

# HUMAN RATIONALITY : Festschrift for Nenad Smokrović

---

**Edited book / Urednička knjiga**

*Publication status / Verzija rada:* **Published version / Objavljena verzija rada (izdavačev PDF)**

*Publication year / Godina izdavanja:* **2022**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:186:273617>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-07-18**



*Repository / Repozitorij:*

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)

# HUMAN RATIONALITY

*Festschrift* for  
Nenad Smokrović

**EDITORS:**

**Boran Berčić**

**Aleksandra Golubović**

**Majda Trobok**

# HUMAN RATIONALITY

*Festschrift* for  
Nenad Smokrović

**EDITORS:**

**Boran Berčić**

**Aleksandra Golubović**

**Majda Trobok**

**Faculty of Humanities and Social Sciences**

**University of Rijeka 2022**

**[www.ffri.uniri.hr](http://www.ffri.uniri.hr)**

*Title*

HUMAN RATIONALITY

*Festschrift* for Nenad Smokrović

*Editors*

Boran Berčić

Aleksandra Golubović

Majda Trobok

*Publisher*

Faculty of Humanities and Social Sciences

University of Rijeka

Sveučilišna avenija 4, 51000 Rijeka

[www.ffri.uniri.hr](http://www.ffri.uniri.hr)

*For the Publisher*

Aleksandar Mijatović

*Reviewers*

Dušan Dožudić, Martina Fuerst, Filip Grgić,

Ljudevit Hanžek, Andrej Jandrić, Tomislav Janović,

Hrvoje Jurić, Guido Melchior, Dragana Sekulić

*Proofreading*

Jelena Kopajtić

Leonard Pektor

*Design & Print*

Grafika Helvetica d.o.o. Rijeka

[www.grafikahelvetica.com](http://www.grafikahelvetica.com)

*Cover*

Luka Aničić: *The Orange End of Nature*, acrylic on cardboard.

[www.anicic.art](http://www.anicic.art)

*Publishing date*

December 2022

© Editors and Contributors

ISBN: 978-953-361-064-1

The CIP record is accessible at the computer catalogue of the University Library in Rijeka under the number 150410040.

This book is published with the support of Ministry of Science and Education of Republic of Croatia; Faculty of Humanities and Social Sciences in Rijeka; research projects of the University of Rijeka *Metametaphysics* Uniri-human-18-239, and *Critical thinking and Society* Uniri-human-18-254.

---

## CONTENT

<b>BORAN BERČIĆ, ALEKSANDRA GOLUBOVIĆ, MAJDA TROBOK</b> Editors' Preface . . . . .	i
<b>MILOŠ ARSENIJEVIĆ</b> The Future Sea Battle and Performing an Infinite Task: Two Remarkable Cases Concerning the Logician Thesis . . . . .	1
<b>IGOR BAJŠANSKI</b> Funkcije rasuđivanja u individualnom i grupnom kontekstu . . . . .	19
<b>GABRIELA BAŠIĆ HANŽEK</b> Teorija i klasifikacija pogreški u argumentaciji: stvarne i manje bitne razlike između dvaju pristupa . . . . .	37
<b>HANOCH BEN-YAMI, EDI PAVLOVIĆ</b> Completeness of the Quantified Argument Calculus on the Truth-Valuational Approach . . . . .	53
<b>BORAN BERČIĆ</b> <i>X is the best, but I prefer Y!</i> On Values and Preferences . . . . .	79
<b>ALEKSANDRA GOLUBOVIĆ, JELENA KOPAJTIĆ</b> Svjetonazor i odgoj kritičkog mislitelja . . . . .	95
<b>MARKO JURJAKO</b> Naturalizam i relativnost u pogledu praktičnih razloga . . . . .	113
<b>PAOLO LABINAZ</b> Argumentation, Knowledge and Reasoning. . . . .	141
<b>NENAD MIŠČEVIĆ</b> How Rational are Human Beings? In Honor of Nenad Smokrović . . . . .	159
<b>INES SKELAC</b> Uloga logike u ljudskom zaključivanju. . . . .	181
<b>MATEJ SUŠNIK</b> Priroda praktičnog zaključivanja . . . . .	195

---

<b>DANILO ŠUSTER</b>	
A Mid-Blue Logic . . . . .	.211
<b>MAJDA TROBOK</b>	
The Role of Argumentation - In Honor of Nenad Smokrović . . . . .	.229
<b>ANDREJ ULE</b>	
Implicit and Explicit Knowledge in Argumentation . . . . .	.241
<b>LINO VELJAK</b>	
O utemeljenju metodologije znanstvenog istraživanja . . . . .	.251
<b>MICHAEL WATKINS</b>	
The Mastery of a Concept: Dispositions and Skills . . . . .	.265
<b>TIMOTHY WILLIAMSON</b>	
Idealized Rationality in Models of Knowledge and Probability . . . . .	.273
<b>NENAD SMOKROVIĆ</b>	
Acknowledgments, Comments and Answers . . . . .	.291
Index of Names. . . . .	.303

---

## CONTRIBUTORS

**MILOŠ ARSENIJEVIĆ**, Department of Philosophy, University of Belgrade  
marsenij@f.bg.ac.rs

**IGOR BAJŠANSKI**, Department of Psychology, University of Rijeka  
sibajsan@ffri.uniri.hr

**GABRIELA BAŠIĆ HANŽEK**, Department of Philosophy, University of Split  
gbgabriellabasic@gmail.com

**HANOCH BEN-YAMI**, Department of Philosophy, Central European  
University, Vienna, benyamih@ceu.edu

**BORAN BERČIĆ**, Department of Philosophy, University of Rijeka  
boran.bercic@ri.t-com.hr

**ALEKSANDRA GOLUBOVIĆ**, Department of Philosophy, University of  
Rijeka, aleksandra.golubovic@ffri.uniri.hr

**MARKO JURJAKO**, Department of Philosophy, University of Rijeka  
mjurjako@ffri.uniri.hr

**JELENA KOPAJTIC**, Department of Philosophy, University of Rijeka  
jelena.kopa@hotmail.com

**PAOLO LABINAZ**, Department of Humanities, University of Trieste  
plabinaz@units.it

**NENAD MIŠČEVIĆ**, Department of Philosophy, University of Maribor  
vismiscevic@ceu.edu

**EDI PAVLOVIĆ**, Faculty of Philosophy, Ludwig-Maximilians-Universität  
München, Edi.Pavlovic@lrz.uni-muenchen.de

**INES SKELAC**, Faculty of Philosophy and Religious Sciences, University of  
Zagreb, ines.skelac@gmail.com

**MATEJ SUŠNIK**, Institute for Philosophy Zagreb  
matej@ifzg.hr

---

**DANILO ŠUSTER**, Department of Philosophy, University of Maribor  
danilo.suster@guest.arnes.si

**MAJDA TROBOK**, Department of Philosophy, University of Rijeka  
trobok@ffri.uniri.hr

**ANDREJ ULE**, Department of Philosophy, University of Ljubljana  
andrej.ule@guest.arnes.si

**LINO VELJAK**, Department of Philosophy, University of Zagreb  
lveljak@ffzg.hr

**MICHAEL WATKINS**, Department of Philosophy, University of Auburn  
watkigm@auburn.edu

**TIMOTHY WILLIAMSON**, Faculty of Philosophy, University of Oxford  
timothy.williamson@philosophy.ox.ac.uk



**BORAN BERČIĆ**  
**ALEKSANDRA GOLUBOVIĆ**  
**MAJDA TROBOK**

## Editors' Preface

### About Nenad

This book is dedicated to professor emeritus Nenad Smokrović, our dear friend and colleague from the Rijeka Philosophy Department.

Smokrović graduated Philosophy and Sociology at the Faculty of Philosophy, University of Zagreb. At the same institution, he defended his Master's thesis on Habermas under the supervision of prof. Hotimir Burger. In 1996 he defended his PhD thesis *Logic and Natural Reasoning*. The thesis was defended at the University of Ljubljana, under the supervision of prof. Frane Jerman. Members of the committee were prof. Andrej Ule from Ljubljana and prof. Nenad Miščević from Zadar.

Nenad wrote two books: *Priroda prirodnog zaključivanja (The Nature of Natural Reasoning)*, Hrvatsko filozofsko društvo, Zagreb, 2004 and *Znanost i metoda (Science and Method)*, KruZak, Zagreb, 2017. Together with Nenad Miščević, he edited two anthologies of texts influential in contemporary philosophy. One on the Action Theory, and another on the Philosophy of Mind: *Namjera i čin (Intention and Action)*, Izdavački centar Rijeka, 1983, and *Kompjutori, mozak i ljudski um (Computers, Brain, and Human Mind)*, Izdavački centar Rijeka, 1989. The second anthology was reprinted in 2001 with significant changes in the content. Smokrović gave significant contribution to *Filozofski leksikon (Lexicon of Philosophy)* of the Leksikografski zavod Miroslav Krleža, Zagreb, 2012. He was contributor and co-editor of this large and important volume. He published over 40 articles or book chapters in national and international journals and publications. He presented his work at over 50 national and international conferences. He organized many conferences, on a dozen of occasions it was annual conference *Contemporary Philosophical Issues* at the Department of Philosophy in Rijeka; he organized many smaller but very good conferences of his *Centre for Logic and Decision Theory*; he regularly participates at the *Mind, World, and Action* Conference of the IUC Dubrovnik; etc. His intellectual integrity has always been impeccable.

Smokrović has taught at the University of Rijeka since 1980. At the *Faculty of Economy* from 1980 until his retirement in 2020, and at the *Faculty of Philosophy* from 1998 until 2020. With other colleagues (Baccarini,

Berčić, Prijjić-Samaržija), Smokrović participated in the epic battle for the establishment of the Department of Philosophy at the University of Rijeka. (Although all the requirements for the establishment of the Department were satisfied, authorities did not allow the Department to begin with its activities.) After two years of struggle, Department normally continued with its activities.

At the Department of Philosophy, Smokrović has taught *Philosophy of Mind, Logic, 20<sup>th</sup> Century Philosophy*, and several elective courses. At the Faculty of Economics he has taught *Decision Theory, Game Theory* and a number of other courses.

Smokrović was Head of the Philosophy Department in two mandates, from 2001 to 2003, and from 2003 to 2005. Since 2004 he has a Chair for the History of Philosophy. At the Department of Philosophy, in 2008 he founded the PhD program *Philosophy and Contemporaneity*. He was its president until 2019. He successfully ran the program and navigated through all the administrative and organizational obstacles, without ever neglecting his teacher's and mentor's duties. This can be seen from the fact that he was PhD mentor or co-mentor to four contributors in this volume (Sušnik, Jurjako, Skelac, Bašić-Hanžek). He also successfully mentored or co-mentored dissertations of several younger scholars at the Faculty of Economy (Debelić, Mance, Vretenar) who worked in the Decision Theory and Game Theory. Nenad has been tutoring students enthusiastically and way beyond what have been his strict professional duties as a mentor. As a result, he has managed to create a group of young enthusiasts who nowadays work on logic, rationality, and other related subjects. He got the honour of professor emeritus in 2020. Currently he is teaching and mentoring at the PhD program *Philosophy and Contemporaneity* at the Department of Philosophy, University of Rijeka.

In 2019 he founded *Center for Logic and Decision Theory* at the University of Rijeka, being very active in organising conferences, round tables, courses and seminars for both students and researchers.

During his long academic career he ran several research projects of the *Ministry of Science Republic of Croatia*, of the *University of Rijeka*, and of *Croatian Science Foundation*. Just to mention the latest one, he was the principal researcher of the Croatian Science Foundation's international project *Rationality: Between Logically Ideal and Commonsensical in Everyday Reasoning*.

Here we have to say that Nenad Smokrović was a key member of an informal group of philosophers that was active in Rijeka in the 80's. Activities and members of this group were the backbone of the Rijeka Department of Philosophy that was formed almost 20 years later. Sometimes we called

it *Kružok*. Besides Nenad Mišćević, Smokrović was the *spiritus movens* of the group. Many lectures and discussions were held in his apartment above the main market, where he has lived all his life. Although the group was informal, it was very ambitious, professional and up to date. Many distinguished philosophers held lectures at these informal meetings (Donald Davidson, Barry Stroud, Roderick Chisholm, ...).

In 80's and 90's Smokrović was active in editing journal *Dometi*. From 1995 to 2000 he was the main editor of the journal *Agora*. Currently he is a member of the Editorial Board of *Croatian Journal of Philosophy*. He is member of several societies for philosophy and logic. Currently he is the President of the *Croatian Society for Analytic Philosophy*.

In 1991/92 Nenad Smokrović fought in the Croatian War of Independence. He was a soldier in the 128<sup>th</sup> Brigade of the Croatian Army.

Smokrović took part in the humanitarian activities. He loaded tons and tons of humanitarian aid for the refugees from Croatia, and for the refugees from Bosnia and Herzegovina. He helped people a lot, in different ways, whenever he could. On innumerable occasions he helped people moving by carrying furniture, firewood, etc.

Nenad Smokrović is a passionate runner. For years he has been running half-marathons and other long distance races. Among other races, he ran *Plitvički maraton* in its full length, 42.195 km. He is also a regular yoga practitioner.

So far, he has owned and loved many cats and one great dog.

His nickname is *Pišta*. We all call him like that. However, there is no consensus about how he got it. Although there are several adventurous stories about it, he denies them all and claims that it is just a matter of a simple fact that as a student he travelled to Hungary.

Nenad Smokrović is also a biker. He rides motorbikes all his life.

For as long as we have know him, he has been married to Tanja, who has been his stronghold and support for all these years.

## About this volume

During his long and rich academic career, Nenad Smokrović has established contacts with many people, in different ways. Contributors in this *Festschrift* reflect that fact. Some people are his friends and colleagues since his student days (Mišćević, Veljak). Some are known for a long time, from conferences and activities in former Yugoslavia (Ule, Arsenijević, Šuster). Some belong to the informal philosophy group that was active even before the establishment of the Philosophy Department in Rijeka (Berčić, Trobok, Bajšanski). Some worked together with him at the Department (Golubović). Some are international contacts established through joint

work in Croatia or wider region (Williamson, Watkins, Ben-Yami, Labinaz). As we already said, some are his graduate students that he mentored or co-mentored (Sušnik, Jurjako, Skelac, Bašić-Hanžek). Some are our former students that are active abroad (Pavlović). Some are still our PhD students (Kopajtić).

In this volume, articles are listed in the alphabetical order. This is not a very creative way of organizing a volume, but we opted for it because there is no obvious way of grouping articles that we received. After all, the work of Nenad Smokrović is a thin red line that keeps all these articles together.

Some articles are in English, some are in Croatian. In initial contacts with possible contributors, we found out that many people already have relevant material in Croatian. For this reason we left the decision about the language up to the authors.

Since human rationality has been the central area of Smokrović's professional interest for over 25 years, all the papers in this collection are either directly or at least loosely related to this topic. For this reason the title of the volume is *Human Rationality*. What human rationality consists in? What is the relationship between its normative aspect (logic) and its descriptive aspect (psychology)? Do people *in fact* reason according to laws of logic? What is the nature and function of reasoning, argumentation, and inference? ... We are publishing this volume in the hope that it will provide further contribution to the research in the field.

In this collection, 19 authors from Croatia, Slovenia, Serbia, Italy, Israel, USA, and GB write about the subject:

**Miloš Arsenijević** (in English) critically analyses Smokrović's thesis concerning normativity of natural language reasoning, more precisely his view that there is an incongruence between descriptive nature of formal logic on one side and the normativity of reasoning in not just everyday argumentation, but also scientific, philosophical and even mathematical dialogues on the other.

**Igor Bajšanski** (in Croatian) concentrates on Smokrović's (2015) paper in which he develops the idea that the argumentative process represents a means for extending knowledge. Bajšanski underlines the importance of inferences that can function, equally well, outside the context of argumentation processes too.

**Gabriela Bašić Hanžek** (in Croatian) in her paper focuses on the differences between the negative argumentative theories (the pragma-dialectic theory of fallacies) and the epistemological approach to argumentation. Bašić Hanžek endorses the view according to which the differences between the two approaches are more evident in the case of positive theories.

**Hanoch Ben-Yami** and **Edi Pavlović** (in English) develop further the Quantified Argument Calculus (Quarc, the formal logic system first developed by Hanoch Ben-Yami in (2014)) by providing a philosophical motivation for the truth-valuational, substitutional approach, by replying to some of the common objections and by resolving the formal gaps left in the original paper.

**Boran Berčić** (in English) endorses the view that statements having the schema: "X is the best, but I prefer Y" contain no contradiction, nor are self-refuting. In order to show this to be the case, Berčić introduces the distinction between values from preferences and rejects the idea that values should be inferred from preferences. The inserted discussion about context relativity and sortal relativity of preferences has the aim to additionally confirm Berčić's initial thesis.

**Aleksandra Golubović** and **Jelena Kopajtić** (in Croatian) analyse the idea of (im)possibility of teaching critical thinking within school and educational systems that are (at least implicitly) impregnated with a specific worldview.

**Marko Jurjako** (in Croatian) investigates the possible implications of connecting naturalism in philosophy with the query whether moral demands give reasons for acting to all rational agents. Jurjako defends the view that both possible options - the one that denies the importance of naturalism in this context and the other according to which naturalism implies relativism for practical reasons - are acceptable to a certain extent.

**Paolo Labinaz** (in English) critically analyses Smokrović's view that reasoning is argumentative in nature and proposes an alternative way to elaborate such a view. The idea is important since it is one of the tenets in Smokrović's theory of argumentation, in which argumentation is taken to be a curiosity motivated and cooperative activity.

**Nenad Miščević** (in English) addresses, in a dialogue with Smokrović, several questions concerning rationality: How rational are human beings? Is there rational thinking in an ordinary, positive way? Is there epistemically virtuous thinking? The author defends the view that these questions go together and their answers should be either jointly positive or jointly negative. He also proposes to distinguish between five degrees or stages of rationality of reasoning.

**Ines Skelac** (in Croatian) deals with discussions about the normativity of logic in everyday reasoning. The author critically analyses different approaches to the normativity of logic and proposes remarks that could or should be addressed as well as several possible problems that are in need of a solution. In particular, the author concentrates on the bridge principle.

**Matej Sušnik** (in Croatian) elaborates the relationship between theoretical and practical reasoning. In particular, Sušnik analyses the elements of theoretical reasoning that Smokrović focuses on, with the aim to establish to which extend (some of) these elements are applicable in the domain of practical reasoning.

**Danilo Šuster** (in English) discusses Smokrović's views on the normativity of logic and elaborates some doubts concerning Smokrović's notion of deductive logic as well as his proposal to model such reasoning in non-monotonic logic. In addition, Šuster presents a looser notion of normativity.

**Majda Trobok** (in English) is particularly interested in the notion of argumentation and argumentative processes and discusses the approach that Smokrović takes with respect to the properties, the role and the importance of argumentation.

**Andrej Ule** (in English) continues the conversation with Smokrović on the notion of argumentation, particularly on the relationship between social and individual-psychological aspects of argumentation. Ule explains the idea that the argumentative process is a continuous process of extending and externalisation of the objective spirit of the reasoning situation.

**Lino Veljak** (in Croatian) analyzes the concept of reductionism, that is grounded in the idea of ontological monism and the unity of science. The consequence of this view is methodological monism - the idea that all the sciences ultimately share the same methodology. Author argues in favor of methodological pluralism, and defends the view that scientific methodology should be freed from the absolutisation of the quantitative.

**Michael Watkins** (in English) discusses Smokrović view that understanding or grasping a concept is constitutively tied to being disposed to assent to certain sentences (the constitutive account), in particular to sentences that we might reasonably take to be (analytically) true. In particular, the author analyses Smokrović's response to Williamson with regard to the constitutive account against purported counterexamples presented by Williamson.

**Timothy Williamson** (in English) discusses different levels at which rationality assumptions, for instance of 'logical omniscience', seem to be built into standard models of epistemic and doxastic logic. Furthermore, the author examines the consequences of the debate between intensionalist and hyperintensionalist theories of attitude ascription, and makes comparisons with corresponding issues about models of probability.

At the end of the volume, **Nenad Smokrović** makes comments on these texts, replies to the criticism and objections, and deals with the remarks and comments.

Nenad deserved this *Festschrift*, as an academic, as a colleague, as a friend. Let us quote Timothy Williamson here:

Nenad's professionalism and well-balanced reasonableness have continued to impress me, not to mention his excessive modesty. He has made a very important contribution to the development of analytic philosophy in Croatia and in the wider region. It is a pleasure to contribute to his much-deserved *festschrift*. (p. 273, this volume)

## Acknowledgements

We owe gratitude for the financial support of the following institutions: Ministry of Science and Education of Republic of Croatia, Faculty of Humanities and Social Sciences in Rijeka, University of Rijeka through the research projects *Metametaphysics* Uni-ri-human-18-239, and *Critical thinking and Society* Uniri-human-18-254.

We apologize for the very slow pace in publishing this collection. Due to the number of people involved, quantity of the work, teaching obligations at the Department, etc. it was hardly possible to finish this book in shorter period of time.

We would like to express gratitude to all the friends and colleagues who made publication of this volume possible. Authors who wrote their texts for this occasion. Reviewers who carefully went through these texts and whose comments led to significant improvement. (Dušan Dožudić, Martina Fuerst, Filip Grgić, Ljudevit Hanžek, Andrej Jandrić, Tomislav Janović, Hrvoje Jurić, Guido Melchior, Dragana Sekulić) Proofreaders who removed many typing and other errors (Jelena Kopajtić and Leonard Pektor). Our former student and painter Luka Aničić who gave us permission to use his acrylic on cardboard *The Orange End of Nature* as the cover of the book. People from *Grafika Helvetica* who had enough patience to make all the changes and corrections that we asked for.

In Rijeka, August 2022





MILOŠ ARSENIJEVIĆ

## The Future Sea Battle and Performing an Infinite Task: Two Remarkable Cases Concerning the Logician Thesis

**Abstract:** In his 2018 article which concerns normativity of natural language reasoning, Smokrović deals with the alleged incongruence between descriptive nature of formal logic and normativity of reasoning in everyday argumentation, juridical debates or scientific, philosophical and even mathematical dialogues. Contrary to the radical stance, according to which there is no possible normative use in the reasoning which is sensitive to propositional content that should be interpreted in connection to the real world, Smokrović supports what he calls logicist thesis, according to which in any case in which there is a discrepancy between standard logic and some argumentation that we consider correct, we ought to explain why it is so and find an appropriate form, i.e. a logic more or less close or remote to standard logic, through which the correctness of the given argumentation could be vindicated. In this article I analyze, following the requirements of the logicist thesis, two remarkable and intriguing philosophical debates, which concerns 1) the future contingencies and the problem of logical determinism, and 2) the impossibility of performing infinite tasks.

**Key words:** logicist thesis, Aristotle, future contingencies, logical determinism, the principle of bivalence, temporal-modal logic, infinite tasks, Grünbaum, anti-infinitism, the impossibility of remote possible worlds.

In his 2018 article “Informal Reasoning and Formal Logic: Normativity of Natural Language Reasoning” Nenad Smokrović deals with the recently raised question concerning the *relation* and *alleged incongruence* between *descriptive nature* of formal logic, which supposedly concerns purely syntactic relations between propositions in view of the preservation of truth within classical propositional and predicate calculi, and *normativity of reasoning* in everyday argumentations as well as in juridical debates or scientific, philosophical and even mathematical dialogues (p. 457).

The debate started with Harman’s (1986) radical view, according to which there is no possible normative use of formal logic in everyday

reasoning, which is sensitive to *propositional content* that should be interpreted in connection to the *real world*. Even in the most simple and obvious cases, such as the following one cited by Smokrović, the necessity of truth preservation is allegedly not secured, in spite of the fact that the argument is put in one of the classically valid forms.

- (p) (According to the time-table) *The 8 a.m. bus from Rijeka to Zagreb starts either from platform 1 or from platform 2;*
- (q) (Actually) *The bus does not start from platform 2;*
- (C) (Therefore) *The bus starts from platform 1.*

What if, for any reason whatsoever, platform 1 is not available at 8 a.m.? Then, the bus will not start from platform 1 either. So, in such a case, though both premises are true, and though we would hardly say that the reasoning was incorrect, the conclusion is false. The point is that even if in many (or even all) previous cases the conclusion was true, it was not *necessarily* true, while according to formal logic it *should be* so. This should be enough for claiming that there is a kind of discrepancy between validity of formal logic and similar forms of reasoning in everyday argumentation.

Reacting to Harman's radical stance, MacFarlane and Hartry Field (cf. Smokrović, *ibid.*, pp. 460-463) suggested different variants of the so-called *bridge principle* between formal logic and everyday reasoning that should give normative force to classically valid logical forms, thus enabling their application in everyday reasoning. Without going into formal details and differences between the ways in which the bridge principles are formulated, let us explain informally how, according to Varga, Stenning and Martigton (2015), appropriate deontic reading of the logical scheme applied in the above example enables us to say that the reasoning in passing from the first two premises, *p* and *q*, to the conclusion *C*, was correct even if *C* turns out false.

Let us suppose that, by looking at the time-table, we have seen by our own eyes that the 8 a.m. bus to Zagreb starts from platform 1 or platform 2. Let us suppose, in addition, that, by passing by platform 2, we have found that, today, the bus on platform 2 is not the bus going to Zagreb but to Pula, and that the bus driver himself has confirmed that, in order to catch the bus to Zagreb, we ought to go to platform 1. There has also been no information that anything has been changed in relation to what stands in the time-table. And then, only much later, we learn that, today, there is no bus to Zagreb. What has happened? Something *abnormal*! So, on one hand, our reasoning can be said to be *correct* just because *C* has turned out false only due to something *abnormal*, while, on the other hand, what has happened *has happened regardless* of whether it is normal or abnormal. The

fact that something abnormal has happened does not make our reasoning incorrect in spite of the fact that  $C$  is false.

In other cases, the explanation of the seeming discrepancy between the validity of a given logical form used in the argumentation and the possible incorrectness of the argumentation may happen to be explainable in some other ways, but anyway, it should be explainable in *some* way. On the basis of this, Smokrović argues in favour of what he calls *logicist thesis*, according to which in *any* case there should be *some* logic, more or less close or remote to standard logic, through which the correctness of argumentation could and should be vindicated.

Varga, Stenning and Martington have argued that in the given example the logic that can resolve the problem is the *default logic*:  $p \wedge q \rightarrow C$  can be used for saying that we correctly pass from the justified belief that  $p$  and  $q$  are true to the belief that  $C$  is true, if  $C$  is true *by default*, i.e., false *only if* something *abnormal* has happened.

One might be tempted to think that the *default logic* could be incorporated in standard logic if the form of the above reasoning, as Varga, Steanning and Martington suggest, were schematized in the following way (see Smokrović 2018, p. 467):

$$p \wedge q \wedge \neg \mathbf{ab} \rightarrow C$$

where  $\neg \mathbf{ab}$  means that *nothing abnormal is the case*. But it is not so. In the first place,  $\mathbf{ab}$  does not stand for any standard proposition as  $p$  and  $q$  do, but is rather a schematic formulation standing for an indefinite number of propositions. There is no way to extensionalize  $\neg \mathbf{ab}$ . Even if we could add an infinite disjunction to  $p \wedge q$  instead of  $\neg \mathbf{ab}$ —which is not possible to do in standard first order propositional calculus—we would not get that the addition of anyone of the disjuncts alone can make  $p \wedge q \vdash C$  wrong, since, for any  $A$ , if  $p \wedge q \vdash C$  is valid, so is *also*  $p \wedge q, A \vdash C$ . In order to get  $p \wedge q \vdash C$  valid and  $p \wedge q, A \vdash C$  non-valid, we need some of Anderson and Belnap's *systems of the logic of relevance*, where  $A, B \vdash C$  may be non-valid in spite of the fact that  $A \vdash C$  supposedly is (Anderson and Belnap 1975).

After all, I accept Smokrović's *logicist thesis*, according to which in any case in which there is a discrepancy between standard logic and everyday argumentation which we supposedly consider correct, we ought to explain why it is so and find an appropriate *logical form* in accordance with which the given argumentation will be correct. Now, as it is said in the introduction, Smokrović believes that the *logicist thesis* should hold not only in everyday and juridical debates, but also in scientific, philosophical and even mathematical dialogues. In view of this, in what follows, I will analyze, in brief, some intriguing philosophical debates I have been dealing

with for a rather long period of time, in order to find out whether, in order to settle the dispute, we need to depart from standard logic (or mathematics), and if it turns out that we have to, what are the ways in which we could or should do it.

### Example 1: Aristotle's Future Sea Battle

Curiously enough, the first philosopher to be confronted with the problem concerning the *logician thesis* was the father of traditional logic, Aristotle himself (*De interpretatione* 19 a 23).

If an utterance of the sentence “There will be a sea battle tomorrow” is true, then it seems that it is *determined* that there will be a sea battle tomorrow. For otherwise, how could the utterance be true? If, however, an utterance of the sentence “There will be a sea battle tomorrow” is false, then it seems that it is *determined* that there will be no sea battle tomorrow. For otherwise, how could the utterance be false? Thus, it seems that it is *determined* whether there will be a sea battle tomorrow or not – and so for any other future event whatsoever. This, however, is in conflict with the plausible assumption that at least for some possible future events, it is not predetermined whether they will take place or not.

The kind of determinism implied by the fact that whether something will happen at some time or not is necessarily predetermined by the truth of the corresponding propositions—that it will happen and that it will not happen respectively—is called by Schlick (1931, p. 202) *logical determinism*. Logical determinism is stronger than any other sort of determinism, and as such threatens any interpretation of quantum mechanics that implies indeterminism. But independently of this, it is incompatible with everyday reasoning, according to which what will happen supposedly depends on our deliberations and decisions connected with various coincidences in the real world.

The purpose of this example is not to raise the debate about determinism and indeterminism. The point is only to show that there seems to be an obvious discrepancy between everyday reasoning about future contingencies and obeying some of the basic logical principles. So, according to the *logician thesis*, it seems that, if we want to preserve the validity of everyday reasoning, we have to abandon, change or restrict some of the logical principles. Which one?

It is clear that for Aristotle it cannot be the *principle of contradiction*, for any contradiction (ἀντίφασις) whose form is  $p \wedge \neg p$  is always necessarily *false*, whatever proposition  $p$  stands for. Whatever some interpreters may think, for Aristotle it is also not the *principle of excluded middle*, for he put it explicitly that  $p \vee \neg p$  is always *true*, because the very form in which the

affirmation and the opposite negation stand here secures the truth, and so also even if neither of the two (οὐ...τόδε ἢ τόδε) is already (ἤδη) either true or false. But then, given that it is necessary that  $p \vee \neg p$  will be true tomorrow, it could and should be said to be true today as well. However, neither  $p$  nor  $\neg p$  themselves may happen to be true or false at some earlier time, independently of the fact that  $p \vee \neg p$  is true. So, what has been much later called by Łukasiewicz the *principle of bivalence* (Łukasiewicz 1922, p. 126) is that whose universal validity was challenged and after that restricted by Aristotle with the use of the *Sea Battle* example.

Aristotle's reaction is deeply revisionist in view of the logic that he has established. Namely, it prevents the very formalization of propositional calculus in which, by using the standard definitions of  $\wedge$ ,  $\vee$  and  $\neg$ , the *principle of bivalence* is derivable from the *principle of contradiction* and the *principle of excluded middle*. Namely, if it holds, for *any* proposition, that the conjunction of it and its negation is always false, while the disjunction of it and its negation is always true, then it follows that *every* proposition must have one and only one of the two truth values—*truth* or *falsity*—which is exactly what the *principle of bivalence* claims. So, under Aristotle's revision of his own logic, being either true or false ceases to be a necessary condition for being a proposition and the revised logic contains *truth-value gaps*.

It was only Łukasiewicz who recognized clearly that the *Sea Battle* was directed against one of the *basic* and mutually *independent* logical principles, which he called, in his famous *Rector's Speech* 1922, the *principle of bivalence*. But he reacted in a different way to the problem of the *logician's choice* between alternatives that the *Sea Battle* had imposed. Instead of restricting the *principle of bivalence*, he constructed the *three-valued logic system* (Łukasiewicz 1918, 1920), which, instead of truth-value gaps, contains, in addition to *truth* and *falsity*, the third value: *indeterminacy*. While *true* propositions are about something that *is* and *false* propositions about something that *is not*, the propositions with the third truth value are about something that does not have a real correlate but which is yet *possible*.

Łukasiewicz's three-valued and many-valued logical systems represent a nice piece of formal and philosophical logic, and one may be tempted to think that in order to make the everyday reasoning concerning future contingences correct, there is hardly a better way to formalize it. But it should be noticed that in Łukasiewicz's formalization *time* as such does not play any role, while in the *sea battle* challenge we are dealing with not just a *possible* sea battle, but with a *future* sea battle. So, there are those, to whom I myself belong, who think that we can still save the *principle of bivalence* when trying to make room for everyday reasoning about future contingences *if* we formulate a logical system in which both *time* and

*modalities* are simultaneously taken in account. This is done in the system of *temporal-modal* logic of events TM, whose semantics is given in my 2016 article, whereas the complete axiomatization of it is formulated in a not yet published manuscript, presented at my visiting talk at the University of Siegen on 19<sup>th</sup> Dec. 2019 (Arsenijević and Jandrić, forthcoming, 2023). I will summarize what the solution to the *sea battle* problem looks like according to TM.

TM is formalized within the interval-based system of the time continuum, where  $t_1, t_2, \dots, t_n, \dots$  are constants standing for particular time intervals and  $t_1, t_2, \dots, t_n, \dots$  variables ranging over the set of all time intervals. Now, the elementary well-formed-formulae of TM is any  $E(t_n)$ , where  $t_n$  is replaceable by any time-constant or time-variable and E stands for any elementary event (those which happen uninterruptedly)  $e_1, e_2, \dots, e_n, \dots$ , as well as any formulae preceded by a quantifier  $\forall$  or a quantifier  $\exists$ , or by temporal operators  $\{t_1\}, \{t_2\}, \dots, \{t_i\} \dots$  or  $\{t_1\}, \{t_2\}, \dots, \{t_i\}, \dots$ , or modal operators  $\diamond$  or  $\square$ . Temporal and modal operators can be iterated and combined.

Now, the endless interval-based time continuum in any TM model contains two abutting parts, one real, meaning that on any of its intervals something has happened, and the other, imaginary, on whose intervals nothing has yet happened. The intervals of the first one are called full or actualized, the intervals of the second one empty or non-actualized. For temporal operator  $\{t_j\}$  and any formula A,  $\{t_j\}A$  is true if and only if  $t_j$  is actualized and A true at it, which is the case if A is a logical truth or any of the factual truths about what happened or failed to happen on  $t_j$ , as well as any of all the truths about what happened on intervals that precede it (which are, hence, actualized themselves). But it is important to notice that it can be not only any of the truths about what happened on any of the actual intervals that happened earlier but also about any of the actual intervals that ended later, which seems to lead us back to logical determinism. Yet, I will shortly explain why it is not the case.

In the general semantics of modal logic, formula  $\square A$  is said to be true if and only if A is true in all accessible possible worlds, and  $\diamond A$  is taken to be true if and only if there is an accessible possible world in which A is true. Now, in the standard possible world semantics, the truth of a formula prefixed by a modal operator is assessed from a single world, and, therefore, it is not necessary to point to the world from which the accessible worlds are accessible. However, in the system TM, the real world always contains *an infinite number of actual worlds* (because the real world consists of an infinite number of actualized worlds), so that some possible worlds are accessible from some actual worlds but not from others. If, for instance, an event e happened on an interval  $t_n$ , then on an earlier interval  $t_m$  it was

possible for  $e$  not to occur on  $t_n$ , while on  $t_n$  itself this possibility is precluded. Thus, there is a merely possible world, in which  $e$  does not happen on  $t_n$ , which is accessible from the world actualized on  $t_m$  but not from the world actualized on  $t_n$ .

So, the formulae with a modal operator outside the scope of a temporal one lack a determinate truth value, as in such cases it is not specified which actual world's set of accessible possible worlds is to be taken into account. In TM we can meaningfully talk about possibilities only by bearing in mind what has, up to a certain time, already been actualized. So, the status of the formulae such as, for example,  $\Box e(t_n)$  and  $\Diamond e(t_n)$ , should be understood by the analogy to the well-formed but open formulae in predicate logic, which become definitely true or false only after some further qualification. Formulae with iterated modalities can, accordingly, be true or false only if the sequence of modal operators is, as a whole, subjected to a temporal operator.

Hence, we can also speak of merely possible worlds being accessible from other merely possible worlds but only provided that the first merely possible world in the chain is accessible from some actual world. In other words, the talk of possible possibilities, possible necessities, etc., has to be *anchored* in the real world.

Let us now, in view of what has been previously said, turn to the *sea battle* problem. Let us take that  $t_m$  precedes  $t_n$ , but so that  $t_m$  is actual and  $t_n$  non-actual. This is the case when  $t_m$  refers to *today* and  $t_n$  to *tomorrow*. Let  $e(t_n)$  denote *the sea battle that happens tomorrow*. Then, according to the above definition of the truth of  $\{t_i\}A$  for any  $A$ ,  $\{t_m\}e(t_n)$  is *false* and  $\{t_m\}\neg e(t_n)$  *true*, since on  $t_m$  there is no actualized world in which  $e(t_n)$  is true. But now, the point is that although  $\{t_m\}\neg e(t_n)$  is true,  $\{t_m\}\{t_n\}\neg e(t_n)$  is *false*, since for  $\{t_n\}\neg e(t_n)$  to be true it is necessary that  $t_n$  is actual and  $\neg e(t_n)$  true on it, which, not being the case, renders  $\{t_m\}\{t_n\}\neg e(t_n)$  *false*. In TM, the prefixing of a temporal operator is generally not a trivial matter, since it may affect the truth value of the ensuing complex formula. This is exactly the fact through which *logical determinism* is avoided, for the fact that *there will be the sea battle tomorrow* means that it is true today that it will be true tomorrow that the sea battle happens that day, which is expressed by  $\{t_m\}\{t_n\}e(t_n)$ , which is *false*.

Let us now remember the seemingly threatening case in which  $\{t_m\}e(t_n)$  may be true even if  $t_n$  is later than  $t_m$ . This can be so *only if* some world is already actualized on  $t_n$ . For only then, either  $\{t_n\}e(t_n)$  or  $\{t_n\}\neg e(t_n)$  is true. So, what matters here is, in the first place, whether  $t_n$  is actual or not, and then, if it is, whether  $e(t_n)$  is true on it or not.



Completely in accordance with ordinary language, it is not only false today that the sea battle will happen tomorrow— $\{t_m\}\{t_n\}e(t_n)$ —but also that it will not happen— $\{t_m\}\{t_n\}\neg e(t_n)$ . Both being false,  $\{t_m\}\{t_n\}e(t_n)$  and  $\{t_m\}\{t_n\}\neg e(t_n)$  are not contradictory but only contrary, and that is why both  $\{t_m\}\diamond e(t_n)$  and  $\{t_m\}\diamond\neg e(t_n)$  may be true. Today, it is both possible that the sea battle happens tomorrow as well as that it does not happen. But this does not mean that at  $t_m$  it is true that on  $t_n$  it will be both possible that the sea battle happens and that it does not happen, since  $\{t_m\}\{t_n\}(\diamond e(t_n) \wedge \diamond\neg e(t_n))$  is *false*. However, though today it is only possible that the sea battle will happen tomorrow (as well as that it will not happen), if the sea battle really happens tomorrow, it will be true tomorrow that it was true the day before that it would be true the day after that the sea battle had happened that day. Similarly, if the sea battle does not happen tomorrow, it will be true tomorrow that it was true the day before that it would be true the day after that the sea battle had not happened that day.

Given that TM models are supposedly distributed along one and the same time continuum, the model in which *today* refers to  $t_n$  contains, as a part, the model in which it refers to  $t_m$  but is not a mere extension of it in view of just the factual truths, for there is an infinite number of ways in which it was possible that the real world history could have developed from the state in which *today* referred to  $t_m$  and the state in which it refers to  $t_n$ . All the possibilities are preserved as *modal truths* about what *could have been the case*. So, history is much richer than the set of factual truths. The system TM enables us not only to speak about what *is* and what *can be* the case but also about what *could have been* the case. All this also explains why in ordinary language there is an asymmetry between prediction and retrodiction.

The reason why we cannot know the truth about the future sea battle is not a matter of *epistemological* fact. Namely, in the given case, the impossibility of knowledge is completely based on the *logico-ontological* fact that *there is nothing to be known*, since there are different possible ways that lead to this or that outcome. We may *guess* that it will be so-and-so, but to guess is not to know. Everything depends on deliberations, decisions and coincidences of events that are not yet actual. However, once the sea battle really happens or really fails to happen, there is just *one single path* the history has paved to this, and we can explain (in principle at least) *how* it has come into being by taking into account *actual* deliberations, decisions, coincidences, etc.

If there is just one real world history in view of any given apex as the boundary between the real and the imaginary part of the time continuum, there must be *one privileged model* from the equivalence class of isomor-



phic models that represents the real world in view of a given apex. Though it could have been otherwise, what is, necessarily is. The distribution of elementary events over the *real* part of the time continuum is *necessarily such as it is*. So, for any given instant, there must be a *unique* real world history that ends at it. Then, though any instant is represented through an equivalence class of models regardless of the history that has paved the way to it, there is always a privileged world line that represents *the* history of *the* real world up to the given instant.

The *factual* truths about events in a model which represents only a part of the history are *preserved* in the privileged model that describes the whole real world history up to the given instant. Then, the *development* of the real world history can be viewed as a continuous transition from one privileged model to others such that each of them represents the real world history up to a certain instant *as if* it ended at that instant. *The so-called flow of time* is nothing else but such a continuous transition from one privileged model to others as a consequence of the development of the real world history, where each of previous models represents an earlier development of the real world history up to a certain instant.

The model in which  $t_n$  is actual contains not only factual truths of previous models but also truths about all the possibilities in previous models. If it was possible yesterday that it would rain today,  $\{t_m\} \diamond e(t_n)$ , it remains true today that it was possible yesterday that it would rain today even if  $\{t_n\} \neg e(t_n)$  has become true.

All these facts are in accordance with intuition and ordinary language. In view of the logicist thesis that we are investigating, we can conclude that the crucial questions concerning the *sea battle* can be dealt with within standard classical predicate logic after the introduction of temporal and modal operators and the suitable choice of the system of axioms. The fact mentioned above, that in TM formulae with iterated modalities can have a determinate truth value only if the sequence of modal operators is, as a whole, subjected to a temporal operator is in accordance with everyday language, where we speak of modalities from within the real or imaginary segments of one and the same time continuum. As Nuel Belnap puts it, “If a certain possibility is real, [...], it must be part and parcel of *Our World*” (Belnap 2007, p. 87, n. 2), so that “the brilliantly conceived doctrine of Lewis 1986 (and elsewhere) ought to be rejected”.

## Example 2: Impossibility of performing infinite tasks

In *Physics* 233 a 22 and 263 a 8, Aristotle mentions a variant of Zeno’s second kinematic paradox, constructed by an unknown author, where the runner, in order to reach the goal, has to count distances that become

smaller and smaller according to the geometric progression  $\frac{1}{2}$ ,  $\frac{1}{4}$ , ... As he takes it for granted that it is not possible to finish counting to infinity, Aristotle takes the given example as an argument that a continuum, be it spatial or temporal, does *not consist of actual* parts that could be counted. But what if, independently of the question concerning the structure of the continuum, the spatial and/or temporal parts individuated by a geometric progression are actually physically distinguished, so that, in order to reach the goal, the runner de facto has to perform the task consisting of an infinite number of steps?

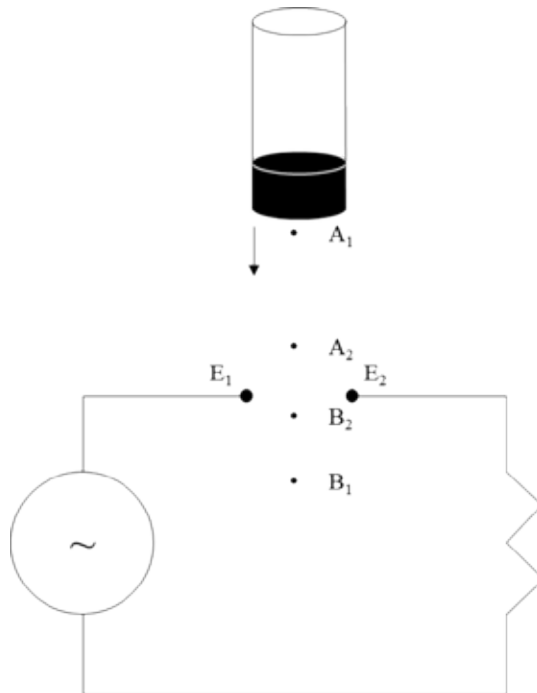
In my *Analysis* paper (Arsenijević 1989) I dealt with the question of whether a limited space can contain an infinite number of physically distinguished parts. Here, we shall deal with the temporal variant of the problem, where the steps of an infinite task are sufficiently well distinguished by the very acts of their performance. At the middle of the last century, this question was discussed by a considerable number of well-known philosophers and mathematicians, who can be divided, independently of additional differences, into two groups: infinitists (Taylor 1951, Watling 1952, Maxwell and Feigl 1961, Grünbaum 1968, 1969, Salmon 1975) and anti-infinitists (Hilbert 1926, Weyl 1949, Black 1951, Wisdom 1952, Schwyder 1955, TeHennepe 1963, Chihara 1965, Hilbert and Bernays 1968). In the context of this paper that concerns Smokrović's *logicist thesis*, we are interested in the relation between the *arguments* used in the debate and *logical forms* in which they are to be put, in order to see whether logic as such can be of use in trying to settle the dispute between the two parties. This will also concern debates in the philosophy of mathematics, which, as mentioned in the introduction, may be involved in the question about *descriptive nature* of formal logic and *normativity of reasoning*.

As the case study, we shall combine two famous examples: *The Thomson Lamp* (Thomson 1968) and *The Staccato Run* (Grünbaum 1968, Arsenijević 1988). The Thomson lamp is just an ordinary lamp, except that it may be on for 1 sec, off for  $\frac{1}{2}$  sec, on for  $\frac{1}{4}$  sec, and so on ad infinitum. If it changes its states in such a way, is it on or off after 2 sec elapse? The *staccato run* of a runner is the motion in which he stops at the half-way point after  $\frac{1}{2}$  sec, rests there for  $\frac{1}{2}$  sec, moves further on for  $\frac{1}{4}$  sec with the same speed, when he stops and rests for  $\frac{1}{4}$  sec, and so on ad infinitum. Where will he be after 2 sec?

Benacerraf remarked in his illuminating 1962 paper that the given description of the way in which the Thomson lamp functions concerns only its being on and off *within* the *open* interval of 2 sec. Nothing is said about its state *after* this time. Both being on as well as being off after 2 sec are *compatible* with what was happening within the open interval of 2 sec. Af-

ter all, the lamp can be destroyed just as 2 sec elapse, so that it makes no sense to say that it is either on or off.

Taken as such, Benacerraf's remark is correct. It *suggests* that, from the mathematical point of view, the performance of the infinite task is possible. However, in order to *prove* that it is physically possible, Grünbaum (1968, p. 97) imagined an electric device in which the state of the lamp *after 2 sec* should be *predictably on as a consequence* of an infinite number of jabbing motions of the button of a lamp. If it is so, then the given infinite task would be proved to be completable, since it would have a direct physical consequence.



Let the button of a lamp be equipped with an electrically conducting base which can close the circuit by fitting into the space between the exposed circuit elements  $E_1$  and  $E_2$  (see the diagram). Let an infinite process begin so that the button, whose base is at  $A_1$ , 1/2 cm above  $E_1E_2$ , is pressed down as to reach in  $(1/2 + 1/4)$  sec the point  $B_1$ , 1/4 cm bellow  $E_1E_2$ . After being at rest for 1/4 sec, the button is raised in  $(1/4 + 1/8)$  sec up to the point  $A_2$ , 1/8 cm above  $E_1E_2$ , being at rest there for 1/8 sec. Then, the button is pressed down in  $(1/8 + 1/16)$  sec as to reach the point  $B_2$ , 1/16 cm bellow  $E_1E_2$ , being at rest there for 1/16 sec. And so on, and so forth, let

the process of downward and upward motions of the button be continued endlessly, by being successively in positions  $A_1, B_1, A_2, B_2, A_3, B_3$ , and so on. Now, it can be argued that, if electric device remains intact after 2 sec, the base of the button of the lamp can be neither above nor bellow  $E_1E_2$ , so that it must be *at*  $E_1E_2$ , which means that the lamp will be *predictably on*. This proves, according to Grünbaum, that, under the given description, it is not only possible that within the open interval of 2 sec the infinite task consisting of downward and upward jabbing motions is *performable* but also that it is *completable*, since it has a definite physical consequence. In the given case, the performance of an infinite task leads to the performance of a *super-task*.

So, at least in the given case, the infinitists seem to be right. However, Allen Janis, Grünbaum's colleague at the University of Pittsburg (see Grünbaum 1968, p. 101, n. 64), concocted the situation ingeniously by suggesting an alternative switching arrangement, where the circuit is *closed* if the button base is at  $E_1E_2$  after having come there from above but *open* if it is at  $E_1E_2$  after having come there from bellow (due to some isolator which covers the button base automatically after it passes through  $E_1E_2$  from above and being removed automatically after it passes through  $E_1E_2$  from bellow). Under this switching arrangement, the lamp should be both on and off, or neither on and off, after 2 sec have elapsed, since the button base should be at  $E_1E_2$  by coming there both from above and from bellow. So, the super-task should be feasible under Grünbaum's original arrangement but not feasible under Janis's alternative arrangement, though the two infinite processes are kinematically identical.

At the conference in Bielefeld 1994, in Grünbaum's presence, I drew attention to the problem that in Janis's arrangement there is a contradiction as the consequence of the fact that there is no last jabbing motion, either from above or from below, which is also the case in the original arrangement, where the infinite task is allegedly feasible. Surprisingly, Grünbaum reacted sympathetically and said that the addition of new conditions can make an otherwise feasible process unfeasible. He added that even his original arrangement may seem suspect due to specifically dynamical difficulties in effecting the infinitude of accelerations, because in ever shorter time intervals of upward and downward motions the acceleration should increase (and decrease) boundlessly. In relation to this question, he pointed to his 1969 paper, where it is shown how the last difficulty can be obviated through Richard Friedberg's re-arrangement, in which intermittent motions can proceed at suitable decreasing average velocities such that both successive peak velocities and accelerations during the decreasing sub-intervals converge to zero as we approach the terminal instant. It seems, however, that Grünbaum missed the point of Janis's re-arrangement.

Janis's re-arrangement is not intended to show how *some* change in the original description can make the originally feasible process unfeasible, but rather to suggest that, given that the introduction of the isolator is something external which does not influence the kinematics of the process so that the continuation of it is feasible up to any point within the open interval of 2 sec, the paradoxical outcome after 2 sec elapse casts doubt to the feasibility of the *kinematically identical* infinite process in the original arrangement. Namely, if one of the two kinematically identical processes is feasible, the other one should also be feasible, and if one of the two is unfeasible, the other one should be unfeasible as well. So, the fact that in the re-arrangement the button base cannot admittedly be at  $E_1E_2$  after 2 sec elapse, it cannot be there in the original arrangement either.

It is important to notice that the outcome in the second case is not analogous to the case in which the apparatus would be destroyed just as 2 sec elapse, for in such a case it would make no sense to ask whether the lamp is on or off. In Janis's re-arrangement the button base *should be* at  $E_1E_2$  but the lamp cannot be *neither on nor* off. It is not so because of the presence of the isolator at  $E_1E_2$  after 2 sec have elapsed, for this would make the circuit open. It is so because the button base didn't reach  $E_1E_2$  from below. Similarly, the circuit is not closed not because of the absence of the isolator at  $E_1E_2$ , but because the button base didn't reach  $E_1E_2$  from above. The conclusion is that after 2 sec the paradoxical outcome is the consequence of the fact that the button base should have reached  $E_1E_2$  *neither from above nor from below*. But then, after 2 sec the button base could not be at  $E_1E_2$  in the original arrangement either.

The dialectic of the above debate makes the problem extremely tricky. On one hand, it is hard to see what else but the impossibility to perform an infinite task can prevent the button base, in any of the two cases, to be at  $E_1E_2$  after 2 sec elapse. It is highly implausible to assume that the difference between the two arrangements is based on the fact that the apparatus "knows" in advance if the outcome would be paradoxical or not, so that it performs the infinite task in the original arrangement only, given that the task is performable in both cases up to any of  $A_n$  and  $B_n$  points. On the other hand, given that the process in both cases is performable up to any of  $A_n$  and  $B_n$  points, the question is where the button base can be after 2 sec elapse but at  $E_1E_2$ .

So, after all, if we reasonably accept that the impossibility to perform the infinite task in Janis's re-arrangement means that it is impossible to perform it in the original arrangement as well, the fact that the infinite process is performable up to any of  $A_n$  and  $B_n$  points should not imply that it can be finished by reaching  $E_1E_2$  when 2 sec elapse. If the process continues

ad infinitum, 2 sec *will not* elapse, and if 2 sec elapse, the process *had to stop* developing further on after some point. If  $f(A_n)$  and  $f(B_n)$  means that the process is performed up to  $A_n$  and  $B_n$  respectively, then  $\forall_n \diamond f(A_n)$  and  $\forall_n \diamond f(B_n)$  are true, but  $\forall_n f(A_n)$  and  $\forall_n f(B_n)$  false.

In many modal logic systems  $\forall_n \diamond f(A_n) \rightarrow \diamond \forall_n f(A_n)$  is not a theorem, and there is a lot of examples in finite models that may illustrate this. For instance, if there are too many pieces of good food on the table, each of them can be eaten during the party, but not all. In our case the truth of  $\forall_n \diamond f(A_n)$  lies in the dynamic character of the process which can develop without end. But it can be developing without end only because it is not unconditionally true that 2 sec will elapse. By dealing with Grünbaum's thought experiment, we smoothly mixed the extremely remote possible world in which the number of downward and upward jabbing motions increases boundlessly within the open interval of 2 sec with the everyday situation in which it seems obvious that the time interval of 2 sec must elapse unconditionally. We overlooked the possibility that the two possible worlds, one in which the process will develop endlessly and the other one in which 2 sec will elapse may be impossible, and therefore impossible.

However, the fact that the two worlds are impossible does not mean that we must choose from the very beginning *which* is the world we are speaking about. We may allow both that the process *may* develop boundlessly *and* that 2 sec *may* elapse, and leave it open whether the process will continue endlessly *or* 2 sec will elapse. What we mustn't suppose simultaneously is that the process *will* develop endlessly and that 2 sec *will* elapse.

## Conclusion

The analyzed examples show how we ought to proceed, following the requirement of Smokrović's logicist thesis, in order to obtain logical forms applicable, respectively, to our ordinary way of speaking about future contingencies and to the argumentation concerning the problem of the possibility to perform an infinite task. Fortunately, it has turned out that in both cases we do not have to depart from standard predicate logic but only to extend it appropriately.

In the first of the two cases, as expected, we have to introduce temporal and modal operators, since we want to speak about possible future events. After the suitable selection of temporal and temporal-modal axioms, we get a system in which neither the *principle of bivalence* nor the *principle of excluded middle* is restricted, but in which, in perfect congruence with everyday reasoning, it is neither true that it will be true tomorrow that the sea battle will happen nor that it is true that it will be true tomorrow that

the sea battle will not happen. But, if the sea battle happens tomorrow, this will mean that it is not possible any longer that it did not happen, for “what is, necessarily is, when it is, and what is not, necessarily is not, when it is not”, as Aristotle put it. This is why in this case it will become true tomorrow that it was true the day before that it would be true the day after that the sea battle had happened that day. At the same time, and again in accordance with everyday reasoning, though it ceased to be possible that the sea battle has not happened, it remains true that it could have been otherwise. And finally, and again in accordance with everyday reasoning, in which the prediction-retrodictation asymmetry plays an important role, the unpredictability of future contingencies is not a matter of epistemological but of logico-ontological fact that there are many and in principle innumerable ways in which the world history can develop, while there is just one, privileged way in which it has de facto developed.

In the second example, the comparative analysis of Grünbaum’s and Janis’s arrangements of the electric device—in which the button base of Thomson’s lamp is allegedly at  $E_1E_2$ , after an infinite number of downward and upward jabbing motions—shows that, if an infinite process is feasible in one of the two arrangements, it must be so in the other one too. But this leads to a contradiction, since the outcome is contradictory in Janis’s arrangement. So given that the presence of the isolator is something external that makes no difference from a kinematical point of view, we need nothing but standard logic to conclude the unfeasibility of an infinite task.

So far, so good! But this is not the end of the story, since the argument against the infinitism does not give the answer to the question concerning the position of the button base after 2 sec *elapse* and the whole device remains *intact*. It seems that, according to standard mathematical analysis, the button base can be just nowhere but at  $E_1E_2$ . However, it is standard mathematical analysis itself which, with the help of modal logic, gives a solution. Namely, within the open interval of 2 sec the infinite process is *performable* endlessly, but this does not mean that it is *completable* within the closed interval of 2 sec. For the 2 sec interval to be closed, there must be the last jabbing motion, which would be possible only if the process ceases to develop endlessly at some point within the open interval of 2 sec. So, there are two possible worlds, one in which the process develops endlessly, and the other one in which it ceases doing that. Since the two worlds are *impossible*, it is *not unconditionally* true that 2 sec will elapse, which seems odd only because the first one is *extremely remote* from the second one, which is the world in which we live. It is very interesting that David Hilbert, one of the greatest mathematicians of twentieth century, who proclaimed emphatically that nobody will push us out from Can-



tor's paradise, claimed that infinite tasks are physically unfeasible (Hilbert 1926 and Hilbert and Bernays 1968, p. 16). What he missed to say is that this is not *accidentally* so but a *consequence* of the *impossibility* of the two worlds, one in which the interval of 2 sec is *open* and the other one in which it is *closed*. The same holds for the so-called remarkable curves, which though mathematically definable are not drawable, not because they are two-dimensional, but because they are non-differentiable at one point at least, which the curve cannot approach from any direction whatsoever (see Arsenijević 1994).

