. (2006). Cognitive Architecture and Descent with Modification. Cognition 101, 443-465.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York: Freeman.

Masin, S.C., Zudini, V. & Antonelli, M. (2009). Early Alternative Derivations of Fechner's Law. Journal of the History of Behavioral Sciences, 2009, 45, 56-65.

Maturana, H. R. & Varela, F. J. (1987). The tree of knowledge: The biological roots of human understanding. Boston: Shambhala Publications.

Maturana, Humberto & Varela, Francisco (1973). Autopoiesis and Cognition: the Realization of the Living. U: Cohen, R. S. & Wartofsky, M. W. (ur.), Boston Studies in the Philosophy of Science 42. Dordecht: D. Reidel Publishing Co.

Maynard Smith, J. (1982). Evolution and the Theory of Games. Cambridge University Press.

McClelland, J.L., D.E. Rumelhart and the PDP Research Group (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models, Cambridge, MA: MIT Press

McClelland, J.L., Patterson, K., 2002. Rules or connections in past-tense inflections: What does the evidence rule out? Trends in Cognitive Sciences 6, 465–472.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. Psychonomic Bulletin and Review, 7, 360-366.

McLaughlin, B. (1997/2008). Emergence and Supervenience. U: Emergence : contemporary readings in philosophy and science, Bedau, M. A. & Humphreys, P. (ur.). The MIT Press

Cambridge, Massachusetts, London, England.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. Psychological Review, 85, 207-238.

Michaels, C.F. & Carello, C. (1981). Direct Perception. Englewood Cliffs: Prentice-Hall, Inc.

Michell, J. (1999). Measurement in Psychology: A Critical History of a Methodological Concept. Cambridge University Press.

Milovanović, G. (1996). Efekat semantičke kongruencije u verifikaciji predikativnih iskaza. Psihologija, XXIX, 2 - 3, pp. 353-372.

Milovanović, G. (2000). Strukturalna kompleksnost semantičkih domena. Diplomski rad. Filozofski Fakultet, Univerzitet u Beogradu.

Milovanović, G. (2010). Science and Knowledge Management: Some Problems Related to the Accumulation of Knowledge in Contemporary Scientific Practice. Wissenschaft und bildung im wandel/Science and Education in Transition/Nauka i obrazovanje u tranziciji. Humboldt-Kolleg, Belgrade, October 28-30, 2010.

Milovanović, G. (2011). Formalizations of Rational Choice in the 20th Century: From Axioms to Preference Conditions. European Society for the History of the Human Sciences, 30th Annual Conference: Faculty of Philosophy, University of Belgrade, Belgrade, Serbia, 5-8 July 2011.

Mirowski, P. (2000). Machine Dreams: Economics Becomes a Cyborg Science. Cambridge University Press, 2001

Monod, J. (1971/1983). Slučajnost i nužnost. Rad, Beograd, 1983.

Morris, M. W., & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. Memory & Cognition, 18, 407-418.

Murphy, G. L. (1993). A rational theory of concepts. U: Nakamura, G. V., Taraban, R. M. & Medin, D. L. (ur.), The psychology of learning and motivation (vol. 29): Categorization by humans and machines (pp. 327-359). New York: Academic Press.

Murphy, G. L. (2002). The big book of concepts. Cambridge, MA: MIT Press.

Murphy, G. L. & Medin, D. L. (1985). The Role of Theories in Conceptual Coherence. Psychological Review, 92, 289 – 316.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 904-919.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 70-84.

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. Journal of Verbal Learning and Verbal Behavior, 21, 1-20.

Murphy, G.L., Hampton, J.A., & Milovanović, G.S. (2012). Semantic Memory Redux: An Experimental Test of Hierarchical Category Representation. Journal of Memory and Language,67, 521-539.

Newell, A. & Simon, H. A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. Communications of the ACM, 19.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 1-55). Hillsdale, NJ: Erlbaum.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. Contemporary Physics 46 (5): 323–351.

Nolan, D. (2005). David Lewis. Chesham: Acumen Publishing, 2005.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(1), 104-114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115(1), 39-57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(4), 700-708.

Nosofsky, R. M. (1991). Relation between the rational model and the context model of categorization. Psychological Science, 2(6), 416-421.

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. U: Healy, A. F. & Kosslyn, S. M. (ur.), Essays in honor of William K. Estes, vol. 1: From learning theory to connectionist theory; vol. 2: From learning processes to cognitive processes (pp. 149-167). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(5), 924-940.

Novick, L.R., & Cheng, P.W. (2004). Assessing interactive causal influence. Psychological Review, 111, 455-485.

Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. Psychological Review, 101, 608-631.

Oaksford, M. & Chater, N. (1998). Rationality in an uncertain world. Psychology Press: Hove, England.

Oaksford, M. & Chater, N. (2006). Bayesian Rationality. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. Trends in Cognitive Sciences. 5, 349-357.

Oaksford, M., & Chater, N. (2009). Precis of "Bayesian rationality: The probabilistic approach to human reasoning." Behavioral and Brain Sciences, 32, 69-84.

Oaksford, M., Roberts, L. & Chater, N. (2002). Relative informativeness of quantifiers used in syllogistic reasoning. Memory and Cognition, 30, 138-149.

Osman, M. (2004). "An evaluation of dual-process theories of reasoning". Psychonomic Bulletin & Review 11 (6): 988–1010.

Pavličić, D. (1997). Individualne preferencije i racionalni izbor. Psihologija, 1997, 1-2, 49-76.

Pearce J.M. & Hall G. (1980) A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review, 87:532-552.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. Psychological Review, 94(1), pp. 61-73.

Pearl, J. (1999). Probabilities of Causation: Three Counterfactual Interpretations and their identification. Synthese, Vol. 121, 93-149, 1999.

Pearl, J. (2000) Causality: Models, Reasoning, and Inference. Cambridge University Press.

Pearl, J. (2010). The Mathematics of Causal Relations. Causality and Psychopathology: Finding the Determinants of Disorders and their Cures, Oxford University Press, 47-65, 2010.

Perales, J. & Catena, A. (2006). Human causal induction: A glimpse at the wholepicture. European Journal of Cognitive Psychology, 18(2), 277-320.

Perales, J. C., Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. Psychonomical Bulletin & Review, 14(4), 577-596.

Perales, J. C., Shanks, D. R. (2008). Driven by power? Probe question and presentation format effects on causal judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 34(6), 1482-1494.

Petitot, J. (1995). Morphodynamics and Attractor Syntax: Constituency in Visual Perception and Cognitive Grammar. U: Port, R. F. & van Gelder, T. (ur.), Mind as Motion. Explorations in the Dynamics of Cognition, Cambridge, MIT Press: 227-281.

Petitot, J., & Varela, F.J., & Pachoud, B., & Roy, J-M. (1999). Naturalizing Phenomenology. Issues in contemporary phenomenology and cognitive science. Stanford: Stanford University Press.

Piaget, J. (1973/1994). Uvod u genetičku epistemologiju. I. Matematičko mišljenje. Izdavačka knjižarnica Zorana Stojanovića. Sremski Karlovci, Novi Sad.

Piatelli-Palmarini, M. (Ur.) (1975). Language and Learning: The Debate Between Jean Piaget and Noam Chomsky. Routledge, 1975.

Pinker, S. & Ullman, M. (2002) The past and future of the past tense. Trends in Cognitive Science, 6, 456-463.

Plott, C. R. & Zeiler, K. (2005). The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect", Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. American Economic Review, 95(3), 530-545.

Popper, K. (1935/2002). The Logic of Scientific Discovery. Routlege: New York and London.

Prelec, D. (1998). The Probability Weighting Function. Econometrica, 66(3), 497-528.

Prigogine, I. & Stengers, I. (1984). Order out of Chaos: Man's new dialogue with nature. Bantam Books.

Prince, A. & Smolensky, P. 1997. Optimality: From neural networks to universal grammar. Science 275, 1604–1610.

Psillos, S, (2007). Past and Contemporary Perspectives on Explanation. U: Thagard, P. (ur.), Philosophy of Psychology and Cognitive Science (Volume 4 of the Handbook of the Philosophy of Science). New York: Elsevier.

Psillos, S. (2000). Carnap, The Ramsey-Sentence and Realistic Empiricism. Erkenntnis, 52: 253–279, 2000.

Quiggin, J. (1982). A theory of anticipated utility. Journal of Economic Behavior and Organization 3(4), 323–43.

Raaijmakers, J.G.W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. Cognitive Science, 27, 431-452.

Radonjić, S. (1967/1994). Uvod u psihologiju. Sedmo izdanje, Zavod za udžbenike i nastavna sredstva, Beograd.

Ramsey, F. P. (1926). Truth and Probability. In Ramsey, 1931, The Foundations of Mathematics and other Logical Essays, Ch. VII, p. 156-198, ed. By R. B. Braithwaite, London:

Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company. 1999 electronic edition.