So, after all, we have found *logical forms* which are in *congruence* both with *everyday reasoning* and *mathematical analysis*.

## References

- Anderson, A. R. and Belnap, N. 1975. *Entailment—The Logic of Relevance and Necessity*. Princeton University Press.
- Aristotle 1831. *De interpretatione*. In: *Opera*, ed. I. Bekkeri. Walter de Gruyter.
- Aristotle 1831. *Physica*. In: *Opera*, ed. I. Bekkeri. Walter de Gruyter.
- Arsenijević, M. 1988. "Solution of the Staccato Version of the Achilles Paradox". In: A. Pavković (ed.). *Contemporary Yugoslav Philosophy: The Analytic Approach*. Kluwer (pp. 27- 55).
- Arsenijević, M. 1989. "How many physically distinguished parts can a limited body contain?". *Analysis* 49 (1) (pp. 36-42).
- Arsenijević, M. 1994. "Mathematics, Infinity and the Physical World". *Dialektik* 3 (pp. 89-107).
- Arsenijević, M. 2016. "Avoiding Logical Determinism and Retaining the Principle of Bivalence within Temporal Modal Logic: Time as a Line-in-Drawing". In: S. Gero-giorgakis (ed.). *Time and Tense*. Munich: Philosophia Verlag (pp. 7-37).
- Arsenijević, M. and Jandrić, A. 2023. (forthcoming) "The Arrow and the Flow of Time in a Non-Deterministic World According to the Temporal-Modal System TM". *Logic and Logical Philosophy*.
- Belnap, N. 2007. "An Indeterministic View of the Parameters of Truth". In: T. Müller (ed.) *Philosophie der Zeit*. V. Klostermann (pp. 87-113).
- Benacerraf, P. 1962. "Tasks, Super-Tasks and Modern Eleatics". *Journal of Philosophy* 59 (24) (pp. 765-784).
- Black, M. 1951. "Achilles and the Tortoise". *Analysis* 11 (5) (pp. 91-101).
- Chihara C. S. 1965. "On the Possibility of Completing an Infinite Process". *Philosophical Review* 74 (pp. 74-87).
- Grünbaum A. 1968. *Modern Science and Zeno's Paradoxes*. London: George Allen and Unwin.
- Grünbaum A. 1969. "Can an Infinitude of Operations be Performed in a Finite Time?" *British Journal for the Philosophy of Science* 20 (pp. 203-218).



- Harman, G. 1986. *Change in View: Principles of Reasoning*. MIT Press.
- Hilbert, D. 1926. "Über das Unendliche". *Mathematische Annalen* 95 (pp. 161-190).
- Hilbert, D. and Bernays, P. 1968. *Grundlagen der Mathematik*. Springer Verlag. Berlin (1<sup>st</sup> ed. 1934).
- Łukasiewicz, J. 1918. Farewell Lecture by Professor Jan Łukasiewicz, delivered in the Warsaw University Lecture Hall on March 7, 1918. In: *Jan Łukasiewicz – Selected Works*. Amsterdam: North-Holland, 1970 (pp. 84-86).
- Łukasiewicz, J. 1920. On Three -Valued-Logic. In: idem, *Selected Works*. Amsterdam: North- Holland, 1970 (pp. 87-88).
- Łukasiewicz, J. 1922. On Determinism. In idem, *Selected Works*. Amsterdam: North-Holland 1970, (pp. 110-128).
- Maxwell, G. and Feigl, H. 1961. "Why Ordinary Language Needs Reforming". *Journal of Philosophy* 58 (18) (pp. 488-498).
- Salmon, W. C. 1975 *Space, Time, and Motion*. Encino CA: Dickenson.
- Schlick, M. 1931. "Die Kausalität in der gegenwärtigen Physik". *Naturwissenschaften* 19 (pp. 145-162).
- Smokrović, N. 2018. "Informal Reasoning and Formal Logic: Normativity of Natural Language Reasoning". *Croatian Journal of Philosophy*. Vol. XVIII, No. 54 (pp. 455-469).
- Schwayder, D. S. "Achilles Unbound" *Journal of Philosophy* 52 (pp. 449-459).
- Taylor, R. 1951. "Mr. Black on Temporal Paradoxes". *Analysis* 12 (2) (pp. 38-44).
- TeHennepe, E. 1963. "Language Reform and Philosophical Imperialism: Another Round with Zeno". *Analysis* (Suppl.) 23 (pp. 43-49).
- Thomson, J. F. 1968. "Tasks and Super-Tasks". *Analysis* 15 (1) (pp. 1-13).
- Varga, A., Stenning, K., and Martignon, L. 2015. "There Is No One Logic to Model Human Reasoning: the Case from Interpretation". In: U. Farbach and C. Shon (eds.). *Proceedings of the first workshop on bridging the gap between human and automated reasoning*. Berlin (pp.32-46).
- Watling, J. 1952. "The Sum of an Infinite Series". *Analysis* 13 (2) (pp. 39-46).
- Wisdom, J. O. 1952. "Achilles on Physical Racecourse". *Analysis* 12 (3) (pp. 67-72).
- Weyl, H. 1949. *Philosophy of Mathematics and Natural Sciences*. Princeton University Press.



IGOR BAJŠANSKI

## Funkcije rasuđivanja u individualnom i grupnom kontekstu

**Sažetak:** Smokrović (2015) razrađuje tezu o argumentaciji kao procesu koji je posebno koristan za stjecanje i proširivanje znanja, oslanjajući se na interaktivnu teoriju rasuđivanja. Tijekom argumentacije kognitivni mehanizam rasuđivanja aktivira se tako da na najbolji način obavlja svoju funkciju, odnosno, ono za što je dizajniran prirodnom selekcijom: za evaluaciju razloga i produkciju argumenta u interaktivnom, argumentacijskom kontekstu. U ovome radu prikazane su i druge potencijalne funkcije rasuđivanja, pri čemu je istaknuto da je rasuđivanje podjednako važno i u ne-argumentacijskim kontekstima, odnosno, da neće optimalno funkcionirati samo tijekom argumentacije. U analizi procesa i funkcija rasuđivanja uspoređeni su teorija dualnih procesa mišljenja i interaktivna teorija i istaknute su sličnosti u koncipiranju nesvjesnih inferencijalnih procesa i procesa racionalizacije. Opisane su funkcije rasuđivanja u individualnom kontekstu, koje uključuju generiranje objašnjenja i učenje. Konačno, istaknuta je uloga metakognitivnih procesa u aktivaciji procesa rasuđivanja.

**Gljučne riječi:** rasuđivanje, interaktivna teorija rasuđivanja, argumentacija, teorija dualnih procesa, racionalizacija.

### Interaktivni pristup rasuđivanju

U članku *Argumentation as a means for extending knowledge* Smokrović (2015) razrađuje tezu o argumentaciji kao procesu koji je posebno koristan za stjecanje i proširivanje znanja. Oslanjajući se na interaktivnu teoriju rasuđivanja<sup>1</sup> (Mercier i Sperber, 2011) i teoriju sigurnoga znanja (Williamson, 2000), Smokrović smatra da proces argumentacije vodi sudionike prema proširivanju znanja. Argumentacija je oblik komunikacije, pri kojem sugovornici iznose argumente, odnosno tvrdnje i razloge koji podupiru tvrdnje, te procjenjuju iznesene argumente. Ako je motivacija sugovorni-

---

<sup>1</sup> Termin *rasuđivanje* koristit ću kao prijevod engl. termina *reasoning*, a termin *zaključivanje* kao prijevod termina *inference*. Često se ne čini razlika između ta dva pojma, međutim, za samu argumentacijsku teoriju rasuđivanja ključna je razlika između procesa izvođenja zaključaka u najopćenitijem smislu koji pokriva sve procese ekstrakcije novih informacija iz danih informacija i procesa rasuđivanja, koji jesu vrsta procesa zaključivanja i koji uključuju pronalaženje i vrednovanje razloga za usvajanje zaključaka te produkciju argumenata.

ka takva da uključuje radoznalost, odnosno želju da se zna je li tvrdnja  $p$  točna, i ako sugovornici posjeduju određena vjerovanja koja se tiču tvrdnje  $p$ , proces argumentacije rezultirat će proširivanjem znanja sugovornika. Takvo optimalno funkcioniranje argumentacije posljedica je toga što se tijekom argumentacije kognitivni mehanizam rasuđivanja aktivira tako da na najbolji način obavlja svoju funkciju, odnosno ono za što je dizajniran prirodnom selekcijom: za procjenu razloga i produkciju argumenta u interaktivnom, argumentacijskom kontekstu.

Ovakva gledišta temelje se na interaktivnoj teoriji zaključivanja (*interactive theory of reasoning*, ITR) koju su u nizu članaka (Mercier i Sperber, 2011, 2021) i u knjizi *The enigma of reason* (2017) prikazali kognitivni znanstvenici Hugo Mercier i Dan Sperber (M&S). Prema ITR-u, kognitivni mehanizam rasuđivanja ustrojen je biološkom evolucijom tako da optimalno funkcionira u argumentacijskom kontekstu. U ovome članku razmotrit ću ključne točke ITR-a i prikazati ograničenja i potencijalne nedostatke teorije. Ukratko, definiranje rasuđivanja kao procjene razloga i produkcije argumenata predstavlja bitan pomak u odnosu na tradicionalno gledanje na rasuđivanje u psihologiji kao na inferencijalni proces izvođenja zaključaka iz skupa premisa. Međutim, smatram problematičnim svođenje rasuđivanja na argumentacijsku funkciju. Pokušat ću iznijeti argumente u prilog gledištu da je rasuđivanje podjednako važno i u ne-argumentacijskim kontekstima, odnosno da neće optimalno funkcionirati samo tijekom procesa argumentacije.

Nadalje, kao osnovne postavke interaktivne teorije možemo izdvojiti 1) definiciju rasuđivanja kao specifičnog oblika intuitivnog zaključivanja, a ne zasebnog reflektivnog i ne-intuitivnog procesa, čija je funkcija vrednovanje razloga i produkcija argumenata, 2) pretpostavku da je mehanizam rasuđivanja evoluirao u socijalnom kontekstu i da je njegova primarna funkcija opravdanje vlastitih vjerovanja i postupaka pred drugima i produkcija argumenata koji služe uvjeravanju drugih i branjenju vlastitih pozicija i vlastitih razloga, 3) tezu da zbog specifičnog evolucijskog ustroja rasuđivanje funkcionira optimalno u argumentacijskom i interaktivnom kontekstu te da je takav kontekst optimalan aktivator rasuđivanja, te 4) određivanje potencijalnih prednosti i nedostataka rasuđivanja (na primjer, pogreške potvrđivanja) s obzirom na kontekst u kojem se koristi. Polazeći od definicije rasuđivanja kao kognitivnog mehanizma čija je funkcija vrednovanje razloga i produkcija argumenata, argumentirat ću 1) da je tako opisan mehanizam rasuđivanja evoluirao jer je imao ključnu ulogu u različitim aspektima mišljenja, kao što su donošenje odluka, učenje i rješavanje problema, 2) da je ono što aktivira mehanizam rasuđivanja meta-kognitivna procjena adekvatnosti sadržaja misli s obzirom na ciljeve, 3) da

se različiti kontekstualni utjecaji na rasuđivanje, kao što je pozitivan efekt grupne rasprave na rješavanje jednostavnih logičkih problema, ne moraju objašnjavati specijaliziranim kognitivnim modulom za rasuđivanje u argumentacijskom kontekstu.

## Mišljenje i rasuđivanje

M&S ističu da se njihov pristup rasuđivanju i ITR-u bitno razlikuju od danas dominantne teorije rasuđivanja i ljudskoga mišljenja općenito – teorije dualnih procesa (*dual process theory*, DPT; Evans, 2003; Evans & Stanovich, 2013; Sloman, 1996). Međutim, oba pristupa u mnogočemu se slažu.

Prema DPT-u, razlikujemo dvije kategorije procesa mišljenja jednostavno nazvanih procesi tipa 1 i procesi tipa 2. Procesi tipa 1 brzi su, automatski, temelje se na heuristikama i stvaraju intuicije. Procesi tipa 2 spori su, serijalni, neintuitivni i oni čine analitičko i reflektivno mišljenje. Procesi tipa 1 ne opterećuju kognitivne resurse pažnje i radnoga pamćenja, dok procesi tipa 2 zahtijevaju ulaganje mentalnog napora. Iako postoje brojne verzije i varijante dualnih teorija, ovo su temeljne univerzalne pretpostavke različitih pristupa. Važna točka slaganja između DPT-a i ITR-a je to da je mišljenje često intuitivno. Što su točno intuicije u ovome kontekstu? Intuicije se produkti nesvjesnih mehanizama zaključivanja ili inferencijalnih procesa. Nesvjesni mehanizmi sveprisutni su u kognitivnim sustavima i oni čine temelj fleksibilnog i adaptivnog ponašanja. Nesvjesno zaključivanje neophodno je funkcioniranje svih kognitivnih procesa, uključujući percepciju, učenje, pamćenje, kategoriziranje, jezik i mišljenje. Naizgled jednostavne kognitivne operacije kao što su percepcija dubine, binokularni vid ili učenje putem klasičnog uvjetovanja, nemoguće su bez inferencijalnih mehanizama.

Dio nesvjesnih mehanizama zaključivanja proizvodi intuicije. Intuicije su prosudbe ili odluke koje prihvaćamo bez znanja o razlozima zašto ih prihvaćamo. Intuicije se javljaju u svijesti bez iskustva kako i zašto nastaju i uvijek su praćene specifičnim metakognitivnim iskustvom, osjećajem uvjerenosti u točnost ili općenito u kvalitetu i snagu intuicije. U skladu s tezom o modularnosti uma, M&S smatraju da postoji puno specijaliziranih kognitivnih mehanizama koji proizvode intuitivne zaključke. Slično, jedan od važnih zastupnika DPT-a, Keith Stanovich (2011) naziva procese tipa 1 skupom autonomnih sustava. Prema tome, danas se uglavnom svi slažu da intuicije generira niz različitih i specijaliziranih inferencijalnih mehanizama.

Psihologiju mišljenja zanima podskup intuicija koje nastaju kada izvodimo nova vjerovanja i donosimo odluke. Dio nesvjesnih inferencijalnih procesa kao input uzima vjerovanja, koja se mogu formirati na temelju

opažanja, mogu biti aktivirana iz dugoročnoga pamćenja ili mogu biti prenesena jezičnom komunikacijom, a kao output generira druga vjerovanja ili odluku. Takva se nova vjerovanja ili odluke manifestiraju kao intuicije u svijesti i uvijek su praćene metakognitivnim iskustvima.

Razmotrimo nekoliko uobičajenih primjera. Zadatak *palice i loptice* najpoznatiji je zadatak iz kratkoga Testa kognitivne refleksije (Frederick, 2005) koji je korišten u stotinama istraživanja i koji se tretira kao ključni test za ispitivanje intuitivnoga i reflektivnoga mišljenja. Zadatci u testu koncipirani su tako da kod većine ispitanika induciraju pogrešan odgovor koji je praćen snažnim osjećajem točnosti. Zadatak glasi ovako: *Palica i loptica zajedno koštaju 110 kuna. Palica košta 100 kuna više od loptice. Koliko košta loptica?* Većina ispitanika daje odgovor *10 kuna*, iako je točan odgovor *5 kuna*.

Ako ispitanicima prezentiramo silogizam oblika *Svi A su B. Svi B su C. Dakle, svi A su C.*, gotovo će se svi ispitanici složiti da je silogizam *u redu*, odnosno da zaključak slijedi iz premisa, što je normativno točan odgovor. Ako im prezentiramo silogizam oblika *Svi A su B. Svi C su B. Dakle, svi A su C.*, mnogi će se složiti i da u ovome slučaju zaključak slijedi iz premisa, iako to nije normativno točan odgovor (Evans i sur. 1999).

Konačno, prisjetimo se i problema Linde (Tversky i Kahneman, 1983), u kojem je Linda opisana kao inteligentna mlada žena koja je diplomirala filozofiju i u studentskim je danima sudjelovala u antiratnim prosvjedima. Kada im se opiše Linda, većina ljudi zaključuje da je veća vjerojatnost da je Linda bankovna službenica koja je i feministica, nego da je bankovna službenica, kršeći pri tome elementarne zakonitosti vjerojatnosti konjunkcije događaja.

U svim navedenim zadacima odgovori se u svijesti ispitanika javljaju spontano i brzo. Takvi odgovori praćeni su osjećajem točnosti: ispitanici će lako procijeniti koliko im se ispravnim čini odgovor do kojega su došli. Ispitanici naknadno mogu pokušati objasniti zašto su odabrali određeni odgovor, ali sam odgovor prethodi takvim potencijalnim objašnjenjima i razlozima. Takvi brzi odgovori javljaju se bez bilo kakvog introspektivnog uvida u proces kojim nastaju. Takvi odgovori jesu intuicije.

Velika se rasprava vodila o tome koliko su intuicije točne. Starije verzije DPT-a, na tragu pristupa heuristika i pristranosti, isticale su netočnost intuicija. Točno mišljenje zahtijevalo bi pažljivu primjenu procesa tipa 2 i inhibiciju pogrešnih intuicija. Međutim, novija istraživanja i novije revizije DPT-a jasno ističu efikasnost intuitivnog mišljenja. Na primjer, u nizu eksperimenata Wim De Neys (2016, 2012, 2014) pokazao je da ljudi, i kada daju pogrešne intuitivne odgovore, svejedno pokazuju osjetljivost na konflikt između pogrešnih intuicija i normativno točnih odgovora koja se ma-

nifestira nižim metakognitivnim procjenama točnosti odgovora i sporijim rješavanjem zadatka. U skladu s tim, u skorašnjijim radovima u području DPT-a govori se i o logičkim intuicijama. Mogućnost točnog i efikasnog intuitivnog mišljenja predstavlja veliki odmak u odnosu na ranije verzije DPT-a i samu teoriju približava glavnom *protivniku* – pristupu ekološke racionalnosti u kojem je velik naglasak stavljen na *snagu intuicije* (Gigerenzer 2007).

Prema DPT-u, kvalitativno drugačiji procesi mišljenja u odnosu na intuitivno mišljenje prisutni su kada se aktiviraju procesi tipa 2. U okviru DPT-a jasno je postavljena razlika između intuitivnog i analitičkog zaključivanja, a ključni kriterij razgraničenja je aktivacija kognitivnih resursa potrebnih za serijalno svjesno mišljenje, a to su pažnja i radno pamćenje. Ilustrirat ću djelovanje dva tipa procesa na problemu palice i loptice. Tipičan pogrešan odgovor (*10 kuna*) ishod je nesvjesnih heurističkih procesa tipa 1. Za normativno točan odgovor potrebno je pokrenuti spore procese tipa 2. Možemo pretpostaviti da se manji broj ljudi neće osloniti na ishode intuitivnog procesiranja, već da će pokušati postaviti problem u obliku matematičkog zadatka (npr. *neka je  $x$  cijena loptice, onda je  $x + x + 100 = 110$ ,  $2x$  je 10, dakle  $x$  je 5*). DPT će ovo nazvati sporim analitičkim mišljenjem ili analitičkim rasuđivanjem. Ovakvi procesi aktivirat će se nakon što brzi procesi generiraju intuitivan, u ovome slučaju, pogrešan odgovor. Na sličan se način objašnjava i izvedba na prethodno spomenutim problemima silogističkog zaključivanja i prosudbe vjerojatnosti konjunkcije događaja. Što određuje hoće li se pokrenuti spori procesi tipa 2? Prema metakognitivnoj teoriji rasuđivanja (Thompson 2009; Thompson i sur. 2011), ono što će potaknuti sporo promišljanje je nizak metakognitivni osjećaj točnosti koji prati pogrešne intuitivne odgovore, a prema pristupu logičkih intuicija i istraživanjima detekcije konflikta između intuitivnog i analitičkog mišljenja (De Neys 2006; 2014), procesi tipa 2 pokrenut će se kada prilikom rješavanja zadatka dođe do konflikta između heurističkih procesa koji će generirati pogrešan odgovor i analitičkih procesa koji će generirati točan odgovor.

ITR rasuđivanje koncipira na drugačiji način nego DPT. Rasuđivanje bi se odnosilo na traženje razloga za ponuđene odgovore, kao i na produkciju argumenata. U slučaju problema palice i loptice, rasuđivanje bi odgovaralo promišljanju zbog čega mislim da je 10 točan odgovor. Kako mogu to provjeriti? Odgovori na ova pitanja mogu biti različiti. Dio rješavatelja će jednostavno odgovoriti da im taj odgovor izgleda točno i bit će zadovoljni s njime (*Zašto mislim da je 10 točno? Pa brzo sam riješio, očigledno je točan odgovor! Što bi drugo moglo biti?*). Dio će vjerojatno zastati i pažljivije razmisliti o razlozima (*Izgleda mi da je 5... Idem ipak probati riješiti... Da*

*vidimo,  $X + X + 100$  je 110, dakle,  $X$  mora biti 5! 5 je točan odgovor*). I u jednom i u drugom slučaju radi se o rasuđivanju. Mogli bismo reći da je za intuitivnog ispitanika u prvom slučaju prihvatljiv razlog to što je brzo došao do odgovora i to što mu drugi odgovor ne pada na pamet. Za reflektivnog ispitanika razlog za odbacivanje intuitivnog odgovora je odgovarajući matematički izvod<sup>2</sup>. Nakon što su razlozi generirani možemo ih vrednovati, tako što ćemo, na primjer, procijeniti da to što smo brzo riješili zadatak ne mora biti dobar razlog za to da rješenje smatramo točnim. Rasuđivanje nije bilo koji inferencijalni proces jer najveći dio inferencijalnih procesa koji su u podlozi intuicija ne uključuje razmišljanje o razlozima niti produkciju argumenata.

Prema tome, bitna je sličnost između ITR-a i DPT-a ta što smatraju da se mišljenje primarno odvija putem intuitivnih mehanizama. Ono što je različito između dvije teorije je koncipiranje rasuđivanja. Prema ITR-u, rasuđivanje je također intuitivan proces koji producira i evaluira razloge i argumente, dok se prema DPT-u analitičko rasuđivanje (ono koje provode procesi tipa 2) odnosi na primjenu pravila na serijalan način. Međutim, treba primijetiti da je u okviru DPT-a jasno prepoznata uloga rasuđivanja u procesiranju razloga. Od samih početaka pretpostavljalo se da su spori analitički procesi primarno odgovorni za racionalizaciju – izmišljanje razloga zašto imamo neko vjerovanje ili zašto smo donijeli određenu odluku. Stoga, prema DPT-u primjena pravila u rješavanju problema, kao što je primjena algebre u rješavanju problema palice i loptice, te generiranje i vrednovanje razloga, dvije su različite funkcije procesa tipa 2.

Kombinirajući DPT i ITR možemo razlikovati nekoliko osnovnih formi mišljenja. Prvo, imamo intuitivne heurističke procese zaključivanja koji će, ako je to moguće, generirati brze odgovore. Drugo, kada nemamo jasnih intuicija, kada ignoriramo intuiciju ili kada je odbacimo kao pogrešnu te ako imamo odgovarajuće znanje koje je relevantno za problemsku situaciju, možemo koristiti određenu proceduru rješavanja i kroz niz transformacija zadanih informacija doći do rješenja o kojima nemamo nikakvih intuicija (očigledan primjer jesu složeniji matematički problemi). Takvo specijalizirano znanje Stanovich (2018) naziva *mindware*. *Mindware* će utjecati na točnost intuitivnog mišljenja, ali će nam i pružiti odgovarajuće procedure koje možemo primijeniti na spori, tip 2 način. Osim primje-

---

<sup>2</sup> Zanimljivo je primijetiti da je ovakav tip refleksije, barem u zadatku palice i loptice, kao i u drugim najčešće korištenim zadacima u istraživanjima u okviru DPT-a (kategorički silogizmi, Linda problem, zadatci kondicionalnog zaključivanja i razni zadatci procjena vjerojatnosti), vrlo rijedak. Malobrojni ispitanici koji daju točne odgovore najčešće to rade brzo: (rijetki) točni odgovori najčešće nisu rezultat pažljive refleksije, već točnih intuicija!



ne specijaliziranog znanja, mogu se koristiti i opće heuristike, kao što je heuristika pokušaja i pogrešaka ili raščlamba kompleksnijeg problema na jednostavnije probleme. Ono što je zajedničko za ovakvo mišljenje, koje možemo nazvati analitičkim, je da se odvija na serijalan način i da zahtijeva usmjerenu pažnju i resurse radnoga pamćenja. Treće, rasuđivanje bi se odnosilo na vrednovanje razloga i produciranje argumenata. Možemo se složiti s ITR-om da ovako koncipirano rasuđivanje nije isključivo serijalni analitički proces koji bi se mogao izjednačiti s procesima tipa 2, iako se i rasuđivanje može odvijati na spori analitički način. Svakodnevno mišljenje će predstavljati kombinaciju različitih procesa. Rješavanje praktičnog problema kao što planiranje budžeta ili donošenje odluke kao što je odabir fakulteta mogu se oslanjati na različite kombinacije intuicije, analitičkog mišljenja i rasuđivanja.

### **Rasuđivanje i racionalizacija**

Ako je mišljenje kombinacija intuitivnog procesiranja, sporog analitičkog mišljenja i rasuđivanja, pitanje je koja je uloga rasuđivanja i kako ono funkcionira. Uloga intuitivnog i analitičkog mišljenja čini se jasnom. Intuitivno mišljenje je brzo, ono će se temeljiti na specijaliziranim evolucijskim dispozicijama i na stečenom znanju koje je pohranjeno tako da može biti lako dostupno. Analitičko mišljenje služi nam za primjenu pravila tako što korak po korak vršimo odgovarajuće kognitivne transformacije ulaznih informacija u situacijama kada kognitivni zadatak ne inducira zadovoljavajuće intuitivne solucije. Zašto i kako rasuđujemo?

ITR i DPT slažu se da je ono što rasuđivanje radi pronalaženje opravdanja za postojeće intuicije ili za donesene odluke, odnosno racionalizacija. Kao primjer racionalizacije možemo navesti istraživanja s metodom dva odgovora (Thompson i sur. 2011) kod kojih se ispitanicima dva puta prezentira isti zadatak, tipično jednostavan logički problem. Nakon prve prezentacije ispitanik treba riješiti zadatak tako što će dati prvi odgovor koji mu padne na pamet. Nakon toga slijedi druga prezentacija zadatka, a ispitanik treba pažljivo promisliti o problemu, bez vremenskog pritiska, i nakon toga ponuditi svoje konačno rješenje. Tipičan je rezultat ovakvih istraživanja taj da ispitanici rijetko mijenjaju pogrešne prve odgovore. Osim toga, kada u drugom pokušaju, nakon pažljivog promišljanja, daju drugi odgovor, taj je drugi odgovor često praćen visokim osjećajem sigurnosti da je odgovor točan. Ovakvo ustrajanje na pogrešnom odgovoru uz visok osjećaj sigurnost uzima se kao indikator procesa racionalizacije početnog pogrešnog odgovora uz pretpostavku da ispitanik nije u drugom pokušaju ulagao napor u traženje točnog odgovora, već u traženje dodatne evidencije koja ide u prilog pogrešnom odgovoru. To da ljudi traže oprav-

danja i izmišljaju objašnjenja za svoja vjerovanja i postupke, nije nova spoznaja. Slični primjeri su i racionalizacija moralnih prosudbi (Haidt 2001), kognitivna disonanca (Festinger 1957) i racionalizacija uzroka akcije (Nisbett i Wilson 1977). Ovoj dugoj listi naknadnih racionalizacija, Mercier i Sperber (2017) dodaju brojne primjere nepoželjnih učinaka rasuđivanja na naše odluke: donosimo one odluke koje je lakše opravdati, a često će više rasuđivanja negativno utjecati na kvalitetu odluke. Kako god definirali pojam racionalnoga mišljenja, ovakva obilježja ljudskoga mišljenja zasigurno ne možemo smatrati racionalnima.

Prema tome, jasno je da je rasuđivanje kognitivni mehanizam koji ima ulogu u racionalizaciji. Ono što je manje jasno je ima li rasuđivanje produktivnu ulogu, odnosno može li rasuđivanje biti temelj za formiranje novih spoznaja i za donošenje odluka, a ne samo za naknadno opravdavanje. Osim racionalizacije, DPT pretpostavlja upravo takvu, produktivnu ulogu rasuđivanja. Međutim, kao što je već spomenuto, analitičko mišljenje, barem kako je istraživano u okviru DPT-a, rijetko pridonosi mijenjanju brzih pogrešnih intuitivnih odgovora. Objašnjenje kako analitičko mišljenje obavlja produktivnu funkciju razlikuje se između različitih dualnih pristupa. Na primjer, prema Evansu (2019) standardna je funkcija procesa tipa 2 racionalizacija, odnosno pronalaženje razloga koji podupiru intuicije koje generiraju procesi tipa 1. Ako pronađeni razlozi nisu dovoljno dobri, rasuđivanje će se aktivirati u traženju drugačijih zaključaka ili odluka. Pri tome, ljudi će biti ustrajni u traženju opravdanja za postojeće intuicije, što je izvor kognitivne pristranosti. Suprotno pojednostavljenom gledištu, prema kojem su krive intuicije glavni razlog pogrešaka i pristranosti u zaključivanju, i analitičko mišljenje snosi dio odgovornosti. Pristrano traženje evidencije u prilog vlastitome gledištu (pristranosti potvrđivanja ili pristranosti vlastitoga gledišta) rezultat je funkcioniranja procesa tipa 2. Stanovich (2018) smatra da će doprinos procesa tipa 2 produktivnom mišljenju ovisiti o stupnju relevantnog znanja, pri čemu će ti procesi igrati ključnu ulogu kod umjerene razine znanja, odnosno kada znanje nije bogato i automatski dostupno, niti kada relevantnog znanja nema. Proces tipa 2 potrebni su kada se javlja konflikt između pogrešnih intuicija i odgovora koji su u skladu sa znanjem, a njihova je funkcija inhibicija i nadvladavanje intuicije te pronalaženje rješenja primjenom odgovarajućih pravila.

S druge strane, ITR pretpostavlja da će rasuđivanje rijetko imati produktivnu ulogu. Rasuđivanje primarno služi opravdavanju vlastitih gledišta i odluka. Međutim, temeljni kontekst funkcioniranja rasuđivanja je socijalna interakcija ili rasprava. Konstruiramo i vrednujemo razloge i argumente ne zato jer oni imaju produktivnu ulogu, već primarno zato da bi se opravdali pred drugima i da bi druge uvjerali. Kao što kažu M&S, *razlozi*

su za društvenu potrošnju. Pri tome postoji temeljna asimetrija u produkciji i vrednovanju argumenata: ljudi će vrlo lako pronalaziti razloge za svoja vjerovanja, neće biti kritični prema svojim argumentima, dok će biti vrlo oprezni i vrlo kritični prema tuđim razlozima i argumentima. U argumentacijskom će se kontekstu katkad pojaviti i produktivna uloga rasuđivanja, kada sugovornici iznose različita gledišta, razloge i argumente koji podupiru ta gledišta, kritiziraju tuđa gledišta, što će često rezultirati boljim rješenjima i odlukama. Smokrović (2015) naglašava upravo ovakvu produktivnu ulogu rasuđivanja u argumentacijskom kontekstu. Evidencija koja ide u prilog tezi o argumentacijskom kontekstu produktivne funkcije rasuđivanja je uvjerljiva. Pa tako, između ostalog, u argumentacijskom kontekstu ljudi produciraju bolje argumente, prihvaćaju jake argumente čak i ako su prethodno čvrsto vjerovali u suprotno te bolje rješavaju inače vrlo teške logičke probleme (za pregled vidi Mercier i Sperber 2017, poglavlje 15). Isto tako, negativni efekti argumentacijskog konteksta kao što su grupno mišljenje, do kojega dolazi kada nema suprotstavljenih gledišta, i grupna polarizacija, do koje dolazi kada su prisutna dva snažno suprotstavljena stava, također se mogu objasniti argumentacijskom funkcijom rasuđivanja i asimetrijom u vrednovanju svojih i tuđih razloga i argumenata.

## **Evolucija i funkcija rasuđivanja**

Evidencija navedena do sada ukazivala bi na to da rasuđivanje nema neku posebnu funkciju izvan argumentacijskog konteksta, da nam ono primarno služi za opravdavanje i uvjeravanje i da korištenje rasuđivanja izvan argumentacijskog konteksta uglavnom proizvodi loše ishode, kao što su pristranost prema vlastitim gledištima i racionalizacija. Međutim, pokušat ću prikazati ograničenja ovako shvaćenoga rasuđivanja i navesti potencijalne funkcije rasuđivanja u individualnom mišljenju.

Da bi mogli odrediti koja je funkcija rasuđivanja ili čemu rasuđivanje služi, potrebno je prvo odrediti što uopće znači funkcija kognitivnog sustava. Strogo gledano, u skladu s evolucijskom biologijom i s evolucijski orijentiranim pristupima u kognitivnim znanostima, funkcija bilo kojeg podsustava u živome organizmu odnosi se na to kako taj podsustav promiče reproduktivnu uspješnost. Funkcije adaptacija, odnosno sustava unutar organizama koji su dizajnirani prirodnom selekcijom, očituju se u onome što ti sustavi rade i kako pridonose vjerojatnosti preživljavanja i reproduktivnoj uspješnosti.

Međutim, ovakav tip adaptacijske analize, odnosno dovođenja u vezu evolucijske prošlosti, procesa prirodne selekcije i funkcije, u kognitivnim znanostima suočava se s brojnim problemima usprkos velikoj popularnosti evolucijske psihologije. Osim tradicionalne kritike da adaptacijska objaš-

njenja u psihologiji nisu više od pukoga pričanja priča, čak i ako prihvatimo adaptacionistička funkcionalna objašnjenja u psihologiji, činjenica je da ćemo jako teško pronaći uvjerljivu empirijsku evidenciju u prilog naše funkcionalne hipoteze, a protiv neke druge hipoteze. Mercier i Sperber (2017) nude blaži kriterij za identifikaciju funkcije kognitivnog mehanizma, a to je ispitivanje podudarnosti između funkcije organa (onoga što organ radi) i njegove strukture i njegovih učinaka. Na primjer, iako ne znamo kako je vidni sustav evoluirao kroz tisuće generacija, kako su se postupno gomilale genetske promjene koje su u podlozi dizajna oka, vidnih puteva i vidnoga korteksa, možemo jasno uočiti da to kako je vidni sustav ustrojen odgovara funkciji procesiranja svjetlosnih signala i ažuriranja mentalnog modela vanjskoga svijeta na temelju tih signala. Različite komponente toga sustava jasno pridonose toj funkciji. Obrnutim zaključivanjem, kada uočimo takvu jasnu vezu između strukture i funkcije, jedino moguće objašnjenje je da je takav sustav evoluirao prirodnom selekcijom. Prethodno spomenuta evidencija, o slabosti individualnog rasuđivanja i efikasnosti rasuđivanja u interaktivnom kontekstu, za ITR predstavlja jasnu evidenciju da je kognitivni mehanizam rasuđivanja evoluirao zbog svojih socijalnih funkcija. Kao jedinu alternativu takvom adaptacionističkom objašnjenju M&S nude ono što nazivaju intelektualnom pozicijom, prema kojoj rasuđivanje služi stjecanju točnoga znanja i donošenju ispravnih odluka. Kratki pogled na zadnjih pedesetak godina eksperimentalnih istraživanja ljudskoga mišljenja i na beskrajnu listu pristranosti i pogrešaka koje karakteriziraju rasuđivanje inteligentnih ljudi ukazuje na to da je intelektualnu poziciju jako teško obraniti.

U raspravi o funkciji rasuđivanja korisno je razmotriti i druge potencijalne funkcije. Pri tome nećemo ulaziti u raspravu koja je izvorna funkcija rasuđivanja i zbog čega je rasuđivanje evoluiralo. Opredijelit ćemo se za vrlo slabi kriterij u identificiranju potencijalnih funkcija, a to je bilo koja vrsta poželjnog ishoda, bez obzira za stvarni evolucijski scenarij. Na primjer, predložene su brojne hipoteze o tome zašto je evoluirao jezik, međutim, poželjne ishode kognitivnog mehanizma jezika pronalazimo u različitim domenama za koje jezik, očigledno, nije evoluirao, kao što su čitanje i pisanje. Ako primijenimo ovakav slabi kriterij, moguće je identificirati i druge potencijalne posljedice rasuđivanja osim socijalnih i intelektualno-individualnih.

Prethodno je navedeno da rasuđivanje može imati dvije funkcije – naknadno opravdanje postojećih intuicija (1) i produktivnu funkciju u generiranju vjerovanja i donošenju odluka (2). Nadalje, u skladu s uobičajenom klasifikacijom na epistemičku i praktičnu racionalnost, rasuđivanje bi moglo imati epistemičke funkcije, odnosno moglo bi pridonositi generiranju i

opravdavanju ispravnih vjerovanja, a moglo bi imati i praktičnu funkciju u donošenju i opravdavanju odluka. Osim toga, rasuđivanje bi moglo funkcionirati u socijalnom ili u individualnom kontekstu, što je podjela koju jasno ističu M&S.

Tradicionalno shvaćenom rasuđivanju odgovarala bi individualno-epistemička funkcija. Individualno-praktična funkcija odnosila bi se na potencijalnu ulogu rasuđivanja pri donošenju odluka. Iako smo prikazali brojna ograničenja individualnog mišljenja i upitnu vrijednost rasuđivanja kod generiranja novih, točnih vjerovanja i kod odlučivanja, čini se neobičnim pretpostaviti da rasuđivanje nema nikakvu ili pak da ima samo štetnu ulogu u individualnom mišljenju.

Što se tiče klasičnih problema korištenih u psihologiji rasuđivanja, kao što su zadaci prosudbe vjerojatnosti ili deduktivni problemi, istina je da individualno rasuđivanje kod takvih zadataka malo pridonosi davanju normativno točnih odgovora, a slično je i s istraživanjima donošenja odluka. Zasiurno dio poteškoća s takvim zadacima proizlazi i iz same prirode zadataka, njihove artifičnosti, poteškoća s praćenjem uputa u istraživanju, nepodudarnosti između interpretacije logičkih veznika i njihove formalne definicije, te iz niza drugih faktora. Međutim, ovakvo isticanje ograničenja laboratorijskih istraživanja mišljenja ne znači da gomila evidencije zaista ne pokazuje da ljudi imaju poteškoća s donošenjem optimalnih zaključaka i odluka te da zbog toga ne trpe štetne posljedice u stvarnome životu. Otkrivanje procesa u podlozi ovakve svakodnevne iracionalnosti, od donošenja pogrešnih odluka, prihvaćanja očigledno pogrešnih objašnjenja pojava u svijetu, usvajanja raznih zabluda o osnovnom funkcioniranju materijalnog i socijalnog svijeta, veliko je postignuće psihologije mišljenja. Možda intelektualna pozicija i jest pogrešna, u smislu da rasuđivanje nije garancija točnoga mišljenja, ali isto tako rasuđivanje nije nešto što je rezervirano samo za socijalni kontekst. Kada izvode nova vjerovanja, makar pogrešna, i kada donose odluke, makar one zbog kojih će kasnije požaliti, ljudi rasuđuju.

Nadalje, osim uvjerljive evidencije o negativnim posljedicama individualnog rasuđivanja, može se pronaći i indirektna evidencija koja bi ukazivala na pozitivne efekte rasuđivanja u individualnom kontekstu. Na primjer, istraživanja koja koriste metodu glasnog mišljenja i analizu verbalnih protokola prilikom rješavanja kompleksnih problema pokazuju da ispitanici stalno rasuđuju i da ih rasuđivanje vodi prema rješavanju problema (Newell i Simon 1972). U takvim istraživanjima pokazuje se da ispitanici razmišljaju o tome koji bi postupak mogao biti ispravan, zašto metoda koju su isprobali ne funkcionira, što će se dogoditi ako isprobaju određeni potez, i tome slično. Sloman i Fernbach (2017) ističu da je ljudsko mišljenje često

usmjereno na traženje kauzalnosti i da se prilikom pokušaja razumijevanja kako svijet funkcionira trudimo pronaći kauzalna objašnjenja. Mogli bismo dodati da, osim kauzalnih objašnjenja, često tražimo i druge razloge koji služe kao objašnjenja. Ovo su samo neki od primjera kod kojih se čini da rasuđivanje pridonosi uspješnosti individualnog mišljenja. Za određivanje uloge rasuđivanja u individualnom mišljenju treba gledati dalje od jednostavnih problema korištenih u klasičnoj psihologiji zaključivanja i odlučivanja – rješavanje kompleksnih problema, donošenje kompleksnih odluka, planiranje – to su kognitivne aktivnosti koje je teško zamisliti bez rasuđivanja.

U dosadašnjim istraživanjima i teorijama rasuđivanja i ljudskog mišljenja općenito, gotovo da je potpuno zanemarena potencijalna uloga rasuđivanja u učenju. Jednom generirani ili usvojeni razlozi kodiraju se u dugoročno pamćenje, postaju dio naših mentalnih reprezentacija – modela, propozicijskih mreža, kognitivnih shema. Kod procjene točnosti tvrdnji i kod donošenja odluka oslonit ćemo se na prethodno usvojene razloge. Iako ćemo često izmisliti razloge za neko vjerovanje ili odluku, što će biti racionalizacija, takvi razlozi, pohranjeni u pamćenju, mogu u kasnijim situacijama biti iskorišteni u produktivnom smislu kao premise za izvođenje novih zaključaka, bilo intuitivno ili reflektivno. U originalnom pristupu fenomenu racionalizacije vlastitih odluka Cushman (2019) ovakvo generiranje razloga naziva korisnom fikcijom. Radi se o fikciji jer pripisujemo razloge odlukama i postupcima koji su bazirani na ne-racionalnim procesima (kao što su instinkti i navike), a ta je fikcija korisna jer se izmišljeni razlozi mogu koristiti u kasnijem rasuđivanju. Pritom pripisujemo razlog odluci tako što konstruiramo razlog koji bi mogao imati racionalni subjekt kada bi donio istu tu odluku.

Socijalno-praktična funkcija odgovara središnjim tezama ITR-a, iako je sam ITR znatno kompleksniji i suptilniji. Rasuđivanje služi opravdavanju pred drugima i uvjeravanju drugih čim se ostvaraju različiti ciljevi povezani s kompleksnom ljudskom socijalnom motivacijom. Druga je mogućnost socijalno-epistemička funkcija. Rasuđivanje funkcionira optimalno u socijalnom argumentacijskom kontekstu kada sugovornici imaju različita vjerovanja i stavove, argumentacija je proces koji im omogućuje da poboljšaju svoje znanje, što je pozicija koju zastupa Smokrović (2015). Rasuđivanje je kognitivni mehanizam koji procesira razloge i argumente, a socijalni kontekst u kojem se javljaju tuđi razlozi i argumenti zasigurno aktivira takav mehanizam. U kojoj će mjeri takav mehanizam proizvoditi poželjne ishode ovisit će o kvaliteti inputa, odnosno o dostupnim tvrdnjama, razlozima i argumentima u samoj raspravi, kao i o nizu faktora koji sa samim rasuđivanjem nemaju puno veze, kao što su ciljevi sugovornika, njihove relacije i osobna povijest.



Ono što je ovdje ključno istaknuti je to da se rasuđivanje javlja u svim navedenim oblicima, bez obzira što više pravaca istraživanja u psihologiji pokazuje ograničenja i negativne učinke rasuđivanja u nekima od njih. Ovi negativni učinci ne javljaju se samo u individualnom kontekstu, kako ističe ITR, već se javljaju u socijalnom kontekstu. Možda polarizirane i homogene grupe nisu optimalni konteksti za ostvarivanje poželjnih učinaka rasuđivanja, ali činjenica je da takve grupe postoje i da upravo rasuđivanje bitno pridonosi fenomenima grupnoga mišljenja i polarizacije.

Čini se neuvjerljivim, ako je rasuđivanje evoluiralo da bi služilo socijalnim funkcijama, da bi bilo tako lako aktivirati ga u svim ostalim *neprirodnim* kontekstima. Za pretpostaviti je da rasuđivanje nije evoluiralo samo zbog socijalnih funkcija, već da je ono sastavni dio ljudskoga mišljenja bez obzira na kontekst. Suprotno pretpostavkama ITR-a, tako ustrojeno rasuđivanje uključuje se kod vrlo različitih kognitivnih i socijalnih aktivnosti, uključujući argumentaciju. Pri tome, argumentacija i rasuđivanje nisu jedno te isto jer argumentacija uključuje niz drugih procesa osim samoga rasuđivanja, na primjer, opažanje neverbalne komunikacije sugovornika, odabir strategija pristojnosti i druge specifičnosti jezičnog uobličavanja poruka.

### **Aktivacija i ishodi procesa rasuđivanja**

Što je onda rasuđivanje i što radi? Možemo se složiti s ITR-om, da je rasuđivanje kognitivni mehanizam koji generira i vrednuje razloge i argumente. Takav je mehanizam uključen u brojne oblike mišljenja i socijalne interakcije, kod jednostavnih situacija kada imamo konfliktne intuicije izazvane logičkim zagonetkama, kada racionaliziramo odluke koje su se pokazale neadekvatnima, kada korak po korak rješavamo kompleksne probleme, kada izmišljamo kauzalna objašnjenja, kada se želimo opravdati, kada želimo uvjeriti druge da smo u pravu.

Ovakva sveprisutnost rasuđivanja pokazuje da rasuđivanje nije evoluiralo samo za funkcioniranje u socijalnom kontekstu, već da se radi o mehanizmu s generalnijom ulogom. Rasuđivanje će se javiti kada pokušavamo objasniti ili opravdati neko vjerovanje i odluku i kada pokušavamo sagledati što slijedi iz onih informacija koje su nam dostupne u danom trenutku.

Što aktivira rasuđivanje? Svaki kognitivni mehanizam aktivira se kad se pojave određeni uvjeti, odnosno kada mu drugi kognitivni sustavi prosljede odgovarajuće informacije. Prema ITR-u, aktivator rasuđivanja je konflikt ideja, detekcija različitih mišljenja. Ako ideju da je aktivator sukob povežemo s koncepcijom pronalaženja sukoba između intuicija razvijenom u okviru DPT-a, možemo prepoznati različite situacije u kojima će se



pokrenuti rasuđivanje: postojanje sukobljenih intuicija, nepostojanje intuicija ili postojanje intuicija praćenih niskim osjećajem pouzdanosti, zastoj u pokušaju da se ostvari cilj (kao kod rješavanja kompleksnih problema), suočavanje s novim tvrdnjama koje teško možemo uklopiti u postojeća vjerovanja, uočavanje pojava i događaja za koje nemamo dostupna objašnjenja (uključujući i vlastite postupke, odluke i emocije), želja da ostvarimo određene ciljeve tijekom socijalne interakcije, itd. Razloge ćemo, u skladu s time, generirati i u individualnom i socijalnom kontekstu. Produktivna konstrukcija argumenta primarno će se javljati u komunikacijskom kontekstu, ali i u kompleksnim problemskim situacijama. Zajednička karakteristika svih ovih situacija i stanja je niska metakognitivna procjena adekvatnosti trenutnog sadržaja misli s obzirom na trenutne ciljeve. Neusklađenost između ciljnoga stanja (npr. rješenje problema, objašnjenje postupka, ostvarenje nekog interpersonalnog cilja) i mentalnog sadržaja koji bi trebao omogućiti ostvarenje cilja je aktivator rasuđivanja u najširem smislu.

Nadalje, jednom pokrenuti mehanizam rasuđivanja djelovat će pristrano. Nepregledni niz eksperimenata pokazuje da ljudi lako pronalaze razloge koji idu u prilog njihovim vjerovanjima i koja opravdaju njihove odluke, a znatno im je teže pronaći razloge koji bi bili protivni. Isto tako, uspješniji smo u kritiziranju tuđih razloga i argumenata. Zašto bi to bilo tako, zašto rasuđivanje nije objektivnije?

Bez obzira na to što su istraživanja često fokusirana na greške i pristranosti koje generira intuitivno mišljenje, intuicije su često ispravne. Svakodnevno iskustvo gotovo uvijek pruža potporu za intuitivna vjerovanja i odluke. To je i očekivano: ako su intuicije rezultat kognitivnih mehanizama koji se temelje na evolucijskim regularnostima, kao i na regularnostima u osobnim iskustvima, takvi bi mehanizmi u ogromnoj većini slučajeva trebali generirati koristan output. Um je evoluirao jer je omogućavao efikasno funkcioniranje u realnoj okolini i u realnom vremenu. Možemo pretpostaviti da je često bolje poduzeti bilo kakvu, ili u skladu s modelima ograničene racionalnosti, dovoljno dobru akciju, nego u nedogled razmatrati razloge za i protiv. Slično je i s epistemičkim aspektom mišljenja – bolje je usvojiti vjerovanje koje ima određeni stupanj korespondencije s realnošću, nego pažljivo graditi što točniji model vanjskoga svijeta. Jednostavni razlozi funkcioniraju i tipično su dovoljno dobri za neometan nastavak kognitivnih, komunikacijskih i drugih aktivnosti. Jednostavni razlozi i jednostavna objašnjenja ne moraju zadovoljavati stroge kriterije točnosti i normativne opravdanosti, ali mogu poslužiti kao korisni modeli određenog mikro segmenta materijalnog ili socijalnog svijeta. Dok takvi modeli funkcioniraju, nećemo tražiti protudokaze. Nadalje, pri traženju razloga

za svoja gledišta oslonit ćemo se na dostupnu evidenciju, u prvom redu na informacije pohranjene u dugoročnom pamćenju. Konstrukcija razloga, objašnjenja i argumenata oslanjat će se na te lako dostupne informacije, odnosno na one iste informacije na temelju kojih smo u startu formirali svoja gledišta. Takvo će se rasuđivanje, koje podupire vlastita gledišta, odvijati bez velikog mentalnog napora i bit će praćeno visokim stupnjem metakognitivne uvjerenosti u njegovu točnost. Prema tome, izvore pristranosti potvrđivanja treba tražiti u kombinaciji efikasnosti intuicija i jednostavnih mentalnih modela u stvarnome životu i lakoće rasuđivanja o tim istim intuicijama i modelima.

Prema tome, pristranost potvrđivanja, sklonost traženja evidencije koja ide u prilog vlastitim gledištima, ne mora biti korisna samo u argumentacijskom kontekstu. M&S ističu da je pristranost potvrđivanja, u argumentacijskom kontekstu kada branimo svoju poziciju, pozitivna karakteristika mišljenja. Ako smo pristrani prema svojim gledištima, veća je šansa da ćemo ih i obraniti. Kada su sugovornici s različitim gledištima pristrani prema svojim gledištima, onda na taj način dolazi do *podjele posla* – jasnije se ističu različite pozicije i argumenti za i protiv tih pozicija. Ovo može biti pozitivan ishod pogreške potvrđivanja, ali pogreška potvrđivanja funkcionirat će i u individualnom kontekstu, dok god naši razlozi i naša objašnjenja ne proizvode ishode koje ćemo ocijeniti štetnima.

Zašto smo, s druge strane, pažljiviji i bolji u vrednovanju tuđih razloga? ITR daje poprilično uvjerljiv odgovor na ovo pitanje: iako tijekom komunikacije možemo doći do važnih i korisnih informacija, isto tako možemo biti izmanipulirani ili zavedeni. U skladu s time, evolucija je dizajnirala specijalizirane mehanizme za podešavanje povjerenja. Takve mehanizme M&S nazivaju epistemička budnost. Kada nas sugovornici pokušavaju uvjeriti, procijenit ćemo trebamo li im vjerovati i je li poruka prihvatljiva, odnosno, trebamo odrediti kome vjerovati i što vjerovati. U takvim ćemo situacijama biti vrlo pažljivi i vrlo efikasni u procjeni tuđih argumenata, maksimalno se oslanjajući na rasuđivanje. Iako Mercier (2020) ističe da mehanizme epistemičke budnosti nije lako zavarati, treba primijetiti da se ocjena točnosti tvrdnji koje primamo tijekom komunikacije oslanja na heurističke procese (Lewandowsky i sur. 2012; Schwartz i sur. 2016). Prilikom procjene prihvatljivosti tvrdnje, kriteriji na koje se možemo osloniti jesu konsenzus (vjeruju li drugi), podrška (koliko je jaka evidencija koja ide u prilog), konzistencija (slaže li se s onim što vjerujem), koherentnost (uklapaju li se elementi u priču ili argument), kredibilitet (mogu li vjerovati izvoru). Ovakve procjene izvršavamo intuitivno, oslanjajući se na dostupne informacije. Na primjer, procijenit ćemo visoku koherentnost ako smo argument ili priču razumjeli bez zastoja i poteškoća, a konsenzus ćemo

procijeniti oslanjajući se na osjećaj poznatosti tvrdnje. Sav ovaj kognitivni rad tipično se neće pokretati izvan komunikacijskog konteksta kada se bavimo svojim uobičajenim aktivnostima. Nove tvrdnje i argumente tipično ćemo susretati u komunikacijskom kontekstu, bilo tijekom komunikacije uživo ili putem medija. Ako se radi o tvrdnjama za koje procjenjujemo da, na primjer, dolaze od nepouzdanog izvora, ili da evidencija za njih nije uvjerljiva, aktivirat će se mehanizmi epistemičke budnosti, uključujući rasuđivanje i vrednovanje argumenata. Epidemiologija zabluda, teorija zavjere, lažnih vijesti i sličnih epistemički sumnjivih vjerovanja ukazuje na to da ovi procesi nisu uvijek pretjerano efikasni.

## Zaključak

U ovome radu pokušao sam prikazati evidenciju i razloge za gledište koje je jednostavno sažeo Evans (2011) – rasuđivanje je za mišljenje, ne samo za argumentaciju. Prema tradicionalnom intelektualnom gledištu, razum, ili naša sposobnost rasuđivanja, je ona sposobnost koja bi trebala jamčiti formiranje točnih vjerovanja i donošenje optimalnih odluka. Iako ovo gledište nije moguće obraniti u kontekstu ogromne istraživačke evidencije koja pokazuje ograničenja toga istoga razuma, to ne mora značiti da je rasuđivanje evoluiralo samo zato jer ima socijalne funkcije i jer dobro funkcionira samo u socijalnom kontekstu. ITR je zasigurno u pravu kada pokazuje kako rasuđivanje dobro funkcionira u kontekstu argumentacije i kada ističe važnost socijalnog konteksta za funkcioniranje ljudskoga uma. Ignoriranje socijalnog konteksta jedan je od velikih nedostataka klasične psihologije mišljenja. Međutim, argumentacija nije jedina funkcija rasuđivanja, već rasuđivanje ima i druge kognitivne funkcije, od donošenja odluka do rješavanja kompleksnih problema, a te se funkcije ne manifestiraju samo u socijalnom kontekstu.

Konačno, Smokrović (2015) je analizirao jednu od situacija u kojoj će rasuđivanje bitno pridonositi efikasnom mišljenju, a to je specifičan argumentacijski kontekst u kojem su sugovornici iskreno motivirani radoznalošću, odnosno, željom da saznaju što je ispravno. Pri tome, identificirao je preduvjete za takvu specifičnu argumentacijsku situaciju i pokazao da u takvoj situaciji rasuđivanje zaista može pridonijeti proširivanju znanja. Tijekom argumentacije možemo objektivnije sagledati naša vjerovanja i usporediti ih s tuđim vjerovanjima, što u konačnici može rezultirati prihvaćanjem točnih i odbacivanjem netočnih vjerovanja. Međutim, pitanje je u kojoj mjeri ovakva idealizirana situacija tipična, a koliko se radi samo o idealizaciji. Stvarna komunikacija može uključivati elemente ovakve idealizirane argumentacije, ali, očigledno, uključuje i mnoge druge manje pozitivne aspekte.

## Literatura

- Cushman, F. (2019). Rationalization is rational. *Behavioral and Brain Sciences* 43, 1–59.
- De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science*, 17, 428-433.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7, 28-38.
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, 20, 169-187.
- Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7 (10), 454-459.
- Evans, J. St. B. T. (2011). Reasoning is for thinking, not just for arguing. *Behavioral and Brain Sciences*, 34 (2), 77–78.
- Evans, J. St. B. T. (2019). Reflections on reflection: The nature and function of type 2 processes in dual-process theories of reasoning. *Thinking & Reasoning*, 25 (4), 383-415.
- Evans, J. St. B. T., Handley, S. J., Harper, C. N. J., i Johnson-Laird, P. N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25 (6), 1495 – 1513.
- Evans, J. St. B. T. i Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8, 223-241.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19 (4): 25–42.
- Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. Viking Press.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814.
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, 13 (3), 106-131.
- Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.
- Mercier, H. i Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34 (2): 57–74.
- Mercier, H. i Sperber, D. (2021). Bounded reason in a social world. U: Viale R. (Ur.), *Routledge handbook of bounded rationality* (str. 257-267). Routledge.
- Mercier, H. i Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Newell, A. i Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice Hall.
- Nisbett, R. E. i T. Wilson. 1977. Telling more than we can know. *Psychological Review*, 84 (1), 231–259.