Raskin, D. J. (2002). Constructivism in Psychology: Personal Construct Psychology, Radical Constructivism, and Social Constructionism. American Communication Journal 5(3).

Read, D. (2009). Experimental tests of rationality. U Anand, P., Pattanaik, P. K. & Puppe, C. (Ur) Oxford Handbook of Rational and Social Choice. Oxford University Press.

Rehder, B. (2003a). Categorization as causal reasoning. Cognitive Science, 27, 709-748.

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 1141-59.

Rehder, B. (2006). Human Deviations from Normative Causal Reasoning. U: Sun, R. & Miyake, N. (Urednici), Proceedings of the 28th Annual Conference of the Cognitive Science Society (p. 2596). Mahwah, NJ: Erlbaum.

Rehder, B. (2007). Essentialism as a generative theory of classification. U: Gopnik, A. & Schultz, L. (ur.), Causal learning: Psychology, philosophy, and computation, pp. 190-207. Oxford, UK: Oxford University Press.

Rehder, B. & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. Journal of Experimental Psychology: General, 130, 323-360.

Rehder, B. & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 659-683.

Rehder, B., & Milovanovic, G. (2007). Bias toward sufficiency and completeness in causal explanations. U: MacNamara, D. & Trafton, G. (ur.), Proceedings of the 29th Annual Conference of the Cognitive Science Society (p. 1843).

Rescher, N. (1989). Cognitive Economy: The Economic Dimension of the Theory of Knowledge, Pittsburgh: University of Pittsburgh Press.

Rescorla, R.A. & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, U: Black A.H. & Prokasy, W.F. (ur.), Classical Conditioning II, pp. 64–99. Appleton-Century-Crofts.

Rodet, L. & Schyns, P. G. (1994) Learning features of representation in conceptual context. In: Proceedings of the XVI Meeting of the Cognitive Science Society, 766–71. Erlbaum.

Rogers, T. T. & McClelland, J. L. (2004). Semantic Cognition: A Parallel Distributed Processing Approach. Cambridge, MA: MIT Press.

Rosch, E. (1975). Cognitive Reference Points. Cognitive Psychology, 7, 532-547.

Rosch, E. (1978). Principles of categorization. U: Rosch, E. & Lloyd, B. B. (ur.), Cognition and categorization (pp. 27-48). Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Rot, N. i Kostić, A. (1993a). Uticaj kvaliteta i veridikalnosti na proces suđenja. Psihologija, 1-2, 51-64.

Rot, N. i Kostić, A. (1993b). Uticaj kvaliteta i veridikalnosti na vreme verifikacije predikativnih i perceptivnih sudova. Psihologija, 3-4, 275-291.

Rot, N. i Kostić, A. (1995). Uticaj konteksta na verifikovanje predikativnih iskaza. Psihologija, 3-4, 225-256.

Rucker, R. (1982/2005). Infinity and the Mind – The Science and Philosophy of the Infinite. Prošireno izdanje Princeton Science Library, sa novim predgovorom autora, 2005. Princeton University Press, Princeton and Oxford.

Rumelhart, D.E., J.L. McClelland and the PDP Research Group (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, Cambridge, MA: MIT Press.

Ryle, G. (1949/1975). The Concept of Mind. Barnes & Noble, 1975.

Savage, L. J. (1954/1972). The Foundations of Statistics. Second Revised Edition. Dover Publications Inc., New York, 1972.

Savion, L. & Morado, R. (2002). Rationality, Logic, and Heuristics. Conference Proceedings of the Special Session on Knowledge Representation and Reasoning, June 2002 International Conference on Artificial Intelligence (ICAI'02, Vol II p. 791-797), CSREA Press.

Schiffer, S. (1982). Intention-Based Semantics. Notre Dame Journal of Formal Logic, 23(2), pp. 119-156.

Schmeidler, David (1989). Subjective Probability and Expected Utility without Additivity. Econometrica 57, 571–587.

Schmidt, U. (2004). Alternatives to Expected Utility: Formal Theories. U: Handbook of Utility Theory, Vol II, Hammond, P, Barbera, S. & Seidl, C. (ur.). Kluwer Academic.

Schneider, S. (2005). Direct Reference, Psychological Explanation, and Frege Cases. Mind & Language, 20(4), 423–447.

Schooler, L. & Anderson, J. R. (1991). Does memory reflect statistical regularity in the environment? In Proceedings of the 13th Annual Conference of the Cognitive Science Society, 227-232.

Schooler, L. J. & Anderson, J. R. (1993). Recency and Context: An Environmental Analysis of Memory. In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, pp. 889-894.

Schooler, L. J. & Anderson, J. R. (1997). The role of process in the rational analysis of memory. Cognitive Psychology, 32(3), 219-250.

Schrödinger, E. (1944). What Is Life? The Physical Aspect of the Living Cell. Zasnovano na predavanjima održanim pod patronatom Dablinskog instituta za napredne studije na Triniti koledžu, Dablin, u Februaru 1943. URL:http://whatislife.stanford.edu/LoCo_files/What-is-Life.pdf

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. Journal of Experimental Psychology: General, 110, 101-120.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory & Cognition, 23, 681–696

Schyns, P. G., Goldstone, R. L., & Thibaut, J-P (1998). Development of features in object concepts. Behavioral and Brain Sciences, 21, 1-54.

Schyns, P., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), The Psychology of Learning and Motivation, Vol. 31 (pp. 305-349). New

York: Academic Press.

Shanahan, Murray, "The Frame Problem", The Stanford Encyclopedia of Philosophy (Winter 2009 Edition), Edward N. Zalta (ed.), URL: http://plato.stanford.edu/archives/win2009/entries/frame-problem/

Shanks, D. R. (2004). Judging covariation and causation. U: Koehler, D. J. & Harvey, N. (ur.). Blackwell Handbook of Judgment and Decision Making (pp. 220-239). Oxford: Blackwell.

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

Shapiro, L. (2011). Embodied Cognition. Rotulege: New York, USA.

Shaw, R. E., McIntyre, M.(1974). Algoristic foundations to cognitive psychology. U: Palermo, D. & Weimer, W., Cognition and the Symbolic Processes., 305 - 362. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shepard, R. (2001). Perceptual-Cognitive Universals as Reflections of the World. Behavioral and Brain Sciences (2001), 24. Reprodukovano iz Psychonomic Bulletin & Review, 1994, 1, 2-28.

Shepard, R. N. (1987); "Toward a universal law of generalization for psychological science." Science, 237, 1317—1323.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. Psychonomic Bulletin and Review, 4(2), 145-166.

Shiffrin, R.M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. U: M. Oaksford & N. Chater (ur.), Rational Models of Cognition, pp. 73-95 Oxford, U.K.: Oxford University Press.

Šikić, Z. (ur.) (1987). Novija filozofija matematike, Nolit, 1987.

Simon, H. A. (1955a). A Behavioral Model of Rational Choice. The Quarterly Journal of Economics, Vol. 69, No. 1. (Feb., 1955), pp. 99-118.

Simon, H. A. (1955b). "On a class of skew distribution functions". Biometrika 42 (3–4): 425–440

Simon, H. A. (1956). "Rational choice and the structure of the environment". Psychological Review, Vol. 63 No. 2, 129-138.

Simon, H. A. (1972). Theories of Bounded Rationality. U: McGuire & Rander, R. (ur.), Decision and Organization. North Holland Publishing Group, 1972.

Sloman, S.A., & Lagnado, D. (2005). Do we 'do'?. Cognitive Science, 29, 5-39.

Slovic, P., Griffin, D. & Tversky, A. (1990). Compatibility Effects in Judgment and Choice. U Hogarth, R. M.(ur.), Insights in Decision Making: Theory and Applications. Chicago: The University of Chicago Press.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81, 214-241.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. Journal of Experimental Psychology: Learning, Memory, & Cognition, 26, 3-27.

Smith, J. D., & Minda, J. P.(1998). Prototypes in the mist: The early epochs of category learning. Journal of Experimental Psychology: Learning, Memory, & Cognition, 24,1411-1430.

Smith, J.Q. (2010). Bayesian Decision Analysis: Principles and Practice. Cambridge University Press.

Smolensky, P. (1987). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. Southern Journal of Philosophy 26 (Supplement), 137–163.

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F. & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. Proceedings of the National Academy of Sciences of the United States of America, 106(13), 5035-40.

Speaks, Jeff, "Theories of Meaning", The Stanford Encyclopedia of Philosophy (Summer 2011 Edition), Edward N. Zalta (ed.), URL: http://plato.stanford.edu/archives/sum2011/entries/meaning/

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. Psychological Science, 7, 337-342.

Sprons, O & Edelman, G. (1993). Solving Bernstein's Problem: A Proposal for the Development of Coordinated Movement by Selection. Child Development, 64, 960-98.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. Journal of Experimental Psychology: General, 127, 161-188.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Individual differences as essential components of heuristics and biases research. U: Manktelow, K., Over, D. & Elqayam, S. (ur.), The science of reason: A festschrift for Jonathan St. B. T. Evans (pp. 335-396). New York: Psychology Press.

Stanovitch, K. E. & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences, (2000) 23, 645–726.

Stewart, N. (2009). Decision by sampling: the role of the decision environment in risky choice. Quarterly Journal of Experimental Psychology, Vol.62 (No.6). pp. 1041-1062.

Stewart, N., Chater, N. & Brown, G. D. A. (2006). Decision by sampling. Cognitive Psychology. 53, 1-26.

Steyvers, M., Griffiths, T.L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. Trends in Cognitive Sciences, 10(7), 327-334.

Stich, S. & Nichols, S. (2003). Folk Psychology. U: The Blackwell Guide to Philosophy of Mind, Stich, S. & Warfield, T. A. (ur.), Oxford: Basil Blackwell.

Stojanović, B. (2005). Teorija igara: elementi i primena. Službeni glasnik. Beograd.

Sun, R. (2001). Hybrid systems and connectionist implementationalism. Encyclopedia of Cognitive Science, MacMillan Publishing Company, 2001.

Suppes, P. (2002). Representation and Invariance of Scientific Structures. CSLI Publications Center for the Study of Language and Information, Leland Stanford Junior University.

Tangen, J. M., & Allan, L. G. (2003). The relative effect of cue interaction. Quarterly Journal of Experimental Psychology, 56B, 279-300.

Tenenbaum, J. B, & Griffiths, T. L. (2001a) Generalization, similarity, and Bayesian inference, Behavioral and Brain Sciences, 24 pp. 629-641.

Tenenbaum, J. B. & Griffiths, T. L. (2001b). Structure learning in human causal induction.. Advances in Neural Information Processing Systems 13. Leen, T., Dietterich, T., and Tresp, V., Cambridge, MIT Press, 2001, 59-65.

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences, 10(7), 309-318.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. Science 331 (6022), 1279-1285.

Tenenbaum, J.B., Griffiths, T. L., and Niyogi, S. (2007). Intuitive theories as grammars for causal inference. U: Gopnik, A., & Schulz, L. (ur.), Causal learning: Psychology, philosophy, and computation. Oxford University Press.

Todd, P.M., and Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. Current Directions in Psychological Science, 16(3), 167-171.

Todd, P.M., Gigerenzer, G. (2000). Précis of Simple Heuristics that make us smart. Behavioral and Brain Sciences, 23, 727-780.

Tooby J. and DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. U: Kinzey, W. (ur.), Primate Models of Hominid Behavior. New York: SUNY Press.

Tooby, J. & Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. U: Buss, D. M. (ur.), The Handbook of Evolutionary Psychology (pp. 5-67). Hoboken, NJ: Wiley.

Townsend, J. T., & Ashby, F. G. (1983). The Stochastic Modeling of Elementary Psychological Processes. Cambridge: Cambridge University Press.

Turing, A. (1950). Computing Machinery and Intelligence. Mind 49, pp 433-460.

Tverksy, A. & Thaler, R. (1990) Anomalies: Preference Reversals. The Journal of Economic Perspectives, 4(2), 201-211.

Tversky, A. (1967). Utility theory and Additivity Analysis of Risky Choices. Journal of Experimental Psychology 75, 27–36.

Tversky, A. (1969). Intransitivity of preferences. Psychological Review, 76:31-48.

Tversky, A. (1977). Features of Similarity. Psychological Review, 84, 327-352.

Tversky, A. & Fox, C. R. (1995). Weighing risk and uncertainty. Psychological Review, Vol 102(2), 269-283.

Tversky, A. & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, New Series, Vol. 185, No. 4157 (Sep. 27, 1974), 1124-1131.

Tversky, A. & Kahneman, D. (1982). Judgments of and by representativeness. U: Kahneman, D., Slovic, P. & Tversky, A. (ur.), Judgment under uncertainty: Heuristics and biases. Cambridge, UK: Cambridge University Press.

Tversky, A. & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference Dependent Model. Quarterly Journal of Economics 106, 1039-1061.

Tversky, A. & Kahneman, D. (1992). Cumulative Prospect Theory: An Analysis of Decision under Uncertainty. Journal of Risk and Uncertainty, 5, 297-323.

Tversky, A. and Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review 90 (4): 293–315.

Tversky, A., and Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. Psychological Review, 101, 547–567.

Tversky, A., Slovic, P. & Kahneman, D. (1990). The causes of preference reversals. American Economic Review, 80(1), 204-217.

Tversky, A.; Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. Cognitive Psychology 5(2), 207–232.

van Gelder, T. (1996). Dynamics and cognition. U: Haugeland, J. (ur.), Mind Design II. (pp. 421-450).

van Hamme, L.J. & Wasserman, E.A. (1994). Cue competition in causality judgements: The role of nonpresentation of compound stimulus elements. Learning and Motivation, 25, 127–151.

Varela, F., Thompson, E., & Rosch, E. (1991). The embodied mind: Cognitive science and human experience. Cambridge MA: MIT Press.

Viscusi, W. K. (1989). Prospective Reference Theory: Toward an Explanation of the Paradoxes. Journal of Risk and Uncertainty 2(3), 235-63.

von Neumann, J. (1931/1987). Formalističko zasnivanje matematike. Prevod originalnog rada von Neumann, J. (1931). Die formalistische grundlegung der mathematik. Erkenntnis 2 (1) (1931), u zborniku "Novija filozofija matematike", Šikić, Z. (ur.), Nolit, Beograd, 1987.

von Neumann, J. & Morgenstern, O. (1944). Theory of Games and Economic Behavior. Princeton, NJ. Princeton University Press. 1944. Drugo izdanje 1947.

Vuong, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica 57, 307–333.

Wakker, P. (2008). Explaining the Characteristics of the Power (CRRA) Utility Family. Health Economics 17, 1329-1344.

Wakker, P. (2010). Prospect Theory For Risk and Ambiguity. Cambridge University Press, Cambridge, UK.

Wakker, P. P. & Tversky, A. (1993). An Axiomatization of Cumulative Prospect Theory. Journal of Risk and Uncertainty 7, 147-176.

Waldmann, M. R. (1996). Knowledge-based causal induction. U: Shanks, D. R., Holyoak, K. J. & Medin, D. L. (ur.), The psychology of learning and motivation, Vol. 34: Causal learning (pp. 47-88). San Diego: Academic Press.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 216-227.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. Journal of Experimental Psychology: General, 121,222-236.

Waldrop, M. (1993). Complexity. Simon & Schuster, New York.

Wason, P. C. (1968). Reasoning about a rule. Quarterly Journal of Experimental Psychology, 20, 273-281.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties and concept naturalness. Cognitive Psychology, 18, 158-194.

Wermter, S. & Sun, R. (2000) An overview of hybrid neural systems. U: Wermter, S. & Sun, S. (ur.) Hybrid Neural Systems. Springer-Verlag, Heidelberg. 2000.

Whitehead, A. N. & Russell, B. (1910, 1912, 1913). Principia Mathematica, 3 vols. Cambridge University Press, 1910, 1912, and 1913.

Wiener, N. (1954). Kibernetika i društvo, Nolit, Beograd, 1973.

Wisniewski, E. J. & Medin, D. L. (1994). On the interaction of theory and data in concept learning. Cognitive Science, 18, 221-281.

Wittgenstein, L. (1921/1987). Tractatus Logico-Philosophicus. Veselin Masleša – Svjetlost, Sarajevo.

Wolfram, S. (1983). Statistical Mechanics of Cellular Automata. Reviews of Modern Physics, 55 (July 1983), 601-644.

Wolfram, S. (2002). A New Kind of Science. Wolfram Media Inc. 2002.

Wood, J. C. & Wood, M. C. (2005). Joseph M. Juran: critical evaluations in business and management. Routledge.

Wright, C. and Bechtel, W. (2007). Mechanisms and psychological explanation. U: Thagard, P. (ur.), Philosophy of Psychology and Cognitive Science (Volume 4 of the Handbook of the Philosophy of Science). New York: Elsevier.

Wu, G. & Gonzalez, R. (1996). Curvature of the probability weighting function. Management Science, 42, pp. 1676–1690.

Wu, S.W., Delgado, M. & Maloney, L. T. (2009). Economic decision-making compared to an equivalent motor task. Proceedings of the National Academy of Sciences, USA, 106(15), 6088-6093.

Zalta, Edward N., "Gottlob Frege", The Stanford Encyclopedia of Philosophy (Spring 2012 Edition), Edward N. Zalta (ed.), forthcoming URL: http://plato.stanford.edu/archives/spr2012/entries/frege/

# Prilog A

## Formalna analiza ograničene racionalnosti pod Viskuzijevom teorijom i teorijom poverenja

Prilog predstavlja formalnu analizu bitnih fenomena ograničene racionalnosti i diskutuje neke posledice teorije poverenja koja je uvedena u V delu ove teze.