- Schwarz, N., Newman, E. i Leach, W. (2016). Making the truth stick and the myths fade: Lessons from cognitive psychology. *Behavioural Science and Policy*, 2, 85-95.
- Slovan, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119 (1), 3-22.
- Slovan, S. A. i Fernbach, P. *The knowledge illusion: Why we never think alone*. Riverhead books.
- Smokrović, N. (2015). Argumentation as a means for extending knowledge. *Croatian Journal of Philosophy*, 15 (44), 223-231.
- Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford University Press.
- Stanovich (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, 24 (4), 423-444.
- Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. u: J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195). Oxford, UK: Oxford University Press.
- Thompson, V. A., Prowse Turner, J. A. i Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, 63, 107-140.
- Tversky, A. i Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 91, 293-315.
- Williamson, T. 2000. *Knowledge and its limits*. Oxford: Oxford University Press.

GABRIELA BAŠIĆ HANŽEK

## Teorija i klasifikacija pogreški u argumentaciji: stvarne i manje bitne razlike između dvaju pristupa

**Sažetak:** U ovome se tekstu promatraju razlike u negativnim teorijama argumentacije (teorijama pogreški – *fallacies*) koju nude pragmatijalektika, s jedne strane, i epistemološki pristup argumentaciji, s druge strane. Argumentira se u korist tvrdnje da je razlika između dvaju pristupa vidljivija u pozitivnim teorijama, a da su razlike u negativnoj dimenziji manje uočljive, odnosno prisutne u objašnjenju zašto je nešto loš argument.

**Gljučne riječi:** *fallacies*, pragmatijalektika, epistemološki pristup argumentaciji, pozitivna i negativna teorija argumentacije.

### Uvod

Da je klasifikacija, teorijska razrada i usustavljanje logičkih pogreški, odnosno pogreški u zaključivanju i argumentiranju (raspravljanju)<sup>3</sup> jedan od glavnih zadataka sada gotovo već sasvim samostalne akademske discipline koju nazivamo teorija argumentacije, dobro je poznato svima koji dijele taj istraživački interes. Osim specijaliziranoga rada na bilo normativnome, bilo deskriptivnome istraživanju pojedinih pogreški (dakako, i kombiniranim pristupom), literatura se iz ovoga područja popratno obogaćuje novim pokušajima da se takva pojedinačna, specijalizirana istraživanja obuhvate, odnosno teorijski ujedine i podvedu pod općenitiji okvir istraživanja argumentacije.<sup>4</sup>

Preokupacija problematikom klasifikacije i definicije pogreški u zaključivanju i argumentiranju, prema mišljenju nekih autora, funkcionira ponekad – reći će, i to pogrešno – kao zasebna disciplina unutar teorije

---

<sup>3</sup> U ovom se tekstu opredjeljujem za novije i sadržaju primjerenije razlikovanje pogreški u zaključivanju od pogreški u argumentiranju (raspravljanju) kao prijevode engleskoga latinizma *fallacies* na hrvatski jezik (usp. D. Sekulić 2016 za korisno razjašnjenje terminoloških neujednačenosti u domaćim, ali i međunarodnim uporabama i prijevodima).

<sup>4</sup> Kratak povijesni pregled i detaljnija analiza novijega stanja u istraživanju pogreški u zaključivanju (premda se one promatraju manje-više isključivo u argumentativnom, odnosno socijalnom kontekstu) na hrvatskome jeziku može se naći u (Kišiček 2010). Valja pri tome napomenuti kako je termin *noviji* ovdje ipak relativan s obzirom na količinu produkcije tekstova od datuma objavljivanja citiranoga članka.

argumentacije. Točnije, usvojimo li terminološku distinkciju pozitivne teorije argumentacije – teorije koja nas opskrbljuje odgovorima što *jest* dobar argument i koje su norme *dobre* argumentacije – naspram negativne teorije argumentacije – teorije koja nam pruža odgovore što *nije* dobar argument, odnosno u čemu se sastoji *loša* argumentacija – moguće je primijetiti da brojni radovi, pa čak i čitavi pristupi u istraživanju pogreški u zaključivanju i argumentiranju, ne uspostavljaju jasnu vezu između navedenih dvaju aspekata cjelovite teorije argumenta i argumentacije (pozitivne i negativne dimenzije).

Posljedica te nejasnoće, odnosno gubitka veze između dvaju aspekata, očituje se najčešće kao izolirano proučavanje pojedinih tipova pogreški, tj. relativno arbitrarno određivanje kriterija što neko zaključivanje, odnosno potez u raspravi, čini instancom neke pogreške. Ili, u drugoj varijanti, ako se prihvati topički, neutralan stav spram pogreški kao čestih obrazaca (shema) mišljenja i uvjeravanja drugih (npr. Walton, Reed i Macagno 2008), u kojim se sve kontekstima, tj. pod kojim sve uvjetima (što opet povlači za sobom neizbježnu i neželjenu problematičnost njihovih tipova i specificiranja) jedna instanca neutralne sheme smatra dobrom (ispravnom, valjanom), a druga lošom, odnosno pogreškom. Idealno zamišljena teorija argumentacije pokazuje svoja dva aspekta na sljedeći način: primat ima pozitivna teorija, tj. prvenstveni je posao teoretičara argumentacije osigurati definiciju (dobrog) argumenta i argumentacije (kao socijalne aktivnosti iznošenja i razmjene argumenata s ciljem uvjeravanja drugih), a do negativne teorije argumentacije dolazimo posljedično, odnosno do odgovora što je loš argument i loša argumentacija doći ćemo negacijom jedne ili više stavki definicija i tvrdnji iz pozitivne teorije. Teorija i klasifikacija pogreški u argumentaciji trebala bi zapravo tako biti negativni aspekt teorije argumentacije, a ne neovisan pothvat.

U ovome tekstu neću se zbog ograničenog prostora ni usuditi baviti tako opširnom i zahtjevnom temom kao što je klasifikacija pogreški u argumentaciji općenito. Ono što ću pokušati jest smjestiti taj problem u raspravu između pragmadijalektičkoga pristupa argumentaciji s jedne strane i epistemičkih/epistemoloških teorija s druge strane. Preciznije rečeno, nastavljajući se na raniju raspravu između ova dva pristupa, prikazat ću kako se jedan i drugi nose s ovako zamišljenim odnosom pozitivne i negativne teorije argumentacije, odnosno koliko se njihove teorije pogreški u argumentaciji međusobno *stvarno* razlikuju i time daju prednost jednom ili drugom pristupu. Na početku ću vrlo sažeto predstaviti osnovne ideje svakog pristupa (uz nužno nepravedne generalizacije nad različitostima pojedinih predstavnika obiju strana), zatim preciznije njihove teorije pogreški u argumentaciji i zaključno iznijeti sud o njihovu međusobnom odnosu.



## 1. Pragmadijalektika

Uspoređujući je s ostalim uvriježenim i dobro zastupljenim pristupima argumentaciji, koje ugrubo možemo podijeliti u retoričke (nefilozofske) i epistemičke/epistemološke (filozofske), pragmadijalektika se smješta negdje po sredini toga kontinuuma i možda baš zato obično smatra najraširenijom argumentacijskom teorijom (usp. Lumer 2010: 41-42). Za razliku od eksplicitno filozofskoga pristupa argumentaciji, koji se manje-više bavi prvenstveno epistemičkom (spoznajnom) dimenzijom argumentacije<sup>5</sup>, bilo da je riječ o autorima u središnjim filozofskim strujama (Alvin Goldman, Richard Feldman i ostali u sklopu socijalne epistemologije), bilo unutar teorije argumentacije kao drukčijega pristupa argumentaciji od onog logičkog i epistemološkog i koji je, kao takav barem donedavno, bio na samim marginama interesa (Harvey Siegel i John Biro, Christopher Lumer, Scott Aikin i dr.), pragmadijalektika uvijek promatra argumentaciju kao eksplicitan ili implicitan dijalog između dviju ili više osoba (preciznije rečeno, uloga). Dijalogom se zajednički pokušava razriješiti razlika u mišljenju koja se ne ograničava na teorijsku argumentaciju. Osim toga, istaknula bih i kako su upravo predstavnici pragmadijalektičkoga pristupa, surađujući sa srodnim pristupima i pokretima (Douglas Walton i cjeloukupna kanadsko-sjevernoamerička škola neformalne logike) vjerojatno najzaslužniji za zasebno institucionalno postojanje teorije argumentacije.

Pragmadijalektički<sup>6</sup> pristup argumentaciji, kako se vidi iz samoga njezi-

<sup>5</sup> Usp. sljedeće: [...] *argumentativni proces prirodno vodi sudionike proširenju njihova znanja budući da a) biološka je funkcija zaključivanja takva da optimalno funkcionira u argumentaciji i b) sami je argumentativni proces tako strukturiran da optimalno podupire radoznalost sudionika da spoznaju je li tvrdnja P (predmet rasprave) istinita.* (Smokrović 2015)

<sup>6</sup> Ovaj je pristup prvotno razvijen na Odjelu za komunikacijske studije Sveučilišta u Amsterdamu još sedamdesetih godina prošloga stoljeća samostalnim i zajedničkim djelovanjem Fransa Hendrika van Eemeren (1946.) i p. Roba Grootendorsta (1944. - 2000.). Prvi rezultati te dugogodišnje suradnje, koji predstavljaju početak pragmadijalektičke teorije argumentacije, objavljeni su prvo na nizozemskome, a zatim i međunarodno (na engleskome jeziku) u knjizi *Speech Acts in Argumentative Discussions* iz 1984. (usp. prikaz razvoja u Van Eemeren, Garssen, Krabbe et al. 2014: 517-518). Nakon Grootendorstove smrti Van Eemeren nastavlja djelovati (i dan danas) prije svega surađujući s Peterom Houtlosserom (1956.-2008.), Bartom Garssenom i drugim suradnicima. Početak Van Eemerenove i Houtlosserove suradnje ujedno se u literaturi smatra i početkom nove, tzv. proširene (*extended*) varijante ili faze pragmadijalektičke teorije (Zenker 2007b) u odnosu na standardnu, koju se također promatra i kao ustupak i približavanje retoričkom pristupu argumentaciji (no ne nužno i izjednačivanje s njim). Glavna promjena, odnosno proširenje (ne revizija), prethodnih teorijskih stavova u ovoj fazi jest uvođenje pojma strateškoga manevriranja (*strategic manoeuvring*). Strateško je manevriranje krovni pojam kojim su Van Eemeren i Houtlosser nastojali



nog naziva, temelji se na integraciji deskriptivnoga proučavanja argumentativnoga diskursa u komunikaciji i interakciji među sudionicima rasprave (*pragma-*) s jedne strane te normativnoga proučavanja argumentacije u reguliranim raspravama s druge strane (*-dijalektika*). Općenito govoreći, njegova šira filozofska teorijska polazišta moguće je pronaći u formalnoj dijalektici Else Barth i Erika Krabbea (*From Axiom to Dialogue*, 1982.), zatim kritičkome racionalizmu Karla Poppera i Hansa Alberta, teoriji govornih činova Johna Austina i Johna Searlea te teoriji racionalnih govornih razmjena Paula Gricea.<sup>7</sup> Uobičajeno je pragmadijalektiku klasificirati u skupinu konsenzualističkih argumentacijskih teorija (usp. Lumer 2005a), među kojima se opet pravi razlika između teorija kvalificiranog (npr. Habermasova teorija komunikativnoga djelovanja u kojoj se također problem rješava u idealno zamišljenoj (nehijerarhijskoj) govornoj situaciji, i slični pristupi) i teorija nekvalificiranoga konsenzusa, a to znači da kao primarni cilj ili standardnu funkciju argumentacije postavlja (nekvalificirani) konsenzus, tj. usuglašavanje sudionika. Iako bi bilo preciznije reći da je termin *standardna funkcija argumentacije* karakterističan za epistemičke/epistemološke teorije argumentacije, stoga je ovdje prikladnije govoriti o glavnome cilju ili svrsi argumentacije.<sup>8</sup> Ovakvu klasifikaciju vrlo je lako opravdati već pregledom pragmadijalektičke definicije argumentacije (Van Eemeren i Grootendorst 2004: 1) u kojoj se bez problema uočava konsenzualistički moment (pri čemu barem zasada taj predikat ostavljam lišenim normativnoga naboja prisutnog u brojnim kritičkim opaskama).

---

obuhvatiti sve varijacije načina i postupaka kojima se sudionici argumentacije služe u nastojanju da pojačaju svoj argumentativni uspjeh, odnosno budu uvjerljiviji i učinkovitiji u različitim tipovima rasprava.

<sup>7</sup> Kao četiri metateorijska polazišta ove integracije pragmatike (teorije govornih činova) i dijalektičkih sustava (doduše, s naglašeno neformalnim pristupom) u jedinstven teorijski pristup redovito se navode: (i) funkcionalizacija (na temelju teorije govornih činova identificiraju se stajališta o kojima je riječ u raspravi, kao i komunikacijske i interakcijske funkcije govornih činova kojima su iskazana), (ii) socijalizacija (argumentacija se uvijek promatra kao eksplicitan ili implicitan dijalog između dviju ili više osoba, preciznije rečeno, uloga), (iii) eksternalizacija (rekonstrukciji se usmjerava na obveze sudionika ostvarene govornim činovima u raspravi, a ne mentalna stanja, bilo da se radi o obvezama koje su sami sudionici izrazili, bilo o onima koje je moguće na temelju toga rekonstruirati kao opće prihvaćene i razumljive) i (iv) dijalektifikacija (kao i u drugim dijalektičkim sustavima, skup pravila regulira ponašanje sudionika u raspravi, onemogućujući pogreške u argumentaciji).

<sup>8</sup> Naravno, iz toga ne slijedi da prema pragmadijalektici argumentacija ne može imati i neke druge, sporedne funkcije osim razumnog usuglašavanja sudionika, no svaka se od njih osniva na onoj glavnoj (Van Eemeren i Grootendorst 2004: 15). Da bi opisali odnos sporednih prema glavnoj funkciji, Van Eemeren i Grootendorst koriste Searleov termin parazitiranja (Van Eemeren i Grootendorst 2004).

Ono što nas ovdje najviše zanima jest pragmadijalektička negativna teorija argumentacije, odnosno kako se ona odnosi prema uobičajenim popisima pogreški te koliko se uspješno veže s pozitivnom teorijom dobre argumentacije i argumenta. Pritom valja spomenuti da se u odluci o polazištu za normiranje argumentacije pragmadijalektičari kreću unutar pojmovnih distinkcija koje je uveo Steven Toulmin u *Knowing and Acting* (1976.), odnosno njegovu distinkciju geometrijske i antropološke perspektive na racionalnost obogaćuju novom kategorijom kritičke perspektive. Kao razloge za takav terminološki odabir i samu poziciju koju odabiru u raspravi između dviju navedenih krajnosti Van Eemeren i Grootendorst citiraju Popperov kritički racionalizam s epistemološkom tezom o falibilizmu kao središnjom, odnosno postavljaju kriticizam (kritičko testiranje) kao središnji epistemički cilj.<sup>9</sup>

Posljedično se ovaj odabir očituje kao izrazito dijaloško razumijevanje racionalnosti, odnosno u kritičkoj je perspektivi žarište postavljeno na raspravu kao *mjesto* u kojemu se očituje racionalnost. Zbog toga se kao središnji problem za teoriju argumentacije postavljaju norme racionalne argumentacije kao aktivnosti i rasprave nauštrb normi valjanog argumenta (apstraktnog entiteta), odnosno njihovo se normiranje prepušta (bez izražene brige oko povezivanja) drugim disciplinama kao što je logika i dr. U te su svrhe i konstruirani Procedura za kritičku raspravu i Kodeks ponašanja razumnih sudionika (tzv. Deset zapovijedi), skupovi pravila kojih se razumni sudionici trebaju pridržavati kako bi izbjegli pogreške u argumentaciji, a u jačoj varijanti i razriješili razliku u mišljenju (problemska valjanost). Stoga je mjera racionalnosti, točnije razumnosti (gdje su pojmom razumnosti pragmadijalektičari obuhvatili uporabu *ratia* na prikladan način) u modelu kritičke rasprave intersubjektivnog i proceduralnog karaktera.<sup>10</sup>

Usporedbom s ostalim sustavnim teorijama pogreški posthamblinovskoga tipa<sup>11</sup> – dakle, pristupima koji smjeraju razotkrivanju teorije po-

<sup>9</sup> Upravo je postavljanje kritičkoga testiranja kao središnjeg epistemičkoga cilja jedna od glavnih točki prijepora s epistemičkim/epistemološkim teorijama argumentacije koje, očekivano, biraju klasičnije ciljeve poput znanja ili opravdanog vjerovanja bilo u monološkoj, bilo u dijaloškoj dimenziji argumentacije.

<sup>10</sup> Pregled teorija racionalnosti, odnosno klasifikaciju pogleda na ljudsku racionalnost s obzirom na standard racionalnosti i mogućnost poboljšanja ljudskih performansi vidi u: Smokrović, 2004.

<sup>11</sup> Prema standardnoj definiciji ili tretmanu pogreški u zaključivanju i argumentiranju, gdje je *standardni tretman* pojam koji dugujemo australskom filozofu Charlesu Hamblinu (Hamblin 1970.), pogreškom se smatrao, odnosno na popis pogreški uvrštavao se svaki tip argumenta za koji se činilo da je valjan, no on to nije. Novi interes za neformalnu logiku, koji se javio polovicom prošloga stoljeća, među ostalim je i tu defi-

greški koja stoji u pozadini naslijeđenih kataloga, odnosno popisa pogreški poznatih još od Aristotela naovamo, a ne tek njihovu popisivanju – u pragmadijalektičkome se pristupu lako uočava prije svega dijalektičko polazište u definiciji i tretmanu pogreški (zajedničko s npr. Hintikkinom analizom po kojoj je pogreške, odnosno njihovo mjesto na tradicionalnim popisima moguće razumjeti samo kao poteze u igrama pitanja i odgovora, usp. Hintikka 1989) s jedne strane te također, barem početno, odbijanje monističke pretpostavke o jedinstvenoj definiciji svih pogreški s druge strane.<sup>12</sup> Ova prva karakteristika (dijalektičko polazište) oblikuje pragmadijalektičku teoriju pogreški u pravcu teorije pogrešnih poteza u argumentaciji kao raspravi.

Druga pak karakteristika, za razliku od standardnoga tretmana pogreški o kojemu je govorio Hamblin, znači da se u pragmadijalektici polazi od toga da ne krše svi nevaljani potezi jedno te isto pravilo dobre argumentacije. Točnije, da su nevaljani argumenti koji se čine valjanima, pri čemu se i dalje naglašava varljivi karakter pogreški iz standardnoga tretmana, odnosno njihova uvjerljivost.<sup>13</sup> Zadržavajući se na varljivoj naravi nevaljanih argumentacijskih poteza, pragmadijalektika se bavi upravo istaknutim dijelom standardnoga tretmana pogreški (čini se valjanim), no ovaj put primijenjenim na argumentacijske poteze, tj. poteze u raspravi, a ne argumente kao apstraktne predmete.

Tako se obično navodi nekoliko razloga koji su zaslužni za uspjeh nevaljanih poteza (njihovu uvjerljivost i slabo uočavanje), odnosno kršenja pravila kritičke rasprave (navodim prema Van Eemeren 2010: 196 – 200, Van Eemeren, Garssen, Krabbe et al. 2014: 565):

- i. Opća je tendencija sudionika u raspravi ona da izbjegnju prigovor druge strane zbog nerazumnosti, odnosno cilj je sakriti odstupanje

---

niciju doveo u pitanje, prije svega zbog nemogućnosti da se njome objasni prisutnost nezanemariva broja pogreški na popisu koje ne zadovoljavaju navedenu definiciju (npr. *petitio principii*).

<sup>12</sup> Sve pokušaje teorije i klasifikacije pogreški koji dijele tvrdnju (bilo kao polazišnu, bilo kao zaključnu, na temelju pojedinačnih istraživanja) da nije moguća jedna jedinstvena definicija svih pogreški, nazivam ovdje, u skladu s uvriježenom terminologijom, pluralističkima (usp. npr. radovi Johna Woodsa i Douglasa Waltona). Nadalje, međusobno se razlikuju u pogledu toga je li njihov popis otvoren (stav koji zastupaju i pragmadijalektičari) ili je pak konačan i zatvoren. Odgovor na prethodno pitanje obično je motiviran time pretpostavlja li pristup o kojemu je riječ određenu danost pogreški, odnosno njihovu neovisnost u odnosu na teorijski tretman ili ih pak približava statusu teorijskih entiteta (usp. Woods 1988 i Woods 1994)

<sup>13</sup> Uvjerljivost, odnosno varljivost pogreški u argumentiranju Dragana Sekulić objašnjava na sljedeći način: loši argumenti nalikuju dobrima jer ako su preloši, malo je vjerovatno da bi netko počinio takvu pogrešku, odnosno bio njome zavaran (Sekulić 2016).

- vlastitih argumentacijskih poteza od normi kritičke rasprave te ih prikazati kao usklađene s istima.
- ii. Valjani i nevaljani potezi u strateškome manevriranju istorodni su (isti tipovi govornih činova) i naizgled isti.
  - iii. Pojedine inačice strateškoga manevriranja predstavljaju kontinuum koji se proteže od jasno valjanih do jasno nevaljanih poteza, čime se javlja poteškoća određivanja statusa pojedinih slučajeva između dviju krajnosti.
  - iv. Utjecaj poznavanja, tj. vladanja tipom komunikacijske aktivnosti na prepoznavanje pogreški očituje se u ponešto različitim standardima strateškoga manevriranja od jednoga do drugoga tipa (npr. korištenje argumenta iz autoriteta drukčije je u sudskome postupku u odnosu na korištenje istoga tipa argumenta u znanstvenome tekstu), zbog čega neupućenome sudioniku rasprave, odnosno svakome tko evaluira raspravu, mogu promaknuti pojedini nevaljani potezi.

Sve se pogreške u pragmatijalektici mogu obuhvatiti podosta općenitom definicijom kao nerazložni potezi u raspravi kojima se ometa njezino razrješenje. Općenito govoreći, svi se potezi sudionika koji su u skladu s pravilima za vođenje razložne rasprave: Procedure za kritičku raspravu (*Critical Discussion Procedure*) i Deset zapovijedi (*Ten Commandments*) klasificiraju kao ispravni (valjani), a pogreške (*fallacies*) jesu različiti prekršaji pojedinih pravila dvaju skupova.<sup>14</sup> Procedura i Kodeks (Zapovijedi) pred-

---

<sup>14</sup> Klasična formulacija Procedure za kritičku raspravu (*Critical Discussion Procedure*) sadrži petnaest pravila koja obvezuju sve sudionike rasprave te je sadržajnija od svoje skraćene, i manje tehničkim rječnikom opterećene varijante, Kodeksa ponašanja razumnih sudionika (*Code of Conduct for Reasonable Discussants*). Kodeks pak obuhvaća deset jezgrovitijih pravila, u literaturi poznatih i kao Deset zapovijedi (*Ten Commandments*). Glavni je povod takvome nazivu iskazivanje zapovijedi u obliku zabrana u skraćenoj varijanti. Oba skupa pravila predstavljaju općenite norme pouzdanosti argumentativne prakse, tj. odnose se na svaku kritičku raspravu, neovisno o njezinom institucionalnome kontekstu, i kao takva pokrivaju sva pravila relevantna za razrješenje razlike u mišljenju: *Norme su za kritičku raspravu univerzalne u smislu da konstituiraju ideal kritičke rasprave koji je primjenjiv na argumentaciju u svakom okruženju.* (Garssen i Van Laar 2010: 127). Premda se isprva čini da su Deset zapovijedi, kako je prethodno rečeno, tek skraćena inačica Procedure za kritičku raspravu (što je tvrdnja koja se daje mjestimično iščitati iz pragmatijalektičke literature), odnos dvaju skupova pravila za pravo nije sasvim jasan. Obično pragmatijalektičari tvrde da je postupanje u skladu s Deset zapovijedi instrumentalno pri zadovoljenju Procedure (što bi vjerojatno trebalo biti ekvivalentno tvrdnji da je izvršenje Deset zapovijedi nužan uvjet ispunjavanja Procedure), no sama pragmatijalektička literatura nije eksplicitna ni konzistentna u odgovorima na ovo i daljnja pitanja. Za raspravu vidi Bašić Hanžek 2020, Zenker 2007a i Zenker 2007b za moguća tumačenja odnosa Procedure i Kodeksa (Deset zapovijedi).

stavljaju sažetak pragmadijalektičkoga modela kritičke rasprave, odnosno u sebi sadržavaju teoriju racionalnosti koja uspostavlja standard dobrog argumenta i argumentacije. Rasprava o snazi odnosno slabosti pragmadijalektičke teorije racionalnosti dovoljno je velika sama za sebe i ovdje joj može biti posvećeno minimalno prostora, odnosno u onoj mjeri u kojoj utječe na razlike spram teorije pogreški konkurirajućega epistemološkoga pristupa.<sup>15</sup>

Kako je kao cilj argumentacije uspostavljeno razložno (razumno) usuglašavanje sudionika, tako se svi argumenti, odnosno potezi u raspravi koji doprinose razumnome razrješenju razlike u mišljenju klasificiraju kao valjani dok se pak svi potezi koji to razrješenje onemogućuju ili ometaju smatraju pogreškama (*fallacies*) ili nevaljanima, neovisno o tome koja ih je strana počinila (protagonist, antagonist ili oboje) i u kojoj fazi rasprave.

Stoga je uporaba pojma pogreške u pragmadijalektici sustavno povezana s pravilima za kritičku raspravu. Pogreška se definira kao potez u raspravi koji krši na neki način pravilo kritičke rasprave u pojedinoj fazi rasprave. (Van Eemeren, Garssen, Krabbe et al. 2014: 545)

Što se tiče samoga popisa, odnosno broja (tipova) pogreški koje navodi, valja primijetiti da pragmadijalektički popis pogreški ipak nadilazi uobičajene popise ili kataloge pogreški, dodajući postojećima nove pogreške nastale zaobilaženjem kojeg od pravila Procedure. Tako je više tradicionalnih pogreški, odnosno njihovih varijanti klasificirano kao prekršaji različitih pravila, tj. različite pogreške, primjerice, dvije inačice argumenta *ad populum*, od kojih se jedna odnosi na nepoštivanje pragmadijalektičkoga pravila relevantnosti, a druga na pravilo argumentativnih shema. Isto tako, u rekonstrukciji argumentativne građe javljaju se i suprotni slučajevi – da se ono što je u tradicionalnome popisu svrstano kao dvije različite pogreške u pragmadijalektičkoj analizi pokazuje kao jedna te ista vrsta ukoliko je slučaj da se radi o nepoštivanju jednog te istog pravila Procedure za kritičku raspravu (Van Eemeren 2001).

Što se tiče odgovora na pitanje je li popis pogreški konačan, odnosno zatvoren, ono što se može iščitati iz pragmadijalektičke literature (pogotovo starije, odnosno standardne faze) jest da je taj popis konačan te da su Procedura i Kodeks, odnosno pridržavanje sudionika uz te skupove pravila nužni (a u snažnijoj varijanti i dovoljni) uvjeti za eliminaciju pogreški.

---

<sup>15</sup> Bašić Hanžek (2020) raspravlja upravo o mogućnosti dobronamjernog tumačenja pragmadijalektičke pozitivne teorije argumentacije, odnosno nadopune samog za sebe uzeto zaista problematičnog utemeljenja teorije argumentacije u za filozofske standarde nedostatno ozbiljno shvaćenoj raspravi o racionalnosti, razumnosti (*reasonableness*) itd.

## 2. Epistemički/epistemološki pristup argumentaciji<sup>16</sup>

Svi epistemološki orijentirani teoretičari argumentacije<sup>17</sup> (neovisno o međusobnim različitostima subjektivističke, objektivističke i drugih varijanti), dijele osnovnu tvrdnju o standardnoj funkciji, svrsi ili *outputu* argumentacije kao znanju ili opravdanome vjerovanju u epistemološkome smislu – dakako, uz dopuštenje drugih, nestandardnih funkcija. Kao osnovni zadatak teorije argumentacije – smještajući je time u socijalnu epistemologiju – postavljaju određenje (prešutnih) standarda dobre argumentacije u argumentativnoj praksi te njihovo poboljšanje i eventualnu nadopunu novima u svrhu osiguranja epistemološki poželjnih ciljeva. Jedna od najvažnijih razlika u odnosu na konkurirajući pragmatodijalektički pristup jest različito postavljanje prvenstva između pojmova argumenta i argumentacije; preciznije rečeno, epistemološke su teorije karakteristične po postavljanju argumenta kao središnjega, a argumentacije kao izvedenoga pojma.

Druga terminološka distinkcija koja uočljivo izostaje u pragmatodijalektici, a na kojoj inzistiraju epistemološki teoretičari argumentacije, jest argument kao apstraktan entitet sastavljen od opet apstraktnih jedinica (propozicija s dvjema različitim funkcijama; premise ili konkluzija) te s druge strane, argument kao složen govorni čin monolektičnoga ili dijalektičnoga tipa.<sup>18</sup> Upravo zbog interesa za argument kao apstraktni entitet

<sup>16</sup> Odsad pa nadalje koristim jedinstveni termin *epistemološke teorije argumentacije*, nauštrb razlikovanja epistemoloških teorija argumentacije u slabome i jakome smislu (usp. Lumer 2005a: 4–5) za sve normative pristupe argumentaciji, za razliku od onih koji tek naglašavaju važnost slušateljeva epistemičkoga stanja za argumentativni uspjeh.

<sup>17</sup> Pregledan prikaz s detaljnom tipologijom različitih pozicija vidi u: Lumer 2005a. Članak donosi korisnu mapu epistemološki orijentiranih teoretičara argumentacije s klasifikacijom koja razdjeljuje sve takve autore na temelju kriterija dobre argumentacije koju pojedini autori (pozicije) prihvaćaju. Pritom Lumer pronalazi u njihovim stajalištima četiri skupine kriterija dobre argumentacije: gnostički ili slabi epistemički kriteriji (adresat argumentacije opravdano vjeruje u razloge (premise) argumenta, nema nikakvih daljnjih informacija koje bi pobile taj argument i razložno je da prijeđe s vjerovanja u razloge (premise) za vjerovanje u tezu argumenta), plauzibilistički (strukturnalni kriteriji koji se odnosi samo na argument i bazu podataka, bez reference na adresata argumenta; dobar je argument onaj koji je načelno u stanju ispuniti standardnu funkciju argumentacije), prozbatički (pravila upotrebe argumenta u određenoj situaciji koja jamče da adresat argumentacije može samostalno spoznati prihvatljivost teze argumenta) i responzibilistički (reguliraju sasvim drugačiji način standardnoga funkcioniranja argumentacije gdje do opravdanoga vjerovanja adresat dolazi prepoznavanjem autoriteta osobe koja iznosi argument, a ne samostalnim izvođenjem konkluzije (teze) iz premisa (razloga)).

<sup>18</sup> Važno je naglasiti da jedino argument kao govorni čin ima navedenu sposobnost pobuđivanja znanja ili opravdanoga vjerovanja u konkluziju posredstvom opravdanoga vjerovanja u premise. Jedino kada se koristi u određenoj situaciji ili kontekstu (*argument in use*), ima smisla reći za argument da može ili ne može to učiniti.



epistemološki se pristup argumentaciji ne iscrpljuje u proučavanju argumentacije kao rasprave (društvene aktivnosti korištenja argumenta), već pokriva i šire područje zaključivanja. Glavni razlog za ovu odluku jest briga oko normativnog utemeljenja teorije argumentacije, koja se jedino može (i mora) temeljiti u široko shvaćenoj epistemologiji (uključuje deduktivnu i induktivnu logiku, odnosno probabilistički račun), a koje, smatraju predstavnici ovoga pristupa, u pragmatijalektici izostaje.<sup>19</sup> Idealno zamišljen odnos pozitivne i negativne teorije argumentacije o kojemu je bilo riječi na početku ovoga teksta također je zajedničko polazište različitih epistemoloških teoretičara argumentacije, čime zapravo dobivamo nacrt teorije pogreški:

Bilo koja teorija pogreški koja se oslanja na neepistemičke pojmove, bili ovi retorički, dijalektički, psihološki ili koji god drugi, rezultat će prekidanjem veze između pogreški i argumentacije shvaćene kao po svojoj biti sredstva za racionalnu promjenu vjerovanja, ili pak zadržavajući vezu s argumentacijom, pretvaranjem same takve promjene vjerovanja u nešto neracionalno. (Siegel i Biro 1997: 289)

Različiti epistemološki autori, dakako, nude različite teorije pogreški, no svi se slažu u tome da se ona izvodi iz pozitivne teorije argumenta i argumentacije, odnosno normativno se utemeljuje u logici i epistemologiji. Pritom valja napomenuti da je većina tih autora sklonija standardnome tretmanu pogreški negoli pragmatijalektičari, odnosno ponekad i izričito prihvaćaju takvu definiciju pogreške. Monistička perspektiva o jedinstvu svih pogreški funkcionira na općenitoj razini da su pogrešni argumenti (ne argumentativni potezi) loši jer se radi o epistemičkim pogreškama, odnosno neuspjehu prijenosa prihvatljivosti koje bi premise trebale pružiti konkluziji<sup>20</sup>:

Ključno je pitanje pružaju li premise argumenta o kojemu je riječ jamstvo za konkluziju. Ono u čemu ustraje epistemički tretman jest da je njihov neuspjeh da to učine i nužan i dovoljan uvjet neispravnosti. (Siegel i Biro 1997: 280).

---

<sup>19</sup> Općenita kritika na račun pragmatijalektike argumentacije jest što konsenzualistički određena svrha argumentacije ima epistemološki nepoželjne posljedice, poput npr. potiskivanja, skrivanja ili zanemarivanja dokazne građe kako se ne bi narušio konsenzus i dr.

<sup>20</sup> Najsnažnija varijanta monističke perspektive nalazi se u Sigelovoj i Biroovoj objektivističkoj teoriji argumentacije u kojoj se eksplicitno tvrdi da jedino pretpostavka o pogreškama kao prirodnoj vrsti može dati zadovoljavajuće objašnjenje ispravnosti, odnosno neispravnosti pojedine instance nekog argumenta.

Ovdje ću posebno razmotriti vjerojatno najsustavniju teoriju pogreški unutar epistemološkoga pristupa – onu koju je u svojoj praktičnoj, instrumentalističkoj teoriji argumentacije, razvio Christopher Lumer i koja je izrazito redukcionističkoga karaktera (*logički* svodi sve tipove pogreški na određen broj osnovnih). Cilj je Lumerova redukcionističkoga programa jedinstvena taksonomija pogreški utemeljena na objašnjenju kršenja pojedinih uvjeta kriterija dobrih argumenata i njihove uporabe za svaki tip pogreške (Lumer 2000, Lumer i Dove 2011). Na taj se način uspostavlja jasna veza s pozitivnom teorijom, odnosno izbjegava *ad hoc* objašnjenje zašto je neki argument dobar ili loš.

Lumerova varijanta epistemološkoga pristupa karakteristična je po razlikovanju (argumentativne) valjanosti (strukturni kriteriji kvalitete argumenta) i situacijske primjerenosti – dvaju skupova kriterija potrebnih za dobru argumentaciju, tj. ispunjavanje međusobno neovisnih uvjeta prihvatljivosti i epistemičke dostupnosti. Ovisno o tome koji se kriteriji dobrog argumenta krše, Lumer sve pogreške dijeli prvo u tri glavne skupine pogreški argumentativne valjanosti – sa šest podskupina – i skupinu pogreški primjerenosti – s pet podskupina, ovisno o podležecem epistemološkom načelu argumenta (deduktivni, induktivni i praktični argumenti). Takva redukcija tipova pogreški omogućuje relativnu potpunost njihova popisa s obzirom da

logički nema ograničenja u pronalaženju sve više i više specifičnih tipova pogreški. Pragmatički, treba definirati i pronaći imena za posebne tipove pogreški samo ako je njihov opseg dovoljno velik ili ako objašnjava od kojega kršenja pogreška potječe. (Lumer 2000: 539)

Nasuprot pragmadijalektičarima koji se pozivaju na nemogućnost objašnjenja pragmatičkih aspekata pogreški u protivnome slučaju, Lumer ograničava pojam pogreški (*fallacies*) na skup nevaljanih argumenata, odnosno drži se u ovoj točki standardnoga tretmana. S druge strane, ne smatra kvalifikaciju čini se valjan bitnom osobinom pogreški. Nedopuštene argumentacijske poteze u dijalogu (koje pragmadijalektičari svrstavaju pod *fallacies*) smatra neispravnim debatiranjem, čiji samo jedan dio potpada pod *fallacies*. Dakle, u Lumerovoj klasifikaciji postoje neispravni argumentativni potezi koji ne zadovoljavaju definiciju *fallacies* (uglavnom jer se ne radi o argumentima), odnosno pojmovi *fallacies* i pogreški u debatiranju imaju tek djelomično zajednički opseg. Primjeri tradicionalnih pogreški (ali i pogreški u pragmadijalektičkome tretmanu) koje Lumer svrstava u skupinu neispravnih dijaloških poteza, no ne i pogreški (nevaljanih argumenata), jesu *argumentum ad baculum* i jednostavni oblik *argumentum ad hominem* (obična uvreda). Također, dijalošku inačicu *tu quoque argumenta*



*ad hominem* drži valjanim potezom (oblikom ispravnog debatiranja), za razliku od *tu quoque* kao argumenta, odnosno ne smatra je pogreškom jer prema definiciji i nije argument. Neutralan stav uspostavlja spram *argumenta ad verecundiam* i *ad misericordiam*. Dakle, ne smatra ih *nužno* pogreškama, već argumentima čije određene forme mogu biti nevaljane, tj. pogreške (Lumer 2000: 539). Odnosno, smatra ih argumentativnim shemama s ispravnim i neispravnim instancama čija se kvaliteta provjerava ovisno o zadovoljenosti situacijskih kriterija (primjerenosti).

Nakon što su iznesene osnovne crte ovih dviju različitih teorija pogreški (pragmadijalektičkog i epistemološkog) i njihov odnos prema standardnome tretmanu istih, nije teško ustanoviti da se međusobno razlikuju zbog različito određenih polazišta u vlastitim pozitivnim teorijama argumentacije. Odnosno, da razlike koje među njima postoje (ograničavanje odnosno neograničavanje skupa pogreški na argumente kao apstraktne entite, zadržavanje odnosno napuštanje momenta varljivosti iz standardnoga tretmana kao sastavne jedinice pogreški) nisu neočekivane prisjetimo li se razlika među dvama pristupima koje su se ticale distinkcije argumenta kao apstraktnog entiteta od argumenta kao govornoga čina. Tu distinkciju ističu i prihvaćaju epistemološki orijentirani teoretičari argumentacije, a pragmadijalektičari odbijaju zadržavajući višeznačnost pojma argumentacije kao apstraktnog entiteta, govornog čina, procesa, proizvoda tog procesa itd. Tako smo mogli uočiti da je pragmadijalektička teorija pogreški, za razliku od epistemoloških, izrazito dijaloška, odnosno da smatra kako se sve pogreške mogu razumjeti jedino promatrane kao neispravni potezi u raspravi. Suprotno tome, vidjeli smo da Lumer ograničava pojam pogreške na loše argumente.

Međutim, promotrimo li s druge strane ono što je zajedničko i jednom i drugom pristupu, mislim da se može argumentirati u prilog tvrdnji da pragmadijalektičari i epistemološki teoretičari zapravo i dijele osnovne pretpostavke o pogreškama i njihovoj vezi s pozitivnom teorijom argumentacije. Te zajedničke pretpostavke, po mojemu mišljenju, oblikuju istraživanje pogreški unutar jednog i drugog pristupa u sličnome pravcu nevezano za deklarativne i stvarne razlike među njima koje postoje u definiciji (dobrog) argumenta i argumentacije. Preciznije rečeno, i pragmadijalektičari i epistemološki teoretičari argumentacije, unatoč različitim objašnjenjima zašto je neki tip argumenta (argumentativnoga poteza) pogrešan, i proklamiranim težnjama da se kritički preispita tradicionalni katalog pogreški, zapravo prihvaćaju tvrdnju da je broj pogreški konačan (usp. Lumerovu tvrdnju da se logički sve pogreške svode na nekoliko tipova u Lumer 2000) i uglavnom preuzimaju popis pogreški iz tradicije. Ekstenzionalno gledano, skup pogreški u pragmadijalektici nadilazi uobičajeni popis (vidjeli smo

da na temelju Procedure za kritičku raspravu izvode nove pogreške, npr. proglašenje stajališta nepovredivim), kao što se i pojedine forme smatraju i u pragmatijalektici i unutar epistemološkoga pristupa neutralnima, a ne nužno pogreškama. No možemo primijetiti da značajnijeg odmaka u samom skupu pogreški od tradicije (ono što John Woods (2013) naziva *Gang of Eighteen*), kao i međusobno, nema. Najveći odmak od tradicije zapravo pronalazimo u tome što, kako je već navedeno, i jedan i drugi pristup za pojedine tradicionalne pogreške rezerviraju status argumentativne sheme, odnosno općenite forme argumenta koja ima dobre i loše instance. Tako da možemo reći i za jedan i za drugi pristup da su tek blago revizionistički raspoloženi spram ocjene tradicionalnih pogrešaka.

Stoga mislim da je, naspram značajnijih razlika koje postoje u pozitivnim teorijama argumentacije koje pronalazimo u pragmatijalektici i epistemološkim pristupima (a o kojima dovoljno svjedoči i žustra rasprava o problemima konsenzualističkog određenja argumentacije, pozitivnom opravdanju, mogućnosti *sui generis* normativnosti u argumentaciji i dr., a koja se javila od samog ulaska epistemološkog pristupa u središnje časopise teorije argumentacije *Argumentation* i *Informal Logic*), razlika između dvaju pristupa daleko manje vidljiva u teoriji pogreški, barem što se tiče stava prema tradicionalnom popisu pogreški, njihovu broju i odnosu prema kriterijima dobrog argumenta. Mislim da ova tvrdnja dobiva na snazi i to što pri rekonstrukciji i vrednovanju argumentacije (što ponekad izrije-kom priznaje i Van Eemeren) razlike između dvaju pristupa često nisu na razini je li neki argument ili argumentativni potez dobar ili loš, već *zašto* je dobar ili loš, odnosno kako ga je moguće objasniti pozivajući se na kršenje kriterija dobre argumentacije.

Dakle, razlika između pragmatijalektičke i epistemoloških teorija pogreški (njihovih negativnih teorija argumentacije) mnogo je jača na razini objašnjenja zašto su pojedine pogreške u argumentaciji loši argumenti odnosno ondje gdje je čvršća veza s njihovim pozitivnim teorijama argumentacije. Pogotovo ako zamijetimo Lumerovu razliku između *fallacies* (pogreški) i neispravnog debatiranja, možemo zaključiti da zapravo neispravnim tipovima argumenata i argumentativnih poteza smatra sve ono što i pragmatijalektičari koji, po mojemu sudu, slabije razrađuju pozitivne teorije argumentacije, obuhvaćaju jedinstvenim pojmom *fallacies*, no promatrajući samo dijaloški kontekst. Ta slabija teorijska razrada, odnosno manje strogo povezivanje s logikom, epistemologijom, psihologijom i dr. oblicima istraživanja racionalnosti, izostavlja iz pragmatijalektike onaj dio istraživanja zaključivanja i pogreški u zaključivanju koji nije dijaloški situiran, što nije slučaj s epistemološkim pristupom argumentaciji.

Nastavljajući se na raspravu o prednostima jednog ili drugog pristupa argumentaciji (Bašić Hanžek 2020)<sup>21</sup>, odnosno na tvrdnju da je epistemološki pristup argumentaciji u prednosti pred pragmatodijalektičkim zbog bolje pozitivne teorije argumentacije i jasnije veze negativne teorije s pozitivnom, no da isto tako ne iscrpljuje mogući istraživački prostor u teoriji pogreški<sup>22</sup>, ovdje zaključujem tvrdnjom da je u samoj negativnoj teoriji argumentacije razlika među dvama pristupima manje vidljiva, tj. da se jasna razlika uočava prvenstveno na razini objašnjenja zašto je neki argumentativni potez pogrešan, a manje je li takav ili nije te da je revizionizam spram tradicionalnoga popisa pogreški kod obaju pristupa vrlo umjeren.

## Literatura

- Barth, E.M. i Krabbe, E.C.W. (1982). *From Axiom to Dialogue. A Philosophical Study of Logics and Argumentation*. Berlin/New York: Walter de Gruyter.
- Bašić Hanžek, Gabriela (2020) *Teorija argumentacije: između pragmatodijalektike i epistemoloških pristupa*. doktorski rad. Filozofski fakultet, Sveučilište u Rijeci
- Biro, J. i H. Siegel. (2006). „In defense of the objective epistemic approach to argumentation.“ *Informal Logic* 26 (1), str. 91-101.
- Hamblin, Ch. L. (1970). *Fallacies*. London: Methuen
- Kišiček, Gabrijele (2010). „Pogreške u argumentaciji – dvojbe u tumačenju i klasifikaciji.“ *Govor* 27 (2), str- 129 – 143.
- Lumer, Ch. (2010). „Pragma-Dialectics and the Function of Argumentation.“
- Lumer, Christopher. (2005a). „The Epistemological Approach to Argumentation – A Map.“ *Informal Logic* 25, str. 189-212..
- Lumer, Ch. (2000). „Reductionism in Fallacy Theory.“ *Argumentation* 14, str. 405-423.
- Lumer, Ch. i Ian J. Dove (2011). „Argument schemes – an epistemological approach.“ *OSSA Conference Archive*, članak 17.

---

<sup>21</sup> Nasuprot oštroj kritici pragmatodijalektičkoga pristupa iz pravca epistemoloških teorija argumentacije, u procjeni dosega utjecaja problematičnoga filozofskog utemeljenja pragmatodijalektike na ostatak teorije iznosi se stav da (i) ta filozofska pozadina zapravo i nije više od tek djelomično razmotrene i artikulirane filozofske rasprave o središnjim pojmovima racionalnosti, razloga, opravdanja, odnosno ponajprije epistemoloških i logičkih problema koji svojim dosegom i složenosti uvelike nadilaze ambicije teorije argumentacije kako su je pokušali osmisliti Van Eemeren i Grootendorst te (ii) da se pragmatodijalektika jednim dijelom može braniti kao legitiman pristup ako se odreče pretenzije na cjelovitost vlastitoga pristupa normiranju argumentacije, odnosno ako prihvati legitimnost i nužnost strože filozofske rasprave o pojmovima racionalnosti i dr. koje preuzima. Stoga ju je najbolje opisati kao teoriju pogreški u argumentaciji (*fallacies*) s posebnim interesom za dijaloški (socijalni) vid argumentacije.

<sup>22</sup> Jer još uvijek nema konsenzusa, unatoč pokušajima redukcije na osnovni skup pogreški kao što to čini Lumer, da je moguć jedinstveni tretman pogreški. Usp. Woods 1994, Woods 2013 i Zenker 2016.

- Sekulić, D. (2016). *Psihologija zaključivanja i logika*. Zagreb: KruZak.
- Siegel, H. i Biro, J. (1997). „Epistemic Normativity, Argumentation, and Fallacies.“ *Argumentation* 11, str. 277–292.
- Smokrović, N. (2015). „Argumentation as a Means for Extending Knowledge.“ *Croatian Journal of Philosophy* sv. 15, br. 44, str. 223 – 231.
- Smokrović, N. (2004). *Priroda prirodnog zaključivanja*. Zagreb: Hrvatsko filozofsko društvo, Biblioteka Filozofska istraživanja.
- Toulmin, S. E. (1976). *Knowing and Acting. An Invitation to Philosophy*. New York: Macmillan.
- Van Eemeren, F. H. (2001). „Fallacies.“ U F. H. van Eemeren (ur.), *Crucial concepts in argumentation theory* (str. 135–164). Amsterdam: Amsterdam University Press
- Van Eemeren, F. H. et al. (2014). *Handbook of Argumentation Theory*. Springer Netherlands
- Van Eemeren, F. H. i R. Grootendorst (2004). *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge University Press.
- Walton, D. N., Reed, C., i Macagno, F. (2008) *Argumentation Schemes*. Cambridge: Cambridge University Press.
- Woods, J. (2013). *Errors of Reasoning: Naturalizing the Logic of Inference* (Studies
- Woods, John. (1994). „Is the theoretical unity of the fallacies possible?“ *Informal Logic* 16 (2), str. 77-85.
- Woods, J. (1988). „Buttercups, GNP's and Quarks: Are Fallacies Theoretical Entities?“ *Informal Logic* 10 (2), str. 67-76.
- Zenker, F. (2016). „The polysemy of ‘fallacy’ – or ‘bias,’ for that matter“. OSA
- Zenker, F. (2007a). „Changes in Conduct-rules and Ten Commandments: Pragma-dialectics 1984 vs. 2004“. U F. H. van Eemeren (ur.), *Proceedings of the International Society for the Study of Argumentation* (ISSA) Amsterdam, str. 1581–1589.
- Zenker, F. (2007b). „Pragma-Dialectic's Necessary Conditions for a Critical Discussion“ U J. A. Blair, H. Hansen, R. Johnson i C. Tindale (ur.), *Proceedings of the Ontario Society for the Study of Argumentation* (OSSA), Windsor: OSSA.



HANOCH BEN-YAMI, EDI PAVLOVIĆ

## Completeness of the Quantified Argument Calculus on the Truth-Valuational Approach

**Abstract:** The Quantified Argument Calculus (Quarc) is a formal logic system, first developed by Hanoch Ben-Yami in (Ben-Yami 2014), and since then extended and applied by several authors. The aim of this paper is to further these contributions by, first, providing a philosophical motivation for the truth-valuational, substitutional approach of (Ben-Yami 2014) and defending it against a common objection, a topic also of interest beyond its specific application to Quarc. Second, we fill the formal lacunae left in the original presentation, which did not incorporate identity systematically into Quarc, and although it proved the soundness of the system did not prove its completeness.

**Key words:** Quantified Argument Calculus; Truth-Valuational Semantics; Natural Deduction; Henkin Completeness.

### 1 Introduction

This paper has been in circulation for the better part of the last decade<sup>23</sup> and even though it has often been cited in the literature on Quarc (Pavlović 2017; Pavlović and Gratzl 2019b, 2021; Ben-Yami 2020, 2021), here it finally, for the first time, appears in print.

The Quantified Argument Calculus (Quarc) is a recently developed formal logic system, arguably closer in syntax and logical properties to natural language(s) than is the Predicate Calculus, its generalized quantifiers version included. The formal system was presented in (Ben-Yami 2014), based on a logical analysis of natural language found in earlier works (Ben-Yami 2004, 2009a, 2009b), and has subsequently seen further developments (Raab 2016; Lanzet 2017; Pavlović and Gratzl 2019b, 2021) and applications (Raab 2018; Pavlović and Gratzl 2019a; Ben-Yami 2020, 2021).

Quarc's basic departure from the Predicate Calculus is in quantifiers not being sentential operators but combining with unary predicates to form quantified arguments. For instance, as 'Alice is polite' is formalized ' $(a)P$ ' (the argument being written to the left of the predicate), so 'All students are polite' is formalized as ' $(\forall S)P$ '. But Quarc also departs from the Predicate Calculus in other features. Like natural language and Aristotelian logic, it

---

<sup>23</sup> Some of the more technical parts of this paper can be found in (Pavlović 2017).

has modes of predication, such as negative predication and modal predication. For instance, ‘Alice isn’t polite’ is formalized as ‘ $(a)\neg P$ ’, and ‘Some students can be polite’ is formalized as ‘ $(\exists S)\diamond P$ ’ (while ‘Possibly, some students are polite’ is formalized as ‘ $\diamond((\exists S)P)$ ’, as one would expect). In addition, again like natural language, Quarc incorporates ways of reordering relation terms, paralleling English active–passive voice distinction and converse relation terms. And it incorporates anaphora as well, the role of which in Quarc is partly related to that of variables in the Predicate Calculus. (Further examples below.)

(Ben-Yami 2014) contains a natural deduction system for Quarc and a truth-valuational, substitutional semantics. While it does not incorporate identity systematically into Quarc, in its last section it contains an outline of a modal extension of it with identity. It also contains a soundness proof but not a completeness one. A sequent calculus for Quarc with identity has been developed in (Pavlović and Gratzl 2019b), and various properties of the system, such as cut-elimination, subformula property, consistency and others, were proven there, and more recently further extended in (Pavlović and Gratzl 2021). Also, (Lanzet 2017) extended Quarc to a three-valued system with defining clauses, using, unlike (Ben-Yami 2014), model-theoretic semantics, and proving the soundness and completeness of the extended system. (Lanzet 2017) also embedded the first-order Predicate Calculus in the extended Quarc, a result obtained earlier for a precursor of Quarc in (Lanzet and Ben-Yami 2004). These results thus establish that the expressive power of Quarc is at least as strong as that of the first-order Predicate Calculus. Quarc has also been used to investigate Aristotelian logic, both assertoric and modal, in works mentioned above as well as in (Pavlović 2017, ch.6.) and (Raab 2018).

Although, as can be seen, some research on Quarc has by now been done (and more can still be done and some is in progress), the first presentation of the system, in (Ben-Yami 2014), left a few formal lacunae. One aim of this paper is to fill these. First, we incorporate identity into Quarc (formation rules for formulas, rules for truth-value assignments, derivation rules). Secondly, as mentioned, that work uses truth-valuational semantics, and it states without proof (Ben-Yami 2014, 129) that each truth-value assignment assigns to any formula a unique truth-value; the proof is provided below. Lastly, and most importantly, a completeness proof for Quarc with identity for truth-valuational semantics is provided.

Truth-valuational, substitutional semantics attracted interest around the sixties and seventies of the previous century, and some prominent logicians have done important work developing it (see (Barcan Marcus 1962; Dunn and Belnap 1968; Kripke 1976; Leblanc 1968, 1976, 1983)). It did

not, however, become a standard part of the logic curriculum, and many logicians have only partial acquaintance with it. As (Ben-Yami 2014) does not motivate the approach or answer any objections to it, we shall do that in this paper. We do not intend to use this paper for a full development and defense of the truth-valuational approach, but rather to say enough to make it attractive and defensible. This is the second aim of this paper.

We proceed as follows. The next section introduces the principles of the truth-valuational approach and defends it against a common objection. It uses Quarc as illustration, but without committing the truth-valuational approach to it. Section 3 introduces the formal system of Quarc, incorporates identity in it, and proves the uniqueness of truth-value assignment mentioned above. Section 4 then proves the completeness of Quarc.

## 2 The Truth-Valuational Approach

The basic idea of the truth-valuational approach is that it is a theory of *truth-value relations*, not a theory of *truth*. In this it is like the standard semantics of the Propositional Calculus, and unlike model-theoretic semantics.

A truth-valuational approach chooses for each calculus a set of basic formulas – different sets for different calculi – and *without* offering an account of the conditions in which they are true, assigns to each of these formulas the value *true* or *false*. Additional rules then determine the truth-value of any non-basic formula. In the Propositional Calculus, this is done by assigning to each propositional variable  $p$ ,  $q$ , etc. the value  $\top$  or  $\perp$ , without asking what it takes for them to be true, and then further rules determine the truth-values of  $\neg p$ ,  $p \vee q$ , and so on. By contrast, model-theoretic semantics introduces the idea of a *domain of objects* and some sort of *reference relation* between symbols of the language and objects in the domain, and says what it takes for a formula like  $P(a)$  to be true on an interpretation. Only then does it determine the truth-values of additional formulas. A truth-valuational approach eschews such concepts as domain of objects or reference, which it takes to be *extra-logical*.

Logic deals with concepts like tautology, contradiction, validity, consistency, soundness and completeness: all these relate to possibilities of truth-value assignments to formulas and to possible relations of truth-values between formulas. The question of the nature of truth is an additional, different kind of question, and if it can be avoided – as it is when the Propositional Calculus is investigated – this will serve the purposes of logic better.

Moreover, it seems that sentences of the same syntax can be true for very different reasons. Consider, for instance, the two true sentences,



**Example 1** *Brutus killed Caesar,*

**Example 2** *Hamlet killed Polonius.*

The former is true because a certain man called ‘Brutus’ stabbed another, called ‘Caesar’, and in this way killed him – here, the account of truth that model-theory provides applies. The latter, by contrast, is true because this is what is written in a Shakespeare’s play: no domain with objects to which the names refer seems to be involved, and model-theory’s picture of truth seems inapplicable. Only if we are already committed to this picture shall we insist that we are referring to abstract objects, say, which stand in the appropriate relation (which? not that of one *killing* the other); we might alternatively suggest that the sentence isn’t really true; or we might hypothesize an implicit ‘in the play’ operator: whatever the merits of these ideas, if Logic can stay away from these debates while investigating the questions that fall within its purview, it surely gains.