Teorija poverenja predstavlja proširenje Viskuzijeve teorije perspektivne reference (Viscusi, 1989) u uslovima rizika koje se odnosi na tačno određene psihološke pretpostavke u homeomorfnom modeliranju odlučivanja (Wakker, 2010) i ne menja strukturu teorije perspektivne reference u domenu koji se odnosi na same procese izbora. U takvoj situaciji, Viskuzijeva teorija perspektivne reference - uz male formalne izmene koje su čisto tehničkog karaktera - predstavlja primenu teorije očekivane korisnosti na *a posteriori* verovatnoće do kojih donosilac odluka dolazi primenom bejzijanske inferencije. Teorija poverenja je u pravom smislu reči „inženjerisana": ona je razvijena imajući na umu nameru da se što više empirijskih efekata koji predstavljaju fenomene ograničene racionalnosti izoluju od teorijske strukture koja opisuje same procese odlučivanja. Tako teorija poverenja u suštini predstavlja teoriju formiranja verovanja, jer daje deskripciju procesa koji prethodi odlučivanju i određuje koje vrednosti *a priori* verovatnoća kognitivni sistem koristi u bejzijanskoj inferenciji. Pošto se proces revizije verovanja bejzijanskom inferencijom smatra normativnim, a odlučivanje po aksiomatici teorije očekivanje korisnosti takođe, primena teorije poverenja obezbeđuje da celokupna konstrukcija teorije odlučivanja ostane racionalna u teorijskom smislu koji diskutujemo u ovoj tezi.

Analize koje slede se oslanjaju na formalnu ekspoziciju teorije poverenja i Viskuzijeve teorije perspektivne reference date u V delu. Za razumevanje ovih analiza, čitaocu se savetuje da prouči originalnu Viskuzijevu ekspoziciju iz 1989. (Viscusi, 1989), posebno dodatak A.1 u Viskuzijevom radu; aksiomatska analiza neće biti poduzeta u narednim redovima. Analize nekih fenomena koje je već poduzeo Viskuzi u razvoju teorije perspektivne reference ovde nećemo ponavljati; fokusiramo se samo na detalje koji učvršćuju bitne zaključke i fenomene koji nisu diskutovani u originalnoj teoriji. Na primer, u razvoju teorije perspektivne reference već je pokazano da ona prirodno objašnjava nalaze poput Aleovog paradoksa (i samim tim sve formalno slične probleme odlučivanja, poput paradoksa zajedničke proporcije). Cilj formalnih opservacija i analiza koje slede jeste da ukažu na uslove pod kojima teorija poverenja može da obuhvati robustne empirijske efekte koji predstavljaju evidenciju o ograničenoj racionalnosti. Pošto pokažemo da teorija poverenja može da obuhvati bitna standardna odstupanja, na ove analize se oslanjamo u VI delu naše teze u formulaciji dva bitna argumenta: (*i*) moguća je racionalna teorijska konstrukcija koja objašnjava fenomene ograničene racionalnosti, i (*ii*) postoji široka klasa problema odlučivanja na osnovu kojih nije moguće eksperimentalno razgraničiti racionalne i ograničeno racionalne teorije odlučivanja: bez obzira na značajno različitu semantiku ovih teorija, njih često odlikuju iste bihejvioralne reference, te su njihove teorijske konstrukcije u tom smislu razmenljive u objašnjenju odlučivanja.

Redom diskutujemo subjektivni tretman verovatnoća (B.1), averziju prema riziku i četvoročlanu strukturu stavova prema riziku (B.2), tretman kršenja deskriptivne invarijantnosti (B.3) i generalno pitanje formalnog odnosa teorije poverenja i kumulativne teorije izgleda (B.4).

### B.1 Subjektivni tretman verovatnoća

Kao što je već diskutovano u V delu ove teze, Viskuzijeva teorija perspektivne reference obezbeđuje objašnjenje ponderisanja verovatnoća, odn. fenomena koji se u kumulativnoj teoriji izgleda objašnjava višim ponderima odlučivanja za niske i nižim ponderima odlučivanja za visoke verovatnoće. Ponderi odlučivanja bi, prema standardnim empirijskim nalazima, generalno trebali da budu niži od objektivnih verovatnoća za visoke verovatnoće ishoda, i viši od objektivnih verovatnoća za niske verovatnoće ishoda, gde moramo biti veoma oprezni u upotrebi termina „objektivne verovatnoće". Naime, najčešće se taj termin interpretira kao da se odnosi na konkretnu verovatnoću pridruženu nekom ishodu na lozu. Pod kumulativnom teorijom izgleda (i svim RDU modelima u koje ona spada), ta interepretacija je složenija. Ponderiše se marginalni doprinos u jedinicima funkcije ponderisanja $w$, što znači da je ponder odluke koji se koristi uz neki ishod razlika ($i$) transformacije funkcijom ponderisanja $w$ sume verovatnoća svih ishoda većih od ishoda koji posmatramo $i$ verovatnoće ishoda koji posmatramo i ($ii$) transformacije funkcijom ponderisanja $w$ dekumulativne distribucije verovatnoće ishoda koji posmatramo, odn. sume svih ishoda većih od njega. U teoriji modela zavisnih od ranga (RDU), suma verovatnoća svih ishoda većih od ishoda koji posmatramo se naziva rangom verovatnoće tog ishoda, pa ako je on dat sa verovatnoćom $p$, kažemo da je njegov rang $p^r$. Sa ovakvom terminologijom, formulišemo ponder odluke kao $\pi(p) = w(p + p^r) - w(p^r)$. Proces je ilustrovan na Slici 4a. Ponderi odlučivanja koji su rezultat ovog procesa generalno nisu aditivni, odn. nisu verovatnoće; funkcija $w$ ponderisanja verovatnoća ima osobine $w(0) = 0$, $w(1) = 1$ i striktno je rastuća na intervalu [0,1] - to je sve što nam aksiomatska struktura modela zavisnih od ranga obezbeđuje da znamo. Detalji, poput inverznog-S oblika (Slika 4b), jesu posledice izbora određene parametarske forme za ovu funkciju.

U teoriji poverenja, za razliku od kumulativne teorije izgleda, transformišu se direktno objektivne verovatnoće kako su navedene na lozu, a ponderi odluke su *a posteriori* verovatnoće i jesu aditivne - odn. jesu prave verovatnoće. U razumevanju teorije poverenja veoma je bitno držati na umu ovu distinkciju između nje i modela zavisnih od ranga. Sam proces je opisan u V delu rasprave: ($i$) polazeći od dekumulativne forme distribucije verovatnoća monetarnih dobitaka i gubitaka, primenom Lusovog aksioma za podskup ishoda koje sadrži određeni loz, formiraju se *a priori* verovatnoće za svaki ishod na lozu; ($ii$) primenom jednostavne Dirišle-multinomijalne bejzijanske inferencije izračunavaju se *a posteriori* verovatnoće; ($iii$) skaliranje doprinosa *a priori* verovatnoća u procesu bejzijanske inferencije kontroliše faktor $N$, čije se izračunavanje oslanja na relativnu neizvesnost (entropiju) koju nosi *a priori* distribucija verovatnoća za određeni loz, dok je doprinos verovatnoća datih na lozu („objektivnih verovatnoća") uvek skaliran faktorom 100 zbog korespondencije sa skalom procenata koja se uobičajeno koristi u opisu rizičnih lozova. Podsetimo se, tek ovo „fiksiranje" skale intenziteta doprinosa objektivnih verovatnoća procesu bejzijanske inferencije omogućava jednoznačnu ocenu parametara modela teorije poverenja; ako bi oba parametra (obe *skale*, drugim rečima) bila slobodna, moguće bi bilo oceniti samo njihov odnos, ali ne i karakteristične vrednosti za svaki.

Viskuzi, naravno, diskutuje mehanizam koji vodi ponderisanju verovatnoća u teoriji perspektivne reference (Viscusi, 1989); mi ovde formalno specifikujemo uslove pod kojima se javljaju precenjivanje odn. potcenjivanje objektivno datih verovatnoća tek da bismo omogućili lakše praćenje formalnih inferencija koje slede. Pokazali smo u V delu da se vrednost *a posteriori*

verovatnoće za neki ishod $x$ na lozu dobija kao:

$$p''_x = \frac{Np'_x + 100p_x}{N + 100}$$

i sada primećujemo da iz prethodnog izraza, posle algebarskog sređivanja, sledi

$$\frac{p_x - p''_x}{p''_x - p'_x} = \frac{N}{100}$$

i pošto je izraz $\frac{N}{100}$ uvek pozitivan, brojilac i imenilac moraju biti istovremeno pozitivni ili negativni; sledi da jedan od naredna dva uslova mora da bude zadovoljen:

(*oba izraza pozitivna*)

$$p_x > p''_x > p'_x$$

(*oba izraza negativna*)

$$p_x < p''_x < p'_x$$

Prvi od dva navedena uslova (oba izraza pozitivna) očigledno opisuje uslove pod kojima dolazi do (a) *potcenjivanja* objektivnih verovatnoća (*a posteriori* verovatnoća je niža od objektivno date verovatnoće i mora biti viša od *a priori* verovatnoće) i (b) *precenjivanja* objektivnih verovantoća (*a posteriori* verovatnoća je viša od objektivno date verovatnoće i mora biti niža od *a priori* verovatnoće). Uzimajući u obzir i specijalan slučaj u kome faktor $N$ ima vrednost nula, koji se javlja samo za sigurne ishode posmatrane kao distribucije verovatnoća na kojima se sva masa nalazi na jednom ishodu, relaciju „>“ („<“) zamenjuje „≥“ („≤“).

### B.2 Četvoročlana struktura stavova prema riziku

Već više puta diskutovan empirijski efekat četvoročlane strukture stavova prema riziku uzima se kao „empirijski potpis“ kumulativne teorije izgleda (Tversky & Kahneman, 1992); proučićemo pod kojim uslovima on važi pod racionalnim modelom koji diskutujemo. Četvoročlana struktura stavova prema riziku podrazumeva sledeće odnose:

(*i*) sklonost prema riziku za niske verovatnoće dobitaka;
(*ii*) averzija prema riziku za srednje i visoke verovatnoće dobitaka;
(*iii*) averzija prema riziku za niske verovatnoće gubitaka;
(*iv*) sklonost prema riziku za srednje i visoke verovatnoće gubitaka.

Averzija prema riziku, ponovimo, implicira preferenciju sigurnog ishoda u visini očekivane vrednosti rizičnog loza nad samim lozom; drugim rečima, očekivana korisnost loza je niža od očekivane korisnosti sigurnog ishoda u visini očekivane vrednosti loza. Najlakše ćemo demonstrirati averziju prema riziku pod teorijom poverenja ako posmatramo ne-pozitivan loz oblika $L$: ($x$,$p_x$;$0$,$1$-$p_x$). Pošto je $u(0)$=0 po definiciji, imamo da je očekivana korisnost loza pod teorijom poverenja

$$EU(L) = p''_x \cdot u(x)$$

sa *a posteriori* verovatnoćom $p''_x$ određenom kao

$$p''_x = \frac{Np'_x + 100p_x}{N + 100}$$

Averzija prema riziku implicira

$$p''_x \cdot u(x) < p''_{p_x} \cdot u(p_x \cdot x)$$

i sada, zbog specifičnosti analize pod teorijom poverenja, moramo da budemo oprezni: desni izraz u nejednačini sadrži korisnost sigurnog ishoda u visini očekivane vrednosti loza L, odn. izraz $u(p_x \cdot x)$, kao i *a posteriori* verovantoću $p''_{p_x}$. Ta *a posteriori* verovatnoća mora da bude uključena u evaluaciju vrednosti sigurnog ishoda, jer pod teorijom poverenja i siguran ishod visine $p_x \cdot x$ ima određenu *a priori* verovatnoću da bude osvojen, bez obzira na to što je u samom problemu odlučivanja koji analiziramo on ponuđen sa verovatnoćom od jedan; očigledno je da će posle odgovarajuće bejzijanske inferencije taj siguran ishod imati verovatnoću osvajanja manju od jedan. Verovatnoća $p''_{p_x}$ je data preko izraza

$$p''_{p_x} = \frac{Np'_{p_x} + 100}{N + 100}$$

Posle ovoga, jasno je da je averzija prema riziku pod teorijom poverenja zadovoljena ako i samo ako

$$\frac{Np'_x + 100p_x}{Np'_{p_x} + 100} < \frac{u(p_x \cdot x)}{u(x)}$$

Nažalost, sve ovakve i slične analize su pod teorijom poverenja složenije nego pod kumulativnom teorijom izgleda, te najčešće vode ka formalnim uslovima koji su opet suviše komplikovani da bismo na osnovu njih mogli da donesemo jednostavne zaključke o tome kada će teorija predvideti, a kada ne, pojavu određene empirijske strukture u odlukama ispitanika. Analiza četvoročlane strukture stavova zahteva da uvedemo dodatne oznake. Neka je $p^s_x$ neka „mala“ verovatnoća sa kojom je na lozu $L_S$: $(x,p^S_x;0,1-p^S_x)$ ponuđen ishod $x$; neka je $p^H_x$ neka „srednja do velika“ verovatnoća sa kojom je na lozu $L_H$: $(x,p^H_x;0,1-p^H_x)$ ponuđen ishod $x$. Lozovi $L_S$ i $L_H$ se, dakle, razlikuju samo po verovatnoći sa kojom je ponuđen ishod $x$ na njima, tako da $p^H_x > p^S_x$. Neka je $p'_x$ *a priori* verovatnoća osvajanja ishoda $x$. Očekivana vrednost loza $L_S$ je $p^S_x \cdot x$, dok je očekivana vrednost loza $L_H$ $p^H_x \cdot x$; očigledno, biće nam potrebne *a priori* verovatnoće za osvajanje ovih očekivanih vrednosti lozova, koje označavamo sa $p'^S_{p_x \cdot x}$ i $p'^H_{p_x \cdot x}$. Na osnovu analize averzije prema riziku znamo da će se za loz na kojem je ishod $x$ ponuđen sa nekom srednjom do visokom verovatnoćom $p^H_x$ ona javiti ako i samo ako

$$p''^H_x \cdot u(x) < p''^H_{p_x \cdot x} \cdot u(p^H_x \cdot x)$$

gde smo sa $p_x^{''H}$ označili *a posteriori* verovatnoću osvajanja ishoda visine $x$ ako je on ponuđen sa verovantoćom $p_x^H$ na lozu. Simetrično, imamo da će se sklonost ka riziku za loz $L_S$ na kome je ishod $x$ ponuđen sa malom verovatnoćom $p_x^S$ javiti ako i samo ako

$$p_x^{''S} \cdot u(x) > p_{p_x^S \cdot x}^{''S} \cdot u(p_x^S \cdot x)$$

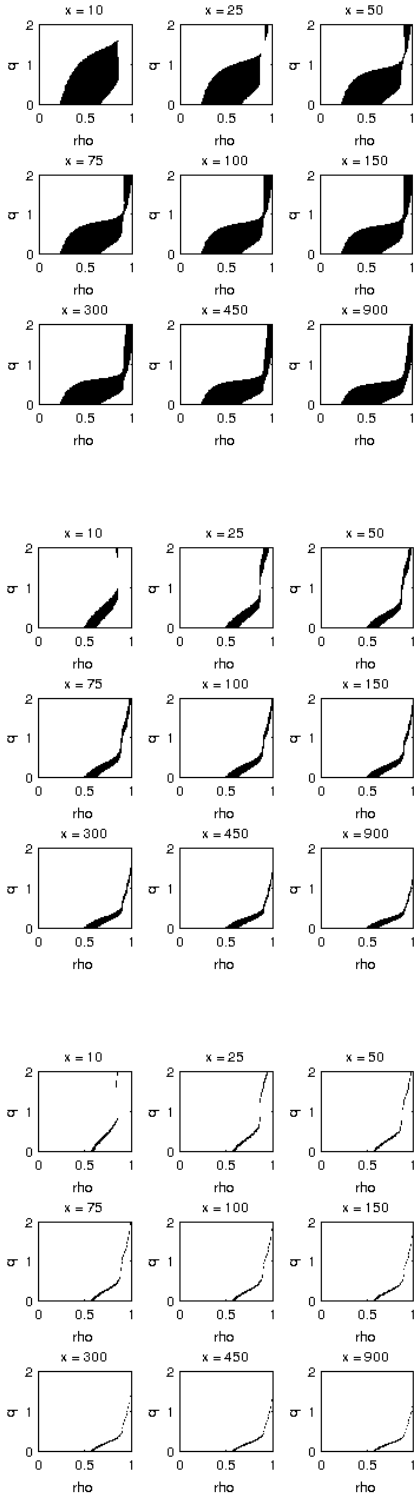
gde sa $p_x^{''S}$ označavamo *a posteriori* verovatnoću osvajanja ishoda visine $x$ ako je on ponuđen sa verovantoćom $p_x^S$. Posle algebarskog sređivanja ovih odnosa, možemo da izvedemo uslov njihovog simulatnog važenja koje je neophodno za egzistenciju četvoročlane strukture stavova prema riziku:

$$\frac{u(p_x^S \cdot x)}{u(x)} < \frac{N \cdot p'_x + 100 p_x^s}{N \cdot p'^S_{p_x \cdot x} + 100} < \frac{N \cdot p'_x + 100 p_x^H}{N \cdot p'^H_{p_x \cdot x} + 100} < \frac{u(p_x^H \cdot x)}{u(x)}$$

Lako se pokazuje da redosled dva srednja izraza u nizu nejednakosti koji predstavljamo jeste uvek zadovoljen, kao i činjenica da će odnosi korisnosti (spoljni levi i spoljni desni izraz) koje uključuje gornji niz nejednakosti uvek biti kakav je ovde dat. Međutim, ako posmatramo dve leve i dve desne nejednakosti, uviđamo da će njihov odnos zavisiti od složene interakcije visine ishoda $x$ koji se nalazi na lozu, verovatnoće sa kojom je $x$ ponuđen na $L$, forme funkcije korisnosti i forme funkcije koja reprezentuje *a priori* verovatnoće; ništa preciznije, nažalost, nije moguće reći o tome pod kojim uslovima će leva i desna nejednakost u gornjem izrazu biti zadovoljene. Isti analitički problem se javlja za ne-pozitivne lozove.