Furthermore, let us examine the following two valid inferences:

**Example 3**

*All philosophers are mortal.*

*Socrates is a philosopher.*

---

*∴ Socrates is mortal.*

**Example 4**

*All virtues are rare.*

*Wisdom is a virtue.*

---

*∴ Wisdom is rare.*

Both inferences have the same grammatical form and are valid for that reason. A translation of their sentences into a formal language – be it the Predicate Calculus, Quarc, or any other – should preserve this uniformity and should make them valid because of their form. However, the way they get their meaning is very different for each case. If we think that ‘Socrates’ refers to an object in a domain, do we have a meaningful concept of a domain that would also make wisdom an object in a domain? Or do we have a concept of reference that applies to both ‘Socrates’ and ‘wisdom’? We could say that in both cases we do have a domain, but with the difference that one is of concrete, flesh-and-blood objects, while the other contains abstract ones; and that in both cases, the words are used as tags for these objects. But making the *descriptions* of the uses of these words similar in this way cannot make the uses themselves any more like one another. Again, if Logic can avoid any commitment to a uniform theory of meaning or truth-conditions while accounting for the validity of such inferences, it would surely gain. And unlike model-theory, the truth-valuational approach provides Logic with the means to avoid them.

The way the truth-valuational approach is applied to quantified logic is by relating the truth-values of quantified sentences to those of their instances. In Quarc, an instance of a sentence like  $(\forall S)P$  or  $(\exists S)P$  is an  $(a)P$  such that  $(a)S$  is true. And on a given truth-value assignment,  $(\forall S)P$  is true just in case all instances  $(a)P$ , for any  $a$  for which  $(a)S$  is true, are also true; while  $(\exists S)P$  is true just in case some of these instances are true. As will be shown below, this approach to the relation of the truth-value of a quantified sentence to those of its instances is sufficient for the construction of a sound and complete formal system. And, like the standard approach to Propositional Logic, it involves no commitment to what it takes for basic formulas – in this case  $(a)P$ ,  $(a,b)R$ , and so on – to be true or false.

A common objection to this approach to quantification, going back to Quine,<sup>24</sup> is that it presupposes that all the particulars which are instances of the quantified statement can be named. Otherwise, for example, if there is an unnameable particular that constitutes a counter-instance to a universally quantified statement, its name cannot be substituted for the quantified phrase and generate the false statement that falsifies the universal one. But surely, the objection continues, this nameability assumption is unrealistic, and if it is rejected, the substitutional account suggested above fails.

We answer this objection in two stages. First, with appropriate changes, the problem also besets the objectual, model-theoretic conception of quantification. Roughly, model-theoretic semantics must assume that all particulars in the domain *can be the values of interpretation functions*, an assumption which is equally in need of justification as the nameability one, this time with respect to interpretation functions instead of names. To be more accurate, the common assumption of objectual semantics is even more elaborate: for every interpretation function  $i$ , for every variable  $v$  and for every particular  $p$  in the domain, there is an interpretation function  $i^*$ , different from  $i$  at most in its value for  $v$ , for which  $i^*(v) = p$ . Wherever the model-theoretic logician finds all his interpretation functions, there can the truth-valuational one find all her names. The Quinean observation is no more an issue for the substitutional approach than it is for the objectual one.

But secondly, both approaches can resolve the issue in the same fashion. What the formal approach does is an idealization, for the purposes of a formal system, of an informal mode of account of the meaning of quantified sentences. And both idealizations are equally justified, if they serve their formal purpose. The limitations of actual discourse need not be imported into the ideal one.

---

<sup>24</sup> Discussions of this issue are found in Quine's papers, 'Ontological Relativity' and 'Existence and Quantification', reprinted in (Quine 1969).

The claimed insight behind the truth-valuational approach is that logic examines the properties of a language when used with certain constraints on the relations of the truth-values of its sentences, independently of what it takes or what it even means for these sentences to be true. This insight sheds light on the workings of Henkin's completeness proof.

According to the truth-valuational approach, the possibilities language offers when used with the mentioned constraints determine whether it is, logically, sound, complete, and so on. We need not examine the way language relates through some reference relation to models with objects of this or that sort. Now Henkin, when proving the completeness of the first-order Predicate Calculus, was forced, by the Tarskian semantics he used, to make use of models. He managed to avoid this spurious need and focus only on language by using language itself as its model. In this way, despite the forced detour through models, he in fact examined only the properties of language itself. As shall be seen below (Section 4), this detour is avoided by the truth-valuational approach, which directly assigns truth-values to the sentences of the language it investigates. It replaces Henkin's *construction* – the stage in which a model is devised from language's sentences in order to assign truth-values to them – by a Henkin *assignment*, directly assigning truth-values to sentences.<sup>25</sup>

As was said in the Introduction, this section was intended to introduce the principles of the truth-valuational approach, defend it against a common objection and show it is of logical interest. In the hope that this has been accomplished we shall stop here, leaving additional questions for a different occasion. We turn to the formal treatment of Quarc.

### 3 The Formal System

In this section we turn to the formal presentation of Quarc. As noted in the previous section, this presentation follows (Ben-Yami 2014), with the addition of identity to the vocabulary of the language, and the corresponding adjustments to truth-value assignments and addition of derivation rules.

#### 3.1 Vocabulary of Quarc

**Definition 5** (Vocabulary of Quarc) The language of Quarc contains the following symbols:

---

<sup>25</sup> Another, minor simplification in the proof is that the treatment of identity does not require, unlike the way it does in Henkin's original proof (Henkin 1949, 165), the formation of equivalence classes of individual constants which are the values of these constants. We do not assign anything to constants, but only require that if both  $t = c$  and  $A[t]$  are assigned truth, so is  $A[c]$ . See below.

1. Predicates:  $P, Q, R, \dots$ , denumerably many and with a fixed arity, including a binary predicate of identity,  $=$ .
2. Singular arguments (SAs):  $a, b, c, \dots$ , denumerably many.
3. Anaphors:  $\alpha, \beta, \gamma, \dots$ , denumerably many.
4. Sentential operators:  $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$ .
5. Quantifiers:  $\forall, \exists$ .
6. Numerals used as indices, comma, parentheses.

If  $P$  is a unary predicate, then  $\forall P$  and  $\exists P$  will be called *quantified arguments* (QAs). An argument is only a singular argument or a quantified one (anaphors are not considered arguments).

An occurrence of an argument  $A$  is the *source* of the anaphor  $\alpha$  if  $A$  is to the left of  $\alpha$ ,  $\alpha$  is written as a subscript to  $A$  (i.e.  $A_\alpha$ ), and is not a subscript to any argument occurring between it and  $A$ . In  $A_\alpha$ , only  $A$  is considered an argument.

For every  $n$ -ary predicate  $R$ ,  $n > 1$ ,  $R^n$ , where  $n$  is any permutation of  $1, \dots, n$ , is an  $n$ -ary predicate, called a *reordered form* of  $R$ .

### 3.2 Formulas of Quarc

**Definition 6** (Formula) The following rules specify all the ways in which formulas can be generated.

1. (Basic formula) If  $P$  is an  $n$ -ary predicate and  $t_1, \dots, t_n$  SAs, then  $(t_1, \dots, t_n)P$  is a formula, called a *basic* formula.
2. (Reorder) If  $P$  is a reordered  $n$ -ary predicate,  $n > 1$ , and  $t_1, \dots, t_n$  SAs, then  $(t_1, \dots, t_n)P$  is a formula.
3. (Identity) If  $a$  and  $b$  are singular arguments, not necessarily different, then  $a = b$  is a formula.  $a = b$  is an alternative way of writing  $(a, b)=$ , which is a basic formula. We shall not use the latter form of the formula.
4. (Negative predication) If  $P$  is an  $n$ -ary predicate or a reordered  $n$ -ary predicate and  $t_1, \dots, t_n$  SAs, then  $(t_1, \dots, t_n)\neg P$  is a formula. Negative predication of identity is written as  $t_1 \neq t_2$ .
5. (Sentential operators) If  $A$  and  $B$  are formulas, so are  $\neg(A)$ ,  $(A) \wedge (B)$ ,  $(A) \vee (B)$ ,  $(A) \rightarrow (B)$  and  $(A) \leftrightarrow (B)$ . The parentheses surrounding formulas are called *sentential* parentheses. They can be omitted if no ambiguity arises.
6. (Anaphora) If  $A$  is a formula containing, from left to right,  $t_1, \dots, t_n$  ( $n > 1$ ) occurrences of the SA  $t$ , none of which is a source of any anaphor, and  $A$  does not contain  $\alpha$ , then  $A[t_\alpha/t_1, \alpha/t_2, \dots, \alpha/t_n]$  is a formula.

7. (Quantification) If  $A$  is a formula containing an occurrence of an SA  $t$ , and substituting a QA  $qP$  for  $t$  will result in  $qP$  governing  $A$  (see Definition 7 below), then  $A[qP/t]$  is a formula.

### 3.2.1 Governance

To make full sense of the rules for quantifiers, a notion of governance, related to that of scope in the Predicate Calculus, needs to be defined:

**Definition 7** (Governance) An occurrence  $qP$  of a QA governs a string of symbols  $A$  just in case  $qP$  is the leftmost QA in  $A$  and  $A$  does not contain any other string of symbols ( $B$ ) in which the parentheses are a pair of sentential parentheses, such that  $B$  contains  $qP$  and all the anaphors of all the QAs in  $B$ .

Let us unravel this definition with a few examples of governance and non-governance.

**Example 8** The formula (say, “All cats are agile, but Spot is lazy.”)

$$((\forall C)A) \wedge ((s)L)$$

is not governed by the quantified argument  $\forall C$ , since there is a string of symbols, namely  $((\forall C)A)$ , the parentheses of which are sentential, and which contains all the anaphors of  $\forall C$ . On the other hand, the formula (say, “All cats are agile but lazy.”)

$$((\forall C_\alpha)A) \wedge ((\alpha)L)$$

is governed by the QA  $\forall C$  since, even though it is contained in another string of symbols the parentheses of which are sentential, namely  $((\forall C_\alpha)A)$ , this string does not contain all the anaphors of  $\forall C$  (the only QA occurring in it), namely it does not contain ‘ $\alpha$ ’ in  $(\alpha)L$ .

**Example 9** The formula (say, “Charlie either ate a shark, or was eaten by it”)

$$(c_\alpha, \exists S_\beta)E \vee (\alpha, \beta)E^{2,1}$$

is governed by the QA  $\exists S$  since there is no other appropriate string of symbols that contains all the anaphors of all QAs occurring in it. On the other hand, it does not govern (“Every person either ate a shark or was eaten by it.”)

$$(\forall P_\alpha, \exists S_\beta)E \vee (\alpha, \beta)E^{2,1}$$

since it is not the leftmost QA in that formula. In this case, the formula is governed by  $\forall P$ , which is leftmost, and no other string of symbols within sentential parentheses contains all the anaphors of all the QAs in it – anaphors of  $\forall P$  and  $\exists S$  both occur in the second disjunct.

For further examples of formulas and explanations of the ideas behind some of the definitions, see (Ben-Yami 2014).

### 3.3 Truth-Value Assignments

We next list the rules for assigning truth-values to formulas.

**Definition 10.** (Truth-Value Assignments) The following holds for any truth-value assignment:

1. (Basic formula) Every basic formula is assigned the truth-value of true or false, but not both.
2. (Reorder) Let  $P$  be an  $n$ -ary predicate and  $\pi = \pi_1, \dots, \pi_n$  a permutation of  $1, \dots, n$ . Then, the truth-value assigned to  $(t_{\pi_1}, \dots, t_{\pi_n})P$  is that assigned to  $(t_1, \dots, t_n)P$ .
3. (Law of Identity) Every formula of the form  $t = t$  is true.
4. (Indiscernibility of Identicals) If  $t = c$  is true and the formula  $A[t_1, \dots, t_n]$  is a basic formula containing the instances  $t_1, \dots, t_n$  of an SA  $t$ , then  $A[c/t_1, \dots, c/t_n]$  is true if  $A[t_1, \dots, t_n]$  is true.<sup>26</sup>
5. (Instantiation) For every unary predicate  $P$  there is some SA  $t$  such that  $(t)P$  is true.<sup>27</sup>
6. (Sentential operators) Let  $A$  and  $B$  be formulas. Then,  $\neg(A)$  is true just in case  $A$  is false. Etc.
7. (Negative predication) Let  $P$  be an  $n$ -ary predicate or a reordered  $n$ -ary predicate, and  $t_1, \dots, t_n$  SAs. The truth-value of  $(t_1, \dots, t_n)\neg P$  is that of  $\neg(t_1, \dots, t_n)P$ .
8. (Anaphora) If  $A$  is a formula containing, from left to right, occurrences  $t_1, \dots, t_n$  of the SA  $t$ , none of which is the source of any anaphor, and  $A$  does not contain  $\alpha$ , then the truth-value of  $A[t_\alpha/t_1, \alpha/t_2, \dots, \alpha/t_n]$  is that of  $A$ .
9. (Quantification) Let  $A[\forall P]$  ( $A[\exists P]$ ) be a formula governed by an occurrence of  $\forall P$  ( $\exists P$ ). If for every (some) SA  $t$  for which  $(t)P$  is true,  $A[t/\forall P]$  ( $A[t/\exists P]$ ) is true, then  $A[\forall P]$  ( $A[\exists P]$ ) is true. If for

---

<sup>26</sup> The *only if* part follows from Definition 10.3 and Definition 10.4. This rule could have been defined for only single occurrence of  $t$ , but our formulation will make some derivations shorter.

<sup>27</sup> As explained in (Ben-Yami 2014, 130), the semantic rule of Instantiation is imposed on truth-value assignments in order to have a simple two-valued semantics. If eliminated, several options become possible, both two-valued and three-valued or gappy, some of which have already been pursued in subsequent publications. (Lanzet 2017; Pavlović and Gratzl 2019a, 2019b, 2021).

some (every)  $t$  for which  $(t)P$  is true  $A[t/\forall P]$  ( $A[t/\exists P]$ ) is false, then  $A[\forall P]$  ( $A[\exists P]$ ) is false.

Again, for a discussion of these definitions, see (Ben-Yami 2014).

### 3.3.1 Uniqueness of Truth-Value Assignments

On any assignment, every basic formula is assigned a truth-value, and any formula that is not basic is assigned a truth-value according to the way it is generated. It follows that on any assignment, every formula has a truth-value. However, since some formulas can be generated in more than a single way, we have to show that on any assignment all these ways determine the same truth-value.

We could use different formula rules, as in (Lanzet 2017), to make sure every formula is generated in a unique way (a result we would then have to prove). Instead, we follow here the rules of (Ben-Yami 2014) and show that on any assignment, different ways of generating a formula determine the same truth-value.

Any element of a formula is introduced by a unique rule. Accordingly, any formula is generated by a unique combination of rules. However, some formulas can be generated by applying the rules in more than one order. Consider for instance that

**Example 11.** The formula

$$(a_{\alpha}\alpha)R \wedge (b)Q$$

can be generated from  $(a_{\alpha}\alpha)R$  and  $(b)Q$  according to Definition 6.5 (Sentential operators), or from  $(a,a)R \wedge (b)Q$  by Definition 6.6 (Anaphora).

In general, if two consecutive applications of rules can be transposed without affecting the generated formula, then a concatenation of such rules can sometimes yield more than two ways of generating the same formula. For instance,

**Example 12.** The formula

$$(a_{\alpha}\alpha)R \wedge (b_{\beta}\beta)Q$$

can be generated by two applications of Definition 6.6 (Anaphora) and one of Definition 6.5 (Sentential operators) to the formulas  $(a,a)R$  and  $(b,b)Q$  in five different sequences, its immediate predecessor(s) being,

- (i)  $(a,a)R \wedge (b_{\beta}\beta)Q$ ,
- (ii)  $(a_{\alpha}\alpha)R \wedge (b,b)Q$  or
- (iii)  $(a_{\alpha}\alpha)R$  and  $(b_{\beta}\beta)Q$ ,



and the former two a result of either the Anaphora or the Sentential operators formation rule.

An examination of Definition 6 shows that the order of two applications of Anaphora can be transposed, and that the order of applications of Sentential Operators or Quantification can sometimes be transposed with applications of Anaphora. The applications of any two other rules can never be transposed.

We will prove the following:

**Theorem 13.** If a formula can be generated by applying definitions in different sequences, then for any truth-value assignment each sequence assigns to that formula the same truth-value.

Note that extended sequences, including either several consecutive applications of the rules or multiple predecessors (both of these cases were illustrated in Example 12) can be broken down into a number of pairwise equivalences. If Anaphora cannot be transposed with the consecutive rule, then that rule must be the Quantification rule, replacing the source of the anaphor, and then it cannot be transposed with any subsequent rule either. Therefore we can, w.l.o.g., prove the following lemma:

**Lemma 14.** If a formula can be generated by applying two formula-formation rules to different predecessor formulas, then any truth-value assignment assigns to that formula a unique truth-value regardless of the rule used.

*Proof.* By induction on the number of formula formation rules used.

Basic case. Trivial, as basic formulas do not allow for sequences of formation rules.

Inductive case. Likewise trivial when only one sequence is possible. We examine the remaining cases.

As an illustration, let us take a look at the case for Definition 6.6 (Anaphora) and Definition 6.8 (Quantification), and more specifically, for the case of the universal quantifier. Consider the formula

$$A[a_\alpha, \dots, \alpha_{n-1}][\forall P],$$

where the governing QA  $\forall P$  is not the source of anaphor  $\alpha$ . This formula can be either

(i)  $A[a_\alpha, \dots, \alpha_{n-1}][\forall P]$ , obtained from  $A[a_\alpha, \dots, \alpha_{n-1}][c/\forall P]$  by Definition 6.8, or

(ii)  $A[\forall P][a_\alpha/a_1, \dots, \alpha/a_n]$ , obtained from  $A[\forall P][a_1, \dots, a_n]$  by Definition 6.6.

By Definition 10.9, (i) is true just in case for every  $c$  for which  $cP$  is true, the formula  $A[\alpha_0, \dots, \alpha_{n-1}][c/\forall P]$  is true, and by Definition 10.8, this is true just in case  $A[a_1, \dots, a_n][c/\forall P]$  is true.

Similarly, (ii) is true just in case  $A[\forall P][a_1, \dots, a_n]$  is true, and this is true (note that  $\forall P$  still governs this formula) just in case for every  $c$  for which  $cP$  is true,  $A[a_1, \dots, a_n][c/\forall P]$  is true.

Both  $cP$  and  $A[a_1, \dots, a_n][c/\forall P]$  have less formation rules applied, so by inductive hypothesis they have a unique truth-value on any assignment. It follows that on any assignment, (i) is true just in case (ii) is. Similarly for the particular and sentential operators and simple for two anaphora rules.

### 3.4 Derivation Rules

We proceed to define the proof system of Quarc. We use a natural deduction system, based on the one introduced by Stanisław Jaśkowski in (Jaśkowski 1934) and further developed and streamlined by Frederic Fitch in (Fitch 1952), as well as others. Proofs are written as follows:

**Definition 15.** (Proof) A proof is a list of lines of the form  $\langle L, (i), A, R \rangle$ , where  $L$  is a list (possibly empty) of line numbers of premises;  $(i)$  the line number;  $A$  a formula; and  $R$  the justification, an element of a set of derivation rules, such that the line is written in accordance with it.  $A$  is said to depend on the premises listed in  $L$ . The line numbers in  $L$  are written without repetitions and in ascending order. The formula in the last line of the proof is its conclusion.

We next list the derivation rules of the system.

**Definition 16.** (Derivation Rules)

1. (Premise) At any stage in a proof, any formula can be written, depending on itself, its justification being Premise:

$$i \quad (i) \quad A \quad \text{Premise}$$

2. (Propositional Calculus Rules, **PCR**) We allow the usual derivation rules of the Propositional Calculus.

3. (Sentence negation to Predication negation, **SP**) Let  $P$  be an  $n$ -ary predicate or a reordered  $n$ -ary predicate, and  $t_1, \dots, t_n$  singular arguments.

$$\begin{array}{ll} L & (i) \quad \neg(t_1, \dots, t_n)P \\ L & (j) \quad (t_1, \dots, t_n)\neg P \quad \text{SP, i} \end{array}$$

4. (Predication negation to Sentence negation, **PS**) Let  $P$  be an  $n$ -ary predicate or a reordered  $n$ -ary predicate, and  $t_1, \dots, t_n$  singular arguments.

$L$		(i)	$(t_1, \dots, t_n) \neg P$	
$L$		(j)	$\neg(t_1, \dots, t_n) P$	PS, i

5. (Reorder, **R**) Let  $P$  be an  $n$ -ary predicate and  $\pi = \pi 1, \dots, \pi n$  and  $\rho = \rho 1, \dots, \rho n$  two permutations of  $1, 2, \dots, n$  (the identity permutation included).

$L$		(i)	$(t_{\pi 1}, \dots, t_{\pi n}) P^\pi$	
$L$		(j)	$(t_{\rho 1}, \dots, t_{\rho n}) P^\rho$	R, i

6. (Anaphora Introduction, **AI**) Let  $A$  be a formula containing, from left to right, occurrences  $t_1, \dots, t_n$  of the singular argument  $t$ , none of which has any anaphors, and suppose  $\alpha$  does not occur in  $A$ .

$L$		(i)	$A$	
$L$		(j)	$A[t_\alpha/t_1, \alpha/t_2, \dots, \alpha/t_n]$	AI, i

7. (Anaphora Elimination, **AE**) Let  $A$  be a formula containing, from left to right, occurrences  $t_1, \dots, t_n$  of the singular argument  $t$ , none of which has any anaphors, and suppose  $\alpha$  does not occur in  $A$ .

$L$		(i)	$A[t_\alpha/t_1, \alpha/t_2, \dots, \alpha/t_n]$	
$L$		(j)	$A$	AE, i

8. (Universal Introduction, **UI**) Let  $A[\forall P]$  be a formula governed by  $\forall P$ . Assume that neither  $A[\forall P]$  nor the formulas in lines  $L$  apart from the formula in line  $i$  contain any occurrence of the singular argument  $t$ .

$i$		(i)	$(t)P$	Premise
$L$		(j)	$A[t/\forall P]$	
$L-i$		(k)	$A[\forall P]$	UI, i, j

Where  $L-i$  is the (possibly empty) list of line numbers occurring in  $L$  apart from  $i$ .

9. (Universal Elimination, **UE**) Let  $A[\forall P]$  be a formula governed by  $\forall P$ .

$L_1$		(i)	$A[\forall P]$	
$L_2$		(j)	$(t)P$	
$L_1, L_2$		(k)	$A[t/\forall P]$	UE, i, j

Where  $L_1, L_2$  is the list of line numbers occurring either in  $L_1$  or in  $L_2$ .

10. (Particular Introduction, **PI**) Let  $A[\exists P]$  be a formula governed by  $\exists P$ .

$L_1$		(i)	$A[t/\exists P]$	
$L_2$		(j)	$(t)P$	
$L_1, L_2$		(k)	$A[\exists P]$	PI, i, j

11. (Instantial Import, **Ins**)<sup>28</sup> Let  $q$  stand for either  $\forall$  or  $\exists$ , and  $A[qP]$  be governed by  $qP$ . Assume  $t$  does not occur in  $A[qP]$ ,  $B$  or  $L_1$  and in no formula in  $L_2$  apart from  $j$  and  $k$ .

$L_1$	(i)	$A[qP]$	
$j$	(j)	$(t)P$	Premise
$k$	(k)	$A[t/qP]$	Premise
$L_2$	(l)	$B$	
$L_1, L_2-j-k$	(m)	$B$	Ins, i, j, k, l

Additionally, the system under consideration includes the rules for identity:

12. (Identity Introduction, =I) In any line of the proof a formula of the form  $t = t$  can be written, depending on no premises, with its justification being =I.
- (i)  $t = t$  =I
13. (Identity Elimination, =E) Let  $A$  be a basic formula containing occurrences  $t_1, \dots, t_n$  of the singular argument  $t$  ( $A$  may also contain additional occurrences of  $t$ ).

$L_1$	(i)	$A$	
$L_2$	(j)	$t = c$	
$L_1, L_2$	(k)	$A[c/t_1, \dots, c/t_n]$	=E, i, j

As an illustration of the functioning of the system, let us examine several proofs. First, that ‘all’ implies ‘some’:

**Example 17.**  $(\forall S)P \vdash (\exists S)P$

Proof.

1	(1)	$(\forall S)P$	Premise
2	(2)	$aS$	Premise
3	(3)	$aP$	Premise
2,3	(4)	$(\exists S)P$	PI, 2, 3
1	(5)	$(\exists S)P$	Ins, 1, 2, 3, 4

---

<sup>28</sup> In (Ben-Yami 2014, 133) this rule was called ‘Instantiation.’ ‘Instantial Import’, however, is preferable since in this way the ambiguity of ‘Instantiation’ is avoided, as it is used only for the truth-value assignment rule in Definition 10.5. In addition, ‘Instantial Import’ hints at a relation of this rule to the Predicate Calculus’ existential import. The short form, *Ins*, follows the one in (Ben-Yami 2014) and in some later publications, although *Imp* has also been used for this rule in (Ben-Yami 2020, 2021).

Next, syllogism Bocardo:

**Example 18.**  $(\exists M)\neg P, (\forall M)S \vdash (\exists S)\neg P$

Proof.

1	(1)	$(\exists M)\neg P$	Premise
2	(2)	$(\forall M)S$	Premise
3	(3)	$aM$	Premise
4	(4)	$a\neg P$	Premise
2,3	(5)	$aS$	UE, 2,3
2,3,4	(6)	$(\exists S)\neg P$	PI, 4,5
1,2	(7)	$(\exists S)\neg P$	Ins, 1, 3, 4, 6

And finally, an instance of the DeMorgan laws:

**Example 19.**  $\neg((\forall S_\alpha)P \wedge (\alpha)Q) \vdash (\exists S_\alpha)\neg P \vee (\alpha)\neg Q$

Proof.

1	(1)	$\neg((\forall S_\alpha)P \wedge (\alpha)Q)$	Premise
2	(2)	$\neg((\forall S_\alpha)\neg P \vee (\alpha)\neg Q)$	Premise
3	(3)	$aS$	Premise
4	(4)	$\neg aP$	Premise
4	(5)	$a\neg P$	SP, 4
4	(6)	$a\neg P \vee a\neg Q$	$\vee$ I, 5
4	(7)	$a_\alpha\neg P \vee (\alpha)\neg Q$	AI, 6
3,4	(8)	$(\exists S_\alpha)\neg P \vee (\alpha)\neg Q$	PI, 3, 7
2,3	(9)	$\neg\neg aP$	$\neg$ I, 4, 2, 8
2,3	(10)	$aP$	$\neg$ E, 9
11	(11)	$\neg aQ$	Premise
11	(12)	$a\neg Q$	SP, 11
11	(13)	$a\neg P \vee a\neg Q$	$\vee$ I, 12
11	(14)	$a_\alpha\neg P \vee (\alpha)\neg Q$	AI, 13
3,11	(15)	$(\exists S_\alpha)\neg P \vee (\alpha)\neg Q$	PI, 3,14
2,3	(16)	$\neg\neg aQ$	$\neg$ I, 11, 2, 15
2,3	(17)	$aQ$	$\neg$ E, 16
2,3	(18)	$aP \wedge aQ$	$\wedge$ I, 10, 17
2,3	(19)	$a_\alpha P \wedge (\alpha)Q$	AI, 18
2	(20)	$(\forall S_\alpha)P \wedge (\alpha)Q$	UI, 3, 19
1	(21)	$\neg\neg((\exists S_\alpha)\neg P \vee (\alpha)\neg Q)$	$\neg$ I, 2, 1, 20
1	(22)	$(\exists S_\alpha)\neg P \vee (\alpha)\neg Q$	$\neg$ E, 21

For further examples of proofs, see (Ben-Yami 2014, sec. 3.5.)

## 4 Completeness of Quarc

Completeness is a metalogical property of a formal system, a relation between validity and derivability. Therefore, before proving completeness, let us define these two:

**Definition 20.** (Validity) An argument whose premises are all and only the formulas in the set of formulas  $\Gamma$ , and whose conclusion is the formula  $A$  is valid (written  $\Gamma \models A$ ) just in case every truth-value assignment that makes all the formulas in  $\Gamma$  true also makes  $A$  true, even if we add or eliminate singular arguments from our language (of course, only singular arguments not occurring in  $\Gamma$  and  $A$  can be eliminated).<sup>29</sup>

**Definition 21.** (Derivability) Formula  $A$  is derivable from a (possibly empty) set of formulas  $\Gamma$  (written  $\Gamma \vdash A$ ) just in case there is a proof whose conclusion is  $A$ , such that  $A$  depends only on premises  $\gamma_1, \dots, \gamma_n$ , where  $\{\gamma_1, \dots, \gamma_n\} \subseteq \Gamma$ .

We can now define completeness:

**Definition 22.** (Completeness) A formal proof system is complete just in case for every valid argument  $\Gamma \models A$ , it holds that  $\Gamma \vdash A$ . Completeness can also be formulated as  $\Gamma \models A \Rightarrow \Gamma \vdash A$ .

We proceed to prove the completeness of Quarc. The proof is an adaptation of Leon Henkin's proof (Henkin 1949). A standard Henkin-style proof consists of the following stages: adding witnessing constants; constructing the Henkin theory; defining the Henkin construction; proving the elimination theorem; and some final steps. The structure of the proof below, which applies to the truth-valuational approach, departs from the standard structure in its replacement of the Henkin construction, which is the step in which models are introduced, by the Henkin *assignment*.

### 4.1 Henkin Theory

The Henkin Theory is a set of formulas that fall under given schemas, which we will use to establish a connection between Quarc and the Propositional Calculus, a calculus that we know to be complete. The Henkin Theory uses a language  $L_H$ , which is an extension of the language of Quarc with new singular arguments, the *witnessing constants*. We first define these.

---

<sup>29</sup> The need to make validity on the truth-valuational approach independent of a specific individual constant list has long been recognized. For more detail see (Dunn and Belnap 1968, 180), (Kripke 1976, 336) and (Leblanc 1983). This definition is also used in (Ben-Yami 2014, 131).

**Definition 23.** (Witnessing Constant) For every formula of Quarc of the form  $A[\exists P]$ , where  $\exists P$  governs the formula, we introduce the witnessing constant  $w_{A[\exists P]}$ .

Extending our language with witnessing constants will generate a language  $L_1$ , which will contain new formulas, some of them once again of the form  $A[\exists P]$ , where  $\exists P$  governs the formula. We repeat the same process for  $L_1$  to obtain  $L_2$ , and so on for any subsequent language  $L_n$ .

**Definition 24.** (Henkin Language  $L_H$ ) The Henkin language  $L_H$  is the union of all the languages  $L_i$ ,  $i \in \mathbb{N}$  produced by the extensions of Quarc by witnessing constants.

**Definition 25.** (Date of Birth) Date of birth of a witnessing constant is the  $i$  of the  $L_i$  in which it was introduced.

With these definitions in place, we can now define the Henkin Theory (where  $P$  is a unary predicate and  $R$  an  $n$ -ary predicate,  $R^\pi$  and  $R^e$  reorderings of  $R$ ,  $c$  and  $t$  singular arguments, with  $c$  with numbered indices the same argument, but  $t$  a possibly different singular argument,  $C$  a formula of Quarc in which the anaphor  $\alpha$  does not occur, and where  $c_1 \dots c_n$  are not the source of any anaphor,  $B$  a basic formula, and  $A[qP]$  a formula of Quarc governed by the quantified argument  $qP$ ).

**Definition 26.** (Henkin Theory) Henkin Theory  $H$  consists of all the formulas that fall under one of the following schemas:

- H1.  $A[\exists P] \rightarrow ((w_{A[\exists P]})P \wedge A[w_{A[\exists P]}/\exists P])$
- H2.  $(tP \wedge A[t/\exists P]) \rightarrow A[\exists P]$
- H3.1  $\neg A[\forall P] \leftrightarrow (\exists P_\alpha P \wedge \neg A[\alpha/\forall P])$ , if the governing occurrence of  $\forall P$  in  $A[\forall P]$  is source of no anaphors
- H3.2  $\neg A[\forall P] \leftrightarrow (\exists P_\alpha P \wedge \neg A[\alpha/\forall P_\alpha])$ , if the governing occurrence of  $\forall P$  in  $A[\forall P]$  is source of anaphor  $\alpha$
- H4.  $c = c$
- H5.  $c = t \rightarrow (B[c_1, \dots, c_n] \rightarrow B[t/c_1, \dots, t/c_n])$
- H6.  $(t_1, \dots, t_n)\neg R \leftrightarrow \neg(t_1, \dots, t_n)R$
- H7.  $C[c_1, c_2, \dots, c_n] \leftrightarrow C[c_\alpha/c_1, \alpha/c_2, \dots, \alpha/c_n]$
- H8.  $(t_{\pi 1}, \dots, t_{\pi n})R^\pi \leftrightarrow (t_{\rho 1}, \dots, t_{\rho n})R^e$
- H9.  $(\exists P)P$

We next define the Henkin Assignments and examine their properties.



## 4.2 Henkin Assignment

**Definition 27.** (Henkin Assignment) A Henkin Assignment  $\chi$  is a truth-value assignment that assigns truth-values to all formulas of  $L_H$ , such that  $\chi$  respects the truth-value assignment rules for the sentential operators of the propositional calculus, while also making all the formulas of the Henkin Theory true.

We now prove the following important Lemma:

**Lemma 28.** Any Henkin assignment  $\chi$  respects all the truth-value assignment rules of Quarc.

*Proof.* First note that the truth-value assignment rule for basic formulas (Definition 10.1), that each is either true or false, is respected by  $\chi$ , since any formula is either true or false on  $\chi$ . By definition, it also respects all the rules for the sentential operators (Definition 10.6). We now prove, in this order, that it respects the rules for the particular quantifier, negative predication, anaphora, reorder, instantiation, the universal quantifier and identity.

**Particular Quantifier.** Suppose that a formula of the form  $A[\exists P]$  governed by  $\exists P$  is true. By H1 it follows that so is a formula of the form  $A[t/\exists P]$ , namely  $A[w_{A[\exists P]}/\exists P]$ , and moreover, that the formula  $(w_{A[\exists P]})P$  is true.

Likewise, if a formula of the form  $A[t/\exists P]$  and the formula  $(t)P$  are true, then by H2  $A[\exists P]$  is true as well. Therefore,  $\chi$  satisfies the truth-value assignment rule for the particular quantifier.

**Negative Predication.** Let  $(t_1, \dots, t_n)\neg R$  be true (false). Then, by H6, the formula  $\neg(t_1, \dots, t_n)R$  is true (false). Therefore, the truth-value assignment for  $(t_1, \dots, t_n)\neg R$  is the same as that for  $\neg(t_1, \dots, t_n)R$ , and  $\chi$  satisfies the truth-value assignment rule for negative predication.

**Anaphora.** Suppose  $C[c_1, c_2, \dots, c_n]$  is true (false). Then by H7, the formula  $C[c_\alpha/c_1, \alpha/c_2, \dots, \alpha/c_n]$  is true (false), and  $\chi$  satisfies the truth-value assignment rule for anaphora.

**Reorder.** As above, since  $\chi$  satisfies H8, it satisfies the truth-value assignment rule for reorder.

**Instantiation.** Since  $\chi$  satisfies H9,  $(\exists P)P$  is true, and since it satisfies H1,  $(w_{(\exists P)P})P$  is true. Therefore, for every unary predicate  $P$  there is a singular argument  $t$  for which  $(t)P$  is true, namely  $w_{(\exists P)P}$ , and  $\chi$  satisfies the truth-value assignment rule for instantiation.

**Universal Quantifier.** Let  $A[\forall P]$  be a formula governed by the quantified argument  $\forall P$ , which does not contain the singular argument  $c$ . Assume

first the governing occurrence of  $\forall P$  in  $A[\forall P]$  is not a source of any anaphor.

Assume (i) that for every singular argument  $t$  for which  $(t)P$  is true (since  $\chi$  satisfies instantiation, we know such  $t$  exists),  $A[t/\forall P]$  is also true. We need to show that  $A[\forall P]$  is then true. Now assume for reductio that (ii)  $\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P]$  is true.

Since it has already been established that  $\chi$  satisfies the truth-value assignment rule for particular quantification, it follows from (ii) that for some  $a$ ,  $a_{\alpha}P \wedge \neg A[\alpha/\forall P]$  is true, and from that, since it has already been established that  $\chi$  satisfies the truth-value assignment rule for anaphora, that  $aP \wedge \neg A[a/\forall P]$  is true. But this is contrary to assumption (i). So,  $\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P]$  is false, and by H3.1,  $\neg A[\forall P]$  is also false, and therefore  $A[\forall P]$  is true.

Next, assume (i) that for some singular argument  $t$  for which  $(t)P$  is true,  $A[t/\forall P]$  is false. We need to show that  $A[\forall P]$  is then false. Same *mutatis mutandis* as above.

Now assume the governing occurrence of  $\forall P$  in  $A[\forall P]$  is a source of anaphor  $\alpha$ . The case is *mutatis mutandis* same as above, using H3.2. Therefore,  $\chi$  satisfies the truth-value assignment rule for the universal quantifier.

**Identity.** Suppose  $c = t$  is true and  $B[c_1, \dots, c_n]$  a basic formula containing occurrences  $c_1, \dots, c_n$  of a singular argument  $c$ . Since  $\chi$  satisfies H5, the formula  $B[t/c_1, \dots, t/c_n]$  obtained by substituting some or all occurrences of  $c$  by  $t$  in  $B$  is true if  $B[c_1, \dots, c_n]$  is true. Moreover,  $\chi$  satisfies H4. Therefore,  $\chi$  satisfies the truth-value assignment rules for identity, which concludes the proof of Lemma 28.

Now take any argument  $\Gamma \vDash A$ , valid in Quarc, where  $A$  and all the formulas of  $\Gamma$  belong to  $L$  (i.e. they contain no witnessing constants). Since  $L_H$  is simply an extension of Quarc with additional singular arguments, by the definition of validity,  $\Gamma \vDash A$  is valid in Quarc with  $L_H$  as its language as well. And, since adding premises to a valid argument does not affect its validity, the argument  $\Gamma, H \vDash A$  is likewise valid in Quarc with  $L_H$  as its language. Now, given that any truth-value assignment that respects the rules of the Propositional Calculus and makes all the formulas of  $\Gamma$  and  $H$  true will be a Henkin Assignment, it follows by Lemma 28 that it will likewise respect all the other truth-value assignment rules of Quarc, and therefore that it will make  $A$  true as well. It follows that  $\Gamma, H \vDash A$  also in the Propositional Calculus.

Given the completeness of the Propositional Calculus, this yields that  $\Gamma, H \vdash A$  in the Propositional Calculus. What needs to be shown now is that, if  $\Gamma, H \vdash A$  in the Propositional Calculus, then  $\Gamma \vdash A$  in Quarc – the Elimination Theorem.

### 4.3 Elimination Theorem

Before proceeding with the proof of this theorem, several preliminary results need to be established. The proofs that involve only the derivation rules of the Propositional Calculus will be omitted.

Let  $\Gamma$  be a set of formulas of Quarc and  $A, B$  and  $C$  formulas of Quarc.

**Theorem 29.** (Deduction Theorem) If  $\Gamma, A \vdash B$  then  $\Gamma \vdash A \rightarrow B$ .

**Proposition 30.** If  $\Gamma, A_1, \dots, A_n \vdash B$  and for every  $i \leq n$ ,  $\Gamma \vdash A_i$  then  $\Gamma \vdash B$ .

**Lemma 31.** If  $\Gamma \vdash A \rightarrow B$  and  $\Gamma \vdash \neg A \rightarrow B$ , then  $\Gamma \vdash B$ .

**Lemma 32.** If  $\Gamma \vdash (A \rightarrow B) \rightarrow C$ , then  $\Gamma \vdash \neg A \rightarrow C$  and  $\Gamma \vdash B \rightarrow C$ .

**Lemma 33.** If  $A[\exists P]$  is governed by  $\exists P$  and  $t$  is a singular argument occurring nowhere in  $\Gamma, A[\exists P]$  or  $B$ , then if  $\Gamma \vdash (tP \wedge A[t/\exists P]) \rightarrow B$ , then  $\Gamma \vdash A[\exists P] \rightarrow B$ .

Proof. Since  $\Gamma \vdash (tP \wedge A[t/\exists P]) \rightarrow B$ , then there is a finite number of formulas of  $\Gamma$ , say  $k$ , such that  $C_1, \dots, C_k \vdash (tP \wedge A[t/\exists P]) \rightarrow B$ . Let us now write the proof of  $(tP \wedge A[t/\exists P]) \rightarrow B$ , with  $C_1, \dots, C_k$  as premises. Suppose that the proof is  $i$  lines long. Our proof then continues as follows:

	...		
$L$	(i)	$(tP \wedge A[t/\exists P]) \rightarrow B$	
$i+1$	(i+1)	$A[\exists P]$	Premise
$i+2$	(i+2)	$A[t/\exists P]$	Premise
$i+3$	(i+3)	$tP$	Premise
$i+2, i+3$	(i+4)	$tP \wedge A[t/\exists P]$	$\wedge I, i+2, i+3$
$L, i+2, i+3$	(i+5)	$B$	$\rightarrow E, i, i+4$
$L, i+1$	(i+6)	$B$	Ins, $i+1, i+2, i+3, i+5$
$L$	(i+7)	$A[\exists P] \rightarrow B$	$\rightarrow I, i+1, i+6$

Since all the formulas in  $L$  are in  $\Gamma$ , it follows that  $\Gamma \vdash A[\exists P] \rightarrow B$ .

Next we prove Lemma 34, needed so we can eliminate the formulas containing witnessing constants from  $\Gamma, H \vdash A$ :

**Lemma 34.** If  $A[\exists P]$  is governed by  $\exists P$  and  $t$  is a singular argument occurring nowhere in  $\Gamma, A[\exists P]$  or  $B$ , then if  $\Gamma, A[\exists P] \rightarrow (tP \wedge A[t/\exists P]) \vdash B$  then  $\Gamma \vdash B$

Proof. By applying Theorem 29 to  $\Gamma, A[\exists P] \rightarrow (tP \wedge A[t/\exists P]) \vdash B$ , we obtain  $\Gamma \vdash (A[\exists P] \rightarrow (tP \wedge A[t/\exists P])) \rightarrow B$ . Then by Lemma 32 we get (i)  $\Gamma \vdash \neg A[\exists P] \rightarrow B$  and (ii)  $\Gamma \vdash (tP \wedge A[t/\exists P]) \rightarrow B$ . Applying Lemma 33 (and given our assumptions) to (ii) we get (iii)  $\Gamma \vdash A[\exists P] \rightarrow B$ . Finally, applying Lemma 32 to (i) and (iii), we get  $\Gamma \vdash B$ .

The final preliminary consideration we need for the proof of the Elimination Theorem is to show that the formulas H2–H9 of Henkin Theory are theorems of Quarc. Let us consider the straightforward cases first.

H2 corresponds to the derivation rule PI and can be proved by it and the derivation rules of the Propositional Calculus; H4 corresponds to the rule =I; H5 corresponds to =E; the two directions of H6 correspond to the rules PS and SP; the two directions of H7, to AI and AE; and H8 to Reorder. This leaves us with H3.1, H3.2 and H9 to prove.

We first prove H3.1. Let the governing occurrence of  $\forall P$  in  $A[\forall P]$  be the source of no anaphor.

(i) $\vdash \neg A[\forall P] \rightarrow (\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P])$			
1	(1)	$\neg A[\forall P]$	Premise
2	(2)	$\neg(\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P])$	Premise
3	(3)	$tP$	Premise
4	(4)	$\neg A[t/\forall P]$	Premise
3,4	(5)	$tP \wedge \neg A[t/\forall P]$	$\wedge$ I, 3, 4
3,4	(6)	$t_{\alpha}P \wedge \neg A[\alpha/\forall P]$	AI, 5
3,4	(7)	$\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P]$	PI, 3, 6
2,3	(8)	$\neg\neg A[t/\forall P]$	$\neg$ I, 4, 2, 7
2,3	(9)	$A[t/\forall P]$	$\neg$ E, 8
2	(10)	$A[\forall P]$	UI, 3, 9
1	(11)	$\neg\neg(\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P])$	$\neg$ I, 2, 1, 10
1	(12)	$\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P]$	$\neg$ E, 11
	(13)	$\neg A[\forall P] \rightarrow (\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P])$	$\rightarrow$ I, 1, 12

In line (3) we choose a  $t$  that does not occur in the string  $\neg A[\alpha/\forall P]$ . This is necessary for the use of UI in line (10).

(ii) $\vdash (\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P]) \rightarrow \neg A[\forall P]$			
1	(1)	$\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P]$	Premise
2	(2)	$A[\forall P]$	Premise
3	(3)	$tP$	Premise
4	(4)	$t_{\alpha}P \wedge \neg A[\alpha/\forall P]$	Premise
4	(5)	$tP \wedge \neg A[t/\forall P]$	AE, 4
4	(6)	$\neg A[t/\forall P]$	$\wedge$ E, 5
2,3	(7)	$A[t/\forall P]$	UE, 2, 3
3,4	(8)	$\neg A[\forall P]$	$\neg$ I, 2, 6, 7
1	(9)	$\neg A[\forall P]$	Ins, 1, 3, 4, 8
	(10)	$(\exists P_{\alpha}P \wedge \neg A[\alpha/\forall P]) \rightarrow \neg A[\forall P]$	$\rightarrow$ I, 1, 9

We now prove H3.2. Let the governing occurrence of  $\forall P$  in  $A[\forall P]$  be the source of anaphor  $\alpha$ . The proof differs only slightly from the one for H3.1 and will be omitted here.

And finally, we prove H9:

1	(1)	$tP$	Premise
	(2)	$(\forall P)P$	UI, 1, 1
1	(3)	$(\exists P)P$	PI, 1, 1
	(4)	$(\exists P)P$	Ins, 2, 1, 1, 3

We should keep in mind that whatever is derivable in the Propositional Calculus is derivable in Quarc as well, as the derivation rules or the latter include those of the former.

With all this preliminary work done, we can now proceed with the proof of the Elimination Theorem itself. We have previously indicated what the Elimination Theorem is, but let us now formulate it precisely:

**Theorem 35.** (Elimination Theorem) If  $A$  is a formula of Quarc derivable from formulas of Quarc  $C_1, \dots, C_n$  together with formulas of Henkin Theory (written as  $C_1, \dots, C_n, h_1, \dots, h_k \vdash A$ ), and  $A, C_1, \dots, C_n$  belong to  $L$ , then  $A$  is derivable from  $C_1, \dots, C_n$  alone ( $C_1, \dots, C_n \vdash A$ ).

*Proof.* By induction on the number of Henkin formulas from which  $A$  is derivable.

*Basic step.* If  $k = 0$ , then Theorem 35 vacuously holds as there are no Henkin formulas to eliminate.

*Inductive step.* We now wish to show that, if the Elimination Theorem holds for any derivation with  $k$  or fewer formulas of the Henkin theory  $H$ , it also holds for any formula  $A$  derivable from  $k+1$  formulas of the Henkin theory and formulas  $C_1, \dots, C_n$  of  $L$ . There are two cases to consider here.

First, suppose one of the formulas of  $H$  is of the form H2–H9, say  $h_{(k+1)}$ . Since all of the formulas H2–H9 are theorems of Quarc,  $\vdash h_{(k+1)}$  and therefore  $C_1, \dots, C_n, h_1, \dots, h_k \vdash h_{(k+1)}$ . Therefore, by Proposition 30,  $C_1, \dots, C_n, h_1, \dots, h_k \vdash A$ , and by inductive hypothesis  $h_1, \dots, h_k$  can be eliminated.

The second case is that in which all the  $k+1$  Henkin formulas are of the form H1. In that case, we choose one instance of H1 the witnessing constant of which is of the same or greater date of birth than any witnessing constant of any other instance of H1. Since this witnessing constant does not occur in any of the other schema instances, and neither does it occur in  $C_1, \dots, C_n$  or  $A$  (as they are formulas of the non-extended language of Quarc), by Lemma 34 it can be eliminated. So,  $A$  is derivable from  $C_1, \dots, C_n$  and  $k$  formulas of  $H$ , and by inductive hypothesis these  $k$  formulas can be eliminated.

Since Quarc includes all the derivation rules of the Propositional Calculus, we can now conclude that if  $\Gamma, H \vdash A$  in the propositional calculus, then  $\Gamma, H \vdash A$  in Quarc as well. Since the conclusion in any proof depends on a finite number of premises, it follows that  $C_1, \dots, C_n, H \vdash A$ . If  $\Gamma, A$  are all formulas of Quarc that belong to  $L$ , then by the Elimination Theorem,  $C_1, \dots, C_n \vdash A$ , and therefore  $\Gamma \vdash A$  in Quarc.

To summarize, what we have established is that, if  $\Gamma \vDash A$  in Quarc, then  $\Gamma, H \vDash A$  in Quarc as well. But, given the definition of a Henkin assignment, it follows  $\Gamma, H \vDash A$  in the Propositional Calculus, and since this calculus is complete,  $\Gamma, H \vdash A$  in it. We have further established by means of the Elimination Theorem that if  $\Gamma, H \vdash A$  in the Propositional Calculus, then  $\Gamma \vdash A$  in Quarc. This finally enables us to conclude that if  $\Gamma \vDash A$  in Quarc, then  $\Gamma \vdash A$  in Quarc, and therefore that Quarc is complete on the truth-valuational approach.

## 5 Conclusion

In this paper we presented and discussed the truth-valuational approach, and its application in semantics in general and to Quarc in particular, and then put that approach to use in presenting a Henkin completeness proof for Quarc, which itself has already been further utilized in literature on Quarc, e.g. in (Ben-Yami 2020). This result is obtained for a version of Quarc containing identity, even though the original formal presentation in (Ben-Yami 2014) does not contain it (although it is discussed e.g. in (Ben-Yami 2021)). However, by the conservativity proof in (Pavlović and Gratzl 2019b), it follows that this suffices to show that Quarc without identity is likewise complete. A different proof of completeness of Quarc, using sequent calculi, can also be found in (Pavlović and Gratzl 2021).

## Acknowledgements

One of the authors had the opportunity to present this paper at the June 2015 conference “Speaking of the Ineffable, East and West” at the University of Rijeka. The authors are grateful to the organizers, and in particular Prof. Nenad Smokrović, for this opportunity and useful feedback.

Finalizing this paper was partially supported by the Academy of Finland, research project no. 1308664.

## References

- Barcan Marcus, Ruth. 1962. "Interpreting Quantification." *Inquiry: An Interdisciplinary Journal of Philosophy* 5 (1-4): 252–259.
- Ben-Yami, Hanoach. 2004. *Logic & Natural Language: On Plural Reference and Its Semantic and Logical Significance*. Routledge. <https://doi.org/10.4324/9781315250267>.
- . 2009a. "Generalized Quantifiers, and Beyond." *Logique et Analyse* 52 (208): 309–326.
- . 2009b. "Plural Quantification Logic: A Critical Appraisal." *The Review of Symbolic Logic* 2 (1): 208–232.
- . 2014. "The Quantified Argument Calculus." *The Review of Symbolic Logic* 7 (1): 120–146.
- . 2020. "The Quantified Argument Calculus and Natural Logic." *Dialectica* 74 (2): 35–70.
- . 2021. "The Barcan Formulas and Necessary Existence: The View from Quarc." *Synthese* 198: 11029–11064.
- Dunn, J. Michael, and Nuel D. Belnap. 1968. "The Substitution Interpretation of the Quantifiers." *Noûs* 2 (2): 177–185.
- Fitch, Frederic B. 1952. *Symbolic Logic: An Introduction*. Ronald Press Co.
- Henkin, Leon. 1949. "The Completeness of the First-Order Functional Calculus." *The Journal of Symbolic Logic* 14 (3): 159–166.
- Jaśkowski, Stanisław. 1934. *On the Rules of Suppositions in Formal Logic*. Studia Logica. Nakładem Seminarjum Filozoficznego Wydziału Matematyczno-Przyrodniczego Uniwersytetu Warszawskiego.
- Kripke, Saul A. 1976. "Is There a Problem about Substitutional Quantification?" In *Truth and Meaning*, edited by Gareth Evans and John McDowell, 324–419. Oxford University Press.
- Lanzet, Ran. 2017. "A Three-Valued Quantified Argument Calculus: Domain-Free Model-Theory, Completeness, and Embedding of FOL." *The Review of Symbolic Logic* 10 (3): 549–582.
- Lanzet, Ran, and Hanoach Ben-Yami. 2004. "Logical Inquiries into a New Formal System with Plural Reference." In *First-Order Logic Revisited*, edited by Vincent Hendricks, Fabian Neuhaus, Uwe Scheffler, Stig Andur Pedersen, and Heinrich Wansing, 173–223. Logische Philosophie. Logos Verlag Berlin.
- Leblanc, Hugues. 1968. "A Simplified Account of Validity and Implication for Quantificational Logic." *The Journal of Symbolic Logic* 33 (2): 231–235.
- . 1976. *Truth-Value Semantics*. North-Holland Publishing Company.
- . 1983. "Alternatives to Standard First-Order Semantics." In *Handbook of Philosophical Logic: Volume I: Elements of Classical Logic*, edited by Dov Gabbay and Franz Guenther, 189–274. Dordrecht: Springer Netherlands.
- Pavlović, Edi. 2017. "The Quantified Argument Calculus: An Inquiry into Its Logical Properties and Applications." PhD thesis, Central European University, Budapest.



- Pavlović, Edi, and Norbert Gratzl. 2019a. "Free Logic and the Quantified Argument Calculus." In *Philosophy of Logic and Mathematics: Proceedings of the 41st International Ludwig Wittgenstein Symposium*, edited by Gabriele M Mras, Paul Weingartner, and Bernhard Ritter, 27:105–115. Walter de Gruyter GmbH.
- . 2019b. "Proof-Theoretic Analysis of the Quantified Argument Calculus." *The Review of Symbolic Logic* 12 (4): 607–636.
- . 2021. "Abstract Forms of Quantification in the Quantified Argument Calculus." *The Review of Symbolic Logic*, 1–31. <https://doi.org/10.1017/S175502032100006X>.
- Quine, Willard Van Orman. 1969. *Ontological Relativity and Other Essays*. John Dewey Essays in Philosophy. Columbia University Press.
- Raab, Jonas. 2016. "The Relationship of QUARC and Classical Logic." Master's thesis, Munich: LMU.
- . 2018. "Aristotle, Logic, and QUARC." *History and Philosophy of Logic* 39 (4): 305–340.



BORAN BERČIĆ

## *X is the best, but I prefer Y!* On Values and Preferences<sup>30</sup>

**Abstract:** Author argues that statements of the form *X is the best, but I prefer Y* are not contradictory or self-refuting. To show that, one has to distinguish values from preferences. The relation *X is better than Y* is understood as purely objective two-place relation, while the relation *A prefers X to Y* is understood as three-place relation with an irreducibly subjective element. Author rejects the idea that values should be inferred from preferences. Author also discusses context relativity and sortal relativity of preferences.

**Key words:** Value, preference, paradox.

Nenad Smokrović dedicated most of his professional career to rationality. He worked a lot on logical fallacies, paradoxes, cognitive shortcuts, collective choices, etc. Therefore, on this occasion, I would like to pay attention to one seeming paradox in the domain of practical reasoning. One can come across it in the everyday conversation. People sometimes utter statements of the form *X is the best, but I prefer Y*. Of course, there is a very strong tension between *X is the best* and *I prefer Y*, and for this reason statements of this form might seem self-refuting or paradoxical.<sup>31</sup> In this article I would like to show that this tension can be solved, and that statements of this form are not really paradoxical.

*Values* and *preferences* are obviously closely related. *Better than* and *preferable to* might even be used interchangeably. Chisholm and Sosa say: “We use P for the relation of better than, or preferable to.” (1966: 245) Allan Gibbard starts his article with a “vague truism” that “goodness is a matter of preferability.” (1998: 239) Elisabeth Anderson sums up the idea of the rational desire theory: “One thing is better (for a person) than another if and only if she would rationally prefer it.” (1993: 129) Sven Ove Hansson says that “‘Logic of preference’ and ‘logic of betterness’ are in practice taken

---

<sup>30</sup> This work has been fully supported by the University of Rijeka, research project Uni-ri-human-18-239.

I wish to thank the reviewer whose comments and criticism led to the substantial improvement of the article.

<sup>31</sup> Statements of this form *prima facie* resemble the Moore’s paradox: *p is true, but I do not believe it; X is the best, but I do not prefer it.*

as synonyms.” (2001: 320) Preference can be seen as a kind of evaluation. Hansson and Grüne-Yanoff “discuss the notion of preference as *subjective comparative evaluation*, of the form “Agent A prefers X to Y””. (2021) These short quotations are of different authors, from different contexts, with different motives. Nevertheless, they all assume that there is a very strong and important link between values and preferences. Independently of any philosophical analysis, expressions like *valuable*, *preferable*, *desirable*, ... stand for a nexus, or a family of interrelated concepts. They are semantically close.<sup>32</sup> Here the question is: What exactly is the nature of the relationship between *values* and *preferences*? More precisely, what is the relationship between our beliefs about values and our preferences? We may tentatively put forward the following principle:

P: A believes that X is better than Y iff A prefers X to Y.<sup>33</sup>

Several important questions arise here. First: What is the logical status of this principle? Should we understand it as a conceptual truth about *being better than* and *being preferable to*? Is it an *a priori* truth of practical reason, in the sense that practical rationality demands that one always strives at the best? Is it a factual claim in the sense that systematic socio linguistic survey would show that people’s values and preferences in fact coincide? Or should it be understood in some other way? Second: Should some *provisos* be added to the principle? Should we understand it as a claim about ideally rational agents? Should we understand it as a normative claim about what agents should prefer? Does it hold within a specific domain, or it holds all things considered? As we will see in the rest of the paper, this question might be relevant for the solution of what appears to be a paradox. Third: What comes first? What grounds what? We have a nice instance of Euthy-

---

<sup>32</sup> Here we have to mention the following possibility: one might claim that values and preferences hold for *different* and *mutually* exclusive domains. Say, that evaluations are appropriate for ethical assessments, while preferences are appropriate for aesthetic and gustatory experiences. Also, that evaluations essentially purport to be universally valid and have justification, while preferences do not (*de gustibus non disputandum est*). According to this view there would be an essential difference between *better* and *preferable*. On this view, principle P, as well as S statements, would both be instances of *category mistake*. Inferences from *better* to *preferable*, or other way around, would not possible because they would belong to different domains or different categories. Although this view has some ground in the way in which we in fact use these terms, strict separation between values and preferences would certainly be unnatural and violent.