Empirijski mi već znamo da teorija poverenja objašnjava pojavu četvoročlane strukture stavova prema riziku, pošto je kombinacija precenjivanja niskih i potcenjivanja visokih verovatnoća sa konkavnom funkcijom korisnosti dovoljna da objasni ovaj fenomen: precenjivanje niskih verovatnoća radi „protiv“ averzije prema riziku sa konkavnom funkcijom korisnosti, kao što potcenjivanje visokih verovatnoća doprinosi averziji prema riziku - opet sa konkavnom funkcijom korisnosti; simetričan odnos obezbeđuje deo četvoročlane strukture stavova prema riziku koja se odnosi na domen gubitaka.

Sledeća demonstracija efekta četvoročlane strukture stavova prema riziku pod teorijom poverenja je ilustrativna. U odnosu na upravo diskutovane uslove pod kojima se ovaj efekat javlja u teoriji poverenja izveli smo pretragu kroz prostor parametara koji definišu donosioca odluka za različite vrednosti iznosa $x$ na lozu $L$: ($x$,$p_x$;*0*,*1-*$p_x$) i različite kombinacije vrednosti $p_x^S$ i $p_x^H$. Podrazumevajući stepenu funkciju korisnosti, varirali smo vrednost eksponenta stepene funkcije na rasponu od 0 do 1 (neutralnost prema riziku) sa korakom .01; podrazumevajući standardnu Pareto dekumulativnu funkciju tipa II (koju smo koristili u V delu u razvoju teorije poverenja) za reprezentaciju *a priori* verovanja o mogućim monetarnim dobicima, varirali smo vrednost njenog parametra $q$ na rasponu od 0 do 2, takođe sa korakom .01. Varirali smo visinu ishoda $x$ na lozu $L$ kao: 10, 25, 50, 75, 100, 150, 300, 450, 900.




Slika B.2.1. *Pretraga prostora parametara teorije poverenja i provera važenja formalnih uslova za četvoročlanu strukturu stavova prema riziku na domenu dobitaka.* Objašnjenje u tekstu.

Za svaku kombinaciju vrednosti parametara, visine ishoda $x$ i tri sheme dodele konkretnih „niskih“ ($p_x^S$) i „srednjih do visokih“ ($p_x^H$) verovatnoća ishodu $x$ na lozu $L$, proverili smo kada

su formalni uslovi za važenje efekta četvoročlane strukture stavova prema riziku zadovoljeni na domenu dobitaka. Slika B.2.1 prikazuje rezultate ove pretrage. Na gornjem panelu slike B.2.1 (prvih devet grafikona) definišemo „niske“ i „srednje do visoke“ verovatnoće kao $p_x^S = .01$ i $p_x^H = .99$; na srednjem panelu, $p_x^S = .25$ i $p_x^H = .75$; i konačno, na donjem panelu, $p_x^S = .45$ i $p_x^H = .55$. Na svakom od tri panela vrednost ishoda $x$ raste kroz grafikone kako je opisano. Abscise svih grafikona predstavljaju eksponent stepene funkcije korisnosti na rasponu od 0 do 1 sa korakom .01, a ordinate parametar $q$ standardne Pareto dekumulativne distribucije tipa II na rasponu od 0 do 2 sa korakom .01. Senčeni regioni na grafikonima predstavljaju tačke u kojima su formalni uslovi za četvoročlanu strukturu stavova na domenu dobitaka zadovoljeni.

Ne treba gubiti iz vida da je ponderisanje sigurnih dobitaka (ili gubitaka) pod teorijom poverenja donekle arbitrarna odluka. Ako biramo da budemo potpuno teorijski konzistentni, u prirodi je ovakve teorije da postulira ocenu stepena poverenja i u siguran dobitak; s druge strane, nema posebnih formalnih prepreka koje bi ograničavale tretman sigurnih dobitaka istovetan onom koji daju teorija očekivane korisnosti i teorija izgleda. U ove dve teorije, sigurni dobici i gubici se tretiraju *prima facie*, bez uticaja subjektivnih parametara u njihovoj percepciji. Ukoliko se odlučimo za takav tretman sigurnih dobitaka u teoriji poverenja, uslovi za važenje četvoročlane strukture stavova prema riziku na domenu dobitaka postaju

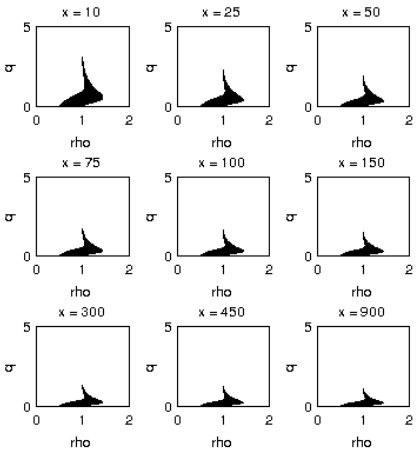
$$p_x^{''H} \cdot u(x) < u(p_x^H \cdot x)$$

na rasponu srednjih do visokih verovatnoća, odn.

$$p_x^{''S} \cdot u(x) > u(p_x^S \cdot x)$$

na rasponu niskih verovatnoća. Slika B.2.2 ilustruje pretragu prostora parametara na sličnim rasponima (parametar $q$ pretražen je na rasponu od 0 do 5 sa korakom .01) kao na slici B.2.1 i definicijom „niskih“ i „srednjih do visokih“ verovatnoća za ishod $x$ na lozu $L$ kao $p_x^S = .10$ i $p_x^H = .50$ (ovaj raspon su kao interesantan prvi odabrali Kaneman i Tverski u radu iz 1992, Kahneman & Tversky, 1992, a mi smo ga se pridržavali u Tabelama 5a i 5b u V delu) u potrazi za zadovoljenjem upravo navedenih formalnih uslova pod kojima se četvoročlana struktura stavova prema riziku javlja kada u teoriji poverenja *ne utičemo* na subjektivnu percepciju izvesnih ishoda.

Javljanje ili odsustvo četvoročlane strukture stavova prema riziku je, pod teorijom poverenja, posledica složenih interakcija velikog broja parametara, kako onih koji odlikuju donosioca odluka, tako i onih koji odlikuju eksperimentalni nacrt pod kojim pokušavamo da taj efekat ustanovimo. Uzimajući u obzir već osvedočenu eksplanatornu moć teorije poverenja, kao i nestabilnost empirijskih efekata kroz ispitanike i eksperimente koja karakteriše odlučivanje u uslovima rizika i neizvesnosti baš kao i svaku drugu oblast eksperimentalne kognitivne psihologije, smatramo da nema boljeg trenutka od ovog da podsetimo u kojoj meri je značajno razumeti da neće svaki eksperimentalni nacrt biti pogodan za selekciju unutar nekog skupa modela kognitivnih funkcija, baš kao što ni svaki pojedinačni ispitanik neće nužno ispunjavati uslove da pruži podatke koji bi omogućili takvu selekciju.




Slika B.2.2. *Pretraga prostora parametara teorije poverenja i provera važenja formalnih uslova za četvoročlanu strukturu stavova prema riziku na domenu dobitaka u verziji teorije poverenja koja ne uvodi subjektivne (a priori) verovatnoće u ocenu izvesnih ishoda.* Objašnjenje u tekstu.