<sup>33</sup> I do not claim that anybody ever defended this principle in this form. This principle should rather be taken as an approximation and idealization for the sake of the argument.

phro dilemma here: Do we prefer *X* to *Y* because we believe that *X* is better than *Y*, or we believe that *X* is better than *Y* because we prefer *X* to *Y*?<sup>34</sup> Also, do we have a one way order of determination here, or perhaps rather an interplay between values and preferences? Fourth: What sort of entities do we evaluate and prefer? Things, properties, facts, states of affairs, propositions, possible worlds? At this point we may remain neutral on this matter and proceed with the analysis.<sup>35</sup> Although, as we will see in part 5, this question might be relevant in the discussion.<sup>36</sup>

Fifth: There is a category of statements that *prima facie* contradicts the principle formulated above. This is the problem that I would like to discuss in this paper. These statements are of the following form:<sup>37</sup>

S: *X* is the best, but I prefer *Y*.

Here are the examples: Toyota is the best car in the world, but I prefer Alfa Romeo.<sup>38</sup> Kim Koter is the most beautiful woman, but I prefer Lisa Bonet. Brad Pitt is the most handsome man in the world, but I prefer Jonny Lee

---

<sup>34</sup> Max Scheler had interesting insights on this matter. On the one hand, it might seem that he believed that preferences come first and that values depend on preferences. "Which value is "higher" can be comprehended only through the acts of preferring and placing after." (1973: 89) "It is in preferring that the synthetic relations of higher and lower values are constituted. (1973: 306) On the other hand, he describes perception of values in realistic terms. He talks about *value-ception* and *value-deception*, and says that "Value-ception ... *precedes* all representational acts." (1973: 201) In my opinion his view amounts to the following: values exist previously and independently of preferences, but their ranking depends on preferences. He suggests that values are eternal while their ranking is historical. The consequence of this view is that values by themselves do not contain grounds for their ranking. I find this consequence odd.

<sup>35</sup> Sven Ove Hansson says that "the relation of the preference relation ... may be physical objects, types of properties of such objects, states of affairs, possible worlds - just about anything. (2001: 320)

<sup>36</sup> One of the options is that we evaluate *particular properties* while we prefer *whole things*.

<sup>37</sup> Statements that would obviously contradict principle *P* would be of the form: "A believes that *X* is better than *Y* and A prefers *Y* to *X*." or perhaps: "Although A believes that *X* is better than *Y*, A prefers *Y* to *X*." Although the difference between *better than* and *the best* is not elaborated here, I believe that statements of the form "*X* is the best, but I prefer *Y*." can be taken as statements that contradict principle *P*. After all, statements of this form are colloquial, people sometimes say so.

<sup>38</sup> Dean Stanisavljević in a *caffè bar Bard*, a couple of years ago. Of course, the immediate reaction was *Why don't you drive a Toyota then?* The thing is that Alfa Romeo is a sports car. It has different suspension, double spark plugs, etc. People who drive Alfa are often organized in clubs - so called *Alfisti*. Therefore, it is not surprising that a fan of Alfa recommends something else to other people.

Miller. Pršut is the best for sandwich, but I prefer parizer.<sup>39</sup> Michelangelo is the best painter of all times, but I prefer Wassily Kandinsky. Luka Modrić is the best football player, but I prefer Ante Rebić. Hortensia is a beautiful flower, but I find it disgusting.<sup>40</sup> The list goes on.

Of course, these statements belong to completely different domains, and it is reasonable to assume that they have different underlying psychological mechanisms. Analyses of particular cases will probably differ. However, they have the same logical form and to this extent they can and should be analyzed together.

Now, the question is how to deal with the *S*-statements? In principle there are two main options here. First, we can reject them as contradictory and self-refuting, and preserve the principle *P*. Second, we can accept them as legitimate and grammatical, and weaken the principle *P*.

What does it mean to weaken the principle? Principle *P* states that believing that *X* is better than *Y* is a *necessary and sufficient condition* for preferring *X* to *Y*. Instead, we may claim, that if *A* believes that *X* is better than *Y*, then *A* has a *strong reason* for preferring *X* to *Y*. Or, if *A* believes that *X* is better than *Y*, then *A* is *strongly disposed* to prefer *X* over *Y*. Or, if *A* believes that *X* is better than *Y*, then *A* *will very likely* prefer *X* over *Y*. So, instead of necessary and sufficient conditions, we should deal with reasons, dispositions, or likelihood.<sup>41</sup> In this article I will not try to say what is the true nature of the relationship between our values and our preferences. I will try to decide whether the relationship is necessary or not. And this will *ipso facto* show whether *S*-statements are legitimate or not.

In the rest of the paper, I would try to briefly sketch several possible reactions to the tension in the *S*-statements. As we will see, some of them are well known philosophical arguments. 1. *S*-statements are contradictions. 2. *S*-statements are instances of the weakness of will. 3. *S*-statements should be understood in the inverted commas sense. 4. *S*-statements should be analyzed as context dependent. 5. *S*-statements are sortal relative. 6. *S*-statements are not contradictory because values are objective and preferences subjective. In my opinion, *S*-statements are not contradictory or paradoxical. Reaction 6. is the right one. Let's go through the options.

---

<sup>39</sup> *Pršut* comes from Italian *prosciutto*. It is dried and smoked pork ham, very delicious and expensive. On the other hand, *parizer* is cheap, waterish, and light chicken or turkey salami. Kids like it. It is a very usual thing. It can be found in any Croatian grocery.

<sup>40</sup> Vlatko Polović in a *caffè bar Art* in Kastav, this year.

<sup>41</sup> Weakening the principle has similar effects like adding provisos (second question in the text above). However, there is a difference. If we add provisos, we limit the scope of the necessary relation. If we weaken the principle, we deny the necessity of the relation.

## 1 S-statements are contradictions<sup>42</sup>

Statements of the form *X is the best, but I prefer Y* are contradictory and self-refuting.<sup>43</sup> People who utter such statements are irrational and they simply do not know what they are talking about. They contradict themselves in a single sentence. If you are rational and consistent, and if you really believe that *X* is the best, then you should prefer *X*. On the other hand, if you prefer *Y*, then what you *really believe* is that *Y* is the best. Values and preferences simply entail each other: *X* is valuable iff we should prefer *X*; *X* is more valuable than anything else iff *X* is preferable to anything else; etc. How can you prefer *Y* over *X* when at the same time you claim that *X* is better than *Y*? If *X* is really the best, then you should prefer *X*, not something else. Why do you prefer a thing that is in your own opinion of lesser quality?

This insight lies at the base of the *internalism about moral motivation*.<sup>44</sup> Plato was led by this, or a very similar, piece of reasoning when he claimed that *virtue is knowledge*. His idea was that if one knows that *X* is good, then by the same token one does *X*.<sup>45</sup> On the other hand, if one does not do *X*, then one does not know that *X* is good, or at least one does not really know that *X* is good. If one really knew that *X* is good, one would do *X*. If everything is OK with our cognitive and motivational apparatus, the insight that *X* is good makes us do *X*. It is impossible to know that *X* is good

---

<sup>42</sup> I discuss this option in the sense of *conceptual contradiction*. It can be also understood as *practical contradiction* - as a breaching the principle of practical rationality. This depends on the reading of the principle *P*, whether we understand it as a conceptual claim, as a claim of practical rationality, etc.

<sup>43</sup> If principle *P* is a conceptual truth, then S-statements are contradictions. Let me *B* stand for *A believes that X is better than Y*, and *P* for *A prefers X to Y*. Principle *P* states that  $(B \leftrightarrow P)$ , while an S-statement states that  $(B \wedge \neg P)$ . The implication of  $(B \leftrightarrow P)$  is that  $(B \rightarrow P)$ , and that is equivalent to  $\neg(B \wedge \neg P)$ .  $(B \wedge \neg P)$  and  $\neg(B \wedge \neg P)$  are a contradiction. One might wonder how  $\neg P$  should be understood here. (I owe this point to Majda Trobok.) Is *A does not prefer X to Y* equivalent with *A prefers Y to X*? It is equivalent, if we assume that *X* and *Y* belong to the same domain, that the relation of preference covers the whole domain, and that the two entities in the domain cannot be equally preferred.

<sup>44</sup> David Velleman takes it as “The view that something’s being good for a person depends on his capacity to care about it.” (1998: 88)

<sup>45</sup> In *Protagoras* Plato says: “no one freely goes for bad things or things he believes to be bad; it’s not, it seems to me, in human nature to be prepared to go for what you think to be bad in preference to what is good.” (358d) In his *Ethics*, Nowell-Smith says: “There is no need therefore to try to bridge the gap between “this is the best thing for me to do” and “I shall do this.” In deciding that something is the best thing for him to do a man has already decided to do it.” (1954: 102)



and not do  $X$ .<sup>46</sup> In the same way, if everything is OK with our cognitive and motivational apparatus, the insight that  $X$  is the best makes us prefer  $X$ . Everything else is irrational.

However, it seems that  $S$ -statements are not paradoxical. Although there is a *prima facie* strong tension in such statements, it seems that people who utter them want to say something consistent. We will try to find out what.

## 2 $S$ -statements are instances of the weakness of the will<sup>47</sup>

Aristotle believed that the failure to do  $X$  despite the insight that  $X$  is good, is a *volitional* failure, not a *cognitive* one. He believed that such cases are cases of *akrasia* or weakness of the will.<sup>48</sup> One might be tempted to analyze  $S$ -statements as cases of the weakness of the will: *Although I know that  $X$  is the best, I cannot resist  $Y$* . It seems that this analysis might cover some cases of  $S$ -statements, like: *Although I know that I should not eat sweets, I prefer a chocolate mousse for dessert. No matter how much time I will have to spend at the mechanic, I must have a new Alfa. Here a chocolate mousse and Alfa Romeo would be the things one simply cannot resist. However, although this analysis could be appropriate for some cases of  $S$ -statements, it cannot systematically explain away all the cases. Preference is not a matter of irresistible urge. My preference is *mine*. It is not something external to me that takes over me. It is the outcome of *my* evaluational and motivational setup. Some  $S$ -statements can be explained as cases of incontinence (*akrasia*) but not all. A chocolate mousse might be a case of incontinence, but buying Alfa Romeo instead of Toyota, planting a cactus instead of hortensia, or taking a parizer sandwich instead of pršut sandwich, do not look like cases of the weakness of the will. It seems that people who do such things have a normal will of their own. This seems to be something that they wholeheartedly want, not something that they give in to. For this rea-*

---

<sup>46</sup> Here we talk about *actions* not *preferences*. However, one might argue that the relationship between the two is very strong or even constitutive. We can try to extend the concept of preference in such a way that it embraces actions. We can claim that preferences should be inferred from the actions, or that actions are realizations or executions of preferences. Max Scheler says: "As an act, "preferring" must be sharply distinguished from its kinds of realization. The realization may consist in the special activity that we experience in its execution." (1973: 89)

<sup>47</sup> One might wonder whether preference is a matter of will at all. It is not implausible to argue that it is. Max Scheler says: "It is only the consciousness of "being able to prefer something else" that must accompany the act." (1973: 88)

<sup>48</sup> Aristotle discusses weakness of will in *The Nicomachean Ethics*, Book VII, 1-10. (1998: 159-182) In the article "How Is Weakness of the Will Possible?" Donald Davidson says: "An agent's will is weak if he acts, and acts intentionally, counter to his own best judgement." (1980: 26)

son, I believe, most of the S-statements cannot be plausibly analyzed as cases of the weakness of will.

### 3 S-statements should be understood in the inverted commas sense<sup>49</sup>

A good candidate for the analysis of S-statements is the so-called *inverted commas sense*.<sup>50</sup> The idea is that people who utter such statements do not really believe that *X is the best*. What they want to say with *X is the best* is something like *people generally believe that X is the best*, *X is usually regarded as the best*, or *according to accepted standards X is the best*, etc. That is, when they say that *X is the best* that should be understood in the inverted commas sense. They do not really claim that *X is the best*. They rather state what is usually regarded as a standard, not what they believe the right standard is. They do not express their own belief. This is why this sense is called *inverted commas sense*.

Talking about gardens and flowers, a guy said: *Hortensia is beautiful, though I find it disgusting*. Was he contradicting himself?<sup>51</sup> When I asked him to clarify his statement, he said: *Well, I meant that people usually regard it as beautiful. But I find it waterish and artificial*. So, he used *hortensia is beautiful* in the inverted commas sense. People sometimes use *X is the best* in this sense, and inverted commas analysis is probably true for some cases. However, it seems that this analysis does not grasp one important aspect of S-statements. Namely, according to the inverted commas analysis, one who says that *X is the best* does not really believe that *X is the best*.

---

<sup>49</sup> The inverted commas sense is designed to explain away counterexamples to the *internalism about moral motivation* - the view that belief that *X* is good necessarily motivates us to do *X*. My favorite example in this context is *Addams family*. They know what is good and they find it disgusting. They are happy to do bad things just because bad things are bad.

<sup>50</sup> The term was introduced into discussion by R.M Hare. (1963) He relied on the analysis of P.H. Nowel-Smith who called this ambiguity the *Janus principle*. (1954) The idea is that one who is not inclined to do *X* cannot really believe that *X* is good. Such a person can believe that *X* is good only on a declarative level, that is, in the inverted commas sense. In *The Nicomachean Ethics*, Book VII, 3c, Aristotle says: "we must suppose that the use of language by men in an incontinent state means no more than its utterance by actors on the stage." (1998: 166)

<sup>51</sup> Let me quote a classic here. In *Critique of the Power of Judgment* Kant says: "It would be ridiculous if someone who prided himself on his taste thought to justify himself thus: "This object ... is beautiful **for me**." For he must not call it **beautiful** if it pleases merely him. Many things may have charm and agreeableness for him, no one will be bothered about that; but if he pronounces that something is beautiful, then he expects the very same satisfaction of others: he judges not merely for himself, but for everyone, and speaks of beauty as if it were a property of things." (2000: 98).

The implication of inverted commas sense is that other people are simply wrong: other people believe that *X is the best* although in fact it is not. The mob is wrong. *X* is not the best. Hortensia is not beautiful because it looks waterish and artificial. People are wrong when they believe that it is beautiful. It is not! Usual standards are wrong. The mob has no taste.

For this reason, the inverted commas sense is not adequate for the analysis of all *S*-statements. Some *S*-statements certainly can be explained in this way, but interesting cases of *S*-statements are those where people sincerely believe that *X* is the best, but nevertheless prefer *Y*. Inverted commas analysis cannot explain such cases. In fact, it seems that this analysis presupposes that *S*-statements, taken literally, are contradictory and self-refuting. So, we have to look for further explanation.

#### 4 *S*-statements are context dependent

One might try to analyze *S*-statements as consequences of the specific features of the context in which they are made. The idea is that our preferences are context dependent in the sense that they might depend on the contingent circumstances of the context in which we find ourselves. In spite of the belief that *X* is the best, we may opt for *Y* due to different reasons: we do not have enough money for *X*, we do not have enough time for *X*, we have diabetes type 2 or high cholesterol, etc.<sup>52</sup> This analysis may have some appeal, but it is not very subtle. Its adequacy may be questioned on two related grounds.

First, if we choose *Y* because we do not have enough money for *X*, does it mean that we *prefer* *Y*? Is the word *prefer* properly used here? If I take a pizza cut because I do not have enough time for a slow food, does it mean that I *prefer* pizza cut to slow food? According to my linguistic intuitions, here the word *prefer* does not fit perfectly well. It would be better to say that I *opted for*, *have chosen*, *took*, ... That would be better choice of words.

---

<sup>52</sup> Von Wright drew the distinction between *extrinsic* and *intrinsic* preferences. In the case of extrinsic ones “*a judgement of betterness serves as a ground or reason for a preference*” (1963: 14), while in the case of intrinsic ones there is no such judgement. For instance, If I prefer *X* because it is better for my health, then my preference is extrinsic. Von Wright developed his logic of preference for the intrinsic cases. It seems that this distinction grasps something relevant, though, I am not sure as to how much we can do with it. The fact that my doctor forbade me to eat my favourite food certainly is in some sense external to my preferences. However, it is easy to reconstruct my preferences in the ideal case: what would I eat had the doctor not forbidden me. Also, it would be hard to find a case of preference without any underlying judgement. Preferences are not whims. Whenever we prefer something, we are able to say a couple of words about it.

It seems that in these cases my preferences were rather suppressed by the circumstances, not formed by circumstances.<sup>53</sup>

Second, the influence of the contingent circumstances can be excluded by appeal to the ideal circumstances. The question is not what I do prefer in the actual circumstances, but rather what I would prefer in the ideal circumstances. The question is: Would I buy Mercedes if I had enough money for it? Would I take chocolate mousse if I didn't have diabetes? Etc. It seems that our true preferences are better revealed in the ideal circumstances.

Of course, one can just stipulate that *prefers* always means *prefers-in-the-particular-circumstances*. This stipulation would successfully explain away the paradoxicality of S-statements, at least for some cases. Though, for the mentioned reasons I am not inclined to this solution.

## 5 S-statements are sortal relative

In order to work out the difference between values and preferences, we have to pay attention to a number of things. One of the things that seems relevant here is that *our evaluations are often sortal relative*.<sup>54</sup> One might claim that M48 is the best rifle of all times, and at the same time feel deep aversion and scorn toward it because rifles are made for killing people. Is this consistent? It seems that it is. On the one hand, one claims that M48 is the best within the sortal of rifles, according to the standards that hold within the sortal. On the other hand, one claims that the whole sortal has very negative value. One might claim that these tripe are excellent tripe (according to the criteria within the sortal), but a horrible food in general because tripe are horrible as a sortal, no matter whether they satisfy criteria of excellence within the sortal.

At this place one might be naturally tempted to say that *we evaluate things within sortals, while we prefer sortals*. For instance, one might say

---

<sup>53</sup> It seems that contingent circumstances do not necessarily distort our preferences. Compare the following two claims: 1) Although *pršut* is better, I *prefer* *parizer* because I don't have enough money for *pršut*. 2) Although *pršut* is better, I *prefer* *parizer* because I had *pršut* last 10 days in a row and I am fed up with *pršut*. In both cases our choice depends on contingent circumstances. However, in 1) our preferences are suppressed by the contingent circumstances, while in 2) they are formed by them. The fact that I had *pršut* 10 days in a row influences my *appetite*, while the lack of money does not. For this reason, the word *prefer* is properly used in 2), but not in 1).

<sup>54</sup> Aristotle was aware of the fact that evaluations can be sortal relative (intra-sortal as well as inter-sortal). In *Organon, Topica*, Book III, Chapter 2, 117b, he says: "Moreover, if A be without qualification better than B, then also the best of the members of A is better than the best of the members of B; e.g. if Man be better than Horse, then also the best man is better than the best horse." (2016: 252)

that although M48 is the best rifle of all times, I will never let any firearm in my home. Or, although this hortensia is a beautiful specimen of hortensia, I would never plant it in my garden because I hate flowers. This shows that some S-statements can be truly analyzed in this manner. However, this analysis cannot be extended to all the cases because the tension between values and preferences typically arises *within a sortal*. Although I believe that Toyota is better car *as a car*, I prefer Alfa Romeo *as a car*. Although I believe that X is better as an F, I prefer Y as an F. Etc. For these cases we need a different kind of analysis.

## 6 S-statements are not contradictory because values are objective and preferences subjective<sup>55</sup>

The idea behind this reaction is that in S-statements no contradiction can arise at all because our evaluations are about the *world*, while our preferences are about *ourselves*. If I say that *Plavac* is better than *Merlot*, I am saying something about these vines. But if I say that I prefer *Merlot* to *Plavac*, I am saying something about myself. We evaluate when we say that something is *better*. We prefer when we say that something is *better for us*. We should not conflate *better* with *better for me*.<sup>56</sup> They are distinct notions. Besides, evaluating and preferring need not range over the different domains (option discussed in f.n.1), they may range over the same domain. They are just different activities, or rather *different stances* or *different perspectives* that we might have toward the same domain.<sup>57</sup> For this reason, there is no tension in S-statements, and principle P is completely untenable.

This option would efficiently remove the air of paradoxality in S-statements, but its overall tenability is questionable. Certainly, there is a sense in which preferences are subjective, and this might help in our analysis. However, the question is whether evaluations and preferences can be completely separated. It seems that they cannot. We cannot just by *fiat* proclaim that values are objective and preferences subjective. We must consider the

---

<sup>55</sup> Faced with the tension in the S-statements, people's first reaction often is that there can be no real tension there because values and preferences are subjective anyway. The saying that *beauty is in the eye of the beholder* is very popular in general. Even in the contemporary semantical analysis, many authors take seriously the view that disagreement in the matters of taste is *faultless disagreement*. (Kölbel 2004; MacFarlane 2014) However, the view that values and preferences are both subjective cannot be of any help here. For in that case the tension in the S-statements would be internal to the subject, but it would still be a tension, it would not be solved.

<sup>56</sup> We may say that values are *agent-neutral* while preferences are *agent-relative*.

<sup>57</sup> If they necessarily covered different domains, this analysis could not get off the ground. Though, it seems natural to claim that if these two activities are so different, then they should apply to different domains.

way in which we *de facto* understand these activities. In the seminal work von Wright says:

The concept of preference is related to the notion of betterness. It is also related to the notion of choice. Betterness is a typically axiological notion. Choice belongs, again, to the group of concepts which I have called anthropological. Because of its double relationship, to betterness and to choice, the concept of preference itself may be said to stand between the two groups of concepts. (1963: 13)

He nicely grasped our understanding of the concept of preference and related concepts. *Better* and *preferable* are closely related concepts and therefore any position that implies their radical separatedness cannot be taken seriously.

Though, this option can be weakened and made much more plausible. One may argue that evaluating and preferring *largely but not completely coincide* because preferring contains an *irreducibly subjective element* that evaluating lacks.<sup>58</sup> The absence of the complete overlap between the two is what makes S-statements possible. This option looks serious and promising. It grasps well what a charitable person may have in mind when faced with S-statements.

We may say that comparative evaluation is a two-place relation that includes two objects and their relevant properties, while preference is a three-place relation that includes two objects and their relevant properties, as well as the subject and his/her relevant properties. So, *being better of* is a dyadic relation, while *being preferred to* is a triadic relation.

$$\begin{array}{l} B(x, y) \\ P(S, x, y) \end{array}$$

Under this assumption, it is clear *why* S-statements are not contradictory and *why* principle *P* cannot hold. Preferences contain further subjective element that evaluations do not contain, *agent specific reasons* for preferring one thing to another.

In *Stara pošta* inn, in Permani near Rijeka, Croatia, every year there is a competition for the best sausage.<sup>59</sup> The jury tastes sausages of different producers and elects the best one. Their choices are led by objective criteria: quality of meat, ratio of meat and fat, balance of spices, how dry it is, how

---

<sup>58</sup> Evaluation is also subjective in the trivial sense that there must be a subject who is evaluating.

<sup>59</sup> If disagreements about matters of taste are really *faultless* (Kölbel 2004; MacFarlane 2014), it is not clear how the vocation of *sommelier* is possible at all.

smoked it is, etc. The relationship between *good* and *good-making-characteristics* seems to be very simple here.<sup>60</sup> Now we are coming to the central point of this paper. Suppose that a member of a jury at the contest votes for the sausage *X*, according to the enlisted criteria. However, after the contest, he orders a pair of *Y* sausages. *Y* sausages are spicier, and the guy likes spicy food. Is he contradicting himself? According to the above analysis, he is not contradicting himself. The fact that he likes spicy food is *the further subjective element* that is not operational in his evaluation of the sausages, and the guy is aware of that. He is aware of the difference between the objective criteria and the criteria or reasons that are *subject specific* for him.<sup>61</sup>

Let's go back to the example of a guy who would never give up Alfa Romeo in spite of the fact that he believes that Toyota is the best car. What are the subject specific considerations in his case? The fact that he drives Alfa Romeo shows *what kind of person he is*. He is a person of style, fast responses, velocity, action, elegance, sport, etc.<sup>62</sup> This is the reason why he prefers Alfa Romeo although he might sincerely believe that Toyota is the best car. For this reason, his preference diverges from his evaluation, and his corresponding S-statement is not inconsistent.

In these examples the *irreducibly subjective element*, that preferring contains and evaluating lacks, consist in the fact that I like spicy food, or in the fact that I see myself as a person of style and action. These are the characteristics that make the difference between  $B(x, y)$  and  $P(S, x, y)$ .

---

<sup>60</sup> Of course, the question: What is the relationship between *good* and *good-making-characteristics*? is very interesting and deep. I will not discuss it here, but let me mention several well-known accounts. Is it a relationship of *a priori identity* - the view that G.E. Moore criticized in *Principia Ethica?* (1903) Is it a *posteriori identity* as moral realists naturalist (new wave realists or Cornell realists) believed? (Berčić 1995) Is it *supervenience* as R.M. Hare thought? (1952: 145) Is goodness just a matter of *Error*, as J.L. Mackie thought? (1977: 41) Should we accept T.M. Scanlon's *buck passing account of value*, according to which good-making-characteristics provide reasons for reactions, while goodness by itself does not? (2000: 97) My favorite discussion about the relationship between *good* and *good-making-characteristics* is Paul Edwards' book *The Logic of Moral Discourse* (1955). His example, the title of Chapter V, is "The Steak at Barney's Is Rather Nice."

<sup>61</sup> Of course, one may be tempted to deny the subject specificity of reasons and regard them as universal. Fisherman's Friend commercials exploit this tendency. *Sind sie zu stark - bist Du zu schwach!* (If they are too strong - you are too weak!) A guy who claims this is funny just because he refuses to acknowledge the fact that there might be other people who do not share his taste. Instead of accepting his taste *as his*, he blames those who do not share it.

<sup>62</sup> Such self-perception might be the cause of the widespread irony toward Alfa Romeos and their owners. In Croatia there are thousands of jokes about Alfa Romeos and their owners. There are dozens of web sites dedicated only to this kind of jokes.



The proponent of the view that S-statements are contradictory would reject this account on the grounds that only the *actual choice* of a subject can be taken as the reliable indicator of the values that he or she really has. The fact that the guy in *Stara pošta* orders a pair of *Y* sausages shows what his *real values* are. His real values are revealed in his order, not in his voting. What he *really* believes is that *Y* sausages are the best, not that *X* sausages are the best. If he *really* believed that *X* sausages are the best, he would order a pair of *X* sausages. The fact that he has just voted for the *X* sausages should be explained away as a case of inverted commas sense, or in some similar manner, perhaps as learned and purely verbal reaction. I think that his objection is ill founded, but we to take it seriously and perhaps show where it goes wrong.

Why wouldn't we simply believe the guy when he says that he believes that *X* is the best sausage? Why reinterpret his report and make guesses about his real values? The problem is that we do not have direct access to his values. We can either trust his reports or make an inference from his behavior. Values are often not measured independently of preferences but rather inferred from the preferences. Warren, McGraw, and Van Boven state the problem:

The value of some state of the world is typically defined as the extent to which it is considered desirable or undesirable. Economists and behavioral decision theorists often use the word utility rather than value to emphasize value's subjective nature. In classic utility theory, values are not measured, but rather inferred from preferences. Consequently, values and preferences are often used interchangeably - a preference for some state of the world over some other state of the world demonstrates that the former is valued more than the latter. (2011: 193)

However, the fact that there is no direct and independent access to our values does not mean that we cannot meaningfully talk about them and distinguish them from preferences, desires, and other related mental states. Elizabeth Anderson believes that values cannot be inferred from preferences because we can always have different attitudes toward our own choices. We can make our choices wholeheartedly, with resignation, we can regret or not regret them, etc.

Again, no straightforward welfare or legitimacy judgment can be derived from the fact of choice or preference. A person need not endorse all the consequences and meanings of even her most preferred choice. (§2.4) Economists make a fundamental mistake in inferring people's values directly from their choices or preferences. (1993: 202)



In some cases, it may really be wise to reject verbal introspective reports and resort to the behavioral reinterpretation instead. But we cannot do this all the time. After all, behaviorism is abandoned as a general methodology for the study of human nature, and it is abandoned for good reasons. Generally speaking, we can and should trust people's verbal reports. It is a default position. If there is a discrepancy between verbal reports and actual behavior, we should ask for an explanation. Only if explanation is not plausible, we may talk about the *real* values and try to infer them from the behavior. We have conceptual means for distinguishing values from preferences and we should use these means when we need them.

The distinction between values and preferences is not only a matter of conceptual acrobatics. It can have empirical consequences as well. Jonathan Wolff analyzes an imagined example:

Does this mean that the group will vote to permit smoking if and only if a majority prefer there to be public smoking? At first sight this might seem obviously so, but a moment's reflection reveals that it need not be. It will be true that some people will vote as if they are expecting to answer the question 'would you prefer there to be smoking or not?' These people will indeed vote according to their preferences. But others will vote as if the question they are answering is 'do you think that smoking ought to be permitted?' Accordingly, some smokers will vote to deny their own pleasure, arguing that it is wrong for smokers to submit others to the adverse effects of their behavior. Some non-smokers, too, will vote against their own preferences, arguing that smoking is a matter for individual decision. In other words these people are voting in a disinterested fashion, and so do not reveal their interests by their vote. (2006: 73)

Say that 60% of a given population smokes, and that 80% votes to ban smoking in enclosed public places. The assumption that their values differ from their preferences would be a part of the best explanation of this fact.

At the end, let me notice that the analysis of values and preferences has two parts, empirical and conceptual. Psychologists and economists should work on the empirical part of the research. Philosophers and logicians should work on the conceptual or logical part of it. I hope that this article is a small contribution in that direction.

## References

- Anderson Elisabeth: *Value in Ethics and Economics*, Harvard University Press, 1993.
- Aristotle: *The Nicomachean Ethics*, Oxford University Press, 1998.
- Aristotle: *The Organon*, edited by Roger Bishop Jones, www.rbjones.com, 1.1.2016.
- Berčić Boran: *Realizam, relativizam, tolerancija*, HKD Rijeka, 1995.
- Chisholm Roderick and Sosa Ernest: "On the Logic of "Intrinsically Better,"" *American Philosophical Quarterly*, Vol. 3, No. 3, 1966, pp. 244-249.
- Davidson Donald: "How Is Weakness of the Will Possible?" in *Essays on Actions and Events*, Oxford University Press, 2001/1980, pp. 21-42.
- Edwards Paul: *The Logic of Moral Discourse*, The Free Press, 1955.
- Fehige Christoph and Wessels Ulla (ed): *Preferences*, Walter de Gruyter, 1998.
- Gabbay Dov and Guenther Franz (ed): *Handbook of Philosophical Logic, 2nd Edition, Volume 4*, Springer, 2001.
- Gibbard Alan: "Preference and Preferability," in Fehige and Wessels (ed): *Preferences*, 1998.
- Hansson Sven Ove: "Preference Logic," in Gabbay and Guenther (ed.) 2001.
- Hansson Sven Ove and Till Grüne-Yanoff, "Preferences", *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2021/entries/preferences/>>
- Hare R. M: *The Language of Morals*, Oxford University Press, 1952/1972.
- Hare R. M: *Freedom and Reason*, Clarendon Press, Oxford, 1963.
- Kant Immanuel: *Critique of the Power of Judgment*, Cambridge University Press, 1790/2000.
- Kölbel Max: "Faultless Disagreement," *Proceedings of the Aristotelian Society* Vol. 104, 53-73, 2004.
- MacFarlane John: *Assessment Sensitivity - Relative Truth and its Applications*, Oxford University Press, 2014.
- Moore, G. E: *Principia Ethica?* Cambridge University Press, 1903/2000.
- Mackie, J. L: *Ethics: Inventing Right and Wrong*, Penguin Books, 1977/1986.
- Nowell-Smith P. H: *Ethics*, Pelican Book, 1954/1961.
- Plato: *Protagoras*, Clarendon Press, Oxford, 1976/2002.
- Scanlon, T. M: *What we Owe to Each Other*, Harvard University Press, 1998/2000.
- Scheler Max: *Formalism in Ethics and Non-Formal Ethics of Values*, Northwestern University Press, 1973.
- Velleman David: "Is Motivation Internal to Value?" in Fehige and Wessels (ed): *Preferences*, 1998.
- von Wright Georg Henrik: *The Logic of Preference*, Edinburgh University Press, 1963.
- Warren Caleb, McGraw Peter, VanBoven Leaf: "Values and Preferences: Defining Preference Construction," *Mays Business School Research Paper No. 2015-8, Cognitive Science*, Volume 2, Issue 2, pages 193-205, March/April 2011.
- Wolff Jonathan: *An Introduction to Political Philosophy*, Oxford University Press, 2006.



ALEKSANDRA GOLUBOVIĆ, JELENA KOPAJTIĆ

## Svjetonazor i odgoj kritičkog mislitelja<sup>63</sup>

**Sažetak:** U ovom radu krećemo od nekoliko pretpostavki koje ćemo pobliže razmatrati i na koncu dati neke smjernice s obzirom na njihovu ulogu u odgojno-obrazovnom sustavu. Pretpostavke su sljedeće: 1) kritičko mišljenje je poželjno unutar odgojno-obrazovnog procesa; 2) racionalnost podrazumijeva kritičko mišljenje i obratno; 3) svjetonazor može biti prepreka za podučavanje kritičkog mišljenja; 4) kritičko mišljenje je moguće podučavati i vježbati; 5) proces argumentacije doprinosi razvoju kritičkog mišljenja koje bi se trebalo odnositi na sve aspekte života. Odlučile smo se za određene definicije racionalnosti i kritičkog mišljenja od kojih polazi teza ovoga rada. Nadalje, želimo prikazati dva svjetonazora: teistički i ateistički kao moguće prepreke kritičkom mišljenju. Ovdje ne ulazimo u raspravu o tome koji je od spomenuta dva svjetonazora moguće racionalno opravdati, već nas zanimaju isključivo njihovi utjecaji na kritičko mišljenje, odnosno na podučavanje istog. Preciznije rečeno, bavimo se mogućnošću (ili nemogućnošću) podučavanja i njegovanja kritičkog mišljenja u odgojno-obrazovnim institucijama u kojima je određeni svjetonazor bar implicitno prisutan.

**Ključne riječi:** racionalnost, kritičko mišljenje, epistemologija religije, svjetonazor, teizam, ateizam, sekularizam.

### Uvod – Racionalnost

Ljudi će rijetko kada reći za sebe da su iracionalni, odnosno nerazumni. Uglavnom volimo misliti da smo razumni i racionalni, da poštujemo logička pravila i razloge te ono najvažnije, volimo misliti da su naša vjerovanja koherentna. Bihevioralni ekonomist koji je dokazao da to uopće nije tako (bar kada je riječ o ekonomskim odlukama koje imaju priličan utjecaj na naš život) dobio je i Nobelovu nagradu za svoje djelo. Riječ je, naravno, o Richardu Thaleru<sup>64</sup>. No bez obzira na to jesu li ljudi po svojoj prirodi pretežno razumni ili nerazumni, gotovo svi misle da pripadaju prvoj kategoriji. Upravo to uvjerenje i jest najveća prepreka dijalogu između različitih svjetonazora budući da često obje strane osuđuju jedna drugu kao nerazumne.

---

<sup>63</sup> Ovaj rad prilog je zborniku u čast prof. dr. sc. Nenadu Smokroviću, profesoru emeritusu na Filozofskom fakultetu u Rijeci. Rad je također dio projekta Sveučilišta u Rijeci *Kritičko mišljenje i društvo* (UNIRI-HUMAN-18-254) voditeljice prof. dr. sc. Majde Trobok.

<sup>64</sup> Vidi: Thaler (2009) i (2020).

Stoga se postavlja pitanje što uopće znači biti razuman? Ovdje ćemo predstaviti neke filozofske ideje o tome, no valja odmah napomenuti kako to pitanje nema filozofski konsenzus te je stoga, poput većine ključnih filozofskih pitanja, još uvijek otvoreno. Zvonimir Čuljak u *Filozofskom leksikonu* racionalnost definira kao *obdarenost razumom (...) sposobnošću proizvodnje pojmova te prosuđivanja i zaključivanja (...)*<sup>65</sup>. Nadalje, pozivajući se na tradicionalno viđenje racionalnosti, on navodi kako racionalnost *u većoj ili manjoj mjeri odlikuje sva normalno razvijena ljudska bića*<sup>66</sup>. Ako prihvatimo da je ova definicija točna, čini se da je racionalnost urođena odlika *normalno razvijenih* ljudi. No, dalje u svojoj natuknici, Čuljak govori o pokusima kognitivnih psihologa koji baš i nisu poduprli prethodnu tezu te su ih, štoviše, naveli da zauzmu pesimističan stav. Drugi autori zauzeli su manje pesimističnu poziciju opravdavajući loše pokusne rezultate *time što ljudska racionalnost nije tipično sposobnost zaključivanja prema obrascima logike i teorije vjerojatnosti, nego prema tzv. pragmatičnim shemama zaključivanja*<sup>67</sup>.

Ljudskom racionalnošću danas se dominantno bave epistemologija (epistemička racionalnost), psihologija (psihologija rasuđivanja), filozofija uma, iako se sve filozofske discipline u nekoj mjeri (nužno) dotiču teme racionalnosti, kao i razne teorije ponašanja i odlučivanja unutar različitih disciplina. U ovom ćemo se radu usredotočiti na epistemičku racionalnost, odnosno na definicije racionalnosti ponuđene od strane jednog od najpoznatijih epistemologa. Kao što je već rečeno, bavit ćemo se svjetonazorom, posebice svjetonazorom naspram religijskih vjerovanja. Stoga je razvidno zašto ćemo svoje obrazlaganje ograničiti na epistemičku racionalnost budući da je ona, najšire moguće, shvaćena kao racionalnost vjerovanja.

Keith Lehrer (2004: 251) navodi osam vrsta racionalnosti. Budući da nam ograničenost ovog rada ne dopušta da potanko analiziramo svih osam, osvrnut ćemo se samo na praktičnu i teorijsku jer smo mišljenja da se kritičko mišljenje bavi (ili bi se trebalo baviti) jednom i drugom. Praktična racionalnost je *vezana uz ono što pojedinac čini, namjerava učiniti ili, jednostavno, što preferira učiniti. (...) Racionalna osoba je ona koja slijedi razum, a uloga razuma je procjenjivanje onoga što osoba čini ili želi učiniti* (Lehrer 2004: 252). S druge strane, teorijska racionalnost se ne bavi toliko djelovanjem, već *zaključivanjem i prihvaćanjem* (Lehrer 2004: 257). Još jedno važno načelo koje je, prema Lehreru, nužno da bi osoba bila racionalna jest načelo dinamike kojim se izbjegava upadanje u paradoks dija-

<sup>65</sup> *Filozofski leksikon* (2012: 960).

<sup>66</sup> Ibid.

<sup>67</sup> Ibid.

kroničkog razuma.<sup>68</sup> Riječ je o tome da se naša racionalnost djelomično sastoji od toga da budemo racionalni u *mijenjanju onoga što preferiramo, prihvaćamo i kako zaključujemo* (Lehrer 2004: 264). Dakle, da bismo bili racionalni, ključno je da smo skloni mijenjati načine mišljenja, ispravljati vlastite pogreške te učiti iz iskustva.<sup>69</sup>

## Proces argumentacije

U svom članku *Može li se naučiti kritički misliti?* Nenad Smokrović (2021: u tisku) ističe:

Danas je govor o argumentiranoj raspravi, školskim kurikulumima kao i rješavanju različitih vrsta dijaloških sporova postao nezamisliv bez spominjanja kritičkog mišljenja. Problematičnim, međutim, ostaje da je unatoč tome razumijevanje pojma kritičkog mišljenja danas u svom optimizmu nedovoljno elaborirano kako s obzirom na svoje temeljno određenje tako i s obzirom na svoju ulogu i primjenu.

Kritičko mišljenje, ima isto svojstvo koje i Smokrović spominje za argumentaciju (točnije, proces argumentacije), a to je da nakon ostvarenja istog, osoba dolazi do točke kada zna više nego što je znala započevši cijeli proces (Smokrović 2015: 225). Dakle, osim što kritičko mišljenje obogaćuje samo rasuđivanje, ono – na neki način – i proširuje znanje.

Smokrović navodi, na Williamsonovom tragu, da ponekad moramo sagledati vlastita uvjerenja odmaknuvši se korak unatrag kako bismo izbjegli *konfirmacijsku predrasudu* (intelektualnu manu koja označava našu tendenciju da pronalazimo one informacije koje potvrđuju naša uvjerenja). Smokrović se tu nadovezuje na Williamsonovu interpretaciju *koraka unatrag* te tvrdi da upravo sama struktura argumenta, na neki način, tjera sudionike na taj korak, kao i to da pospješuje inferencijske sposobnosti (Smokrović 2015: 227). Argument kao temeljna *građevna* struktura u središtu je Smokrovićevog viđenja procesa argumentacije, kao i samog procesa kritičkog mišljenja.

Također, i proces argumentacije, a i proces kritičkog mišljenja podrazumijevaju određene forme intelektualnih vrlina. Čini se da je njegovanje intelektualnih vrlina i susprezanje intelektualnih te kognitivnih mana ključno za kvalitetno postizanje kritičkoga mišljenja. Relativno lako je zamisliti osobu koja ima vrlo razvijene ilativne sposobnosti, no nedostaju joj intelektualne vrline, odnosno zadržava i ustraje na intelektualnim

<sup>68</sup> Više o tome u: Lehrer 2004: 264.

<sup>69</sup> Lehrerovo načelo dinamike potpuno je u duhu onoga što je John Dewey ponudio kao definiciju odgoja, a prema kojoj je cilj odgoja stalno preuređenje vlastite osobnosti. Usp. Dewey 1970: 39.

manama. Unatoč dobrim (ili čak izvrsnim) ilativnim sposobnostima, teoretičari kritičkoga mišljenja neće takvu osobu nazvati dobrim kritičkim misliteljem. Upravo zato zaključujemo kako su intelektualne vrline nužan aspekt kritičkoga mišljenja. Primjerice, Smokrović (2015: 225) kao jednu od ključnih intelektualnih vrlina za proces argumentacije navodi *propozicijsku radoznalost* (radoznalost koja se odnosi na to da sudionike argumentacije zanima je li neka propozicija P istinita ili nije), a koju neki autori svrstavaju pod epistemičku vrlinu kao *znatiželju* (primjerice Alfano i Whitchomb<sup>70</sup>) ili intelektualnu vrlinu kao *ljubav prema znanju* (primjerice Snježana Prijić-Samaržija<sup>71</sup>). Bez da potanko predstavljamo svaku od intelektualnih vrlina i njihovu važnost za kritičko mišljenje, smatramo dovoljnim naglasiti kako je, primjerice, spomenuta intelektualna vrlina gotovo nužna kao temelj bilo kakve potrage za znanjem, stoga i ne čudi da je Smokrović posebno ističe kao predispoziciju za proces argumentacije. Kao što je *propozicijska radoznalost/znatiželja/ljubav prema znanju* ključna za kritičko mišljenje, tako i ostale intelektualne vrline imaju podjednaku važnost.<sup>72</sup>

Međutim, osim posjedovanja, odnosno razvitka ili makar imanja dispozicija za intelektualne vrline, kritičko mišljenje podrazumijeva još aspekata, što ćemo razmotriti u nastavku.

## Kritičko mišljenje

Učenje općenito najčešće se odvija u grupama, odnosno, u hrvatskom su obrazovnom sustavu to razredni odjeli. Debatne skupine također se baziraju na grupnom radu; dvije grupe učenika raspravljaju o nekoj temi polazeći od dva različita gledišta prema toj temi. Rasuđivanje, kao glavna funkcija razuma, također se učinkovitije događa u grupama, o čemu govori i Nenad Smokrović pozivajući se na istraživanja Merciera i Sperbera.<sup>73</sup> Ako je rasuđivanje doista učinkovitije u grupi, nema razloga da i podučavanje kritičkoga mišljenja u grupi ne bude bolje nego li bi bilo koristeći individualni pristup. U obrazovnim institucijama najčešće se koristi grupni pristup.<sup>74</sup>

<sup>70</sup> Vidi: Turri, Alfano i Greco (2017)

<sup>71</sup> Prijić-Samaržija 2020: 5-26.

<sup>72</sup> Čitatelja kojega zanima koncizno i precizno predstavljanje nekih intelektualnih vrlina upućujemo na: *Ibid.*, str. 7-8.; također vidi: Elder 2008: 24-25.

<sup>73</sup> Smokrović 2015: 224.

<sup>74</sup> Ovdje grupnim pristupom smatramo rad jednog nastavnika s grupom učenika – obično razredom, naspram individualnog pristupa (rada jednog nastavnika s jednim učenicom). Iako je nakon reforme školstva HNOS preporuka bila čim više koristiti grupni rad, mnogi su se nastavnici našli u problemima jer nisu točno znali što se misli pod grupnim radom i kako ga provoditi. Za proces argumentacije dovoljno je najmanje dvoje ljudi koji će zajedno dolaziti do određenih zaključaka. Hrvatski obrazovni sustav već je zasnovan na grupi kao što je razred, međutim mišljenja smo da u našem susta-

U ovom radu u fokusu su upravo odgojno-obrazovne institucije.<sup>75</sup> Dakle, tipična odgojno-obrazovna institucija ima plodno tlo za razvitak i podučavanje rasuđivanja. Drugo je pitanje razvijaju li odgojno-obrazovne institucije moć rasuđivanja pa tako i kritičkog mišljenja, ili se njihova poduka temelji uglavnom na autoritetu, što bi značilo da ne potiču učenike na postavljanje pitanja, da ne koriste često metodu heurističkog razgovora, da sumnja u pruženo znanje nije dobrodošla te da se učenicima informacije najčešće prezentiraju iz one perspektive koja najviše odgovara učitelju itd. Nije potrebno puno objašnjavati da autoritarni način podučavanja pretvara plodno tlo za kritičko mišljenje u jalovo.

U ovom se radu odlučujemo podržati tzv. optimističan stav prema podučavanju kritičkog mišljenja. On se ne odnosi na to da kritičko mišljenje podučava velik broj odgojno-obrazovnih institucija (iako se neke nedvojbeno trude), već se odnosi na samu mogućnost podučavanja kritičkog mišljenja, budući da se i to može dovesti u pitanje. Postoji nekoliko prepreka za učenje, uvježbavanje, pa samim time i podučavanje kritičkog mišljenja. U prvom redu to su logičke pogreške, zatim kognitivne predrasude, intelektualne mane te neznanje.<sup>76</sup> Unatoč tim preprekama, koje predstavljaju teškoće za sve ljude, smatramo da se kritičko mišljenje može podučavati te, stoga, i naučiti, tj. uvježbati. Pitanje je samo u kojoj mjeri.

No što je zapravo kritičko mišljenje? Tko god je pročitao makar jedan *Uvod u kritičko mišljenje* sigurno je pri početku, implicitno ili eksplicitno, primijetio da postoje mnoge definicije kritičkog mišljenja, najmanje dvije interpretacije istog te da ga nije lako (ako ne i nemoguće) definirati, što većina autora naglašava. U svom *Uvodu u kritičko mišljenje* Alec Fisher<sup>77</sup> nudi nekoliko definicija započevši s Deweyevim *reflektivnim mišljenjem* (Fisher

---

vu razredi često sadrže prevelik broj učenika što onemogućuje kvalitetno provođenje nastavne metode heurističkog razgovora koja je izvrsna podloga za vježbanje kritičkog mišljenja. Ako nastavnik pritom inzistira na korištenju argumenata i zajedno s grupom (razredom) provjerava jesu li dani argumenti dovoljno dobri, heuristički razgovor prerasta u proces argumentacije u kojem sudjeluje cijeli razred. Još jednom naglašavamo da su razredni odjeli uglavnom prebrojni da bi se taj proces provodio dovoljno kvalitetno i da bi svi učenici mogli jednako kvalitetno sudjelovati u dijalogu, no to ne znači da se kvalitetniji način provedbe ne može ostvariti u, primjerice, izvannastavnim skupinama.

<sup>75</sup> Iako želimo naglasiti da nam ni na koji način nije namjera umanjiti osobna nastojanja osoba koje same žele *uvježbati* kritičko mišljenje, bilo da se radi o samo-usavršavanju neovisnom o određenoj instituciji ili individualnom interesu unutar neke institucije.

<sup>76</sup> Preuzeto s radionice dr. Geralda Nosicha, održane 15. i 16. svibnja 2021. (online) pod nazivom *Foundations of Critical Thinking: Placing Critical Thinking at the Core of Higher Education*, u organizaciji The Foundation for Critical Thinking, na kojoj je jedna od autorica članka sudjelovala.

<sup>77</sup> Fisher (2011).



2011: 2). Budući da je John Dewey, jedan od najvećih filozofa odgoja i obrazovanja, prvi definirao i uveo pojam *reflektivnog mišljenja*, Fisher ga opravdano smatra *ocem* kritičkoga mišljenja.<sup>78</sup> On navodi Deweyevu definiciju koju kasnije nadograđuje, odnosno proširuje *novijim* definicijama i spoznajama autora koji su se bavili kritičkim mišljenjem. Između ostalih, navodi Glaserovu definiciju (Fisher 2011: 3), Ennisovu (Fisher 2011: 4) te, konačno, Paulovu koja je ponešto drugačija od prethodnih budući da naglašava metakogniciju:

Critical thinking is that mode of thinking – about any subject, content, or problem – in which the thinker improves the quality of his or her thinking by skillfully taking charge of the structures inherent in thinking and imposing intellectual standards upon them. (Paul, Fisher, and Nosich 1993: 4)<sup>79</sup>

Osim što naglašava metakogniciju<sup>80</sup> (razmišljanje o vlastitom razmišljanju), definicija Richarda Paula za naš je rad značajna zato što navodi pridjev *skillfully* (vješto) koji implicira vještinu, a vještina implicira to da se može uvježbati. Također, Paul navodi i upotrebu *intelektualnih standarda* za koje škola (organizacija) nastala na temeljima njegovog učenja (The Foundation for Critical Thinking) smatra da su jedan od tri kamena-temeljca za uspješno učenje i poučavanje kritičkog mišljenja (uz elemente mišljenja i intelektualne vrline)<sup>81</sup>. Ono što je još ključno za rad ove organizacije je naglasak na praktičnoj upotrebi alata za poboljšanje kritičkog mišljenja i na podučavanju uporabe tih alata. Budući da se u ovom radu bavimo upravo podučavanjem kritičkoga mišljenja u praksi, očigledno je zbog čega smo se odlučile baš za definiciju Richarda Paula.<sup>82</sup>

S druge strane, Nenad Smokrović nema toliko optimističan stav poput spomenute Paulove škole ili poput onoga koji mi nastojimo braniti u ovom radu. Njegov stav nije potpuno pesimističan, no uglavnom jest s obzirom na strukturu kritičkog mišljenja koju nudi. Smokrović dijeli kritičko mišljenje na četiri aspekta: 1) sposobnost uočavanja ilativnih odnosa (između razloga i tvrdnje koju oni podupiru); 2) kognitivne sposobnosti koje

<sup>78</sup> Snježana Prijić-Samaržija daje, primjerice, prednost Deweyevoj definiciji naspram Heumerove budući da Deweyeva definicija ne isključuje povjerenje u stručnjake (epistemičke autoritete). Vidi: Prijić-Samaržija 2020: 12.

<sup>79</sup> Citirano prema: Fisher 2011: 5.

<sup>80</sup> U samom procesu argumentacije, Williamsonov i Smokrovićev *korak unatrag* možemo također razumjeti kao meta-kogniciju budući da se od sudionika rasprave očekuje da iz perspektive svjedoka razmotre vlastiti način zaključivanja i iznošenja razloga.

<sup>81</sup> Vidi: Elder 2008: 19.

<sup>82</sup> Time, naravno, ne želimo implicirati da su definicije drugih autora manjkave ili netočne. Samo naglašavamo da Paulova definicija sjajno služi svrsi ovoga članka.

omogućavaju ilativne sposobnosti; 3) epistemičke vrline; 4) dispozicije na kojima se zasnivaju epistemičke vrline (Smokrović 2021: u tisku). Nadalje, Smokrović (2021: u tisku) ističe:

(...) važno je reći da je dominantno mišljenje unutar kognitivne teorije da je nerealno očekivati poboljšanja ili meliorativne pomake unutar inferencijalne domene bilo u deduktivnom bilo u induktivnom području. Pogotovo je to nerealno očekivati kroz učenje 'ispravnih' pravila zaključivanja i ukazivanja na tzv. logičke pogreške, kako se to često naivno pretpostavlja u teorijama kritičkog mišljenja.

Budući da Smokrović navodi kako kognitivne sposobnosti stoje u osnovi ilativnih sposobnosti i da su kognitivne sposobnosti (poput kapaciteta radne memorije, brzine percipiranja, efikasnosti prizivanja informacija iz dugoročne memorije i točnosti razlikovanja) uglavnom nesvjesne i automatske, na temelju toga zaključuje kako je teško *pretpostaviti da ono što se nalazi ispod praga svjesnog i refleksivnog može biti podvrgnuto učenju i poboljšanju kroz učenje* (Ibid.).<sup>83</sup>

Mjesta na kojima Smokrović pokazuje nešto optimističniji stav odnose se na zadnja dva aspekta (njegove) strukture kritičkoga mišljenja: epistemičke vrline i dispozicije za njih. Za njega je najbolji način učenja *epistemičkih dobara* proces argumentacije, dakle praksa argumentiranja te usvajanja onih strategija koje se u praksi argumentacije pokažu uspješnijima u otklanjanju neslaganja (Smokrović 2021: u tisku).

Međutim, iako se možemo složiti da proces argumentacije zacijelo predstavlja ključni način usvajanja kritičkog mišljenja, kao i razvijanja istog u praksi, smatramo da bi kritičko mišljenje trebalo prožimati čitavo mišljenje. Nije važno radi li se samo o tome da želimo prikazati svoj argument boljim i uvjerljivijim od protivnikova, važno je da znamo kritički misliti i kad prosuđujemo argumente u različitim medijima, kao i znanstvene članke (primjerice, možemo se pitati je li uopće novinar/znanstvenik ponudio dobar argument za poziciju koju zastupa). Definicija Richarda Paula navodi da je kritičko mišljenje razmišljanje na određeni način *o bilo kojem sadržaju, temi ili problemu*. Stoga, smatramo da kritičko mišljenje nije samo predmet teorijske, već u jednakoj mjeri i praktične racionalnosti, a Lehrerovo načelo dinamike zapravo je *modus operandi* kritičkog mišljenja.

Dakle, ako osoba želi kritički prosuditi svjetonazor unutar kojeg je od-

<sup>83</sup> Iako u svom članku iz 2015. Smokrović (2015:231) zaključuje: (...) *Using reliable methods, participants are in the position to reach safe knowledge. In this way, the argumentation is a media that increases their deductive competences*. Znači li to da argumentacijom ipak možemo uvježbati ilativne sposobnosti ili je to uvjerenje koje je Smokrović napustio, sudeći po najnovijem članku iz 2021. čija se objava očekuje?

gojena i za kojeg je naučena da ga uzima zdravo za gotovo s drugačijim svjetonazorom, ona to čini zapravo u ulozi svjedoka procesa argumentacije, a ne kao njegov direktni sudionik. Ona to čini kao tražitelj istine, a ne kao onaj koji direktno nudi najbolje razloge za istinu<sup>84</sup>. Mišljenja smo da kritičko mišljenje treba, uz samu argumentaciju, pokrivati i ovaj aspekt – *svjedoka*, odnosno *promatrača* samog procesa argumentacije. Međutim, tu su i mnoge druge situacije svakodnevnog života koje se ne mogu nužno reducirati na proces argumentacije. Primjerice, Paul i Elder govore kako kritičko mišljenje znači koristiti *discipliniranu umjetnost* kojom osiguravamo da koristimo najbolje moguće razmišljanje za koje smo sposobni, u *svim okolnostima* (Paul i Elder 2002: 26). Te se okolnosti odnose, osim na sam proces argumentacije, i na procjenjivanje svakodnevnih situacija, od samo-korekcije (*Kako mogu biti uspješnija u...?*), samo-spoznavje (*Zašto sam reagirala tako kako jesam?*), do životnih odluka (*Trebam li raditi taj posao?*; *Trebam li ostati s tim partnerom?*) i preispitivanja vjerovanja (*Varam li samu sebe kada vjerujem...?*; *Vjerujem li ovo samo zato što želim da bude istina?*). Na temelju ovog što je rečeno, može se zaključiti da kritički mislitelj mora izoštriti vlastiti način razmišljanja koliko je god to moguće i preuzeti ga kao glavnu vodilju kroz život, kao filter kroz koji provlači sve situacije i okolnosti koje doživljava, a za koje je to moguće<sup>85</sup>, kao i vlastita svjesno ili manje svjesno usvojena vjerovanja.

---

<sup>84</sup> Primjerice, osoba odgojena unutar određenog svjetonazora, suočena (preko obrazovanja, društvenog utjecaja ili zbog nekog drugog razloga) s drugačijim ili čak oprečnim svjetonazorom, prati proces argumentacije između dva suprotstavljena svjetonazora. Prosuđuje argumente koji su – po njenom sudu – uvjerljiviji. Prosuđuje razloge koje obje strane nude za svoje stavove i tek nakon upoznavanja s argumentima i razlozima jedne i druge strane, donosi sud o tome koja vjerovanja valja zadržati, koja odbaciti, a koja pospješiti (npr. dodatnim razlozima). Dok prosuđuje, ne sudjeluje u samom procesu argumentacije, već promatra samu situaciju neslaganja između dva svjetonazora kao promatrač. Dakako, opravdano se možemo upitati što bi nekoga tko je naučen prihvaćati bez pitanja određen svjetonazor motiviralo da ga krene propitivati. To može biti sama želja da se postane bolji kritički mislitelj (pa će tako osoba u ulozi *promatrača* prosuđivati razloge jedne i druge strane, nakon čega može jedne razloge usvojiti, a druge odbaciti, ovisno o tome čije razloge i zašto smatra uvjerljivijima), može biti uočavanje nelogičnosti unutar samoga svjetonazora koje je ranije u nekoj mjeri percipirala, ali im se nije posvećivala do određenog trenutka, može biti sumnja koja se javila u određenom trenutku prema čitavom svjetonazoru ili pak samo nekim njegovim dijelovima, a može biti i samopropitivanje u svrhu samorazvoja (npr. *Je li mi ovo vjerovanje nametnuto odgojem?*; *Zašto smatram da je ovo ovako?* itd.).

<sup>85</sup> Dakako, postoje situacije i okolnosti u kojima treba donijeti brzu odluku za koju osoba nema vremena koristiti alate kritičkog mišljenja. Primjerice, kirurg koji brzo treba odlučiti u nekoj hitnoj situaciji i čije bi korištenje alata kritičkog mišljenja vjerojatno značilo pacijentovu smrt. Ili stresne i hitne situacije u vojnoj, policijskoj, vatrogasnoj i sličnim domenama.

## Teistički svjetonazor

U uvodu ovoga rada napomenule smo da nas prvenstveno zanima je li određeni svjetonazor, posebice religijski, prepreka kritičkom mišljenju i njegovom podučavanju. Preciznije, zanima nas je li unutar odgojno-obrazovne institucije u kojoj se zastupaju i prisutni su određeni svjetonazori, moguće ostvariti obrazovanje za kritičko mišljenje. Počinjemo od teističkog svjetonazora. Što sve uključuje teistički svjetonazor? Je li preuzimanje takvog svjetonazora nužno prepreka za razvitak i odgoj kritičkog mislitelja?

Kada je riječ o Bogu i vjerovanju u Boga postoje zapravo tri temeljne pozicije. Netko može biti ili teist ili ateist ili agnostik. Teist, jednostavno rečeno, vjeruje u Boga, tj. smatra da Bog postoji.<sup>86</sup> Ateist smatra da Bog ne postoji, i da je nemoguće, pa čak i besmisleno, govoriti o vjeri ili vjerovanju u Božje postojanje. Agnostik drži da, s obzirom na to kako nije moguće niti dokazati niti opovrgnuti Božje postojanje, valja ostati suzdržan. Ostavimo li agnostičku poziciju po strani, primijetit ćemo da se i teistička i ateistička strana diče pozivanjem na određene *dokaze*, odnosno razloge koji podupiru njihove tvrdnje. Činjenica je, naime, da svatko vjeruje u ono što smatra istinitim, ponekad (ili često) čak i unatoč dokazima i protu-dokazima. Prebacivanje tereta dokazivanja vodi do stava da teist ostaje pri svom uvjerenju, tj. vjeri u Boga čak i onda kada mu ateist ponudi *dokaze* koji bi ga trebali uzdrmati i poljuljati njegovu poziciju. Ateist pak često niti uz predočenje nekih teističkih *dokaza* ne prihvaća makar *mogućnost* Božjeg postojanja.

Čini se da dolazimo do paradoksalne pozicije. Svatko vjeruje u svoje pretpostavke bez obzira na postojanje dokaza i protu-dokaza.<sup>87</sup> Takva pozicija nije zdravorazumska i ne bismo je očekivali od racionalnih ljudi. No, ona je ipak vrlo česta i uvelike zastupljena u našem društvu. Stoga se na kraju možemo pitati, ima li uopće smisla tražiti dokaze kad svatko tvrdoglavo ostaje pri svojim uvjerenjima i stavovima. Koliko smo zapravo racionalni pri takvom odlučivanju i kada je *racionalno* odustati od vlastitih uvjerenja?

---

<sup>86</sup> Kada govorimo o teizmu uglavnom mislimo na vjeru u jednog Boga, dakle, poziciju koju po pitanju vjere i opisima Božanstva zastupaju monoteističke religije. Preciznije definicije teizma, ateizma i agnosticizma vidi u: Mišić (2000).

<sup>87</sup> Puno je knjiga napisano o tome je li teistički svjetonazor racionalno utemeljen. Preporučamo: Lambert (2003). Autor u knjizi govori o mogućnosti susreta znanosti i teologije.

## Dokazivanje i opovrgavanje

Anthony Flew preuzeo je priču Johna Wisdoma i razvio je u narativ koji dobro naznačuje suprotstavljanje između teista i ateista o kojem smo prethodno govorili. U nastavku donosimo vrlo skraćenu verziju te priče, uz isticanje ključnih elemenata, kako bismo postavili okvire za raspravu o mogućnosti formiranja adekvatnih uvjerenja i stavova. Riječ je o Paraboli o nevidljivom vrtlaru.<sup>88</sup> Teist i ateist šetaju džunglom. U jednom trenutku, usred džungle, naiđu na čistinu na kojoj se prostire prekrasan vrt. Teist smatra da vrt ukazuje na vrtlara (ovdje je vrt Božje djelo, a vrtlar je Bog). Ateist odbacuje takvu mogućnost. Što im je činiti? Potrebno je detaljnije istražiti o čemu se radi, odnosno, proučiti evidenciju za obje tvrdnje. Što god njih dvoje poduzimaju (dnevno, noćno praćenje, pas tragač itd.), vrtlar se ne pojavljuje.

Teist tvrdi da je očito riječ o 'nevidljivom' vrtlaru, ali ateist ne prihvaća takvo obrazloženje i traži dokaz. Kako razriješiti ovaj slučaj? Mnogi ljudi koji čuju ovu priču zaključit će kako je ovdje u pitanju ono što najčešće smatramo tipičnom raspravom između teista i ateista (ili dvije suprotstavljene strane). Ateist prihvaća isključivo opipljive<sup>89</sup> dokaze, dok teist smatra da se u Boga može vjerovati i bez dostatne evidencije, odnosno da je moguće ponuditi neku drugu vrstu *evidencije*. Želi se dakle ukazati na široko prihvaćeno vjerovanje, ili točnije, uvjerenje o – s jedne strane, racionalnosti ateista, a s druge, neracionalnosti ili, bolje rečeno, nerazumnosti teista. Pritom se teiste želi prikazati kao *dogmatike* (ili one koji prihvaćaju tvrdnje bez dokaza) ili čak kao *fanatike*, a ateiste kao *racionalne mislitelje* kod kojih uvijek znanost ima posljednju riječ i koji nikada ne bi prihvatili nešto kao istinito bez da je to prethodno potkrijepljeno dokazom i evidencijom.<sup>90</sup>

Iako je samo vjerovanje u Boga ključno za teistički svjetonazor, osim samog vjerovanja (kao epistemološke kategorije), teistički svjetonazor barata još nekim kategorijama. Nicholas Wolterstorff (2004: 375) navodi:

(...) u mnogim religijskim oblicima života (nasuprot, možda, znanosti),

---

<sup>88</sup> Usp: Flew 1963: 96. Također vidi: Davies 1998: 12-13.

<sup>89</sup> U epistemologiji, vjerovanja se formiraju prema prilično preciznim i jasnim kriterijima. Odličan prikaz osnovnih kriterija i procedura, kao i rasprava iz te domene nalazimo u prijevodu koji je pripremio Zvonimir Čuljak. Više o tome vidi u: Čuljak 2003.

<sup>90</sup> Ovakvom gledištu (ali i raspravi općenito) doprinio je najviše Richard Dawkins i pokret tzv. *neoateista* (*novih ateista*) koji je nastao početkom ovog stoljeća i razvio se uglavnom oko publikacija spomenutog Dawkinsa, zatim Christophera Hitchensa, Daniela Dennetta i Sama Harrisra. Za *nove ateiste* tipično je da poistovjećuju pojmove poput praznovjerja, religije i iracionalnosti. No isto tako smatraju svojom dužnošću ukazivati na iracionalnost ljudi religijskih svjetonazora te čak i na smanjenu inteligenciju istih. Vidi: Dennett 2009: 261-267; 291-292.

nada, pouzdanje, žaljenje, prihvaćanje itd. jednako su istaknuti kao i vjerovanje.

Također, Wolterstorff opisuje, na neki način, *promašenost* u epistemološkim nastojanjima u kojima se zahtijeva da vjerovanja nužno trebaju biti *opravdana* i *racionalna* kako bi se smatrala istinitima. Dodaje:

Moja je tvrdnja, nasuprot tome, da postoji velik broj istinosno relevantnih zasluga u vjerovanjima, i da ni „opravdanje“ niti „racionalnost“ ne pogađaju nijednu takvu pojedinačnu zaslugu; oba su pojma veoma dvosmislena, i svaki od njih pogađa niz različitih vrijednosti. (Wolterstorff 2004: 374)

U suvremenoj epistemologiji religije, prilično je poznat pravac reformirane epistemologije čiji su predstavnici, primjerice, Alvin Plantinga i William Alston. Oni smatraju da za vjerovanje u Boga nije potrebna niti evidencija niti argument. Tako Alston u djelu *Percieving God*, ukratko rečeno, tvrdi da u Boga možemo vjerovati na temelju religijskog iskustva.<sup>91</sup> A Plantinga u svom podužem članku *Reason and Belief in God* zastupa tezu da se religijsko vjerovanje može tretirati kao temeljno vjerovanje (*properly basic belief*) za koje nam, kao ni za matematičke aksiome, nije potrebna evidencija. Za njega temeljna vjerovanja služe za opravdanje ostalih vjerovanja.<sup>92</sup>

U ovom članku nije nam namjera ulaziti u dublju raspravu o tome koja je strana u pravu (novi ateisti ili reformirani epistemolozi), već smatramo ključnim naglasiti da su obje strane dužne provoditi kritičko mišljenje i sagledavati argumente protivne strane pritom njegujući intelektualne vrline i osvještavajući vlastite intelektualne mane i predrasude.

## Teistički svjetonazor i kritičko mišljenje

Vidjeli smo da teisti za svoj svjetonazor mogu imati opravdane razloge. Ne vjeruju neutemeljeno, već postoje razlozi za njihova uvjerenja. Bez obzira na to činili se ti razlozi nekome dovoljnima ili ne, razloga ima. S druge strane, bilo kakva čvrsta uvjerenja često nisu spojiva s idealima kritičkog mišljenja koje uvijek pretpostavlja duboko propitivanje svega, pogotovo čvrstih uvjerenja. Stoga se pitamo je li odgoj kritičkog mislitelja uopće moguć pored već formiranih stavova i pozadinskih uvjerenja? Ovdje govorimo, primjerice, o odgoju unutar obitelji, ali i o odgoju u obrazovnim ustanovama u kojima je nedvojbeno prisutan određen svjetonazor, poput primjerice katoličkih škola ili škola čiji su planovi i programi proizašli iz određenih pedagogija koje njeguju određenu svjetonazorsku pozadinu.

---

<sup>91</sup> Vidi: Alston 1993; vidi još: Alston 1983.

<sup>92</sup> Vidi više u: Plantinga 1983; također vidi Wolterstorff 1983.

Neke društvene institucije teističkih svjetonazora prvenstveno povezuju svoja uvjerenja s moralnim i životnim vrijednostima (u smislu načina življenja najboljeg života)<sup>93</sup>, smislom života te, na koncu, s pogledom na stvarnost. Stoga se postavlja pitanje je li netko tko smatra da je njegova istina očita, sposoban odgajati učenike i mlade čim objektivnije, otvorenog uma, pružati različite točke gledišta (čak i one za koje smatra da su potpuno pogrešne!)? Sve navedeno predstavlja odlike kritičkog mislitelja i čini se da su one teško spojive s čvrstim svjetonazorskim uvjerenjima.

Znači li to da spomenute institucije ne mogu odgajati kritičke mislitelje? Ne, svakako ne. Bez obzira na svjetonazor, svaka bi institucija (a pritom mislimo na škole) trebala biti sposobna odgajati kritičkog mislitelja. Pitanje je to ponajprije želje, odnosno volje u prvom redu roditelja koji dijete upisuju u određenu školu. Žele li roditelji – vjernici odgajati dijete s čim manje indoktrinacije<sup>94</sup>? Hrvatska stvarnost pokazuje da to baš i nije čest slučaj, no to svakako ne znači da takvi roditelji nisu sposobni za to ili da ne bi trebali težiti tome žele li odgojiti čovjeka koji misli svojom glavom. Dakako, isto se može reći i za institucije (obitelji, škole) koje njeguju ateistički (agnostički) svjetonazor. Čini se da nikome ne odgovara da odgajanik prijeđe na suprotnu stranu. No, iako takva nelagoda može biti razumljiva, ako želimo odgajati kritičke mislitelje, trebali bismo dozvoliti i to – ako razumno procijene da je druga strana plauzibilnija – treba im dozvoliti pravo da *prijeđu liniju*, da se tako izrazimo.

Dakle, odgoj kritičkog mislitelja trebao bi biti odgoj s čim manje indoktrinacije, čim manje tzv. zdravo-za-gotovo prihvaćenih stavova i uvjerenja, čim manje čvrstih i fiksiranih ideja koje su imune (ili se takvima predstavljaju) na propitivanje. Osobi koju se odgaja treba dozvoliti propitivanje svega: moralnih stavova i uvjerenja, vjerskih stavova i uvjerenja (bili oni teistički ili ateistički), životnih stavova i uvjerenja, političkih stavova i uvjerenja pa čak i znanstvenih stavova i uvjerenja. Za kritičko mišljenje nema gotovih, u kamenu upisanih istina. Propitivanjem se dolazi do potvrda ili opovrgavanja određenih stavova, teza, uvjerenja itd. Stoga bi svi koji žele odgajati kritičke mislitelje trebali poticati propitivanje svega o čemu se uči.

---

<sup>93</sup> Ovdje nije moguće ulaziti u raspravu o povezanosti življenja kvalitetnog života i odabira određenog svjetonazora, no ljudi su svakako skloni vlastiti svjetonazor povezivati s kvalitetom življenja. Više o kriterijima za dobar život vidi: Gavran Miloš, Ana (2021).

<sup>94</sup> Ovdje indoktrinaciju razumijemo kao postupak nametanja određene ideologije bez da su pruženi racionalni razlozi za istinitost te ideologije. U ovom slučaju racionalni razlozi nisu empirijski dokazi budući da sumnjamo da možemo o njima govoriti u domeni religije.



## Ateistički svjetonazor

U prethodno spomenutoj Paraboli o nevidljivom vrtlaru, Flew želi prikazati kako bi svaka rasprava o istinitosti teizma trebala započeti zapravo ateističkom pretpostavkom. Ta bi se početna pretpostavka, po njegovu mišljenju, trebala držati istinitom sve dok teistička strana ne ponudi empirijske dokaze koji bi je opovrgnuli. Osim tog zahtjeva, ono što je također izazvalo reakcije kod filozofâ jest da obje strane čvrsto stoje iza svojeg početnog mišljenja i uvjerenja unatoč svim ponuđenim razlozima i *dokazima*, a to se – tvrdi Flew – odnosi pogotovo na teiste, budući da su oni ti na koje je prebačen teret dokaza (Flew; MacIntyre 1963: 97-99).

Bez obzira koliko teist vjerovao i imao religijskog iskustva, njegovo vjeronaučno neće nikad biti prihvaćeno od svih kao dovoljno opravdano (iako za njega samoga može biti). Isto tako, bez obzira na to koliko ateist ne vjerovao i smatrao *očitim* da ne postoji Bog, činjenica je da on to ne može znati. Može možda, na temelju nedostatka empirijskih dokaza, tvrditi da je 99% siguran kako Boga nema, no – ako je racionalan – trebao bi ostaviti barem taj 1% vjerojatnosti da je u krivu. Richard Dawkins želi prikazati teiste potpunim fanaticima koji tvrdoglavo vjeruju u bajke<sup>95</sup>, no način na koji to čini, ne demonstrirajući apsolutno nikakvu intelektualnu poniznost, počinje također djelovati kao fanatizam, samo s drugog kraja spektra.

## Ateistički svjetonazor, sekularnost i kritičko mišljenje

Danas često poistovjećujemo sekularizam i ateizam, međutim to nisu sinonimi. Sekularizam zanemaruje vjerska uvjerenja i ideje te se zalaže da budu u domeni privatnog. Ako želimo javno surađivati, zanemarivanje vlastitog religijskog svjetonazora, nužno je – tvrdi sekularizam. Religijski svjetonazor ne bi trebao imati svoje mjesto u javnim institucijama. U Hrvatskoj je većina državnih (javnih) odgojno-obrazovnih institucija nominalno sekularna, iako se uči religijsko obrazovanje (vjeronauk) i slave se razni blagdani koji spadaju u religijsku domenu. U nacionalnim kurikulumima (posebice nakon reforme obrazovanja HNOS), često se spominje poticanje kritičkog mišljenja, iako se u praksi to načelo većinom ne sprovodi. Ne ulazeći sada u probleme zbog kojih je to tako, želimo razjasniti je li uopće moguće inzistirati na kritičkom mišljenju u strogo sekularnoj instituciji.

Naglašavamo da strogo sekularnih odgojno-obrazovnih institucija u Hrvatskoj nema. Blagdani Božića i Uskrsa slave se u svakoj. U školama se organiziraju izložbe kruha za Dane kruha, a doneseni kruh nerijetko dolazi blagosloviti katolički svećenik, dok djecu drugih svjetonazora (vje-

---

<sup>95</sup> Usp. Dawkins, Richard 2007: 50-51.



roispovijesti) za to vrijeme čuvaju uglavnom pripravnici ili knjižničari. Sve je to daleko od strogog sekularizma, no također je i daleko od poticanja kritičkog mišljenja. Zašto? Nema ništa loše u poštivanju tradicija, no ako se institucija smatra sekularnom, religijski svjetonazor doista bi trebao biti u domeni privatnog – van institucije.

Budući da je taj zahtjev vjerojatno prejak, trebalo bi makar težiti smanjenju dvostrukih mjerila (naspram ostalih svjetonazora) kao i tzv. unutar-grupnih preferencija<sup>96</sup>. Ako odgojno-obrazovna institucija doista želi podučavati svoje učenike kritičkom mišljenju, trebala bi uzeti u obzir svjetonazor *svih* učenika i obilježavati *sve* veće blagdane. Tako bi se spriječila dvostruka mjerila (koja obično, u Hrvatskoj, idu u korist katolicizmu), ne bi se favoriziralo pripadnike određenih religijskih svjetonazora dok su ostali osuđeni na boravak u knjižnici. Također, ostalim učenicima, koji se (još) nisu opredijelili za određeni (ne)religijski svjetonazor, omogućio bi se uvid u više religija, učenja i kultura. Tako bi se smanjilo i međusobno nerazumijevanje koje je često prisutno, posebice u ruralnim sredinama.

Uvođenje posebnog predmeta u institucije ne-religijskih svjetonazora koji bi obuhvaćao sve religijske svjetonazore (i koji postoje i koji ne postoje u Hrvatskoj)<sup>97</sup>, koji bi podučavao o nastanku, antropologiji i sociologiji tih religija, omogućio bi učenicima puno širu sliku o sofisticiranosti religijskog svjetonazora. Takav bi im predmet također pružio dovoljno duboke uvide u kompleksnost same tematike nastanka, srodnosti i prakticiranja religijskih praksi te bi svakako ti učenici imali bolje temelje za racionalniju odluku o vlastitom (ne)religijskom svjetonazoru.

Trenutno u tzv. sekularnim institucijama nemamo praksu jednakog odnosa prema svim religijskim svjetonazorima. Još uvijek je sporno trebaju li državne škole i vrtići imati križeve u učionicama te se čini da institucija koja želi provoditi obrazovanje za kritičko mišljenje u pogledu religijskog svjetonazora i dijaloga, treba prestati favorizirati jednu religiju ma koliko ona tradicionalno bila prisutna. Ako se već odbija maknuti religijske simbole iz učionica, onda bi ih se trebalo rezervirati za učionicu u kojima se uči o religijama te bi u istoj trebalo biti prisutno čim više religijskih simbola (osim samog križa), a pogotovo bi trebali biti zastupljeni simboli religijskih svjetonazora svih učenika (pa čak i onih koji bi se svrstali u agnostike

---

<sup>96</sup> Jedna od kognitivnih predrasuda (engl. *in-group bias/in-group favoritism*) koja označava našu tendenciju da favoriziramo one koji se nalaze unutar naše grupe za razliku od onih koji su van grupe kojoj pripadamo.

<sup>97</sup> Trenutno je predviđeno u planu i programu nastave Etike u srednjim školama, učenje o svjetskim religijama, no smatramo da je tih nekoliko lekcija nedovoljno da pruži istinsko razumijevanje i dublju informiranost učenika kako bi se mogli sami, primjerice, odlučiti za određenu religiju, ili pak kako bi je mogli odbaciti.

ili ateiste). Budući da to još uvijek nije tako, i pitanje je hoće li ikad biti, *sekularne* obrazovne institucije u Hrvatskoj, u globalnom, multikulturalnom svijetu zapravo demonstriraju jednostranost, dvostruka mjerila te potiču kulturu neznanja.

## Zaključak

Na tragu Nenada Smokrovića, ako ćemo se složiti da se kritičko mišljenje najbolje uči i razvija kroz proces argumentacije, u duhu same škole Richarda Paula i Linde Elder, ključno je da obje strane u procesu budu informacijski dobro potkovane. Dakle, obje bi strane trebale razumjeti s kojim uvjerenjima barataju, koliko su ona utemeljena na činjenicama, a koliko na nečem drugom (primjerice, na autoritetima u području, na tradiciji zajednice, na funkcionalnosti, instrumentalnosti i sl.) te koji su razlozi jedne i druge strane da drže samo svoja uvjerenja istinitima.

Elementi mišljenja, po učenju Richarda Paula, koji su temelj za samo kritičko mišljenje (uz intelektualne vrline i standarde mišljenja) su: svrha/e, ključna pitanja, perspektive, informacije, zaključci, koncepti, implikacije i pretpostavke<sup>98</sup>. Smatramo da je nemoguće sprovesti kvalitetan proces argumentacije bez uključenja tih elemenata mišljenja. Također, smatramo da bi obrazovne i odgojne ustanove trebale omogućiti svojim učenicima kvalitetan uvid u spomenute informacije o uvjerenjima kako bi im omogućile ispravan proces argumentacije koji bi ih mogao inspirirati na još čvršće zauzimanje stava ili revidiranje vlastitog početnog uvjerenja, a sve kroz samu argumentaciju i uz pomoć alata kritičkoga mišljenja<sup>99</sup>.

Isto tako, da bi se kritičko mišljenje ispravno podučavalo, vježbalo, njegovalo, razvijalo i usavršavalo, nužno je usmjeriti pažnju na njegovanje intelektualnih vrlina kao i na osvješćivanje intelektualnih mana. U tom smislu se svakako slažemo sa Smokrovićem da je poboljšanje u ovom dijelu svakako moguće, a mi bismo dodale i ključno, budući da čovjeka koji savršeno koristi deduktivne alate u svom promišljanju no istodobno iskazuje nekoliko intelektualnih mana, ne možemo smatrati istinskim kritičkim misliteljem.

Upravo je možda razvijanje intelektualnih vrlina unutar odgojno-obrazovnih institucija određenog svjetonazora ključni aspekt u približavanju motivacijske atmosfere za njegovanje kritičkog mišljenja. No također je ključno da institucije, odnosno nastavnici – kao njihovi predstavnici – i sami rade na razvijanju i usavršavanju vlastitih intelektualnih vrlina. U domeni religijskih svjetonazora, to bi značilo da svim svjetonazorima

---

<sup>98</sup> Detaljnije vidi u: Paul i Elder 2002: 93-96.

<sup>99</sup> O alatima kritičkoga mišljenja vidi: *Ibid.*; vidi također: Elder 2008.

treba dati jednaku šansu. Sve strane u procesu argumentacije trebaju biti zastupljene bez predrasuda ili favoriziranja, što implicira da se vlastita uvjerenja u određenoj mjeri trebaju suspregnuti kako bi se zaustavila in-doktrinacija i epistemički egoizam te omogućio pravedni i iskreni dijalog.

Razvidno je da to ne odgovara onima koji svoja uvjerenja bespogovorno žele prenijeti na nove generacije, no mi ovdje ne govorimo o takvima. Govorimo o svima koji njeguju određeni svjetonazor i žele dozvoliti odgoj kritičkog mislitelja. Znači li to da tražimo da se gore spomenute institucije trebaju, na neki način, odreći svojih stavova i svjetonazora? Ne. Smatramo da se kritičko mišljenje može učiti i pod svjetonazorskim okriljem, no isto tako smatramo da se taj svjetonazor mora neprekidno zatomljivati ne bi li se osigurao kvalitetan dijalog sa suprotnom stranom.

## Bibliografija

- Alston, William (1983) *Christian Experience and Christian Belief.*, u: Plantinga, Alvin; Wolterstorff, Nicholas (eds.) *Faith and Rationality. Reason and Belief in God.* University of Notre Dame Press. London, str. 103-135.
- Alston, William (1993) *Percieving God.* Cornell University Press. Ithaca and London
- Čuljak, Zvonimir (2003) *Vjerovanje, opravdanje i znanje.* Ibis grafika. Zagreb
- Davies, Brian (1998) *Uvod u filozofiju religije.* Scopus. Zagreb
- Dawkins, Richard (2007) *Iluzija o Bogu.* Izvori. Zagreb
- Dennett, Daniel C. (2009) *Kraj čarolije. Religija kao prirodna pojava.* Jesenski i Turk. Zagreb.
- Dewey, John (1970) *Vaspitanje i demokratija.* Obod. Cetinje
- Elder, Linda (2008) *The Miniture Guide to Critical Thinking Concepts and Tools.* The Foundation for Critical Thinking
- Fisher, Alec (2011) *Critical Thinking. An Introduction. 2nd ed.* Cambridge University Press
- Flew, Anthony (1963) *Theology and Falsification. Essay A.,* u: Flew, Anthony; MacIntyre, Alasdair (eds.) *The New Essays in Philosophical Theology.* William Clowes and Sons. London/Beccles, str. 96-99.
- Flew, Anthony; MacIntyre, Alasdair (eds.) *The New Essays in Philosophical Theology.* William Clowes and Sons. London/Beccles
- Gavran Miloš, Ana (2021) Što filozofi znaju o dobrom životu?, *Prolegomena*, 20, 1, str. 25-51.
- Greco, John; Sosa, Ernest (ur.) *Epistemologija. Vodič u teorije znanja.* Naklada Jesenski i Turk. Zagreb
- Lambert, Dominique (2003) *Znanost i teologija.* Kršćanska sadašnjost. Zagreb
- Lehrer, Keith (2004) *Racionalnost.,* u: Greco, John; Sosa, Ernest (ur.) *Epistemologija. Vodič u teorije znanja.* Naklada Jesenski i Turk. Zagreb, str. 251-267.
- Kutleša, Stipe. gl. ur. (2012) *Filozofski leksikon.* Leksikografski zavod Miroslav Krleža. Čakovec

- Mišić, Anto (2000) *Rječnik filozofskih pojmova*. Verbum. Split
- Paul, Richard W.; Elder, Linda (2002) *Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life*. Financial Times Prentice Hall (e-izdanje)
- Plantinga, Alvin (1983) *Reason and Belief in God.*, u: Plantinga, Alvin; Wolterstorff, Nicholas (eds.) *Faith and Rationality. Reason and Belief in God*. University of Notre Dame Press. London, str. 16-94.
- Plantinga, Alvin; Wolterstorff, Nicholas (eds.) *Faith and Rationality. Reason and Belief in God*. University of Notre Dame Press. London
- Prijčić-Samaržija, Snježana (2020) Isključuje li kritičko mišljenje povjerenje u epistemičke autoritete? *Prolegomena*. 19, 1, str. 5-26.
- Smokrović, Nenad (2015) Argumentation as a Means for Extending Knowledge. *Croatian Journal of Philosophy*, Vol. XIV, 44, str. 223-231.
- Smokrović, Nenad (2021) Može li se naučiti kritički misliti? (u tisku)
- Thaler, Richard (2009) *Poticaj (Nugde). Moguće je donositi bolje odluke o zdravlju, bogatstvu i sreći*. Planetopija. Zagreb
- Thaler, Richard (2020) *Nerazumno ponašanje. Stvaranje bihevioralne ekonomije*. Školska knjiga. Zagreb
- Turri, John; Alfano, Mark; Greco, John (2017) Virtue Epistemology. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/epistemology-virtue/> (04-8-2021)
- Wolterstorff, Nicholas (1983) *Can Belief in God Be Rational If It Has No Foundations?*, u: Plantinga, Alvin; Wolterstorff, Nicholas (eds.) *Faith and Rationality. Reason and Belief in God*. University of Notre Dame Press. London, str. 135-187.
- Wolterstorff, Nicholas (2004) *Epistemologija religije.*, u: Greco, John; Sosa, Ernest (ur.) *Epistemologija. Vodič u teorije znanja*. Naklada Jesenski i Turk. Zagreb, str. 373-399.



MARKO JURJAKO

## Naturalizam i relativnost u pogledu praktičnih razloga

**Sažetak:** Jedno od najznačajnijih pitanja u filozofiji morala jest predstavljaju li moralni zahtjevi ujedno i razloge za djelovanje prema kojima se trebaju ravnati sve racionalne osobe. Prema jednoj koncepciji, moralni apsolutizam tvrdi da sve racionalne osobe imaju dovoljan razlog da poštuju moralne zahtjeve. Prema tom shvaćanju, moralni relativizam je tvrdnja da neće svi racionalni djelatnici imati dovoljan razlog da poštuju iste moralne zahtjeve. Manje je istraženo pitanje što se događa ako pokušamo povezati filozofski naturalizam s pitanjem daju li moralni zahtjevi razloge za djelovanje svim racionalnim djelatnicima. U ovom radu bavimo se potonjim pitanjem. Neki autori smatraju da prihvaćanje filozofskog naturalizma nije važno za ovo pitanje. Drugi autori smatraju da filozofski naturalizam implicira relativizam u pogledu praktičnih razloga koji nije spojiv s moralnim apsolutizmom. U radu argumentiramo za to da su oba gledišta donekle ispravna. Tvrdimo da ako se naturalizam shvati kao ontološka tvrdnja onda se možemo složiti s time da naturalizam ne igra ulogu u razrješenju debate između relativista i apsolutista. Međutim, ako se naturalizam shvati kao metodološko gledište, onda smatramo da naturalizam podržava relativizam u pogledu praktičnih razloga, tj. ideju da moralni zahtjevi ne pružaju razloge za djelovanje svim racionalnim djelatnicima.

**Ključne riječi:** Moralni zahtjevi; moralni relativizam; principi racionalnosti; praktični razlozi.

### 1. Uvod

Poznato je da se Nenad Smokrović u svojoj dugoj akademskoj karijeri, koja sada već prelazi 40 godina, bavio nizom filozofskih tema koja pokrivaju područja filozofije logike, filozofije uma, filozofije psihologije, filozofije jezika, filozofije društvenih znanosti i filozofije matematike.<sup>100</sup> U mnogim od tih područja Nenad je dao svoj originalni znanstveni doprinos. No, osim toga, Nenad je često bio među prvima u Hrvatskoj koji je svojim člancima, knjigama i uredničkim zbornicima upoznao domaćeg čitatelja s najno-

---

<sup>100</sup> Za iscrpni pregled Nenadovih znanstvenih radova, čitatelj se upućuje na njegov profil u bazi CROSBİ: <https://www.bib.irb.hr/pregled/profil/20736>.

vijim filozofskim temama i publikacijama (vidi npr. Mišćević i Smokrović, 2001; Smokrović, 2004). Time je znatno doprinio unaprjeđenju stanja filozofije u Hrvatskoj. Štoviše, češće nego što bi se očekivalo, svoju strast za razmatranjem i korištenjem formalno-logičkih okvira koji prožimaju njegovu filozofsku misao uspio je prenijeti na niz generacija studenata te ih privoljeti da i sami promišljaju uloge koje logika igra u našim intelektualnim životima te kritički preispitavaju njezinu normativnu snagu. S obzirom na filozofske interese koji se odnose na prirodu ljudskog zaključivanja i određivanja normi koje bi trebale regulirati naše mišljenje i ponašanje, nije čudno da se Nenad dosta bavio pojmom racionalnosti.<sup>101</sup> Međutim, manje je poznato da se Nenad također bavio temama iz filozofije morala (točnije metaetike), gdje je dao svoj obol debatama koje se odnose na prirodu razloga za djelovanje te na odnos razloga i moralnih zahtjeva (Smokrović, 2007).

U ovom radu bavit ćemo se jednim aspektom ove teme, a odnosi se na problem određivanja odnosa između pojmova racionalnosti, razloga za djelovanje i moralnih zahtjeva. Jedno od najznačajnijih pitanja u filozofiji morala jest predstavljaju li moralni zahtjevi ujedno i razloge za djelovanje<sup>102</sup> prema kojima se trebaju ravnati sve racionalne osobe. U povijesti filozofije ovo pitanje je dobilo dijametralno suprotne odgovore u radovima Davida Humea (1983/1739) i Immanuela Kanta (2003/1785). Hume je poznat po svojoj sumnji u moć uma da motivira ljude na djelovanje, te je u skladu s time smatrao da niti u moralnoj domeni um ne igra značajnu ulogu (Schaffer, 2015). Kant je nasuprot tome smatrao da postoji praktični um te da će sve racionalne osobe, ako ispravno koriste svoj um, doći do zaključka da imaju razloga ponašati se u skladu s moralnim normama (vidi npr. Korsgaard, 1996). U suvremenoj raspravi filozofi se obično određuju po tom pitanju u odnosu na to prihvaćaju li Humeovo ili Kantovo gledište na

---

<sup>101</sup> Ovdje vrijedi istaknuti Nenadov znanstveni projekt RACIO (<http://racio.uniri.hr>, Hrvatska zaklada za znanost), u sklopu kojeg se bavio različitim aspektima odnosa logike i pravila ispravnog zaključivanja.

<sup>102</sup> U ovom radu, kada govorimo o praktičnim razlozima ili razlozima za djelovanje, onda mislimo na *normativne* ili *opravdavajuće* razloge za djelovanje. Ukratko, to su ona razmatranja koja govore u prilog poduzimanju ili nepoduzimanju neke radnje (Parfit, 2011; Scanlon, 1998). Normativni razlozi se obično razlikuju od motivacijskih razloga (vidi npr. Jurjako, 2010; Smokrović, 2007; Sušnik, 2012). Motivacijski razlozi objašnjavaju zašto je osoba poduzela neku radnju dok normativni razlozi govore zašto bi trebala poduzeti neku radnju. Radi ilustracije, možemo koristiti poznati primjer Bernarda Williama (1981). Ako osoba X popije tekućinu iz čaše jer misli da je u njoj voda, onda znamo razlog koji ju je motivirao na tu radnju. *Mislila* je da se u čaši nalazi voda. Međutim, ta osoba nije imala *normativan* razlog popiti tekućinu iz čaše jer se u njoj ne nalazi voda, već benzin. Štoviše, možemo reći da su činjenice govorile *protiv* toga da popije tekućinu iz čaše.

prirodu praktičnog uma. Humeovci smatraju da su moralni razlozi samo kontingentno povezani s razlozima koje racionalne osobe imaju za djelovanje, dok Kantovci smatraju da sve osobe, bez obzira na svoje početne motivacije, imaju razloga slijediti moralne norme (vidi npr. Jurjako, 2010; Smokrović, 2007; Sušnik, 2009).

U ostatku rada bavit ćemo se jednom zanimljivom instancom ove debate. Gilbert Harman (2000) tvrdi da se glavna razlika između moralnih apsolutista i moralnih relativista sastoji u njihovom stavu prema naturalizmu. Moralni apsolutizam se ovdje shvaća kao gledište prema kojemu postoje univerzalne moralne istine koje nadilaze pojedine kontekste i odnose se na sve racionalne pojedince (vidi npr. Kant, 2003). Nasuprot tome, moralni relativizam se razumije kao gledište prema kojemu su moralne istine relativne u odnosu na povijesne, kulturalne i društvene kontekste (Berčić, 1995; Levy, 2004). Važno je za ovaj rad da Harman definira moralni apsolutizam i relativizam u terminima razloga za djelovanje. Moralni apsolutizam definira kao gledište prema kojemu moralne norme daju svim osobama razloge za djelovanje bez obzira na njihove subjektivne želje, vrijednosti i tome slično, dok moralni relativizam definira kao negaciju tog gledišta. Prema Harmanu, prihvaćanje naturalizma u ovoj raspravi obvezuje na prihvaćanje moralnog relativizma. Michael Smith (2012) odbacuje to gledište i argumentira da je naturalizam nebitan za raspravu između moralnih relativista i apsolutista.

Mi ćemo tvrditi da odgovor na pitanje podržava li naturalizam moralni relativizam ovisi o tome kako se shvati ideja naturalizma u metaetici. Slijedeći Harmana, debatu ćemo razmotriti u kontekstu rasprave o prirodi praktičnih razloga te će nam u fokusu biti pitanje podržava li naturalizam tezu da su praktični razlozi relativni u odnosu na subjektivna motivacijska stanja djelatnika. Nasuprot Smithu, argumentirat ćemo da neke forme naturalizma, naime metodološki naturalizam, podržavaju relativizam u pogledu praktičnih razloga.

Članak je podijeljen na sljedeći način. U drugom odjeljku objašnjavamo kako se moralni apsolutizam i moralni relativizam mogu shvatiti u terminima razloga za djelovanje. U trećem odjeljku predstavljamo Harmanov naturalistički izazov za apsolutizam u pogledu moralnih razloga. U četvrtom odjeljku raspravljamo Smithov (2012) odgovor na Harmanov argument. U sklopu te rasprave uvodimo razliku između ontološkog i metodološkog naturalizma, te tvrdimo da prihvaćanje metodološkog naturalizma podržava Harmanov argument protiv moralnog apsolutizma. U šestom odjeljku, raspravljamo niz prigovora tvrdnji da metodološki naturalizam podržava relativizam u pogledu praktičnih razloga.



Prije glavnog dijela rada, želimo naglasiti doseg naše rasprave. U radu se bavimo pitanjem povlači li naturalizam u filozofiji tezu da su praktični razlozi relativni u odnosu na motivacije i vrijednosti koje prihvaćaju pojedini djelatnici. Kao što smo najavili, branit ćemo potvrđan odgovor na prethodno pitanje. Za nekoga će taj rezultat imati negativne posljedice. Naime, može se argumentirati na način da ako naturalizam povlači to da ne postoje praktični ili moralni razlozi koje bi svako racionalno biće trebalo poštovati, onda se naturalizam treba odbaciti. Također, naturalizam u filozofiji morala bi nekima mogao biti sporan zbog drugih razloga (za oprečna gledišta vidi npr. Cuneo, 2007; Leiter, 2015). Kako god bilo, ta pitanja izlaze izvan okvira naše rasprave te se njima nećemo baviti. Za potrebe ovog rada pretpostavit ćemo da je naturalizam u filozofiji minimalno koherentna pozicija te ćemo nastojati vidjeti što njegovo prihvaćanje implicira za prirodu praktičnih razloga.

## 2. Moralni zahtjevi i praktični razlozi

Harman definira moralni relativizam i apsolutizam u terminima razloga za djelovanja na sljedeći način:

(...) uzimam da je moralni apsolutizam gledište o moralnim razlozima koje ljudi imaju da nešto čine i žele ili se nečemu nadaju. Razumjet ću vjerovanje u apsolutne vrijednosti kao vjerovanje da postoje stvari za koje svatko ima razloga da im se nada ili ih želi. Reći da postoji moralni zakon koji se „odnosi na sve“ je, ovdje stipuliram, reći da svi imaju dovoljnih razloga slijediti taj zakon (Harman, 2000, str. 84)

Ovdje vidimo da Harman stipulira<sup>103</sup> pojam moralnog apsolutizma prema kojemu moralni zahtjevi konstituiraju razloge za djelovanje koji se odnose na sve djelatnike. Drugim riječima, moralni zahtjevi su oni imperativi koje svi imamo dovoljan razlog poštovati. U skladu s prethodnom karakterizacijom, Harman moralni relativizam definira kao negaciju moralnog apsolutizma:

Moralni relativizam negira da postoje univerzalni temeljni moralni zahtjevi i kaže da su različiti ljudi podložni različitim moralnim zahtjevima ovisno o društvenim običajima, praksama, konvencijama, vrijednostima i principima koje prihvaćaju. (Harman, 2000, str. 85)

---

<sup>103</sup> Prateći terminologiju iz suvremene literature koja se bavi filozofskom metodologijom, umjesto „stipulira“, rekli bismo da Harman nudi „eksplicaciju“ pojma moralni apsolutizam (vidi npr. Dutilh Novaes, 2020).

Drugim riječima, moralni relativizam implicira da neće svi imati dovoljan razlog da prihvate moralne zahtjeve i zbog toga se neće svi moralni zahtjevi odnositi na sve djelatnike.

Smith (2012, str. 227) primjećuje dvosmislenost u Harmanovoj karakterizaciji moralnog apsolutizma i relativizma. Nije jasno koji je doseg fraze „moralni zahtjevi se odnose na svakoga“. Na primjer, moguće je da se „svatko“ (engl. *everyone*) odnosi na svaku ljudsku osobu ili svakog mogućeg djelatnika ili na svakog mogućeg racionalnog djelatnika. Budući da Harman ne specificira doseg termina „svatko“, na koji se odnose praktični razlozi, Smith blagonaklono interpretira Harmanovu tvrdnju kao da se odnosi na svako racionalno biće. U tom pogledu slijedit ćemo Smitha te ćemo interpretirati tvrdnju da svatko ima razlog činiti to i to u smislu da svako biće, ako je racionalno, ima razlog učiniti to i to.

### 3. Harmanov naturalistički izazov

Smith iznosi Harmanov argument u prilog moralnom relativizmu na sljedeći način:

Prva premisa: Ako osoba ne namjerava učiniti nešto i to nije zato što je na neki empirijski ustanovljiv način zaključivanjem propustila doći do te odluke – [zbog] nepažnje, nedostatka vremena, nerazmatranja ili neuvažavanja određenih argumenata, nepoznavanja određenih dostupnih dokaza, pogreški u zaključivanju, neke vrste iracionalnosti ili nerazumnosti, ili slabosti volje – tada prema naturalistu osoba ne može imati dovoljan razlog da to učini.

Druga premisa: postoje ljudi, poput određenih profesionalnih kriminalaca, koji se ne ponašaju u skladu s navodnim zahtjevom da ne nanose štetu ili ozlijede drugima, i to nije posljedica bilo kojeg od tih propusta. (Harman, 2000, str. 86–87; citirano u Smith, 2012, str. 230)

Čini se da ako su ove dvije premise istinite onda moralni apsolutizam ne može biti ispravno gledište. Ove premise impliciraju da neki ljudi nemaju dovoljan razlog poštovati moralne zahtjeve koji se, na primjer, odnose na nenanošenje štete ili ozljede drugim ljudima.

Argument može na prvi pogled izgledati neuvjerljivo. Naime, zastupnik moralnog apsolutizma bi mogao reći da se protiv njega pretpostavlja ono što tek treba dokazati. Moralni apsolutist ne mora nužno prihvatiti tvrdnju da njegova pozicija implicira da postoje *dovoljni* razlozi da se poštuju moralne norme. Moglo bi se tvrditi da moralni apsolutizam pretpostavlja da svaka racionalna osoba ima *nekakav* razlog da poštuje moralne norme, no da je moguće da će u nekim konkretnim situacijama postojati neki drugi

razlog koji će prevladati, te u tom smislu neće imati *dovoljan* razlog poštovati moralne norme koje se odnose na nju. Zbog te komplikacije ostavit ćemo po strani pitanje je li moralni apsolutizam točno ili netočno gledište te ćemo se u nastavku rada fokusirati na općenitije pitanje podržava li naturalizam relativnost u pogledu praktičnih razloga.

Prema prvoj premisi argumenta ono što netko ima razloga učiniti svodi se na pitanje što osoba ima razloga namjeravati. Ovdje je ključno pitanje koji je točan odnos između praktičnih razloga i naturalizma. Kako bi odgovorio na to pitanje, Harman pretpostavlja da ako osoba ima razlog nešto učiniti, onda će ona biti u stanju, zaključivanjem, formirati namjeru da to učini, tj. konkluzija njezinog zaključivanja će biti ta namjera (vidi Harman, 1976). Ovdje je vrlo važna ova pretpostavka da postoji veza između razloga i zaključivanja jer čini razloge podlošnima za naturalističko istraživanje. Naime, zaključivanje se ovdje shvaća kao proces koji se može istraživati u sklopu psihologije ili generalnije u kognitivnim znanostima (vidi npr. Sekulić, 2016; Smokrović, 2004). Stoga, ako su razlozi elementi u procesu zaključivanja, onda se i oni mogu znanstveno istraživati.

Nadalje, nije dovoljno pretpostaviti da su razlozi entiteti koji se pripisuju osobi koja ima sposobnost praktičnog zaključivanja, nego je još nužno pretpostaviti da je osoba sposobna zadovoljiti određene *standarde dobrog* zaključivanja (Smokrović, 2001). Na primjer, rekli bismo da nije u potpunosti racionalno ponašanje osobe koja, unatoč tome što želi biti zdrava i u dobroj kondiciji te vjeruje da svakodnevnim vježbanjem po 30 minuta može ostvariti tu želju, uopće ne vježba (vidi npr. Sušnik, 2012). Ta osoba bi prema Harmanu bila iracionalna jer ne odgovara na razloge koji su joj dostupni. Ili da slijedimo alternativnu terminologiju koju nudi Harman, ta osoba ne formira namjeru da vježba putem sposobnosti za praktično zaključivanje jer postoji nekakav defekt u njezinim racionalnim sposobnostima. Slično tome, za neku osobu bismo rekli da je epistemički iracionalna ako se, na primjer, samoobmanjuje da je njezin suprug ne vara unatoč svoj mogućoj dostupnoj građi koja upućuje na suprotno (npr. svaki dan ima ruž na košulji, miriše na tuđi parfem, dolazi kasno doma itd.), i to upravo zato što ne formira vjerovanja na temelju dokazne građe koja joj je dostupna (Jurjako, 2020). Koristeći Harmanovu terminologiju, mogli bismo reći da osoba ne formira prikladno vjerovanje jer ima određeni defekt u sposobnostima za racionalno zaključivanje.

Dakle, prema Harmanu, prva premisa nam kaže da, ako osoba ima dovoljan razlog poduzeti radnju A onda će, pod uvjetom da ne pati od nekakvog racionalnog defekta, formirati namjeru da poduzme radnju A na temelju racionalnog procesa zaključivanja.<sup>104</sup> Kao naturalističko ograničenje

---

<sup>104</sup> Treba uzeti u obzir da proces racionalnog zaključivanja ne mora nužno biti svjestan

na prethodnu karakterizaciju odnosa razloga i racionalnog zaključivanja, Harman u prvoj premisi dodaje da se mogući racionalni defekti moraju moći odrediti putem empirijskih istraživanja. Drugim riječima, da se ovi defekti, kao i razlozi za djelovanje, moraju moći objasniti koristeći razmatranja koja možemo formulirati kroz zdravorazumske ili znanstvene teorije (vidi Smith, 2012, str. 231). U tom pogledu, postoje barem dva načina kako se ideja naturalizma manifestira u Harmanovom argumentu. Prvi se odnosi na poveznicu razloga s procesima zaključivanja, dok se drugi način odnosi na ideju da se iracionalnost u zaključivanju ili ponašanju mora moći utvrditi kroz aposteriorna ili empirijska istraživanja.

Druga premisa Harmanovog argumenta odnosi se na opis tipa osobe za koju se čini na zdravorazumskoj razini da nije iracionalna, no ima neki svoj razlog učiniti nešto što nemaju nužno razloga činiti ostale racionalne osobe. Kako bismo ilustrirali ovu ideju, možemo se poslužiti primjerom koji daje Bernard Williams (1995, str. 39). Williams zamišlja psihopatsku ličnost koja nije dobra prema svojoj supruzi te koja u svojem skupu vrijednosti nema ništa zbog čega bi marila za dobrobit svoje supruge. Pretpostavimo da ta osoba jest psihopat (u smislu kako ga opisuje Nichols, 2002). Dakle, ona nema empatije niti osjeća kajanje za nemoralne stvari koje je u stanju počinuti. Većinu ljudi tretira kao da su predmeti kojima može manipulirati kako želi radi zadovoljenja svojih egoističnih ciljeva. Inače je inteligentna te nema vidljivih defekata zbog kojih bismo mislili da su joj racionalne sposobnosti oštećene. Na sličan način tretira svoju suprugu: nije je nimalo briga za dobrobit vlastite supruge niti za to što će se s njom dogoditi. Prema Williamsu, takva osoba nema razloga da bude bolja prema svojoj supruzi jer, prema pretpostavkama primjera, nije u poziciji da kroz zaključivanje formira želju ili namjeru da se počne bolje ponašati prema njoj. Budući da ta nemogućnost nije posljedica nekog defekta u racionalnim sposobnostima, slijedi da ta osoba nema razloga da bude bolja prema svojoj supruzi. Mi možemo smatrati da to ponašanje nije dobro, da je nemoralno i da bi se osoba trebala bolje odnositi prema svojoj supruzi. Međutim, prema Williamsu (i Harmanu), ono što ne možemo reći za takvu osobu je da *ima razloga* biti bolja prema svojoj supruzi.

Štoviše, ako je Harman u pravu, onda prema njegovom argumentu prihvaćanje naturalizma implicira da takva pozicija točno opisuje prirodu razloga za djelovanje. Iz nje slijedi da je barem pojmovno moguće da postoji

---

da bi Harmanova formulacija odnosa razloga i racionalnog zaključivanja imala smisla. Ono što se pretpostavlja jest da je moguće racionalno rekonstruirati ponašanje osobe *kao* da je prošla proces racionalnog zaključivanja. Dakle, ideja je da ćemo moći osobi *pripisati* obrazac zaključivanja koji racionalizira ili opravdava formiranje namjere da poduzme radnju A (vidi Davidson, 1970).

takva duboko nemoralna osoba koja nema razloga poštovati moralne norme (za daljnju raspravu vidi Street, 2009). Međutim, mnogima se ne sviđa posljedica takvog gledišta te bi htjeli reći da psihopat iz našeg primjera zapravo *ima* razloga da bude bolji prema svojoj supruzi (vidi npr. Scanlon, 1998, Appendix). U ovom radu mi se nećemo direktno baviti tim pitanjem. Za nas je relevantno pitanje implicira li zaista prihvaćanje naturalizma relativizam u pogledu praktičnih razloga. Smith smatra da Harmanov argument u prilog toj tvrdnji nije dobar te da zbog toga prihvaćanje naturalizma ne implicira relativizam u pogledu praktičnih razloga. U nastavku ćemo razmotriti način na koji Smith brani svoju tvrdnju.

#### **4. Smithov prigovor: naturalizam nije relevantan za određivanje principa racionalnosti**

Smith ukazuje na to da istinosna vrijednost druge premise Harmanova argumenta varira ovisno o pozadinskim pretpostavkama argumenta. U tom pogledu navodi sljedeće:

Ono što Harman mora pretpostaviti kako bi druga premisa bila istinita jest ne samo da su određeni profesionalni kriminalci minimalno racionalni, nego i to da nisu neupućeni u što god je relevantno za argumente u prilog tome, ili protiv toga, da netko ima namjere koje ima, niti da su neosjetljivi (...) na te argumente. Drugim riječima, mora pretpostaviti da su, barem u relevantnim aspektima, maksimalno racionalni. No, hoćemo li se složiti s njim oko toga, ne ovisi o našem prihvaćanju (...) naturalizma (...). To ovisi o specifičnim pretpostavkama koje prihvaćamo o tome kako bi izgledao argument u prilog toga da netko ima određenu namjeru. (Smith, 2012, str. 233)

Kako bismo jasnije shvatili Smithov prigovor, razmotrit ćemo detaljnije pojam dovoljnog razloga kako ga shvaća Harman. On navodi da:

netko ima dovoljan razlog nešto učiniti ako i samo ako postoji ispravno zaključivanje koje bi osoba mogla provesti, koje bi je dovelo do odluke da izvrši dotičnu radnju. (Harman, 2000, str. 86)

Prema Harmanu pojam dovoljnog razloga povezan je sa sposobnošću osobe da kroz zaključivanje dođe do odluke da poduzme neku radnju koju taj razlog podržava. Slično gledište na razloge u filozofskoj literaturi izražava se u terminima „pouzdana deliberativne rute“ (Wedgwood, 2011, str. 180; Williams, 1981, 1995). Na primjer, Williams (1995) prihvaća gledište prema kojemu je nužan uvjet za imanje razloga sljedeći: osoba X ima razloga poduzeti radnju R samo ako postoji pouzdana ili racionalna deliberativna ruta od X-ovih trenutačnih motivacijskih stanja do toga da X poduzme radnju R.

Jednom kada se odnos razloga i sposobnosti za racionalnu deliberaciju ili promišljanje odredi na ovakav način, postavlja se pitanje što točno određuje principe ili načela opravdanog zaključivanja? Ili, koristeći alternativnu terminologiju, koji principi određuju to da je neka deliberativna ruta racionalna? Uvjerljivost Harmanovog argumenta ovisi o tome koji ćemo odgovor dati na ovo pitanje.

Kako bi provjerio uvjerljivost Harmanovog argumenta, Smith nudi tri plauzibilna principa racionalnosti koji se razlikuju prema svojem dosegu i normativnoj snazi (vidi Smith, 2012, str. 234–236, vidi također Smith 2009). Najslabiji od triju principa, Smith naziva principom SREDSTVO-CILJ+. Prema njemu racionalnost zahtijeva samo to da osoba ima istinita i epistemički dobro podržana vjerovanja te da ima intrinzične želje<sup>105</sup> koje podržavaju potpune i tranzitivne preferencije.<sup>106</sup> Ovaj je princip formalan jer ne postavlja ograničenja na sadržaje želja koje racionalna osoba može imati. Drugi princip, koji naziva UNIVERZALIZACIJA+ je supstancijalniji. Uključuje principe poput Kantove „Formule humanosti“ koja u jednoj formulaciji glasi: „Djeluj tako da tretiraš ljude, kako sebe tako i druge, uvijek kao cilj, a nikad kao sredstvo!“ (Kant, 2003). Ovaj princip je supstancijalniji od prethodnog jer zahtijeva da osoba ima određene želje koje se mogu univerzalizirati i time postati zajedničke različitim racionalnim djelatnicima. Najsupstancijalniji princip u ovom kontekstu Smith naziva RAZLOZI+. Prema njemu intrinzične prirode, na primjer, štete (engl. *harm*) i ozljeda, same po sebi pružaju svim racionalnim bićima razlog da ne žele da im se naškodi ili nanese ozljeda te da ne nanose štetu i ozljede drugima.

Smith ukazuje na to da istinitost druge premise u Harmanovom argumentu ovisi o tome koji ćemo od navedenih principa prihvatiti kao princip racionalnosti ili racionalnog zaključivanja. Na primjer, ako tvrdimo da je SREDSTVO-CILJ+ jedini princip racionalnosti, onda bi druga premisa mogla biti istinita. Razlog za to je što SREDSTVO-CILJ+ samo zahtijeva da racionalni djelatnici imaju koherentne želje i dobro utemeljena vjerovanja

---

<sup>105</sup> Intrinzične su one želje čije predmete želimo zbog njih samih, a ne zbog toga što bi nam omogućile da dođemo do nekih drugih stvari koje želimo. Na primjer, možemo reći da je želja za srećom intrinzična želja jer je želimo radi nje same. Instrumentalne želje su one koje nam omogućavaju da u konačnici zadovoljimo neku intrinzičnu želju. Na primjer, želja za novcem može biti instrumentalna ako osoba smatra da će joj posjedovanje novca donijeti sreću.

<sup>106</sup> Terminologija dolazi iz formalne teorije odlučivanja. Tamo je osnovna ideja da je osoba racionalna ako ima tranzitivne i potpune preferencije (koje su formalizirana verzija pojma želje). Da su preferencije tranzitivne znači da ako osoba preferira opciju *a* opciji *b* i *b* opciji *c*, onda također preferira *a* opciji *c*. Da su preferencije potpune znači da osoba preferira opciju *a* opciji *b* ili *b* opciji *a* ili je indiferentna prema njima (niti preferira opciju *a* opciji *b* niti obrnuto) (vidi npr. Resnik, 1987).

o tome kako zadovoljiti te želje. To znači da se od racionalnih djelatnika ne zahtijeva da imaju točno određene želje, a još manje da postoje neke želje koje svi racionalni djelatnici moraju imati. Međutim, ako prihvatimo neki od supstancijalnijih principa racionalnosti koji zahtijevaju da racionalna osoba ima određeni tip želja i preferencija, onda će druga premisa najvjerojatnije biti neistinita.