### B.3 Tretman kršenja deskriptivne invarijantnosti

Vratimo se još jednom diskusiji Aleovog paradoksa, možda najznačajnijeg empirijskog fenomena koji je motivisao potragu za dubljim uzrocima odstupanja od teorije očekivane korisnosti i razvoj deskriptivnih teorija odlučivanja poput teorije izgleda i drugih. Sam fenomen koji indukuje postavka Aleovog problema diskutovan je još u I delu ove teze. U slavnim „*Osnovama statistike*“ Leonard Džimi Sevidž (Savage, 1954/1972) diskutuje alternativnu deskripciju Aleovog problema za koju veruje da bi otklonila empirijski fenomen koji protivreči normativnim osnovama teorije očekivane korisnosti. Sevidž polazi od Aleove originalne postavke (ovde koristimo monetarne iznose koje koristi Sevidž u svojoj prezentaciji problema):

*Opcija A*: sa sigurnošću (100%) dobitak od \$500.000
*Opcija B*: sa 89% dobitak od \$500.000, sa 10% dobitak od \$2.500.000, i sa 1% *status quo* (\$0)

*Opcija A1*: sa 89% *status quo* (\$0) i sa 11% dobitak od \$500.000
*Opcija B1*: sa 90% *status quo* (\$0) i sa 10% dobitak od \$2.500.000

posle koje je sam dao lično Aleu odgovor koji protivreči normativnim principima očekivane korisnosti, tj. krši aksiom nezavisnosti: preferencija $A \succ B$ između prva dva, te $B1 \succ A1$ između druga dva loza. Razvijena formu Aleovog problema koju daje Sevidž predstavlja strukturu lozova u problemu na donekle transparentniji način:

Tabela B1. Sevidžova formulacija Aleovog paradoksa u razvijenoj formi (prema Savage, 1954/1972)

| | 1% | 10% | 89% |
|---|---|---|---|
| *A* | $ 500.000 | $ 500.000 | $ 500.000 |
| *B* | $ 0 | $ 2.500.000 | $ 500.000 |
| *A1* | $ 500.000 | $ 500.000 | $ 0 |
| *B1* | $ 0 | $ 2.500.000 | $ 0 |

Sevidž dalje elaborira ovakvu prezentaciju lozova i zaključuje da mu ona omogućava da ispravi prvobitnu „grešku“ u preferencijama - izraz koji sam koristi, primećujući da je neobično iskoristiti termin „greška“ kada se govori o čisto subjektivnom konstruktu preferencije - tako da njegove odluke budu normativne (Savage, 1954/1972). Sve diskusije ovog tipa, treba da bude jasno, izlaze iz okvira same teorije odlučivanja i u principu zahtevaju psihološke teorije šireg obima - teorije koje govore o reprezentaciji problema odlučivanja. Teorija poverenja, u kojoj je evaluacija ishoda na rizičnim lozovima suštinski zavisna od konteksta koji čini celokupan loz i počiva na *a priori* verovanjima donosioca odluka, međutim, predstavlja teoriju u okviru koje je moguće diskutovati probleme deskriptivne invarijantnosti poput ovog, i to bez uvođenja dodatnih teorijskih pretpostavki, odn. *u istom pojmovnom okviru koji objašnjava procese odlučivanja same*. Ovu osobinu garantuje već teorijska struktura Viskuzijeve teorije poverenja, na kojoj teorija poverenja počiva; tom modelu teorija poverenja dodaje izvesnu fleksibilnost tako što nudi mehanizam koji egzaktno određuje *a priori* verovanja u situaciji višestrukih referentnih tačaka kako je diskutovano u V delu teze (up. Viscusi, 1989).

Pogledajmo ponovo Sevidžovu deskripciju Aleovog problema u tabeli B1. U teoriji poverenja, *ako pretpostavimo da ovakvu reprezentaciju loza donosilac odluka ne kolapsira u početnu formulaciju koju daje Ale* (što je, verujemo, morao da pretpostavi i Džimi Sevidž kada je predstavio novu deskripciju problema), evaluacija lozova postaje zavisna od načina na koji su oni predstavljeni. Na primer, zbog efekta konteksta u izračunavanju stepena poverenja $N$ koji je određen kao (reskalirana) relativna entropija loza u teoriji poverenja, nije isto da li loz formulišemo kao „90% *status quo* ($0) i 10% dobitak od $2.500.000“ (Ale) ili kao „1% *status quo* ($0), 10% dobitak od $2.500.000 i 89% *status quo* ($0)“ (Sevidž); u slučaju druge formulacije, na lozu su ponuđena dva *status quo* (odn. $0) ishoda, od kojih oba imaju *a priori* dekumulativnu verovatnoću od 1 na standardnoj Paretovoj distribuciji tipa II koju koristimo u modelu teorije poverenja. Vidimo da će stepen poverenja $N$ za dve formulacije „istog“ loza biti različiti, a pored toga, vrednosti *a posteriori* verovatnoća *za iste visine ishoda* će biti različite; samim tim, biće različite i konačne evaluacije ovih lozova.

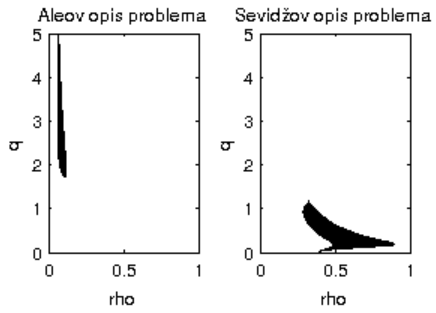
Kao što smo već diskutovali u sekciji B.2, u teoriji poverenja postoje dve mogućnosti tretmana sigurnih ishoda (poput sigurnog ishoda od $500.000 u lozu A originalne Aleove postavke problema). Ukoliko se tretman rizičnih ishoda proširi na sigurne ishode teorijski konzistentno, onda se korisnost

sigurnog ishoda evaluira polazeći od njegove *a priori* verovatnoće bez obzira što je on na lozu ponuđen sa izvesnošću; s druge strane, on može da se tretira i samo preko funkcije korisnosti, odn. *prima facie* - kao što je dat i kao što bi bio tretiran u teoriji očekivane korisnosti ili teoriji izgleda. U zavisnosti od toga za koji od dva moguća tretmana izvesnih ishoda se odlučimo, evaluacije problema odlučivanja koji uključuju izvesne ishode će se pod teorijom poverenja razlikovati.

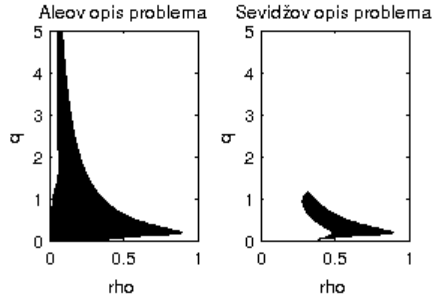
Razvili smo formalne uslove koji moraju da budu zadovoljeni da bi struktura preferencija u Aleovom paradoksu bila moguća pod teorijom poverenja; postupak je sličan prethodnoj analizi četvoročlane strukture stavova prema riziku ali zahteva detaljnu analizu sva četiri loza u problemu, zbog čega izbegavamo prikaz same formalne inferencije. Zatim smo, takođe slično postupku u sekciji B.2, izveli pretragu prostora parametara teorije poverenja - kroz različite vrednosti ($i$) parametra $q$ standardne dekumulativne Pareto distribucije tipa II i ($ii$) vrednosti eksponenta stepene funkcije korisnosti $\rho$ - da bismo pronašli regione u prostoru parametara u kojima se Aleov paradoks javlja. Raspon vrednosti parametara koji pretražujemo je isti kao u prethodnoj analizi četvoročlane strukture stavova prema riziku (slika B.2.2). U prvoj pretrazi koristimo originalnu Aleovu formulaciju problema, dok u sledećoj problem formulišemo kao što je to predložio Sevidž; u oba slučaja, dopuštamo subjektivni tretman izvesnih ishoda (dakle, evaluiramo ih polazeći od njihovih *a priori* verovatnoća), Zatim ponavljamo obe pretrage ali izvesne ishode tretiramo *prima facie* (tj. tretiraju se isključivo kroz funkciju korisnosti, bez ikakve subjektivne korekcije datih verovatnoća). Slika B3.1 prikazuje rezutate prvog para pretraga, dok su rezultati drugog para pretraga dati na slici B3.2.

Teorija poverenja, kao što vidimo na slici B.3.1, ne dozvoljava pojavu Aleovog paradoksa u istom regionu ključnih parametara odlučivanja tj. za iste ispitanike *u dve različite deskripcije problema*: ukoliko dozvolimo da se i sigurni ishodi evaluiraju tek pošto se u obzir uzmu njihove *a priori* verovatnoće, teorija poverenja predviđa da ispitanici koji pokažu preferencije Aleovog paradoksa u originalnoj formulaciji problema *to neće učiniti ako se problem formuliše kako ga je formulisao Sevidž*, i obrnuto. Međutim, teorija poverenja predviđa da je region u kome se paradoks javlja, nasuprot Sevidžovoj intuiciji, još širi ako se problem postavi kako ga on diskutuje nego kao što ga je Ale originalno postavio. Ukoliko se, međutim, pretpostavi da se sigurni ishodi tretiraju *prima facie*, dakle kao što ih tretira teorija očekivane korisnosti, region u kome se paradoks javlja pod opisom koji je originalno dao Ale je daleko veći nego pod opisom koji je dao Sevidž, što vidimo sa slike B.3.2. Ipak, sada se dva regiona donekle preklapaju, tako da je prema teoriji poverenja moguće da se jave ispitanici koji će konzistentno kršiti aksiom nezavisnosti teorije očekivane korisnosti u ovom problemu - bez obzira da li je on dat u originalnoj formulaciji ili u Sevidžovoj razvijenoj formi. Drugi par pretrage prostora parametara sugeriše da je Džimi Sevidž mogao da bude u pravu tvrdeći da je njegova deskripcija Aleovog problema transparentnija.

Ovaj skroman prilog jednoj klasičnoj raspravi u teoriji odlučivanja predstavljamo više kao ilustraciju tretmana deskriptivne invarijantnosti pod teorijom poverenja nego kao vid konačnog rešenja problema koje je nametnuo Aleov problem odlučivanja. Fenomeni vezani za normativni kriterijum deskriptivne invarijantnosti su toliko složeni da njih izvesno neće sve moći da objasni ovaj tretman koji prirodno sledi iz teorije poverenja. Teorija poverenja daje prilično složene predikcije o javljanju ovog čuvenog paradoksa; one očigledno zahtevaju temeljnu eksperimentalnu proveru.




Slika B.3.1. *Pretraga prostora parametara teorije poverenja i provera važenja formalnih uslova za javljanje Aleovog paradoksa u originalnoj (levi panel) i Sevidžovoj postavci problema: verzija teorije poverenja koja koristi a priori verovatnoće u oceni izvesnih ishoda.*




Slika B.3.2. *Pretraga prostora parametara teorije poverenja i provera važenja formalnih uslova za javljanje Aleovog paradoksa u originalnoj (levi panel) i Sevidžovoj postavci problema: verzija teorije poverenja koja ne koristi a priori verovatnoće u oceni izvesnih ishoda.*

### B.4 Teorija poverenja, kumulativna teorija izgleda i teorija očekivane korisnosti

Vreme je da posvetimo pažnju odnosima formalnih deskripcija odlučivanja u uslovima rizika pod teorijom poverenja (CT), kumulativnom teorijom izgleda (CPT) i teorijom očekivane korisnosti (EU). CT, poput CPT, predstavlja određenu generalizaciju EU. Zapitajmo se pod kojim uslovima CT i CPT daju istovetne evaluacije rizičnih lozova? Neka je dat loz $L$: ($x$,$p_x$;*0*,*1-*$p_x$); pod CT:

$$EU(L) = \frac{Np'_{p_x} + 100}{N + 100} \cdot u(x)$$

dok pod CPT, podrazumevajući da u obe teorije koristimo *istu* funkciju korisnosti, dobijamo evaluaciju:

$$EU(L) = \pi(p_x) \cdot u(x)$$

gde je $\pi(x) = w(x)$, specifično za distribuciju ne-negativnog loza sa samo jednim dobitkom i nulom. Očigledno sledi

$$w(p_x) = \frac{Np'_{p_x} + 100}{N + 100}$$

i

$$\frac{p_x - w(p_x)}{w(p_x) - p'_x} = \frac{N}{100}$$

Razvijajući $N$ prema CT kao

$$N = \frac{H(p'_x, 1 - p'_x)}{log\frac{1}{2}}$$

posle algebarskog sređivanja izraza, dobijamo

$$w(p_x) \cdot H(p'_x, 1 - p'_x) + w(p_x) \cdot log\frac{1}{2} = p'_x \cdot H(p'_x, 1 - p'_x) + p \cdot log\frac{1}{2}$$

iz čega očigledno sledi (*i*) $w(p_x) = p'_x$, (*ii*) $w(p_x) = p_x$ i (*iii*) $p_x = p'_x$. Iz (*i*) ili (*ii*) sledi da je funkcija ponderisanja verovatnoće u CPT linearna funkcija identiteta, što znači da se pod CPT realizuje EU kao specijalan slučaj. Iz (*iii*) sledi da je verovatnoća data na lozu identična *a priori* verovatnoći odgovarajućeg ishoda, što ukazuje na to da se pod CT realizuje EU kao specijalan slučaj (up. jednačinu (61), V deo). Drugim rečima, kumulativna teorija izgleda i teorija poverenja, očekivano, vode ka istoj evaluaciji lozova na domenu u kome se realizuje EU kao njihov specijalan slučaj. Tvrdnja se lako generalizuje na slučaj loza sa proizvoljnim brojem ishoda.

Iz (*iii*) sledi dalja specifikcija ovog odnosa. Pošto (*iii*) tvrdi da su *a priori* verovatnoće i objektivne (date) verovatnoće na lozu iste za odgovarajuće ishode, a *a priori* verovatnoće su (posle normalizacije) pod CT proporcionalne dekumulativnim verovatnoćama ishoda (sa standardne Pareto distribucije II tipa), sledi da se EU realizuje kao specijalan slučaj pod CT ako i samo ako su viši ishodi na lozovima konzistentno dati sa verovatnoćama manjim od verovatnoća sa kojima su dati niži ishodi, što sledi iz činjenice da je Pareto dekumulativna funkcija striktno opadajuća. Takve lozove možemo da nazovemo *prirodno sortiranim lozovima*.

Problem odnosa CT i CPT van domena u kome one realizuju EU kao specijalan slučaj je složen. Empirijski znamo da postoji veliki podskup lozova na kojima nećemo moći da razlikujemo predikcije jedne od predikcija druge teorije; ovo sledi iz prethodnih dokaza o mogućnosti da se karakteristični empirijski fenomeni ograničene racionalnosti poput četvoročlane strukture stavova prema riziku ili Aleovog paradoksa (i, samim tim, čitave klase problema odlučivanja koja počiva na formalnoj strukturi Aleovog paradoksa, up. Kahneman & Tversky, 1979) realizuju i pod teorijom poverenja. U konjukciji sa očiglednom konvergencijom ove dve teorije u domenu lozova na kojima obe realizuju EU kao specijalan slučaj, jasno je da postoji ogroman prostor eksperimentalnih podataka u kojima njihove predikcije nije moguće razlikovati. Na osnovu naših eksperimentalnih nalaza dobijenih merenjem monetarnih ekvivalenata (eksperimenti 2a i 2b, V deo), jasno je da je aksiomatsku strukturu CPT neophodno proširiti tako da ona omogući uvođenje više funkcija ponderisanja verovatnoća. Posle takve modifikacije CPT, formalne strukture CPT i CT bi bilo lakše uporediti. Međutim, pošto bi te funkcije ponderisanja verovatnoća u CPT, kao što smo pokazali empirijski u V delu, morale da budu određene nekim parametrom koji bi ih sistematski

doveo u vezu sa visinama ishoda na odgovarajućim lozovima, postavlja se pitanje u kojoj meri bi formalna struktura tako proširene CPT uopšte bila različita od strukture CT.

Model CT koji smo koristili u ovoj raspravi je dat direktno u parametarskoj formi. Jasno je, i bez dublje matematičke diskusije, da se (*i*) dodavanjem aksioma koji bi podrazumevao striktno opadajuće uređenje *a priori* verovatnoća sa porastom visine ishoda Viskuzijevoj teoriji poverenja, te (*ii*) daljim ispitivanjem mogućnosti različitih mehanizama formiranja verovanja o stepenu poverenja $N$, dobija zapravo familija modela odlučivanja sa sličnim osobinama - familija čiji je tek jedan član diskutovan u ovoj tezi.

# Biografija autora

Goran Milovanović je rođen 1974. u Beogradu, od oca Slobodana i majke Milice. Tokom očeve diplomatske službe sedamdesetih godina odrasta u Rimu, Italija. Osnovnu školu i gimnaziju završava u Beogradu. 1992. godine upisuje Matematički fakultet Univerziteta u Beogradu. 1993. godine upisuje psihologiju na Filozofskom fakultetu Univerziteta u Beogradu gde diplomira 2004. godine sa radom „*Strukturalna kompleksnost semantičkih domena*“ pod mentoratom akademika Dr Aleksandra Kostića, neko vreme studirajući uporedo i na Odeljenju za filozofiju.

U gimnaziji se interesuje za kompjuterske nauke i probleme veštačke inteligencije, pohađa naučne programe Istraživačke stanice Petnica u tim oblastima, programira i piše za popularne kompjuterske časopise u bivšoj Jugoslaviji. Od 1993. godine saradnik je Laboratorije za eksperimentalnu psihologiju u Beogradu, gde sprovodi eksperimentalne studije kategorizacije, semantičke verifikacije i rezonovanja. Prvi naučni rad završava 1994. godine, a objavljuje u časopisu „*Psihologija*“ 1996. Redovno izveštava o svom radu na naučnim skupovima, učestvuje u nastavi opšte psihologije i metodologije, drži predavanja po pozivu u više naučnih institucija.

Doktorske studije započinje 2005. na Programu za kogniciju i percepciju Njujorškog univerziteta, Njujork, SAD, pod mentoratom Dr Gregori Marfija, a nastavlja 2009. godine na Filozofskom fakultetu Univerziteta u Beogradu, pod mentoratom Dr Gordane Jovanović.

Urednik je, autor i koautor više knjiga i empirijskih studija iz oblasti razvoja Interneta i informacionog društva, više radova u oblasti eksperimentalne psihologije u domaćim i stranim časopisima i na konferencijama nacionalnog i međunarodnog značaja, predavao je razne oblasti psihologije na srpskom i engleskom jeziku u zemlji i inostranstvu. Predaje kognitivnu psihologiju, učenje i više kognitivne procese na Fakultetu za medije i komunikacije Univerziteta Singidunum u Beogradu.