Smithov prigovor temelji se na tvrdnji da koji god bio konačan odgovor na pitanje o pravim principima racionalnosti, on se može dati neovisno o prihvaćanju naturalizma. Prema Smithu (2012), odnos između racionalnosti i znanstvenog poimanja svijeta određen je apriorno. U tom smislu navodi da kako bismo odgovorili na potonje pitanje prvo moramo odgovoriti na „prethodno pitanje“ koje se odnosi na to „[š]to radimo kada osmišljavamo karakterizaciju racionalnosti ili razumnosti?“ (Smith, 2012, str. 238). Kao mogući odgovor, Smith ističe jedino našu sposobnost za apriorno zaključivanje:

Jedan odgovor, inspiriran mišlju da se norme racionalnosti ili razumnosti mogu odrediti kroz definiciju pojmova vjerovanja i želje, jest da će karakterizacija racionalnosti i razumnosti slijediti iz objašnjenja svega što možemo *a priori* znati o vjerovanju i želji. (Smith, 2012, str. 238–239)

Dakle, prema Smithu, nije točna Harmanova pretpostavka da naturalizam podržava relativizam u pogledu praktičnih razloga. Prihvaćanje naturalizma nas ne sprječava da smatramo da razlozi nisu relativni prema subjektivnim preferencijama racionalnih djelatnika.

## 5. Ontološki i metodološki naturalizam: odgovor na Smithov prigovor

Kako bismo ocijenili Smithov prigovor moramo odrediti koji se pojam naturalizma podrazumijeva u ovoj raspravi. Općenito možemo reći da postoje barem dvije vrste naturalizma u filozofiji: ontološki ili metafizički naturalizam i metodološki naturalizam (Papineau, 2021). Prethodni se „odnosi na sadržaje stvarnosti, tvrdeći da u stvarnosti nema mjesta za ‘natprirodne’ ili druge ‘sablasne’ vrste entiteta“ (Papineau, 2021, uvod). Metodološki naturalizam je gledište koje se odnosi na filozofsku metodologiju te se njime tvrdi da bi metode koje se koriste u filozofiji trebale biti u kontinuitetu s metodama koje se koriste u znanostima. U tom pogledu David Papineau navodi: „metodološka komponenta se odnosi na načine istraživanja stvarnosti, te tvrdi da znanstvena metoda ima određeni znanstveni autoritet“ (Papineau, 2021, uvod). Te kasnije dodaje da



[m]etodološki naturalisti vide filozofiju i znanost kao u suštini iste pothvate, koji slijede iste ciljeve i koriste iste metode. Metodološki antinaturalisti vide filozofiju kao odvojenu od znanosti, s različitim ciljevima i metodama. (Papineau, 2021, odjeljak 2; vidi također Quine, 1969)

Harmanove (2000) tvrdnje u pogledu prihvaćanja naturalizma su dvosmislene. Na primjer, na jednom mjestu navodi da se moralni naturalisti „koncentriraju na pronalazak mjesta vrijednostima i dužnostima u svijetu činjenica kako nam ih znanost otkriva“ (Harman, 2000, str. 79). Ovakva tvrdnja sugerira da možda na umu ima ontološko čitanje naturalizma.

Ako je to čitanje ispravno onda Smith daje uvjerljivu kritiku Harmanovog argumenta. Ontološki naturalizam samo tvrdi da našu stvarnost moramo shvatiti kao da se sastoji od entiteta i zakona koje postuliraju naše najbolje znanstvene teorije. Pojednostavljeno, tu ideju možemo razumjeti kao tvrdnju da postoje samo materijalni predmeti, sile, energije itd. koji su priznati u suvremenoj fizici, a sve ono što druge znanosti otkriju ili postuliraju mora biti kompatibilno s priznatim zakonima fizike. Ako se na taj način shvati naturalizam u Harmanovom argumentu, onda je moguće da je Smith u pravu kada govori da principi racionalnosti nisu određeni prihvaćanjem naturalizma. Ontološki naturalizam postavlja ograničenja na to koji entiteti postoje prema našim najboljim znanstvenim teorijama, a ne kako bismo trebali zaključivati o njima. Stoga se čini da je ontološki naturalizam u principu kompatibilan s prihvaćanjem epistemoloških postavki koje se oslanjaju na apriorne principe zaključivanja.

Međutim, Smith (2012, str. 229) interpretira Harmana kao da prihvaća metodološki naturalizam. Sam Harman također u svojim drugim radovima jasnije sugerira da zapravo na umu ima metodološki naturalizam. Na primjer, u kasnijem radu Harman navodi sljedeće:

[f]ilozofski naturalizam je posebna instanca šire koncepcije filozofije prema kojoj su predmet istraživanja i metode filozofije u kontinuitetu s predmetima istraživanja i metodama drugih znanosti, uključujući prirodne znanosti. (...) Naturalizam često također ima ontološki ili metafizički aspekt (...). Međutim, glavna naturalistička tema je metodološka. (Harman, 2012, str. 10)

Ako Harmanov argument podrazumijeva tu vrstu naturalizma, onda Smithov prigovor postaje manje uvjerljiv. Kao što smo ranije spomenuli, metodološki naturalizam pretpostavlja kontinuitet između filozofije i znanosti te, što je važnije, pretpostavlja općeniti autoritet znanosti. Iz te perspektive, prikladnije je teoretizirati o normama racionalnog zaključivanja kao o nečemu što se može odrediti u odnosu na relevantne znanstvene teorije (vidi



npr. Colyvan, 2008). Stoga se čini da će se prema metodološkom naturalizmu ispravna karakterizacija racionalnosti morati temeljiti na koncepciji racionalnosti kako se ona shvaća u relevantnim znanostima. Ako je tome tako, onda se Smithov argument mora interpretirati kao da se njime tvrdi da pojam racionalnosti koji se podrazumijeva u relevantnim znanostima *ne* favorizira princip SREDSTVO-CILJ+ u odnosu na druge supstancijalne principe. Kako bismo dobili odgovor na to pitanje moramo provjeriti postoje li i kako se određuju principi racionalnosti koji se koriste ili podrazumijevaju u našim najboljim znanstvenim teorijama.

Pregled literature koja se odnosi na metodologiju društvenih znanosti i njihova primjena u istraživanjima u kognitivnim znanostima pokazuje da se jedini priznati principi racionalnosti mogu općenito kategorizirati kao principi instrumentalne racionalnosti (Guala, 2005; Miyazono i Bortolotti, 2021; Stein, 1997). Na primjer, različiti autori u kognitivnim znanostima prave razliku između instrumentalne (ili praktične) i epistemičke (ili teorijske racionalnosti) (vidi npr. Elqayam i Evans, 2011; Smokrović, 2001). Prethodna uključuje teoriju odlučivanja i formalnu teoriju igara, dok potonja uključuje teoriju vjerojatnosti i druge teorije koje se bave različitim formalnim aspektima revizije vjerovanja. Međutim, ovakvo instrumentalno shvaćanje racionalnosti vrlo je nesupstancijalno. Njegovu klasičnu karakterizaciju je dao nobelovac Herbert Simon kada je napisao da je

[u]m u potpunosti instrumentalan. Ne može nam reći kamo da idemo; najviše što nam može reći jest kako da dođemo tamo. On je plaćenik koji se može koristiti u svrhu postizanja bilo kojih ciljeva koje imamo, bili oni dobri ili loši. (Simon, 1983, str. 7–8; citirano u Over, 2004, str. 5)

Krucijalno za našu diskusiju je da se dvije gore spomenute koncepcije racionalnosti mogu općenito svrstati pod normativni princip koji Smith naziva SREDSTVO-CILJ+. Dakle, ako ozbiljno shvatimo naturalistička ograničenja na ono što možemo uzeti u obzir pri karakterizaciji racionalnosti, onda se čini da se principi koji upravljaju racionalnošću mogu svesti na principe instrumentalne racionalnosti i stoga ne postavljaju apriorna ograničenja na razloge koje osoba mora imati kako bi bila racionalna.

Stoga se čini da pretpostavka metodološkog naturalizma ide u prilog tezi o relativnosti praktičnih razloga. Naime, ranije smo vidjeli da prema Smithu druga premisa Harmanovog argumenta pretpostavlja princip SREDSTVO-CILJ+. Smith tvrdi da ta interpretacija druge premise nije zajamčena pretpostavkom naturalizma, već da prihvaćanje naturalizma ostavlja otvorenim koji princip racionalnosti upravlja željama i vjerovanjima racionalnih djelatnika. Međutim, sada vidimo da to nije točno. Kada

se uzme u obzir da Harman pod naturalizmom misli na metodološki naturalizam, onda je uvjerljivo tvrditi da naturalizam ipak opravdava čitanje druge premise kao da podrazumijeva princip SREDSTVO-CILJ+.

## 6. Metodološki naturalizam i principi racionalnosti: prigovori

### 6.1 Prigovor cirkularnosti

Netko bi mogao prigovoriti da je cirkularan prethodni argument prema kojemu metodološki naturalizam podrazumijeva instrumentalnu racionalnost. Na primjer, moglo bi se tvrditi da je koncepcija instrumentalne racionalnosti samo jedna od filozofskih koncepcija racionalnosti koja se nekritički počela uvoditi u različita znanstvena polja. Štoviše, moglo bi se prigovoriti da nekritički prihvaćamo Humeov skepticizam u pogledu dosega praktičnog uma, koji je, *nota bene*, također utemeljen na filozofskim, a ne empirijskim argumentima (vidi npr. Korsgaard, 1986; Millgram, 1995). Drugim riječima, moglo bi se tvrditi da je cirkularno oslanjati se na autoritet znanosti kada nastojimo odrediti koje principe racionalnosti bismo trebali prihvatiti u filozofiji morala.

Ovaj prigovor bi se mogao shvatiti na dva načina. Jedan je da ga se shvati kao instancu općenitijeg prigovora da se u filozofskim raspravama ne smijemo oslanjati na znanstvene teorije i metode istraživanja jer se filozofija bavi temeljnijim pitanjima koja nadilazile svaku znanstvenu spoznaju te, u metafizičkom smislu, prethode onome što nam znanost može otkriti. Stoga je gotovo uvijek cirkularno koristiti znanstvene spoznaje kako bismo došli do filozofskih zaključaka. Drugi način je da ga shvatimo kao da u ovom slučaju postoji nešto specifično zbog čega se ne smijemo oslanjati na znanstvene teorije i koncepcije racionalnosti kako bismo odredili koje principe racionalnosti smijemo prihvatiti kada govorimo o prirodi razloga za djelovanje.

Prvi je način shvaćanja prigovora prejak te ga možemo lako odbaciti. Naime, ta koncepcija odnosa filozofije i znanosti je utemeljena na posebnom shvaćanju filozofije kao apriorne discipline čije su metode i predmet istraživanja u diskontinuitetu s metodama i područjem istraživanja pojedinih znanosti. Upravo takvu koncepciju filozofije odbacuju zastupnici metodološkog naturalizma (vidi npr. Quine, 1969; Leiter, 2015). Stoga prigovor treba shvatiti na drugi način.

Moglo bi se argumentirati da, čak i pod pretpostavkom metodološkog naturalizma, ne smijemo se oslanjati na znanstvene teorije kada nastojimo odrediti koji principi upravljaju željama i vjerovanjima racionalnih djelatnika. No, sad se postavlja pitanje zašto se baš u ovom slučaju ne smijemo oslanjati na spoznaje iz znanosti? Što točno slučaj određivanja principa ra-

cionalnosti čini različitim od drugih filozofskih problema u čije rješavanje bismo bili spremni uključiti znanstvene spoznaje? Nije jasno koji bi bio dobar odgovor na ovo pitanje. Jedan način da se ovo pitanje učini pristupačnim je sljedeći. Možemo razmotriti neki primjer za koji će se mnogi složiti da pokazuje nelegitimnost ili cirkularnost oslanjanja na znanstvene spoznaje i teorije u izvlačenju određenih filozofskih zaključaka. Ako se pokaže da je naš slučaj relativnosti praktičnih razloga dovoljno sličan tom paradigmatom primjeru, onda bismo imali razloga prihvatiti da argument iz metodološkog naturalizma nije uvjerljiv. Radi ilustracije koristit ćemo dva primjera istraživanja moralnog razvoja u psihologiji.

Lawrence Kohlberg (1971) razvio je utjecajnu teoriju moralnog razvoja. Prema Kohlbergu, moralni razvoj se može podijeliti u šest stadija koji se odvijaju na tri razine (svaka razina zahvaća dva stadija razvoja). Prva razina uključuje predkonvencionalni moral, druga razina uključuje konvencionalni moral, dok treća razina uključuje postkonvencionalni moral. Svaka od razina prema Kohlbergu predstavlja jedan korak u postupnom razvoju ispravnog shvaćanja morala sve dok ne dođemo do zadnje razine koja je najsofisticiranija i uključuje razmišljanje u skladu s univerzalnim etičkim principima koji za osobu predstavljaju unutrašnje standarde u odnosu na koje ocjenjuje etičnost pojedinačnih ponašanja i pravičnost društvenih (konvencionalnih) normi.

Kao drugi primjer možemo uzeti istraživanja Eliotta Turiela (1983), koja pokazuju da ljudi općenito razlikuju između dviju vrsta pravila. Jednu vrstu pravila ljudi shvaćaju kao da su univerzalna, da vrijede neovisno o kontekstima i autoritetima koji bi ih mogli provoditi te se poglavito odnose na zabranu nanošenja štete drugim ljudima. Tu vrstu pravila Turiel naziva moralnim pravilima. Druga pravila, koja ljudi doživljavaju kao relativna o kontekstima te čija snaga ovisi o postojanju autoriteta koji ih može provoditi, Turiel naziva konvencionalnim pravilima.

Zamislimo sada da netko tvrdi da Kohlbergova klasifikacija moralnog razvoja ili Turielova razlika između moralnih i konvencionalnih normi empirijski određuju što znači imati sposobnost za ispravno moralno razumijevanje te da filozofske teorije koje daju drukčije shvaćanje moralnog razumijevanja nisu točne. Netko bi s pravom mogao prigovoriti da te osobe čine pogrešku jer zaboravljaju da su ti konstrukti izvorno formulirani unutar neke moralne (filozofske) teorije.<sup>107</sup> Ako želimo razumjeti u čemu se sastoji ispravno moralno razumijevanje, moramo konzultirati same filozofske teorije i argumente, a ne njihove psihološke izvedenice. U suprot-

---

<sup>107</sup> Na primjer, Kohlberg (1971) se u formulaciji svoje teorije moralnog razvoja jako oslanja na kantovsku tradiciju u moralnoj filozofiji.

nom bismo se mogli dovesti u situaciju da na svijet projiciramo filozofski nelegitimne psihološke konstrukte te na temelju njih pokušavamo pokazati da druge, nekompatibilne filozofske teorije nisu istinite.

Primjer konkretne rasprave koja odgovara upravo opisanom obrascu odnosi se na pitanje imaju li osobe s psihopatijom sposobnost ispravnog moralnog razumijevanja (Maibom, 2018). Neke studije su pokazale da psihopati ne prave razliku između moralnih i konvencionalnih normi poput ostalih ljudi (vidi npr. Blair, 1995; cf. Aharoni i ostali, 2012). Na temelju takvih studija neki autori su tvrdili da psihopati nemaju sposobnost moralnog razumijevanja koju je potrebno imati kako bismo ih smatrali moralno odgovornima (za raspravu vidi Jurjako i Malatesti, 2018). Drugi autori pokazuju da takav način argumentiranja možda i nije najuvjerljiviji jer se, na primjer, u kantovskoj moralnoj tradiciji ne pravi striktna razlika između moralnih i konvencionalnih normi kako se to podrazumijeva u Turielovim istraživanjima (vidi npr. Maibom, 2005).

Ovdje se postavlja pitanje predstavlja li slučaj moralnog razumijevanja dobru analogiju sa slučajem određivanja principa racionalnosti? Mi smatramo da ne predstavlja. Analogija ne vrijedi jer je u našem slučaju pozivanje na naturalizam usko povezano s metodologijom koja se koristi u znanostima. Principi racionalnosti koji se koriste u znanostima ne koriste se samo kao standardi za istraživanje toga koliko su ljudi racionalni, a onda i za njihovu normativnu evaluaciju, nego se koriste kao sredstva znanstvenog istraživanja i za proizvodnju znanstvenih alata i tehnologija. Na primjer, u kognitivnim znanostima racionalnost se koristi za predviđanje i objašnjenje ponašanja i kognitivnih procesa kod ljudi (Cardella, 2020; Oaksford i Chater, 2007); u ekonomiji se koristi za analiziranje i predviđanje ponašanja tržišta (Guala, 2005; Lagueux, 2014; za pregled rasprave na hrvatskom vidi Smokrović, 2017); u društvenim znanostima se koristi za objašnjenje evolucije kooperacije (Gaus, 2011, pogl. 3; Gauthier, 1986) i tako dalje. Iz naturalističke perspektive, upravo ta činjenica, da se instrumentalna koncepcija racionalnosti uspješno koristi kao sredstvo za izgradnju znanstvenih modela i teorija u različitim poljima, podržava njezino prihvaćanje kao ispravne koncepcije racionalnosti (Leiter, 2015). Stoga, ako prihvatimo da je znanost naš najbolji način otkrivanja i objašnjenja svijeta, onda principi racionalnosti koje znanstvena metoda pretpostavlja pružaju podršku ili opravdanje koje nije na nelegitimnan način cirkularno (vidi također Jurjako, 2017).

Ovome možemo dodati da nije jasno da pojam racionalnosti koji se koristi u znanosti na ovaj ili onaj način pretpostavlja hjumovsko ili neko drugo supstancijalnije viđenje racionalnosti. Na primjer, ako razmotrimo pojmovni aparat formalne teorije odlučivanja, možemo primijetiti da je to

u suštini matematički okvir koji se može interpretirati na različite načine, ovisno o tome za što ga želimo koristiti (Gintis, 2009, pogl. 1). Stoga nema supstancijalnih ograničenja načina na koji treba interpretirati vokabular teorije odlučivanja. Možemo uzeti dva primjera koja ilustriraju ovu tvrdnju. Teorija odlučivanja može se interpretirati kao da govori o vjerovanjima i željama racionalne osobe (vidi Lewis, 1983, 114). Također se može interpretirati kao da govori o objektivnom dobru koje se može ostvariti poduzimanjem neke radnje (Broome, 1999, pogl. 2). Ono što je važno ovdje istaknuti jest da pojam racionalnosti kako je shvaćen unutar teorije odlučivanja ne daje supstancijalna ograničenja što osoba, životinja ili stroj mora maksimizirati (ili minimizirati), nego samo postavlja strukturalna ograničenja koja moraju biti zadovoljena kako bi se matematički okvir mogao primijeniti. Ta ograničenja su zahvaćena onime što Smith naziva principom racionalnosti SREDSTVO-CILJ+.

## **6.2 Prigovor: instrumentalna racionalnost i svrha pripisivanja krivnje**

Drugi prigovor tezi da metodološki naturalizam favorizira principe instrumentalne racionalnosti bi mogao biti da takva koncepcija racionalnosti ne odgovara dobro našoj intuitivnoj koncepciji razloga za djelovanje (Smith, 2012, str. 244–245). Na primjer, ako profesionalni kriminalac ne mari za moralne norme, svejedno bismo ga držali odgovornim za krivična djela. Međutim, moglo bi se prigovoriti da nije jasno koja je svrha pripisivanja krivnje i držanja odgovornom osobe koja je racionalna i ne mari za ono što nalažu moralni zahtjevi? Ne bi li takvo gledište učinilo bilo koji konzistentan skup preferencija jednako vrijednim i stoga otpornim na racionalnu kritiku?

Na ovo možemo odgovoriti na barem dva načina. Jedan je da možemo razlikovati vrste razloga. Ova ideja je već prisutna u svakodnevnom govoru koji razlikuje prudencijalne, estetske, epistemičke i moralne razloge (Copp, 2007). U skladu s tom idejom mogli bismo reći da moralni razlozi da čini određene stvari objektivno postoje, bez obzira na to što mi mislili o tome. Međutim, iz toga ne slijedi nužno da je svaki moralni razlog ujedno i racionalni razlog. Drugim riječima, nije svaki moralni razlog racionalno obvezujući, tj. osoba neće biti manje racionalna ako ne poštuje moralne zahtjeve ili razloge.

Drugi odgovor je da svrha prakse pripisivanja krivnje i držanja odgovornim ovisi o tome što želimo i koje ciljeve smatramo vrijednima ostvarenja (vidi npr. Jefferson, 2019). Ako kao racionalne osobe i članovi društva želimo živjeti smislene i vrijedne živote, onda će držanje drugih odgovornima za njihove moralne prijestupe zasigurno biti svrhovito. To će mini-

malno uključivati odvratanje drugih od činjenja moralnih prijestupa. Budući da su ljudi u aktualnom svijetu društvena bića koja imaju određene preferencije i vrijednosti neophodne za uspješan život u društvu, onda iz te perspektive svaka osoba i ideja mogu biti podvrgnuti racionalnoj kritici. Jedino ograničenje koje postavlja metodološki naturalizam jest da se ne može tvrditi da će svako racionalno biće imati aprioran razlog za to da živi u određenom društvu i poštuje njegove norme.

### 6.3 Daljnji prigovor: metodološki naturalizam osigurava samo nužan uvjet za principe racionalnosti

Moglo bi se inzistirati na tome da čak i ako se složimo s prethodno napisanim, prihvaćanje metodološkog naturalizma svejedno ne implicira relativizam u pogledu praktičnih razloga. Naime, moglo bi se tvrditi da metodološki naturalizam ukazuje na to da je poštovanje principa SREDSTVO-CILJ+ samo *nužan* uvjet racionalnosti, no ne ukazuje na to da ne postoje i drugi principi koji definiraju što znači imati praktičan razlog za djelovanje. Štoviše, moglo bi se tvrditi da osim proceduralnih moramo pretpostaviti i neke supstancijalne principe racionalnosti. Ovdje nam je opet zanimljiv Smith (2009) koji ukazuje na mogućnost izvođenja supstancijalnih principa racionalnosti poput UNIVERZALIZACIJA+ i RAZLOZI+ na temelju proceduralnih principa racionalnosti poput SREDSTVO-CILJ+. Kada bi ti principi doista bili izvedivi iz principa SREDSTVO-CILJ+, onda prihvaćanje metodološkog naturalizma ne bi podržavalo prihvaćanje relativnosti u pogledu praktičnih razloga jer bi isti razlozi koje imamo za prihvaćanje jednostavnijih principa racionalnosti poput SREDSTVO-CILJ+ podržavali prihvaćanje supstancijalnih principa racionalnosti koji bi propisivali univerzalne razloge za djelovanje. U nastavku ćemo detaljnije razmotriti kako bi se ova argumentacija mogla razviti. Da bismo to postigli, poslužiti ćemo se formulacijama principa racionalnosti koje razmatra Smith u jednom ranijem radu (vidi Smith, 2009).

Smith (2009, str. 119–120) navodi sljedećih šest principa racionalnosti u rasponu od jednostavnijih proceduralnih zahtjeva do zahtjevnijih supstancijalnih principa. Slijedeći Smitha, principe racionalnosti ćemo označavati s R1-R6, te ćemo koristiti skraćenicu RZ za „racionalnost zahtijeva da“:

R1: RZ (Ako osoba ima intrinzičnu želju da  $p$  i vjerovanje da može ostvariti  $p$  tako da učini  $q$ , tada osoba ima instrumentalnu želju da učini  $q$ )

R2: RZ (Ako osoba ima intrinzičnu želju da  $p$  i intrinzičnu želju da  $q$  i intrinzičnu želju da  $r$ , te ako se predmeti želja da  $p$  i  $q$  i  $r$  ne

razlikuju jedan od drugog i od predmeta želje da  $s$ , a da se ne uvode arbitrarne razlike, tada osoba ima intrinzičnu želju da  $s$ )

R3: RZ (Ako osoba ima intrinzičnu želju da  $p$ , slijedi da je ili sam  $p$  prikladno univerzalan ili da je zadovoljavanje želje da  $p$  konzistentno sa zadovoljavanjem želje čiji sadržaji su sami prikladno univerzalni)

R4:  $\exists p \exists q$  RZ (Ako osoba vjeruje da  $p$ , onda ima intrinzičnu želju da  $q$ )

R5:  $\exists p$  RZ (Ljudi ne žele da  $p$ )

R6:  $\exists q$  RZ (Ljudi žele da  $q$ )

Možemo primijetiti da je R1 samo varijanta principa racionalnosti koji smo dosad nazivali SREDSTVO-CILJ+. Njime se navodi samo to da ako osoba ima neku želju i vjerovanje kako je ostvariti, onda slijedi da ima želju upotrijebiti sredstva potrebna za njezino ostvarenje. R2 je proceduralan princip racionalnosti kojim se zahtijeva da osoba, ako je racionalna, neće praviti arbitrarne razlike među predmetima svojih želja. No, princip je i dalje proceduralan jer ne određuje koje želje bismo trebali imati. S obzirom na njihovu instrumentalnu i proceduralnu prirodu, čini se da je prihvaćanje principa R1 i R2 konzistentno s metodološkim naturalizmom. Principi R3-R6 djeluju supstancijalnije te nije jasno jesu li kompatibilni s metodološkim naturalizmom. Na primjer, princip R3 zahtijeva da sadržaji želja koje imamo budu univerzalni, što predstavlja jednu varijantu principa UNIVERZALIZACIJA+. Međutim, ako se može pokazati da se prihvaćanjem principa R1 i/ili R2 obvezujemo na prihvaćanje supstancijalnijih principa R3-R6, onda bi slijedilo da metodološki naturalizam ne isključuje prihvaćanje neinstrumentalnih principa racionalnosti.

Sada se nameću sljedeća pitanja: (i) slijedi li zaista da se prihvaćanjem instrumentalnih principa racionalnosti ujedno obvezujemo na supstancijalnije principe racionalnosti koji bi opravdali tvrdnju da racionalnost obvezuje na prihvaćanje moralnih zahtjeva?; (ii) kako bismo to mogli pokazati? Smith (2009) ukazuje na to da se prihvaćanje principa R4 može koristiti kako bi se derivirali principi tipa R5 i R6. U tom pogledu navodi sljedeće:

(...) R4 je različit od R1 i R2 po tome što usmjeruje prema vjerovanjima i željama s određenim sadržajima (...). Stoga narušava Wallaceov princip želja-kao-input-želja-kao-output: želje koje je za nas racionalno imati u potpunosti ovisi o tome što vjerujemo; nije relevantno s kojim željama započinjemo. Štoviše, ako su dana vjerovanja



dostupna svim racionalnim bićima, tada bi R4 čak zahtijevalo od svih racionalnih bića kao takvih da imaju određene želje. Stoga bi nam princip poput R4 mogao omogućiti da iz njih izvedemo principe poput R5 i R6. (Smith, 2009, str. 121)

U nastavku Smith sugerira način na koji bismo mogli argumentirati za to da prihvaćanje principa R1 i R2 zahtijeva prihvaćanje zahtjevnijih principa poput R4.

Zamislite subjekta koji je dovoljno refleksivan da formira vjerovanje da bi imao želju da  $r$ , pod uvjetom da također ima skup želja i vjerovanja kompatibilan sa svim drugim racionalnim principima koji upravljaju formiranjem želja i vjerovanja (npr. R1 i R4). Budući da je jasno da subjekt ne može istodobno imati to reflektivno vjerovanje te imati averziju ili biti ravnodušan prema  $r$ , onda se čini da ga izbjegavanje psihološke nekoherentnosti obvezuje da želi  $r$ . Međutim, pod pretpostavkom da nekoherentnost mora biti podržana nekim racionalnim principom, slijedi da [zastupnik instrumentalizma] mora odobriti barem sljedeću varijantu principa R4 (...):

R4<sup>refleksivni</sup>: RZ (Ako osoba vjeruje da bi željela da  $r$  kada bi imala skup želja i vjerovanja koje su kompatibilne sa svim (drugim) racionalnim principima koji upravljaju formiranjem želja i vjerovanja, tada bi željela da  $r$ ) (Smith, 2009, str. 122)

Prema Smithu, ako osoba želi izbjeći nekoherentnost među svojim psihološkim stavovima, mora prihvatiti to da postoji princip racionalnosti poput R4<sup>refleksivni</sup> koji nalaže da za svaku racionalnu osobu vrijedi da ako vjeruje da bi imala određenu želju kada bi bila racionalna, onda racionalnost od nje zahtijeva da formira tu želju u skladu s dotičnim vjerovanjem. Kako bi pojačao intuitivnost svoje argumentacije, Smith (2009, fusnota 12) navodi da slično vrijedi u slučaju teorijske ili epistemičke racionalnosti. Kada bi neka osoba vjerovala da bi formirala vjerovanje da  $p$  u okolnostima u kojima poštuje sve ostale norme racionalnosti, onda bi njezina psihološka stanja bila u međusobnom neskladu (tj. bila bi nekoherentna) kada stvarno ne bi formirala vjerovanje da  $p$ . Drugim riječima, ako osoba vjeruje da kada bi bila racionalna, onda bi vjerovala da  $p$  ili željela da  $q$ , onda racionalnost od nje zahtijeva da vjeruje da  $p$  ili želi da  $q$ .

Smithova argumentacija u prilog prihvaćanja principa R4<sup>refleksivni</sup> na temelju instrumentalnih principa R1 i R2 izgleda uvjerljivo. Budući da principi R1 i R2 izražavaju instrumentalne i proceduralne principe racionalnosti koji su kompatibilni s metodološkim naturalizmom, onda se čini da prihvaćanje principa R4<sup>refleksivni</sup> također nije u suprotnosti s metodološkim

naturalizmom. Pitanje na koje sada moramo odgovoriti jest slijedi li iz toga da metodološki naturalizam ne podržava relativnost u pogledu praktičnih razloga?

Dosadašnja argumentacija ne pokazuje to. Čak i ako se prihvati princip  $R4^{\text{refleksivni}}$ , ne slijedi da će sve želje koje racionalne osobe mogu imati biti univerzalne ili da postoje neki supstancijalni razlozi (poput ranije spomenutih moralnih razloga) koje će sve racionalne osobe nužno dijeliti bez obzira na svoje izvorne motivacije. Kako bismo to uvidjeli treba primijetiti da  $R4^{\text{refleksivni}}$  samo propisuje da je osoba u nekoherentnom stanju ako vjeruje da bi željela da  $p$  pod uvjetom da je racionalna i istodobno ne želi  $p$  (ili nakon što usvoji to vjerovanje ne formira želju da  $p$ ). Međutim, iz prihvaćanja tog principa nije jasno hoće li sadržaji odgovarajućih želja biti univerzalni ili na neki drugi način istovjetni sadržaju želja drugih racionalnih djelatnika. To će ovisiti o tome kako interpretiramo tvrdnju da su naše želje „u skladu sa svim (drugim) racionalnim principima koji upravljaju formiranjem želja i vjerovanja“. Ranije smo argumentirali da prema metodološkom naturalizmu jedino možemo očekivati da će proceduralni principi poput SREDSTVO-CILJ+ (ili R1 i R2) biti opravdani. Sada se postavlja pitanje možemo li iz principa racionalnosti poput R1, R2 i  $R4^{\text{refleksivni}}$  izvesti supstancijalnije principe racionalnosti koji bi opravdali tvrdnju da postoje razlozi koje će dijeliti svi racionalni djelatnici?

U povijesti suvremene filozofije bilo je pokušaja da se pokaže kako se iz principa instrumentalne racionalnosti mogu izvesti univerzalna moralna pravila. U tom pogledu, Smith spominje rad Davida Gauthiera (1986). Međutim, uspješnost takvih pokušaja je u najmanju ruku upitna (za dobar pregled rasprave vidi Gaus, 2011, pogl. 2). Drugi pokušaji imaju ishodište u kantovskoj tradiciji (vidi npr. Korsgaard, 1996) čiji sljedbenici, poput Smitha (1994), oslanjanjem na analizu pojma racionalnih sposobnosti nastoje pokazati da su savršeno racionalne osobe one koje prihvaćaju moralne zahtjeve kao principe racionalnosti. Ostaje otvoreno pitanje koliko su uspješni takvi pokušaji te hoće li se u budućnosti pojaviti neki filozof koji će pokazati da je nešto takvo moguće.

Međutim, iz perspektive metodološkog naturalizma imamo razloga smatrati da izvođenje univerzalnih i supstancijalnih razloga za djelovanje nije moguće samo na temelju prihvaćanja proceduralnih principa poput R1 i R2. U nastavku ćemo se osvrnuti na dvije vrste argumentacije za tu tezu.

Prva vrsta argumentacije poziva se na znanstvena istraživanja o tome kako ljudi shvaćaju odnos racionalnosti i moralnih zahtjeva. Iako nema empirijskih istraživanja o tome kako obični ljudi shvaćaju principe racionalnosti i njihovu vezu s moralnim zahtjevima, u najmanju ruku bismo

očekivali da će ljudi, ako prihvaćaju ideju da postoje supstancijalni principi racionalnosti koji se na nekontingentan način povezuju s moralnim zahtjevima, pokazivati tendenciju prihvaćanja moralnih zahtjeva kao objektivno važećih i univerzalnih. To bi nam dalo kakve takve indikacije da je uobičajeni pojam savršeno racionalne osobe određen time da takva osoba prihvaća supstancijalne razloge za djelovanje.

Unatoč očekivanjima, istraživanja pokazuju da kod običnih ljudi značajno variraju intuicije koje se odnose na objektivnost moralnih sudova (Hopster, 2019). Neke moralne tvrdnje ljudi tipično smatraju objektivnima, neke relativnima o kontekstu, a neke u potpunosti ovisnima o preferencijama ljudi, što nam sugerira da su intuicije običnih ljudi u pogledu objektivnosti moralnih sudova pluralističke (Pözlner i Wright, 2020; Wright i ostali, 2013). Nadalje, ta istraživanja nam sugeriraju da pojam racionalnosti koji obični ljudi imaju neće nužno uključivati supstancijalnija razmatranja koja će dijeliti svi racionalni pojedinci. I doista, jedino istraživanje koje nam je poznato, a bavi se time kako obični ljudi shvaćaju pojam racionalnosti, ukazuje na to da se racionalna osoba shvaća kao pojedinac koji je logičan i inteligentan te maksimizira svoje preferencije bez obzira na njihove sadržaje (vidi Grossmann i ostali, 2020). To bi prema našoj terminologiji bilo u skladu sa shvaćanjem racionalnosti koje je određeno principom SREDSTVO-CILJ+.

Druga vrsta argumentacije poziva se na apriorna razmatranja prema kojima iz proceduralnih principa racionalnosti ne možemo izvesti supstancijalnije razloge za djelovanje. Gerald Gaus (2011, odjeljak 13.2) pokazuje da čak i kada se radi o idealno racionalnoj osobi (u proceduralnom smislu) ne možemo očekivati da će takva osoba doseći jedinstven i koherentan skup želja i vjerovanja o tome koje razloge ima za djelovanje. Nasuprot tome, koji će skup razloga ta osoba imati da učini nešto, te koje će principe racionalnosti prihvatiti, ovisit će o njezinim početnim motivacijama, preferencijama i slijedu životnih okolnosti.

Kako bi pokazao da je racionalni izbor ovisan o redoslijedu izbora (engl. *path-dependent*), Gaus (2011, str. 242) daje sljedeći hipotetički primjer. Zamislimo da je Betty osoba koja je idealno racionalna. Ima maksimalno razvijene sposobnosti deduktivnog i induktivnog zaključivanja, vidi sve posljedice svojih vjerovanja, u stanju je izvesti zaključak bilo koje kompleksnosti i može neposredno uvidjeti maksimalno koherentan i konzistentan skup vjerovanja koji će rezultirati iz njezinih zaključaka. Unatoč tome, nemamo razloga očekivati da će Betty neovisno o početnim uvjetima deliberacije i životnim okolnostima doći do istog skupa vjerovanja kao i neka druga osoba koja je na sličan način idealno racionalna.

Kako bismo to uvidjeli, zamislilo da je u vremenu  $t_1$  Betty suočena s izborom radnji iz skupa  $Y$ . Kako bi racionalno izabrala, mora razmotriti koja su relevantna vjerovanja o opcijama s kojima je suočena i koje su relevantne vrijednosti koje ima. Također, mora uzeti u obzir epistemičke i druge norme koje su relevantne za ispravno procesiranje informacija na osnovi kojih može racionalno djelovati. Nakon te idealne deliberacije Betty zaključuje da su joj u tom trenutku relevantna vjerovanja  $v_1$  i  $v_2$ , međutim primjećuje da su ta vjerovanja međusobno proturječna te da može imati još koherentniji skup vjerovanja o tome što vjerovati i učiniti ako odustane od jednog od dvaju vjerovanja. Recimo da odustane od  $v_1$  i umjesto njega prihvati vjerovanje  $v_3$ . U sljedećem trenutku  $t_2$  susreće se sa novim skupom opcija  $Z$  unutar kojega mora odabrati jednu. S obzirom na prethodni izbor, u  $t_2$  Betty ulazi s točno određenim skupom vjerovanja ( $v_2$  i  $v_3$ ) te na temelju njega bira opciju  $Z_1$ . Sada zamislimo da je Betty u  $t_1$  umjesto skupa opcija  $Y$  bila suočena s drugim skupom opcija, nazovimo ga  $X$ , u kojemu su relevantna vjerovanja bila  $v_1$  i  $v_4$  te da u toj situaciji Betty uviđa da iz tih vjerovanja slijedi vjerovanje  $v_5$ . Recimo sada da se u  $t_2$  Betty suočava sa skupom opcija  $Y$  te da opet mora razriješiti sukob između  $v_1$  i  $v_2$ . Međutim, u ovom slučaju Betty uviđa da će imati koherentniji skup vjerovanja ako odbaci  $v_2$  jer se  $v_1$  bolje povezuje s novim vjerovanjem  $v_5$ . Sada možemo zamisliti da kada se u  $t_3$  Betty susretne sa skupom opcija  $Z$  više neće imati razloga odabrati opciju  $Z_1$  jer joj nedostaju vjerovanja  $v_2$  i  $v_3$  te će stoga imati razloga prihvatiti neku drugu opciju  $Z_n$  koja će biti bolje usklađena sa skupom vjerovanja u kojem se nalazi  $v_1$ .

Ovaj primjer nam pokazuje da nije realno očekivati da će savršeno racionalna osoba, samo na temelju činjenice da je savršeno racionalna, doći do supstancijalnih razloga za djelovanje koje će dijeliti sa svim drugim racionalnim djelatnicima. Štoviše, na temelju prethodnog primjera, Gaus tu ideju naziva mitom potpune racionalnosti, a taj se mit svodi na tvrdnju da „bi u potpunosti racionalne osobe birale isti specifičan moral ili iste principe pravednosti (...)“ (Gaus, 2011, str. 243). Bitno za našu raspravu je to što čak i ako zamislimo savršeno racionalnu osobu te razmišljamo što će ona učiniti te do kojih zaključaka će doći o tome koje razloge ima, ti će se zaključci razlikovati ovisno o povijesti njezinih izbora i početnog skupa vjerovanja i vrijednosti koje prihvaća. Ako to vrijedi u slučaju pojedinačne osobe, onda imamo još manje razloga vjerovati da će se različite savršeno racionalne osobe složiti, neovisno o kontingentnim okolnostima i kontekstu odabiranja, oko toga koje supstancijalne razloge ili norme racionalnosti trebaju prihvatiti. Stoga nije realno očekivati da će se samo na temelju prihvaćanja proceduralnih principa racionalnosti poput R1, R2 ili principa SREDSTVO-CILJ+ moći izvesti supstancijalni principi racionalnosti koji bi garantirali nerelativizam u pogledu praktičnih razloga.

## 7. Zaključak

U ovom radu bavili smo se raspravom između Harmana i Smitha o implikacijama metodološkog naturalizma na pitanje prirode praktičnih razloga. Harman (2000) je tvrdio da filozofski naturalizam podržava relativističko gledište u pogledu prirode praktičnih razloga. Smith (2012) tvrdi da Harmanov argument nije dobar jer ne uzima u obzir ideju da se određivanje istinskih principa racionalnosti koji određuju prirodu praktičnih razloga može ostvariti jedino apriornom analizom pojmova koji se odnose na racionalne želje i vjerovanja. Stoga je zaključio da prihvaćanje filozofskog naturalizma ne igra nikakvu ulogu u određivanju toga je li relativizam u pogledu praktičnih razloga točna teorija.

U radu smo argumentirali da je Smithov prigovor uvjerljiv samo ako pod filozofskim naturalizmom podrazumijevamo ontološki naturalizam. Međutim, ako njegov argument analiziramo iz perspektive metodološkog naturalizma, onda Harmanova tvrdnja o relativnosti praktičnih razloga postaje uvjerljivija. Argumentirali smo da prihvaćanje metodološkog naturalizma pruža resurse za iznošenje aposteriornog argumenta u prilog tvrdnji da su praktični razlozi relativni u odnosu na racionalne pojedince i njihova početna vjerovanja, motivacije i vrijednosti. Suština argumenta odnosi se na tvrdnju da metodološki naturalizam daje autoritet znanstvenoj metodi i njezinim nalazima. Međutim, čini se da u relevantnim znanostima u kojima se koristi pojam racionalnosti jedina općenito praktična norma racionalnosti uključuje instrumentalnu koncepciju racionalnosti. Stoga nam ova razmatranja sugeriraju da metodološki naturalizam ide u prilog tvrdnji da su praktični razlozi relativni u odnosu na vjerovanja, motivacije i vrijednosti pojedinih racionalnih djelatnika. Naravno, pitanje na koje nismo dali odgovor jest trebamo li uopće prihvatiti metodološki naturalizam kao uvjerljivu poziciju u filozofiji morala. No, to vrlo važno pitanje ostavljamo za raspravu u nekoj drugoj prilici.

## Zahvale

Ovim putem htio bih se zahvaliti Nenadu Smokroviću, Zdenki Brzović, Martini Blečić, Mateju Sušniku i anonimnom recenzentu na komentiranju ranije verzije rada i savjetima kako ga poboljšati. Rad je nastao u sklopu znanstvenih aktivnosti koje podržavaju Hrvatska zaklada za znanost (projekt HIRE, UIP-2017-05-4308) i Sveučilište u Rijeci (projekt KUBIM, uniri-human-18-265).

## Bibliografija

- Aharoni, E., Sinnott-Armstrong, W. P. i Kiehl, K. A. (2012). Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *Journal of Abnormal Psychology*, 121(2), 484–497. <https://doi.org/10.1037/a0024796>
- Berčić, B. (1995). *Realizam, relativizam, tolerancija*. Rijeka: Hrvatski kulturni dom.
- Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1–29. [https://doi.org/10.1016/0010-0277\(95\)00676-P](https://doi.org/10.1016/0010-0277(95)00676-P)
- Broome, J. (1999). *Ethics out of economics* (1. izd.). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511605888>
- Cardella, V. (2020). Rationality in mental disorders: Too little or too much? *European journal of analytic philosophy*, 16(2), 13–36. <https://doi.org/10.31820/ejap.16.2.1>
- Colyvan, M. (2008). Naturalizing normativity. U D. Braddon-Mitchell i R. Nola (ur.), *Conceptual analysis and philosophical naturalism* (str. 303–311). Cambridge, Mass.: The MIT Press. <https://doi.org/10.7551/mitpress/9780262012560.003.0013>
- Copp, D. (2007). *Morality in a natural world: Selected essays in metaethics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511497940>
- Cuneo, T. (2007). *The normative web: An argument for moral realism*. Oxford: Oxford University Press.
- Davidson, D. (1970). Mental events. U L. Foster i J. W. Swanson (ur.), *Essays on Actions and Events* (str. 207–224). Oxford: Clarendon Press.
- Dutilh Novaes, C. (2020). Carnapian explication and ameliorative analysis: A systematic comparison. *Synthese*, 197(3), 1011–1034. <https://doi.org/10.1007/s11229-018-1732-9>
- Elqayam, S., i Evans, J. S. B. T. (2011). Subtracting “ought” from “is”: Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, 34(5), 233–248. <https://doi.org/10.1017/S0140525X1100001X>
- Gaus, G. F. (2011). *The order of public reason*. Cambridge: Cambridge University Press.
- Gauthier, D. P. (1986). *Morals by agreement*. Oxford: Clarendon Press.
- Gintis, H. (2009). *Game theory evolving: A problem-centered introduction to modeling strategic interaction*. Princeton: Princeton University Press. <https://doi.org/10.2307/j.ctvc4gjh>
- Grossmann, I., Eibach, R. P., Koyama, J., i Sahi, Q. B. (2020). Folk standards of sound judgment: Rationality versus reasonableness. *Science Advances*, 6(2), eaaz0289. <https://doi.org/10.1126/sciadv.aaz0289>
- Guala, F. (2005). *The methodology of experimental economics*. Cambridge: Cambridge University Press.
- Harman, G. H. (1976). Practical reasoning. *The Review of Metaphysics*, 29(3), 431–463.
- Harman, G. H. (2000). Is there a single true morality? U *Explaining value and other essays in moral philosophy*. Oxford: Oxford University Press.
- Harman, G. H. (2012). Naturalism in moral philosophy. U S. Nuccetelli i G. Seay (ur.), *Ethical naturalism: Current debates* (str. 8–23). Cambridge: Cambridge University Press.
- Hopster, J. (2019). The meta-ethical significance of experiments about folk moral objectivism. *Philosophical Psychology*, 32(6), 831–852. <https://doi.org/10.1080/09515089.2019.1627304>



- Hume, D. (1983). *Rasprava o ljudskoj prirodi*. Sarajevo: Veselin Masleša.
- Jefferson, A. (2019). Instrumentalism about moral responsibility revisited. *Philosophical Quarterly*, 69(276), 555–573. <https://doi.org/10.1093/pq/pqy062>
- Jurjako, M. (2010). Uvod u teorije praktičnog uma. *Novi Kamov*, 10(32), 37–63.
- Jurjako, M. (2017). Normative reasons: Response-dependence and the problem of idealization. *Philosophical Explorations*, 20(3), 261–275. <https://doi.org/10.1080/13869795.2017.1381274>
- Jurjako, M. (2020). Samoobmana, namjere i pučko-psihološko objašnjenje djelovanja. *Prolegomena*, 19(1), 91–117. <https://doi.org/10.26362/20200106>
- Jurjako, M. i Malatesti, L. (2018). Neuropsychology and the criminal responsibility of psychopaths: Reconsidering the evidence. *Erkenntnis*, 83(5), 1003–1025. <https://doi.org/10.1007/s10670-017-9924-0>
- Kant, I. (2003). *Osnivanje metafizike čudoređa*. Zagreb: Feniks.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. U T. Mischel (ur.), *Cognitive development and epistemology* (str. 151–235). Cambridge, Mass.: Academic Press. <https://doi.org/10.1016/B978-0-12-498640-4.50011-1>
- Korsgaard, C. M. (1986). Skepticism about practical reason. *Journal of Philosophy*, 83(1), 5–25.
- Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge: Cambridge University Press.
- Lagueux, M. (2014). *Rationality and explanation in economics*. London and New York: Routledge.
- Leiter, B. (2015). Normativity for naturalists. *Philosophical Issues*, 25(1), 64–79. <https://doi.org/10.1111/phs.12046>
- Levy, N. (2004). *Moralni relativizam*. Zagreb: Naklada Jesenski i Turk.
- Lewis, D. K. (1983). *Philosophical papers. Volume I*. Oxford University Press. <https://doi.org/10.1093/0195032047.001.0001>
- Maibom, H. L. (2005). Moral unreason: The case of psychopathy. *Mind and Language*, 20(2), 237–257.
- Maibom, H. L. (2018). What can philosophers learn from psychopathy? *European Journal of Analytic Philosophy*, 14(1), 63–78.
- Millgram, E. (1995). Was Hume a Humean? *Hume Studies*, 21(1), 75–94.
- Miščević, N. i Smokrović, N. (ur.). (2001). *Računala, mozak i ljudski um: Zbornik tekstova iz teorije umjetne inteligencije i kognitivne teorije* (2. izd.). Rijeka: Izdavački centar Rijeka.
- Miyazono, K. i Bortolotti, L. (2021). *Philosophy of psychology: An introduction*. Cambridge: Polity Press.
- Nichols, S. (2002). How psychopaths threaten moral rationalism: Is it irrational to be amoral? *The Monist*, 85(2), 285–303. <https://doi.org/10.5840/monist200285210>
- Oaksford, M., i Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.
- Over, D. (2004). Rationality and the normative/descriptive distinction. U D. J. Koehler i N. Harvey (ur.), *Blackwell handbook of judgment and decision making* (str. 3–18). Cambridge: Blackwell Publishing. <https://doi.org/10.1002/9780470752937.ch1>



- Papineau, D. (2021). Naturalism. U E. N. Zalta (ur.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2021/entries/naturalism/>
- Parfit, D. (2011). *On what matters* (Sv. 1 i 2). Oxford: Oxford University Press.
- Pözlner, T. i Wright, J. C. (2020). Anti-realist pluralism: A new approach to folk meta-ethics. *Review of Philosophy and Psychology*, 11(1), 53–82. <https://doi.org/10.1007/s13164-019-00447-8>
- Quine, W. van O. (1969). Epistemology naturalized. U *Ontological relativity and other essays*. New York: Columbia University Press.
- Resnik, M. D. (1987). *Choices: An introduction to decision theory*. Minneapolis: University of Minnesota Press.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge, Mass.: Harvard University Press.
- Schafer, K. (2015). Hume on practical reason: Against the normative authority of reason. U P. Russell (ur.), *Oxford handbook of David Hume*. Oxford: Oxford University Press.
- Sekulić, D. (2016). *Psihologija zaključivanja i logika*. Zagreb: KruZak.
- Simon, H. A. (1983). *Reason in human affairs*. Palo Alto: Stanford University Press.
- Smith, M. (1994). *The moral problem*. Cambridge: Blackwell.
- Smith, M. (2009). Desires, values, reasons, and the dualism of practical reason. *Ratio*, 22(1), 98–125. <https://doi.org/10.1111/j.1467-9329.2008.00420.x>
- Smith, M. (2012). Naturalism, absolutism, relativism. U S. Nuccetelli i G. Seay (ur.), *Ethical naturalism: Current debates* (str. 226–244). Cambridge: Cambridge University Press.
- Smokrović, N. (2001). Računarska metafora i problem racionalnosti. U N. Mišćević i N. Smokrović (ur.), *Računala, mozak i ljudski um: Zbornik tekstova iz teorije umjetne inteligencije i kognitivne teorije* (2. izd., str. 219–231). Rijeka: Izdavački centar Rijeka.
- Smokrović, N. (2004). *Priroda prirodnog zaključivanja*. Zagreb: Hrvatsko filozofsko društvo.
- Smokrović, N. (2007). Uvod u teoriju normativnih razloga. *Novi Kamov*, 25(4), 5–10.
- Smokrović, N. (2017). *Znanost i metoda: Od opće metodologije do metodologije ekonomije*. Zagreb: KruZak.
- Stein, E. (1997). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford: Clarendon Press.
- Street, S. (2009). In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters. *Philosophical Issues*, 19, 273–298.
- Sušnik, M. (2009). Internalizam razloga: Dvije interpretacije. *Filozofska istraživanja*, 29(2), 349–363.
- Sušnik, M. (2012). Hjumovska teorija motivacije: U obranu dogme. *Prolegomena*, 11(1), 83–105.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.

- Wedgwood, R. (2011). Primitively rational belief-forming processes. U A. Reisner i A. Steglich-Petersen (ur.), *Reasons for belief* (str. 180–200). Cambridge: Cambridge University Press.
- Williams, B. (1981). Internal and external reasons. U *Moral luck* (str. 101–113). Cambridge: Cambridge University Press.
- Williams, B. (1995). Internal reasons and the obscurity of blame. U *Making sense of humanity* (str. 35–45). Cambridge: Cambridge University Press.
- Wright, J. C., Grandjean, P. T. i McWhite, C. B. (2013). The meta-ethical grounding of our moral beliefs: Evidence for meta-ethical pluralism. *Philosophical Psychology*, 26(3), 336–361. <https://doi.org/10.1080/09515089.2011.633751>



PAOLO LABINAZ

## Argumentation, Knowledge and Reasoning

**Abstract:** This paper deals with Nenad Smokrović's account of argumentation as a curiosity-driven, cooperative effort. More specifically, I focus on one of the two assumptions on which his account is based: that reasoning is argumentative in nature. Since Smokrović assumes Mercier and Sperber's argumentative theory as the starting point for developing this account, I point out the difficulties in combining their conception of reasoning as a persuasive device with the picture of the argumentation process he provides. I then suggest that one can assume the argumentative nature of reasoning while dismissing their conception of reasoning. In particular, I propose an alternative way to elaborate the idea that reasoning is argumentative by highlighting its reason-giving function. I argue that this function appears to be better suited to the argumentation process described by Smokrović than the persuasive one suggested by the argumentative theory. As I will try to show, only if we consider reasoning in its basic form, that is as a reason-giving device, can we understand why two or more people driven by their curiosity can get together in a collaborative effort to safely establish whether a certain proposition is true, without manipulating each other.

**Key words:** argumentation, reasoning, reason-giving function, persuasion, cooperation.

### 1 Introduction

Over the last decade, Nenad Smokrović (2011, 2015, 2017, 2018) has written an illuminating series of papers on reasoning and argumentation and how they relate to each other. From among the many suggestions he put forward in these papers, I have chosen to focus here on his claim that argumentation is a particularly effective means for extending knowledge (Smokrović 2015). While the claim in itself is nothing new, having been made by other scholars who have developed epistemological approaches to argumentation (e.g., Biro, Siegel 1992, 1997, 2006; Goldman 1994, 1999, 2003; Lumer 2005), what is original in Smokrović's work is how he argues for it. In fact, his argument rests on two assumptions that, to my mind, have never been combined before: (i) that reasoning is designed for argumentation (Mercier, Sperber 2009, 2011a, b, 2017), and (ii) that knowledge requires avoidance of error (the so-called "safety" condition; see Williamson 2000). In Smokrović's view, since people involved in argumentative exchanges are prompted to use the most reliable methods to

establish whether the proposition under discussion is true, in such circumstances it becomes more likely that they will acquire safe knowledge about it. Here, safety is to be understood as a modal condition, on the basis of which someone obtains (safe) knowledge in a given case insofar as they could not easily have been mistaken in similar situations. This paper sets out to develop the implications of this connection between argumentation and knowledge which Smokrović has made evident, outlining a way of conceiving the function of reasoning as reason-giving (rather than persuasive, as Mercier and Sperber have suggested) which in my opinion offers a better fit with his project.

The paper is organized as follows. Section 2 introduces the epistemic approach to argumentation, focusing in particular on the perspective developed by Goldman. Section 3 presents Smokrović's account of argumentation, and also describes how it differs from Goldman's. Section 4 points out the difficulties in combining Mercier and Sperber's conception of reasoning as a persuasive device with Smokrović's curiosity-driven, cooperative picture of the argumentation process. In conclusion, section 5 suggests a different way of elaborating the idea that reasoning is argumentative by highlighting its reason-giving function. I will argue that the conception of the argumentative nature of reasoning that results fits better with the argumentation process as described by Smokrović than the one to be found in the writings of Mercier and Sperber.

## 2 Argumentation as an epistemic practice

That argumentation plays a particularly significant role in our epistemic lives is widely accepted by philosophers, particularly those working within the analytic tradition. However, not all of them agree that its primary purpose is precisely to pursue valuable epistemic goals, such as maximizing true beliefs (while minimizing false ones) or extending knowledge. Indeed, some argumentation theorists maintain that argumentation has a purely persuasive function, that is, its primary goal is to persuade someone of something (e.g., Perelman, Olbrechts-Tyteca 1958; Hamblin 1970), while others hold that argumentation is designed to produce consensus or agreement among interlocutors supporting conflicting claims (e.g., Habermas 1984; van Eemeren, Grootendorst 2004). In contrast, those supporting epistemological approaches to argumentation claim it to be one of the most powerful epistemic practices at our disposal. So, for them, the expected outcome of argumentation is justified belief or knowledge. On this view, a good argument should provide epistemic justification for its conclusion, thereby making it epistemically rational for the audience to believe the content of that conclusion (Biro, Siegel 1992). And, since an

epistemically justified belief is one that is epistemically likely to be true, argumentation should lead people to acquire more accurate beliefs, thereby improving their epistemic position.

The most prominent epistemological account of argumentation has been proposed by Alvin Goldman (1994, 1999, 2003). In his view, argumentation is a social epistemic practice promoting “[...] the exchange of truths through sincere, non-negligent, and mutually corrective speech” (Goldman 1994: 30). Argumentative discourse is expected to occur when, in order to appraise or convince her audience of the truth of a proposition  $p$ , a speaker asserts not only  $p$ ,<sup>108</sup> but presents reasons or evidence in support of it, because simply asserting  $p$  might not suffice to achieve that goal. Goldman points out that argumentative discourse guarantees the truth of its conclusion only when arguers adhere to certain basic norms. These norms specify that they must believe both the premises and the conclusion(s), that they must be justified in believing that the premises support the conclusion(s), and that the premises must jointly provide strong support for the conclusion(s) (Goldman 1994: 34; see also 1999: 134). While these conditions apply solely to the arguer, norms for good argumentation also involve reference to the audience. Accordingly, Goldman outlines also a series of norms that relates argumentative discourse to its intended audience. These norms include, among others, the requirements that all the premises of the argument are believable to at least some members of the intended audience and that the premises–conclusion(s) relationship is presented in ways that promote its understanding by the audience (Goldman 1999: 135-139). It is to be noted that conformity to them does not guarantee the truth of the conclusion, but only helps convince the audience to accept it.

According to Goldman (1994: 27-29), when all these norms guide an argumentative discussion, participants would be better able to critically examine reasons and evidence presented in support of a certain claim and their relationship with it, thus ending up with more accurate beliefs. This makes it clear that in analyzing argumentative discourse he is not interested in describing how argumentation actually works, but in specifying what an ideal arguer is expected to do in order to appraise or convince her audience of the truth of a proposition (Goldman 1994: 44-45).

### 3 Smokrović on argumentation and its expected outcome

Like Goldman, Smokrović intends to show that argumentation is a social practice having significant epistemic import. In particular, as said before,

---

<sup>108</sup> A situation in which it is sufficient to assert  $p$  to appraise or convince a hearer of its truth is labelled “informative discourse” by Goldman (1994: 30-31).

he focuses on its capacity to extend knowledge. Where they diverge is on how the argumentation process and its expected outcome are characterized. This becomes clear when considering the two assumptions on which Smokrović's account is based: (i) reasoning is for argumentation, and (ii) knowledge requires safety. Let's consider them in more detail.

### **3.1 Reasoning is for argumentation**

The first assumption made by Smokrović concerns the nature of reasoning and its relationship with argumentation, issues about which there has been much debate in the last decade. In particular, he takes side with those scholars claiming that reasoning is argumentative in nature. According to them, reasoning is not geared to solitary use, but adapted to be done interpersonally, since it typically occurs in "broader argumentative" contexts where people interact with each other (Hornikx, Hahn 2012). In developing his account, Smokrović explicitly refers to Hugo Mercier and Dan Sperber's naturalistic, evolutionary-oriented argumentative theory of reasoning (2009, 2011a, b, 2017). Within their evolutionary framework, reasoning is a function (and probably the primary one) of an evolved, cognitive module that deals with reasons and their relationship with the claims they purportedly support. Specifically, Smokrović is interested in the two following theses put forward by Sperber and Mercier:

(i) reasoning is "an aspect of social, and more specifically communicative competence" (Mercier, Sperber 2009: 165). Indeed, as observed by Sperber (2001), we cannot understand the emergence of reasoning without considering its role in the evolution of human communication. He points out that communication may have evolved only if it would have been advantageous to both speakers and their audiences. The problem is that the interests of speakers and their audiences usually diverge. Speakers typically communicate to influence their audience, prompting them to believe, feel and act in specific ways. On the other hand, communication is advantageous to audiences because it provides them with reliable information that they could not have obtained independently. The reason why communication has evolved and stabilized is precisely that it can serve both of these purposes. In particular, Mercier and Sperber (2011a: 57) hold that reasoning makes human communication effective and advantageous for both speakers and their audiences because it enables them "to devise and evaluate arguments intended to persuade". So, reasoning as a socio-cognitive tool has evolved to support people in their attempts to convince a cautious audience and to evaluate possibly valuable information that could not be accepted on trust (Mercier, Sperber 2017: 194);



(ii) people reason better in argumentative contexts. According to Mercier and Sperber (2011a: 61-63, 2017: 263-267), reasoning works better when used in dialogical situations, particularly in group discussions, where participants offer each other arguments in support of their viewpoints to convince one another, compared to when one is thinking on one's own. In support of this claim, they have presented a large amount of evidence collected in psychological experiments based on reasoning tasks. This evidence shows that reasoning tasks in which isolated subjects tend to give wrong answers are more frequently solved correctly when they are approached in groups and discussed collectively. For example, while only few subjects gave the right answer in the standard version of the selection task, more than half responded correctly if asked to discuss its solution in a group (see Moshman, Geil 1998). Similarly, as highlighted by Mercier and Sperber (2011a: 61), although it is empirically demonstrated that people find it very difficult to recognize the *modus tollens* (if  $p$  then  $q$ , not- $q$ , so not- $p$ ), when engaged in argumentative dialogues, in order to criticize the claims of their opponents, they recognize and easily apply this argumentative schema (Thompson, Evans, Handley 2005).

### 3.2 The “safety” condition for knowledge

As for the second assumption, Smokrović relies on Williamson's safety requirement on knowledge (see Williamson 2000, 2009a, b).<sup>109</sup> According to this requirement, a subject  $s$  knows a proposition  $p$  only if she is safe from error. And  $s$  is safe from error when there must be no risk or danger that *she* falsely believes in similar cases. On this view, safe belief can be understood as a kind of reliable belief. Accordingly, if  $s$  truly knows the proposition  $p$  in a given case, then that proposition must be true in every similar case she believes that proposition. In Williamson's words, “[i]f one knows, one could not easily have been wrong in a similar case” (Williamson 2000: 147). It is probably clear from what I have said so far that safety is to be viewed as a modal state, namely a state that concerns what could have happened (Williamson 2000: 123).<sup>110</sup>

Williamson (2009b: 14) characterizes the modal notion of safety in terms of possible worlds. According to this characterization,  $s$  is safe from error in believing the proposition  $p$  on a basis  $b$  (or via a method  $m$ ) in the

<sup>109</sup> Two other prominent epistemologists proposing a safety requirement for knowledge are Duncan Pritchard (2007, 2009) and Ernest Sosa (1999). However, there are substantial differences in how Sosa, Pritchard and Williamson formulate the notion of safe belief. Here I will consider only the account proposed by Williamson.

<sup>110</sup> More generally, Williamson (2000: 123) claims that not only safety but also similar notions such as those of stability and robustness are to be conceived as modal states.

actual world if and only if there is no suitably close world in which one believes  $p$  on  $b$  (or via  $m$ ) and  $p$  is false (Williamson 2000: 126-127; see also Williamson 2009a: 325). Simplifying a bit, if  $s$  safely believes that  $p$  on a basis  $b$  (or via a method  $m$ ), then  $p$  cannot be false in all suitably close worlds in which she believes it on  $b$  (or via  $m$ ). By referring to the possible worlds in which  $s$  truly believes  $p$ , Williamson claims his theory to be able to exclude the kind of epistemic luck we can observe in Gettier cases.

It is to be pointed out that Williamson does not consider safety as a necessary and sufficient condition on knowledge, otherwise it would require omniscience on the part of a knower. At most, since he introduces the safety condition with a conditional using the expression “only if”, it can be conceived as a necessary condition. If so, Smokrović (2015: 230) notes, Williamson’s safety requirement is clearly consistent with a counterfactual such as “if the proposition  $p$  had been false, one would (or might) still have believed  $p$ ”. In other words, safety does not preclude the possibility that one might be wrong in believing that  $p$ .

### **3.3 From curiosity to safe knowledge through the argumentation process**

Smokrović (2015) develops his account by integrating the two theses presented above with Williamson’s safety principle. In doing so, he aims to describe argumentation as a communicative process that puts its participants in an epistemically privileged position to safely establish whether a proposition  $p$  (the object of discussion) is true. Accordingly, the expected outcome of the argumentation process should be that of attaining safe knowledge compared to the initial lack of knowledge (or a safer state of knowledge compared to the initially less safe state) concerning the proposition under discussion. This is because people involved in an argumentative process tend to rely on more reliable methods of inquiry than when they are trying to discover individually whether a proposition  $p$  is true. More specifically, by critically examining their arguments for and against  $p$ , they are more likely to acquire accurate beliefs, and thereby extend their knowledge safely.

Let’s look in more detail at how Smokrović (2015: 226-227) describes the communicative dynamics and the structure of the argumentation process. According to him, its starting point is a situation in which an addresser, while sincerely believing the proposition  $p$ , does not know whether  $p$  is true. Insofar as not even her addressee knows whether  $p$  is true,<sup>111</sup> but

---

<sup>111</sup> Henceforth, whenever the gender of the addresser or of the addressee is not specified, the pronoun “she” (plus “her”, “hers” and “herself”) will be used to refer to the

(propositional) curiosity is present in both of them to know that,<sup>112</sup> they may be prepared to involve themselves in an argumentation process.<sup>113</sup> Smokrović points out that although it is clear that the addresser and her addressee play different roles in this process, their involvement suggests that they are committed to pursuing the same goal because they both want to find out whether  $p$  is true. While the addresser is expected to put forward the proposition  $p$  whose merits are to be established, and provide reasons and/or evidence supporting it, it is up to the addressee to evaluate the acceptability of  $p$  and establish its relationship with those reasons and/or evidence.

It is to be noted that, in arguing for their claims, people may make errors of various kinds, or be influenced by *confirmation bias*.<sup>114</sup> In particular, when confirmation bias is at work, they are more likely to focus only on the reasons and/or evidence supporting their claims, and not consider those contrary to those claims (Smokrović 2015: 227). It is clear that when people reason in isolation, they are unlikely to notice when they have made a mistake or when their reasoning has been influenced by confirmation bias. In an argumentative exchange, on the other hand, things are very different. In fact, an addressee may point out to the addresser that her reasoning is fallacious, or that she has failed to consider specific evidence, or that there are inconsistencies in her speech. Moreover, he can counterbalance the addresser's tendency to confirmation bias by presenting counterexamples to her argument in an attempt to falsify the claim under discussion. More generally, Smokrović (2015: 227-228) maintains that the dialogical structure of the argumentation process typically prompts an effect of "stepping back" on the part of the addresser. Indeed, when appropriately challenged by the addressee, she may be stimulated to reconsider the reason(s) and evidence presented to support her claim and reassess the strength of the relationship between the two. In this way, the dialogical structure of the argumentation process has beneficial consequences for the addresser (but also on her addressee) in that it enhances their reasoning abilities and limits the influence of confirmation bias on their way of reasoning (Smokrović

---

addresser and "he" (plus "him", "his" and "himself") to refer to the addressee.

<sup>112</sup> Smokrović (2015: 225) regards propositional curiosity, namely the wish to know whether a proposition is true or false, as the primary form of curiosity, underlying any other type.

<sup>113</sup> Smokrović (2015: 229) rightly observes that, while not knowing whether  $p$  is true, both the addresser and her addressee may be acquainted with some facts concerning  $p$ .

<sup>114</sup> In the psychology of reasoning, confirmation bias refers to the tendency, usually considered contrary to the norms of good reasoning, to look for (or interpret) evidence or proofs in favor of one's beliefs, expectations or hypotheses rather than those that could falsify them (see Nickerson 1998).

2015: 229). Indeed, by challenging each other, they are prompted to use the most reliable methods to establish the truth of the claim under discussion, thereby increasing their chances of attaining safe knowledge. In Williamson's terms, if they come to establish the truth (or the falsity) of the claim under discussion by using argumentative exchange, it is likely that they would not be wrong about the claim in other similar situations (namely, in worlds suitably close to the actual one).

Smokrović (2015: 229) proposes an interesting example to show how participants in an argumentation process can safely extend their initial knowledge, or revise their false beliefs, by producing and evaluating each other's arguments. Imagine, he suggests, that Hercule Poirot and Colonel Hastings are trying to solve a murder case. Suppose Hastings claims it was the gardener, even if he does not know whether it is true, but he truly believes that to be the case. Poirot himself does not know whether the gardener is the murderer. So, there is a proposition  $p$ , "The gardener is a murderer", of which both of them are curious to ascertain the truth. Poirot and Hastings certainly possess some evidence related to the crime scene, about which (we can say) they may have safe knowledge. The moment Hastings uses some of this evidence to support  $p$  directly or to develop a piece of reasoning having  $p$  as its conclusion, we can consider them to have entered into an argumentative exchange. In doing so, he might be influenced by confirmation bias in that he might have considered only evidence in support of  $p$  or not have noticed other evidence that is contrary to it. Furthermore, his piece of reasoning might be fallacious, e.g., it might start from mistaken or unjustified premises. Clearly, Poirot might consider the evidence sufficient or the piece of reasoning convincing and thus accept Hastings's claim. More probably (knowing Poirot) he might point out that Hastings has failed to consider other evidence or that his reasoning does not work. In other words, as addressee, the fundamental role of Poirot is to correct what is wrong in Hastings' reasoning. In fact, there is no way they could attain safe knowledge about  $p$  by relying on epistemically unsound bases or methods. It should be pointed that they can swap roles, and so after Poirot has pointed out the weaknesses in Hastings' reasoning, Hastings can inform Poirot whether his counter-reasoning actually works. In this virtuous argumentative exchange, it is possible for both of them to overcome the negative influence of confirmation bias and fine-tune their ability to reason. Thanks to these mutual exchanges and role-swapping, it is likely that they will acquire safe knowledge about  $p$  (whether true or false) in a way that makes it unlikely that they would be wrong in other similar cases (Smokrović 2015: 231).

#### 4 Cooperation and adversariality in argumentation

The picture of argumentation provided by Smokrović is a very comforting one, also because he describes it as a process of mutual adjustment between an addresser and her addressee.<sup>115</sup> This process starts because of their shared curiosity about whether the proposition under discussion is true and continues until they safely establish whether it is or not. However, they may come to the conclusion that it is impossible to establish (or find an agreement about) the truth of the proposition. Even in this latter case, however, the fact remains that both participants were prepared to find out something more about *p*. This means that, according to Smokrović, in order to be engaged in an argumentative exchange, there must be a modicum of cooperative attitude on the part of the participants. Indeed, in Gricean terms (1975: 26-30), this desire to know whether the proposition under discussion is true can be seen as the accepted purpose or direction of the talk exchange in which addresser and addressee are engaged. Of course, they may have a whole range of different reasons for aiming to achieve that goal. But the fact remains that wanting to know whether that proposition is true is the mutually accepted goal for both of them. Accordingly, the representation of the argumentation process provided by Smokrović contains an essential cooperative component. This does not preclude the fact that the argumentative exchange must be adversarial to some extent. It is a fact that the absence of conflict would make it impossible (or well-nigh impossible) to achieve safe knowledge. Indeed, it is precisely the adversariality which is present in a cooperative framework that puts its participants in an epistemically privileged position from where to safely establish whether the proposition under discussion is true. However, this way of conceiving adversariality does not seem to fit well with what Mercier and Sperber have described in some of their writings. In particular, in response to comments on their target article in *Behavioral and Brain Sciences*, when clarifying their views on the social origin and functions of reasoning they claim that:

“[t]he main function of reasoning is indeed social, but by serving the social interests of individuals rather than the collective interests of the group” (Mercier, Sperber 2011b: 96);

“[w]hat makes communication advantageous to communicators is that it allows them to achieve some desirable effect in the receivers. For this, the information they emit has to be conducive to this effect, whether it is true or false” (Mercier, Sperber 2011b: 96);

---

<sup>115</sup> As we saw with Poirot and Hastings, the participants can exchange roles: the participant playing the role of addresser can shift to the role of addressee and vice versa.

“[communicators] argue for whatever it is advantageous to them to have their audience believe” (Mercier, Sperber 2011b: 96).

Let me remind you that Smokrović assumes Sperber and Mercier’s theory to be the starting point for his account of argumentation. Accordingly, the problem for him is how to connect these quotations as to the irrelevance of truth and the collective interests of the group to his account. Indeed, insofar as Mercier and Sperber hold that the function of reasoning is to give us what is advantageous to us, they appear to conceive of reasoning as an instrument for achieving one’s interests in communicative contexts. Thus, they trace the argumentative nature of reasoning back to the purely individualistic component of dialogical situations, and namely, the goal of achieving “personal” advantages from these situations. If we follow this line of argument, however, reasoning should be conceived as best adapted to adversarial contexts. Accordingly, the function Smokrović attributes to reasoning in the argumentation process should be seen at best as a by-product of its original “strategic” function.<sup>116</sup> Obviously, the fact remains that reasoning is argumentative by its nature, but given its strategic function in argumentative exchanges, it seems unable to play a significant role in the attainment of safe knowledge. Indeed, if reasoning evolved as a strategic instrument to serve one’s interests, and not to establish the truth, any truth that results from it would be mostly incidental. Moreover, the reason(s) an addresser provides in support of her claim should not be regarded as aimed at truth, but at persuading her addressee of whatever is to her advantage to have him believe. In short, if Smokrović assumes Sperber and Mercier’s theory to be the starting point for developing his account of argumentation, he really needs to explain whether (and if so, how) their way of conceiving reasoning fits into his curiosity-driven, cooperative picture of the argumentation process.

It seems to me that Smokrović has three possible ways of defending his account against this criticism.

First, he could respond that he does not take the entire argumentative theory of reasoning at face value, but only the part regarding the effectiveness of reasoning in dialogic situations. However, since Smokrović has clearly stated that he wants to integrate the argumentative theory with epistemological insights, this way of overcoming the criticism is not viable.

---

<sup>116</sup> In two previous works, I have argued that the argumentative theory of reasoning represents reasoning as a persuasive device and so appears to presuppose an instrumental conception of rationality (see Labinaz 2014, 2020). Catarina Dutilh Novaes (2018) and David Moshman (2018) have also argued that Mercier and Sperber emphasize too much the “strategic” component of reasoning in argumentation.

In fact, he does not seem to distance himself from specific aspects of the argumentative theory. It is also true that if it is his intention to integrate the theory, he is obviously not entirely satisfied with it, and so we would expect him to explain how his revised or extended version is supposed to be better than the original. But there is none of that in Smokrović's article. As far as I can understand, he does not wish to reject specific parts of the argumentative theory, including the one characterizing reasoning as a strategic instrument serving one's own personal interests.

Second, he could point out that the criticism focuses only on one dimension of the reasoning capacity (namely that of producing arguments), and omits the one related to evaluating arguments. That is true. If one considers this second dimension, one could argue that the capacity to evaluate arguments can counterbalance attempts at persuasion that are accomplished through the capacity to produce arguments. However, the fact remains that the quotations presented above suggest that people do not usually get involved in an argumentative process with the aim of knowing whether  $p$  is true, but in order to convince the addressee that they are right, regardless of the truth of what they are claiming. From this perspective, cooperativity should not be considered a relevant property of the argumentation process. However, this conflicts with Smokrović's picture of the argumentative process presented above.

Third, like Goldman, Smokrović could argue that his account describes what happens in an argumentative exchange in a somewhat idealized way. Accordingly, the argumentation process presented above does not need to exactly fit real-life instances of argumentative discourse. Instead, it describes what is supposed to happen in an argumentative exchange when certain conditions hold. More specifically, by considering argumentation as a curiosity-driven, cooperative process, Smokrović focuses on how an addresser and her addressee should behave in an argumentative exchange if they want to acquire safe knowledge. This means that his account concerns a very specific kind of argumentative exchange. In other words, it is the kind of argumentative exchange in which the adversarial component is not predominant. After all, argumentation can occur in various different forms of interaction, including negotiation, persuasion inquiry, deliberation, information-seeking and so on (for a complete list, see Walton 1998: 30-34). Some are highly adversarial (e.g., persuasion), while others are cooperative in nature (e.g., inquiry). But if argumentation is not inherently adversarial (despite the fact that certain forms of interaction in which it is involved are adversarial), it would be wrong to consider the argumentative function of reasoning as essentially persuasive. Indeed, reasoning also serves argumentation in situations where cooperativity is predominant or



adversariality occurs in a cooperative framework, such as in the kind of argumentation process described by Smokrović. What is needed, then, is another way of elaborating the idea that reasoning is argumentative by its nature, which can then be applied to those cases of argumentation lacking in any predominant adversarial component. If we could do this, then we could argue that the reasoning involved in the kind of argumentative process described by Smokrović is not the same kind of reasoning suggested by Sperber and Mercier.

## **5 The reason-giving function of reasoning and its role in the argumentation process**

The claim that reasoning has an argumentative function may be understood in (at least) two different ways: as the attribution of a function which is either reason-giving or persuasive. The reason-giving function refers to one's ability to reason, that is, the ability to make the connection between premises and conclusions. The persuasive function consists instead of the ability to produce convincing arguments, namely to be able to convince one's addressee of something. It should be noted that, on the one hand, one may be very good at making premises-conclusions connections, but not necessarily interested in using this ability to produce arguments to convince other people and, on the other, that it is quite possible to successfully convince other people with what one says without needing to use one's ability to make premises-conclusions connections. In light of this, one could justifiably assume that these abilities are independent of each other. They might have developed at different times and for very different reasons, which are not necessarily connected. By way of contrast, the argumentative theory of reasoning conflates being able to reason and being able to produce convincing arguments with the more general ability to argue. However, reducing the argumentative function of reasoning to producing persuasive arguments does not really explain much about our special attention for (and interest in) reasons and their relationship to the claims in support of which they are presented. Indeed, devising arguments intended to persuade is only one of various things that we can do thanks to our ability to reason. For example, we can provide reasons for a whole series of motives, such as giving meaning to our speech or actions, making explicit the premises of what we say or do, collaborating with others, or even improving our self-image. Of course, we can also give reasons in order to persuade others. But there is no compelling reason why the main function of reasoning should be regarded as primarily persuasive. Instead, it seems more reasonable to consider the reason-giving function of reasoning as more basic than its persuasive function. If we conceive this as the primary

function of reasoning, our interest in reasons may have something to do with accountability, rather than personal advantage. Philip E. Tetlock (together with Jennifer Lerner), who first defined this term, characterizes it as “the implied or clearly expressed expectation that one will be called upon to justify one’s beliefs, feelings or actions to others” (Lerner, Tetlock 2003: 434). It is thanks to reasoning that we can give reasons to explain and justify ourselves. More specifically, insofar as we are held to be accountable for our beliefs, feelings or actions, reasoning enables us to motivate and justify these beliefs, feelings and actions to others, thereby letting people know what to expect of us. Obviously, the reason-giving function of reasoning can be exploited to convince others of a certain claim or opinion and thus become part of a more complex social situation involving adversariality.

If we focus now on our ability to argue, we can observe that this ability depends on our acquaintance with specific social practices, which involve, among other things, the rules and expectations that guide our behavior when engaged in them. As suggested by many argumentation theorists (cf. van Eemeren, Grootendorst 2004; Walton 1998), these rules and expectations involve (among other things) attributions of entitlement, undertakings of commitments, turn-taking, ways of questioning each other’s claims, adoption of standards of precision, and so on. However, there is no single set of rules and expectations that governs any argumentative practice. Even if argumentative exchanges often seem to evolve into a competition among interlocutors, argumentation is not a monolithic enterprise. Indeed, as said above, adversarial interaction is just one of the forms of interaction where argumentation is involved. These forms of interaction can differ because of the initial situation, the goals of the arguers involved, and the aim of the interaction as a whole. Differences on these aspects determine differences related to the rules and expectations governing a certain argumentative exchange. Being able to make premises-conclusions connections, then, is a necessary, but not sufficient, condition to be engaged in argumentative practices. Indeed, one needs to respect the rules characterizing the argumentative situation in which one is involved and be guided by expectations as to how to proceed when engaged in it if one wants to achieve the goal one is pursuing to achieve in that situation.

Returning to the picture of the argumentation process provided by Smokrović, the beneficial epistemic effect to which he refers (which is attaining safe knowledge) may only come about against a backdrop of specific conditions that ensure a certain degree of cooperativity between interlocutors. These conditions apply to both the interlocutors involved in the argumentation process and the interaction that takes place between them. If these conditions obtain, the result will be that no interlocutor emerge as

“loser.” Indeed, those involved in the argumentation process described by Smokrović are expected to benefit from the exchange by acquiring something of value, namely safe knowledge. It is hard to see here how reasoning as a persuasive device can be suitable for the purposes of this argumentative process.<sup>117</sup> If such reasoning were predominant in it, then arguers were involved in an argumentative practice, which may be similar to some degree, but not identical, to that described by Smokrović. If instead we consider reasoning in its basic form, that is as a reason-giving device, we can understand why people can jointly try to find out whether the proposition under discussion is true. On the one hand, for the addresser, it is a question of accountability in front of her addressee: she will want to bring appropriate epistemic support to her claim by providing what she recognizes as valuable epistemic resources in the form of evidence, justification etc. On the other, her addressee may want to evaluate the accountability of the addresser by challenging her claim through appropriate objections, requests for further clarifications, or defeaters. And it is precisely this kind of adversariness occurring in a cooperative framework that puts its participants in an epistemically privileged position to safely establish whether the proposition under discussion is true.

## 6 Concluding remarks

This paper dealt with Smokrović’s thought-provoking account of argumentation as a curiosity-driven, cooperative effort. I have focused mainly on one of the two assumptions on which his account is based: that reasoning is argumentative in nature. Since Smokrović assumes Mercier and Sperber’s argumentative theory as the starting point for developing this account, I have tried to highlight the difficulties one encounters when attempting to put together their way of conceiving reasoning as a persuasive device with the picture of the argumentation process Smokrović provides. I have suggested that one can assume the argumentative nature of reasoning while dismissing that picture of reasoning. In particular, I have proposed an alternative way to elaborate the idea that reasoning is argumentative by highlighting its reason-giving function. Indeed, since devising arguments intended to persuade is only one of the various things that we can do thanks to our ability to reason, we should consider the reason-giving function as more fundamental than that of devising persuasive arguments. I went on to argue that the former function appears to be better suited to the argumentation process described by Smokrović than the latter. Indeed, as

---

<sup>117</sup> Obviously, this does not mean that reasoning aimed at producing persuasive arguments cannot be appropriate in certain interactions involving argumentation.

I have tried to show, only if we consider reasoning in its basic form, that is as a reason-giving device, can we understand why two or more people driven by their curiosity can get together in a collaborative effort to safely establish whether a certain proposition is true, without manipulating each other. If their goal were to use their reasoning capacity to manipulate each other, they would no longer be engaged in the argumentative process described by Smokrović but in a totally different argumentative practice in which the adversarial component would be predominant.

## References

- Biro, John, and Harvey Siegel. 1992. "Normativity, Argumentation, and An Epistemic Theory of Fallacies". In *Argumentation Illuminated*, ed. by Frans H. van Eemeren, Robert Grootendorst, Jonathan A. Blair and Charles A. Willard, 81-103. Amsterdam: SICSAT.
- Biro, John, and Harvey Siegel. 1997. "Epistemic Normativity, Argumentation, and Fallacies". *Argumentation* 11: 277-292.
- Biro, John, and Harvey Siegel. 2006. "In Defense of the Objective Epistemic Approach to Argumentation". *Informal Logic* 26(1): 91-101.
- Dutilh Novaes, Catarina. 2018. "The Enduring Enigma of Reason". *Mind & Language* 33: 513-524.
- van Eemeren, Frans H., and Robert Grootendorst. 2004. *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge: Cambridge University Press.
- Goldman, Alvin I. 1994. "Argumentation and Social Epistemology". *The Journal of Philosophy* 91(1), 1994: 27-49.
- Goldman, Alvin I. 1999. *Knowledge in a Social World*. Oxford: Clarendon Press.
- Goldman, Alvin I. 2003. "An Epistemological Approach to Argumentation". *Informal Logic* 23: 51-63.
- Grice, Paul. 1975. "Logic and Conversation". In *Syntax and Semantics 3: Speech Acts*, ed. by Peter Cole and Jerry Morgan: 41-58. New York: Academic Press. Reprinted in P. Grice (1989), *Studies in the Ways of Words*: 22-40. Cambridge, MA: Harvard UP.
- Habermas, Jürgen 1981. *Theorie des kommunikativen Handelns*. Bd. 1: *Handlungs-rationalität und gesellschaftliche Rationalisierung*. Frankfurt a. M: Suhrkamp. English Translation (1984), *The Theory of Communicative Action*. Vol. I: *Reason and the Rationalization of Society*. Boston: Beacon.
- Hamblin, Charles. 1970. *Fallacies*. London: Methuen.
- Hornikx, Jos, and Ulrike Hahn. 2012. "Reasoning and Argumentation: Towards an Integrated Psychology of Argumentation". *Thinking & Reasoning* 18: 225-243.
- Labinaz, Paolo. 2020. "Reasoning, Argumentation and Rationality". *Etica e Politica* 16: 576-594; reprinted in *Moral Realism and Political Decisions. Practical Rationality in Contemporary Public Contexts*, ed. by Gabriele De Anna and Riccardo Martinelli: 81-103. Bamberg: University of Bamberg Press.

- Labinaz, Paolo. 2020. "In What Sense (If Any) Can Rationality Be Said to Be Universal?". *Isonomia* 10: 75-103.
- Lerner, Jennifer S., and Philip E. Tetlock. 2003. "Bridging Individual, Interpersonal, and Institutional Approaches to Judgment and Choice: The Impact of Accountability on Cognitive Bias". In *Emerging Perspectives on Judgment and Decision Research*, ed. by Sandra Schneider and James Shanteau: 431-457. Cambridge: Cambridge University Press.
- Lumer, Christopher. 2005. "The Epistemological Theory of Argument – How and Why?". *Informal Logic* 25: 213-243.
- Mercier, Hugo, and Dan Sperber. 2009. "Intuitive and reflective inferences". In *In Two Minds: Dual Processes and Beyond*, ed. Jonathan St.B.T. Evans and Keith Frankish: 149-70. Oxford: Oxford University Press.
- Mercier, Hugo, and Dan Sperber. 2011a. "Why Do Humans Reason? Arguments for an Argumentative Theory". *Behavioral and Brain Sciences* 34: 57-74.
- Mercier, Hugo, and Dan Sperber. 2011b. "Argumentation: Its Adaptiveness and Efficacy". *Behavioral and Brain Sciences* 34: 94-111.
- Mercier, Hugo, and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.
- Moshman, David, and Molly Geil. 1998. "Collaborative Reasoning: Evidence for Collective Rationality". *Thinking & Reasoning* 4: 231-248.
- Moshman, David. 2018. "Essay Review of *The Enigma of Reason* by Hugo Mercier and Dan Sperber". *Human Development* 61: 60-64.
- Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises". *Review of General Psychology* 2(2): 175-220.
- Perelman, Charles, and Lucie Olbrechts-Tyteca. 1958. "Traité de l'argumentation. La nouvelle rhétorique". Paris: PUF. English Translation (1969), *The New Rhetoric: A Treatise on Argumentation*. Notre Dame: University of Notre Dame Press.
- Pritchard, Duncan. 2007. "Anti-Luck Epistemology". *Synthese* 158: 277-98.
- Pritchard, Duncan. 2009. "Safety-Based Epistemology: Whither Now?". *Journal of Philosophical Research* 34: 33-45.
- Smokrović, Nenad. 2018. "Informal Reasoning and Formal Logic: Normativity of Natural Language Reasoning". *Croatian Journal of Philosophy* 17(54): 455-471.
- Smokrović, Nenad. 2017. "Normativity of Logic and Everyday Reasoning". *Analiza* 21(2): 45-61.
- Smokrović, Nenad. 2015. "Argumentation as a Means for Extending Knowledge". *Croatian Journal of Philosophy* 15(44): 223-231.
- Smokrović, Nenad. 2011. "Logical Consequence and Rationality". In *Between Logic and Reality: Modeling Inference, Action and Understanding*, ed. by Majda Trobok, Nenad Mišćević and Berislav Žarnić, 173-192. Heidelberg: Springer-Verlag.
- Sosa, Ernest. 1999. "How Must Knowledge Be Modally Related to What Is Known?". *Philosophical Topics* 26 (1&2): 373-384.
- Sperber, Dan. 2001. "An Evolutionary Perspective on Testimony and Argumentation". *Philosophical Topics* 29: 401-143.

- Thompson, Valerie A., Evans, Jonathan St.B.T., and Simon J. Handley. 2005. "Persuading and Dissuading By Conditional Argument". *Journal of Memory and Language* 53(2): 238-257.
- Walton, Douglas. 1998. *The New Dialectic: Conversational Contexts of Argument*. Toronto: University of Toronto Press.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
- Williamson, Timothy. 2009a. "Replies to Critics". In *Williamson on Knowledge*, ed. by Patrick Greenough and Duncan Pritchard: 282-384. Oxford: Oxford University Press.
- Williamson, Timothy. 2009b. "Probability and Danger". *The Amherst Lecture in Philosophy* 4: 1-35. <<http://www.amherstlecture.org/williamson2009/>>.





NENAD MIŠČEVIĆ

## How Rational are Human Beings? In Honor of Nenad Smokrović

**Abstract:** How rational are human beings? Is there rational thinking in an ordinary, positive, not crippled way? Is there epistemically virtuous thinking? The paper addresses these questions in a dialogue with Smokrović. Two questions are raised. First, is there rational thinking in an ordinary, positive, not crippled way? Second, is there epistemically virtuous thinking? To my mind, the questions go together, and their answers should be either jointly positive - people think and infer in an ordinary, positive, not crippled way, or jointly negative, their reasoning is not epistemically virtuous, but rather vicious. We propose to distinguish five degrees or stages of rationality of reasoning: at the top we are ideal and fully rational individuals, and at the bottom epistemically more or less crippled ones. There is a realistic option in between, the stage number three. I propose that ordinary human reasoning belongs to the bottom part of the hierarchy; people do not reason rationally in a positive, complimentary way. Smokrović is a bit more optimistic, and his views are discussed in some detail. At the end we propose employing the framework of virtue epistemology, by taking positive rationality as an epistemic virtue and crippled rationality (or simply irrationality) as an epistemic vice.

**Key words:** reasoning, rationality, degrees, virtue epistemology.

### 1 Introduction

Smokrović is the only Croatian philosopher who published a book on rationality and inference. His 2004 book, *Priroda prirodnog zaključivanja*, (*The nature of ordinary reasoning*) is the only systematic book on the topic in Croatian. It is systematic, excellently informed, and original in its main conclusions; we shall discuss them, in the developed form they take in his more recent work, below. He anticipated them as early as early as 1997, in an article available only in Slovenian translation (see References). The present paper addresses human reasoning from the perspective of study of epistemic virtue and vice. Nenad is the right person to whom one can dedicate this kind of exposition.

Here is the preview. I shall follow Smokrović and discuss the epistemology of rationality of reasoning, setting aside the challenging issues connected to rationality in the choice of ultimate goals in one's life and similar

issues (see Nathanson, (1999) Nozick, (1993)). The term “rationality” will therefore stand for the narrower, but in many respect central meaning of rationality in reasoning. After a brief mention of Smokrović’s kind comment on my claims on the epistemic role of curiosity, the paper turns to his analysis of rationality and the question of how rational human beings are. First, following his (2004) book, it summarizes the main problems that contemporary cognitive research has bequeathed to us, epistemologists. It then turns to the task of characterizing stages of rationality of reasoning, and offers a proposal of five such stages, from a miserable, “crippled” one, through intermediary, less imperfect ones, to a semi-divine idealization of logically ideal reasoning. I am taking this slightly roundabout way in order to offer a framework in which one can discuss Smokrović’s picture of ordinary rationality: once possible pigeon-holes for such a rationality are identified, the discussion becomes clearer and the prospect of finding a good answer appears on the horizon. It ends with longer questions for Smokrović, concerning his optimism or pessimism regarding human reasoning: how perfect is it, and is ordinary human rationality an epistemic virtue or not? How rational are human beings; this is the main topic of the central part of the paper. At the end I briefly return to the possibility of addressing human reasoning from the perspective of study of epistemic virtue and vice, thus proposing the virtue epistemology of rationality as a promising way forward. So, let me conclude the section by expressing gratitude to Smokrović. First, for his discussion of my views concerning the positive epistemic role of curiosity. He notes that positive curiosity is crucial for logical argumentation:

***Curiosity and Argumentative Process***

Connecting the notion of curiosity and argumentative process I will show that there is a particular form, a specific narrow type of argumentation that fits appropriately my claim: if it happens that epistemic curiosity becomes an intrinsic motivating force in the argumentative process, then merely rhetorical “fight” of the hitherto opposing participants can be transformed into joint activity guided by the goal shared by all participants. They are ready to accept the goal and to form the intention to do their best to find out whether a proposition  $p$  is true. (2017: 83).

I completely agree. Let me also thank him for his kind discussion of my views of apriority of logic in his brilliant “Logičko znanje i apriorno opravdanje”.

## 2 Problem of defective rationality – overview

Smokrović entitles the first part of his (2004) book on rationality “The dire perspectives for human rationality” and gives a detailed overview of the problem of defective rationality of human reasoning detected by cognitive scientists. In the second part he is a bit more optimistic, suggesting this already in its title: “Is there nevertheless some place for optimism concerning human rationality?”

So, let us start by taking a look at what Smokrović rightly describes as “the dire perspectives for human rationality”. We can follow the mainstream description and characterize the vice of human reasoning capacity as cognitive miserliness, or laziness, the contrary of positive, epistemically virtuous rationality. Cognitive miserliness (the term has been made famous by K. Stanovich) goes pretty far. It might be constitutive for normal functioning of our cognitive system, in particular in relation to reasoning and inference. Daniel Kahneman, one of the pioneers in research on these topics has proposed his model of heuristics and biases as exemplifying such miserliness (see, for instance, Kahneman (2011)). A few lines below, we shall be retelling Smokrović’s presentation of two famous examples from cognitive research on the topic.

### *Nenad Smokrović on Wason selection task*

(Wason, 1966). Although not strictly a deductive reasoning task, the selection task involves the logic of conditionals and is considered part of the literature on deduction. In a typical abstract version of the problem, participants are shown four cards lying on a table and told that each has a capital letter on one side and a single figure number on the other. The visible sides are B L 2 9. They are told that the following rule applies to these four cards and may be true or false: If a card has a B on one side, then it has a 2 on the other side.

B L 2 9

Which card should you turn to check if the rule applies?

Wason argued that the correct choice is B and 9 because only a card with a B on one side and a number other than 2 on the other side could disprove the rule. (...) Only around 10% of university students typically choose the B and 9. The most common choices are B and 2, or just B. Wason originally argued that this provided evidence of a confirmation bias in reasoning (Wason & Johnson-Laird, 1972). That is, participants were trying to discover the confirming combination of B and 2 rather than the disconfirming combination of B and 9.

Smokrović also usefully presents the other famous example of miserly processing, the Linda story from Tversky and Kahneman. It concerns a woman called Linda; here is Smokrović's retelling of it:

Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations. (Kahneman 2013, 211 ff.) In one version of the study, the researchers then presented eight possible scenarios for Linda's future. They were:

- a) Linda is a teacher in primary school;
- b) Linda works in a bookstore and practices yoga;
- c) Linda supports the feminist movement (abbreviation: (F));
- d) Linda is a social worker and helps people with mental disorders;
- e) Linda is a member of the Women's Electoral League;
- f) Linda is a bank teller;
- g) Linda is an insurance agent;
- h) Linda is a bank teller and is active in the feminist movement .

The participants of the study were asked which of the following two propositions is more probable. "Linda is a bank teller" (T) or "Linda is a bank teller and is active in the feminist movement" ( $T \wedge F$ ).

Smokrović, following the original authors, then notes that the members of the group who did not have any education in probability and statistics judged the proposition "Linda is is a bank teller and is active in the feminist movement" more probable than the proposition "Linda is a bank teller", as he says "in spite of the obvious fact that one cannot be a feminist *and* a bank teller" without being *a bank teller*." (2004: 31, transl. NM)

Daniel Kahneman has, together with Amos Tversky, developed the consequences of numerous experiments, similar to Linda and to Selection tasks. He concludes that our mind is composed of two systems. System-2 is the conscious one, it is „we“; it is cautious, but not very active. The other, System-1 is automatic, unconscious and quick. He writes:

The defining feature of System 2, in this story, is that its operations are effortful, and one of its main characteristics is laziness, a reluctance to invest more effort than is strictly necessary. As a consequence, the thoughts and actions that System 2 believes it has chosen are often guided by the figure at the center of the story, System 1. However, there are vital tasks that only System 2 can perform because they require effort and acts of self-control in which the intuitions and impulses of System 1 are overcome. (2011:33).

System 1 does not reflect, it just proceeds automatically with little effort. Kahneman notes that „ it sometimes answers easier questions than the one it was asked, and it has little understanding of logic and statistics.” (2011: 27)<sup>118</sup>

In the original Linda problem the conjunction appeared plausible, and that sufficed for an endorsement.<sup>119</sup>

Similarly, System 1 sometimes substitutes an easier question for a difficult one (heuristics), overweighs low probabilities, responds more strongly to losses than to gains (loss aversion)\*, frames decision problems narrowly, in isolation from one another and over-exaggerates emotional consistency (halo effect). It reacts more strongly to losses than to gains.

How does all this happen? What is the underlying mechanism? Let me mention one line of answer which I find attractive, it is due to Philip N. Johnson-Laird, and it works with the notion of mental models. To use the case of Linda, here is what the model theory would say:

Remember that participants in many experiments gave a higher ranking to the probability that Linda is a feminist bank teller than to the probability that Linda is a bank teller. Her description is more representative

---

<sup>118</sup> Here is one example due to Kahneman:

Now I will show you a logical argument—two premises and a conclusion. Try to determine, as quickly as you can, if the argument is logically valid. Does the conclusion follow from the premises?

All roses are flowers.

Some flowers fade quickly.

Therefore some roses fade quickly.

A large majority of college students endorse this syllogism as valid. In fact the argument is flawed, because it is possible that there are no roses among the flowers that fade quickly. Just as in the bat-and-ball problem, a plausible answer comes to mind immediately. (2011: 46)

<sup>119</sup> And similar reasoning appears concerning the topic of football:

Italy and France competed in the 2006 final of the World Cup. The next two sentences both describe the outcome: “Italy won.” “France lost.” Do those statements have the same meaning? The answer depends entirely on what you mean by *meaning*. For the purpose of logical reasoning, the two descriptions of the outcome of the match are interchangeable because they designate the same state of the world. As philosophers say, their truth conditions are identical: if one of these sentences is true, then the other is true as well. “Italy won” evokes thoughts of the Italian team and what it did to win. “France lost” evokes thoughts of the French team and what it did that caused it to lose, including the memorable head butt of an Italian player by the French star Zidane. In terms of the associations they bring to mind—how System 1 reacts to them—the two sentences really “mean” different things. The fact that logically equivalent statements evoke different reactions makes it impossible for Humans to be as reliably rational as Econs. 354.

of the former than the latter (as independent ratings showed). The mental model theory would explain this in the following way: the participants, average human beings from our culture, having listened to the description of Linda, have built a mental model of a bank teller always politically engaged for women's rights and quite a feminist. They don't have a model of a bank teller who was fighting for women's rights but is not a feminist anymore; no model of theirs for the lonely claim "Linda is a bank-teller", separately from the conjunction with feminism. So, they chose the proposition for which they have a ready model.

Keith Stanovich has introduced the term "cognitive miser" and "miserly processing, and we shall be using it in what follows. So, we shall accept the main thesis of the researches just mentioned: miserliness is the main characteristics of everyday reasoning and inference. And here, we shall be interested in vices-disabilities, like miserliness. This will be our topic till the end of the paper.

So, to repeat the question from its title, how rational are human beings? Is their thinking rational in an ordinary, positive, non-crippled way? Smokrović seems rather optimistic about the answer. For instance, in his (2004) book, he sides with the optimists:

We have tried to support the claim, that people, in spite of dominant pessimistic views, reason in accordance with logic. We accept the thesis only under the assumption that logical principles can be implemented in human mind, give its limitations and its given resources. In the context the following general principle holds: the richer and more expressive the logic, the smaller its implementability. Therefore, in a limited system such as human mind one can deal only with basic and very limited logical principles. (2004: 177)

His clearest statement comes from his 1997 paper "*Ali ljudje sklepajo racionalno?*" (Do humans infer in a rational way?) Here is his abstract:

In the present paper we analyze two contrasting understandings of rationality, which appear to be mutually irreconcilable. The first, optimistic view, sees human beings as being rational and logical. The second, pessimist view, sees irrationality as the basic state of human cognitive setup. In spite of this situation, I believe that one can envision a resolution of these contraries. If we give up on radical variants on both sides (theory of full rationality vs. eliminativism), we are left with more moderate positions, among which one can find a compromise solution of our problem. My choice is on the side of the thesis of rationality. The basic claim is that human beings are not fully rational, but rationality is still in a way, a part of their reasoning.

The proposal that I find acceptable goes in the direction of weakening the rationality thesis. The claim that logical relations are paradigms of reasoning-inference, and that the later corresponds to some given logical system, requires essential corrections (against the exaggerated postulation that has to be satisfied by logical relations among beliefs). I propose that it is sufficient for rationality, that relations between contents exemplify proto-logical forms. These forms do not constitute a logical system. Proto-logical forms are the innate basis, from which, with the help of learning, logical competences are being developed. No infallibility follows from these assumptions. We gain logical competencies on the ground of innate and universal logical kernel.

Optimism stays with him, throughout his rich investigation. Take his (2018) example with the bus platform, and the person we called Jane, who reflects about train schedule, and successfully performs disjunction elimination. His favorite idea is that such local, unsystematic cases of logical micro-reasoning, somehow make the thinker into a rational person. (1997: 17).

Smokrović appeals to John MacFarlane's "bridge principle" connecting a valid logical form with the corresponding norm of reasoning. In our example, disjunction of elimination would be connected to the norm requiring Jane to pass from her two initial beliefs to the conclusion that the train starts from platform 1. (218: 460). Here is a quote from Smokrović illustrating his appeal to "bridge principle" in ordinary reasoning:

Assigning the directive role to normative rules for reasoning one understands normativity in a stronger sense than taking it to have only evaluative role. If normativity is directive it is in principle also evaluative, while the evaluative role does not imply the directive one. It seems that Harman had in mind the directive role of logic for reasoning when he denied its normative influence. Accordingly, in order to defend the normativity thesis against Harman's objections, the strong, directive meaning of normativity has to be embraced. Summing up the discussion in this chapter and putting together the questions of scope and of ways of understanding normativity, among the possible answers to these questions I'm picking up the restricted, apprehensive scope of rules' application and the directive, guiding role of normative rules. They together determine the desiderata, for, I hope, a promising way to uphold my view of normativity that is going to be exposed in §4. (2018: 465).

Well, people normally do perform such inferences; unfortunately, they are not systematic and cannot be put under the roof of a single, or a small number of logical systems. Smokrović defends the thesis that people normally reason rationally by introducing a strong plurality of systems:



Reasoning as a cognitive activity is not a uniform endeavor and it can't be idealized as having a closed list of characteristics and normative constraints. On the contrary, as it is indicated above, people in real-life situations perform different forms of reasoning, each form guided by a different goal. Being engaged in various forms, accomplishing different goals, they can be normatively warranted from the viewpoint of different logics. (2018: 468)

He proposes a “weakening of rationality thesis”: it is sufficient for rationality that relations between beliefs embody some proto-logical forms, which do not add to a logical system. These forms are the innate universal bases, out of which logical capacities develop. (1997: 13).

### 3 A positive proposal: degrees of rationality

I agree with a lot of what Smokrović has to say. We both think that usual human reasoning does not embody any systematic logical patterns; Smokrović says this in the abstract we just quoted. My question is how optimistic we should be about characterizing this reasoning? Does rationality admit of stages?

Let me introduce, with apology to Smokrović, the idea of stages or degrees of rationality, itself quite close to Smokrović's characterization. I have mentioned at the beginning that this is a slightly roundabout way to discuss his view, but I hope it can be useful, and can make the issues clearer. Here is a fine formulation of the ordinary understanding of rationality due to Julia Staffel:

The scale associated with “rational” in ordinary English clearly has a maximum. This is confirmed by the observation that it makes sense to say “James is completely rational,” or “James' credences are completely rational” (and that, by contrast, it doesn't make sense to say “James is completely tall”). Hence, given the categories introduced above, “rational” is an absolute gradable adjective. Relatedly, the scale for “irrational” also seems to be closed, as it makes sense to call something “entirely irrational” or “completely irrational.”<sup>4</sup> This means that “irrational” behaves similarly to other closed-scale adjectives like “empty,” rather than open scale-ones like “bad.” This also means that if “rational” is an absolute gradable adjective, we should expect positive rationality judgments, such as “Her beliefs are rational” or “She is rational,” or “This is a rational decision” to be correct only if the entity to which rationality is attributed possesses a degree of rationality that is very close to or at the end of the scale. (Staffel 2019: 164)

Very little has been written about the topic; for example, Christopher Cherniak in his (1990) book distinguishes minimal from ideal rationality; what is there between the two is hardly mentioned.<sup>120</sup> I shall opt for five stages.

Here, I agree with Staffel, and disagree with “absolutists” like Robert Weston Siscoe, who in his (2021) paper, “Being Rational Enough” claims that “the only way to be rational enough is to be maximally rational”. So, start with the stage of very miserly reasoning, relying on easy heuristics and unjustifiable biases. Call such a stage Stage One, stage of strongly crippled rationality. (I am borrowing the term “crippled” in this context from Clyde Hardin, see his (2002) and (2009)). We mentioned the Wason selection task; the typical erroneous answer to it exemplifies the Stage One. Other typical mistakes are noted by Kahneman (2011): The subject “responds more strongly to losses than to gains (loss aversion)”, frames decision problems narrowly, in isolation from one another and sometimes substitutes an easier question for a difficult one (heuristics). A typical bias, so called “Myside” bias occurs when people evaluate evidence, generate evidence, and test hypotheses in a manner biased toward their own prior opinions and attitude. The subject characterized by fully crippled rationality never reflects on his ‘conclusions’. If asked for reasons he/she ignores the question.

Many authors, such as Gigerenzer (2008), (2021), Sperber et. al. (1995) and Brogaard, (2016) insist that actual human reasoning is not so strongly crippled. They offer explanations-excuses for illogicalities detected by

---

<sup>120</sup> Here we have two statements that could be useful in developing the idea. First, by Allan Hazlett

Both excellence and individual excellences come in degrees. Given the present assumption, this entails that both virtue and individual virtues come in degrees, which seems right—people are more or less virtuous, more or less courageous, more or less open-minded, and so on. This, as yet, says nothing about the threshold for the attribution of virtue and for the attribution of individual virtues. (2019:230, fn. 10)

And second, by Robert Audi.

*Rigor.* Clear-headedness is connected with rigor: the latter virtue apparently requires the former one, even if the converse need not hold. This is a good point at which to note that virtues may be possessed to differing degrees. Moreover, the highest degree of some virtues might require at least the minimal degree of another virtue. A signally high degree of (intellectual) rigor would require at least minimal clear-headedness. But rigor (as a virtue) implies skill in reasoning, and it may be that someone could be quite clear-headed but lack overall strength in reasoning. Clear-headedness might require, however, abstention from certain kinds of reasoning, as required by a sense of one’s limitations. Rigor might imply some kind of awareness of a similar limitation. As this suggests, those with the virtue of rigor must know how to reason in certain basic ways, will tend to avoid kinds of reasoning beyond their ability, and must know the difference between good and bad reasoning for a significant range of cases. (2019:354).

the pioneers of cognitive psychological rationality research. They claim that the subjects studied reinterpret the question asked; for instance, when asked about the probability that Linda is bank teller they read it as asking about the probability of Linda being bank-teller and *nothing more* (not being a feminist). Other subject interpret conditionals as bi-conditionals, and so on. If we take into account the decades of the debate I think the following summary of it can be accepted: some drastic cases of irrationality might be explained-excused in the ways proposed, but probably not all. And explanation itself points to misunderstandings that are not themselves impressively rational. Miserly processing is the reality of human reasoning.

We could take as the next step a subject who makes similar mistakes, but sometimes reflects about them; the catch is that reflection does not lead to change. Maggie Toplak (2018) puts it in terms of “overriding”: does the subject have enough knowledge or mental capacities (“mindware”) to carry an override of initial miserly response. If the subject does not, then he-she ends up acquiescing in the response. For example, if the subject thinks about his-her answer to Wason selection task, the B and 2 option, she ends up thinking “yes, this is obviously the correct response”. No ability to override.

This leads to a picture of partly crippled rationality. Let me call it Stage Two, characterized by rationality crippled to some extent, quite imperfect. Like Smokrović, I shall borrow my examples from Arthur Conan Doyle, and take his character Dr. Watson as exemplary subject. Smokrović also admits that the unsuccessful Watson is not rational in the complimentary sense (1997: 14). Our Stage Two subject stays, after a short reflection with all the answers that characterize Stage One, responding more strongly to losses than to gains (loss aversion), framing decision problems narrowly, in isolation from one another, sometimes substituting an easier question for a difficult one (heuristics), stays with anchoring, with overweighing low probabilities, with outcome bias and myside bias. Link to epistemic vice is omnipresent.

What about logical rules? Smokrović briefly mentions a character, whom I shall call Jane, who, in the example, wonders about the platform from which the 8 am. bus from Rijeka to Zagreb starts. She believes that the bus starts either from platform 1 or from platform 2. Believing that the bus does not start from the platform 2, she “infers to the conclusion”: therefore, it starts from platform 1. (2018: 457). Smokrović relies on John MacFarlane’s “bridge principle” that connects a valid logical form with the corresponding norm of reasoning. In the above mentioned example, disjunction of elimination should be connected to the norm that requires Jane to pass from initial beliefs to the conclusion that the train will start from the Platform 1. (218: 460).

I would say that thinkers with partly crippled rationality (maybe even those with completely crippled one) would be capable of performing such simple logical inferences; unfortunately, they will do it only within a framework of extremely unreliable cognitive tendencies. One possible hypothesis would be that completely crippled rationality goes with minimal mastery of only some logical constants, where is it the task of cognitive research to find out which individual can use which constants (see Cherniak's (1990) book). The partially crippled rationality would go with the use of all basic constants (see S. Evnine (2008), chapter "Personhood and Logical Ability"). I find the hypothesis worthy of investigation.

Back to the tendencies of miserly reasoning. Of course, there are debates about all of them. Take overweighing low probabilities. This can be useful in particular circumstances, e.g. when the danger of improbable outcome is very great, and is in such circumstances "ecologically rational", as Gigerenzer and his followers would put it. It does not follow that it is rational as a general strategy, of the kind that is encountered across very different subjects and circumstances, as described by Stanovich and his collaborators.

Back to the stages of rationality. What would the next stage look like? It would be positive, with some logical reasoning and with successful avoidance of the worse traps of miserly reasoning.

Think of people whom you find impressively intelligent and reasonable, on various challenging subjects. They would be living examples of this first higher degree of positive rationality, our Stage Three. No wonder that Smokrović takes the successful reasoning of Sherlock Holmes as his paradigm of rationality, (1997:14): Holmes is impressive because people are rarely as capable of rational reasoning as he is.

The wound upon the dead man was, as I was able to determine with absolute confidence, fired from a revolver at the distance of something over four yards. There was no powder-blackening on the clothes. Evidently, therefore, Alec Cunningham had lied when he said that the two men were struggling when the shot was fired. Again, both father and son agreed as to the place where the man escaped into the road. At that point, however, as it happens, there is a broadish ditch, moist at the bottom. As there were no indications of boot-marks about this ditch, I was absolutely sure not only that the Cunningham had again lied, but that there had never been any unknown man upon the scene at all.

Smokrović invites us to imagine a conversation:

Let us imagine a dialogue between Sherlock Holmes and Dr. Watson in which Dr. Watson reasons in an erroneous way about the murder

discussed, and Holmes is correcting him. Here is how our imagined dialogue might proceed:

Holmes: "Tell me, doctor, why do you think that the bullet was fired from distance less than four yards?"

Dr. Watson: "Well, I looked at the murdered person and I noticed that the bullet has entered his body under an unusual angle. Therefore, I concluded that the murderer was shooting from very close, certainly from distance less than four yards"

Holmes: "Dear Doctor, I have to tell you that your inference is, to say the least, incautious.

First, you started from irrelevant facts, and then from them derived the conclusion that does not follow. I have started from the following assumption: if the bullet was shot from distance of four yards or less, there will be traces of gunpowder on the clothes. Given that there are no such traces. I have correctly inferred the bullet was fired from a distance greater than four yards."

Leaving aside the question of why had Dr. Watson made such grave mistake (let us imagine, in our invented scenario, that the good doctor was just too tired), we can note that in this dialogue Holmes is evaluating Watson's and his own reasoning, in a way in which any grown up, normal participant in a conversation could to. (2014: 39)

Watson is not quite rational, Sherlock Holmes is. So, note that Stage Three is the first one that corresponds to the ordinary notion of "being rational".

We can proceed to the next stage, Stage Four, of Full rationality. A stage Four thinker is never a victim of miserly processing, her conclusion all hold water, and she can go quite far in rational reasoning. If you need an illustration, think of top logicians, people like Kurt Gödel. Smokrović links correct reasoning to epistemic virtue (in his "*Može li se naučiti kritički misliti?*" paper presented on 12. 11. 2020. (*Konferencija: John Locke i suvremenost. Institut za filozofiju*), arguing that acquiring critical thinking requires virtue:

#### ***Can critical thinking be learned?***

The goal of this paper is, in addition to the analysis of the concept of "critical thinking" its contemporary usage, to analyze the possibility of acquiring knowledge or learning, of the ability to think critically. There is no doubt that critical thinking can be improved in some of its segments (similarly to general intelligence). The question is whether it can be learned (as one can learn a foreign language). It is unrealistic to expect that thinking might

become “critical” through acquiring of formal or informal logical rules. If there is at all a possibility of learning in the domain of critical thinking, then this can be done through acquisition of epistemic virtues.

Here, I would agree with Smokrović, that we have to count on a plurality of logics. Almost a century ago Carnap wrote:

In logic there are no morals. Everyone is at liberty to build his own logic, i.e. his own language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (Carnap 1937, §17)

As Stewart Shapiro puts it more recently “there are different, mutually incompatible articulations or sharpenings of the intuitive notion or notions of logical consequence. At least some of the supposedly rival explications of logical consequence are (more or less) accurate theories of different aspects of consequence, different members of the cluster. (2014: 25).

The final step is to the Stage Five, with idealization of logical omniscience, often assumed in contemporary views of rationality in regard to logic and computation. For example,, Michael G. Titelbaum talks about “logical omniscience requirements in formal theories of rationality.” He introduces them in the following way

The best-known such requirement comes from Bayesian epistemology, which takes Kolmogorov’s probability axioms to represent rational requirements on agents’ degrees of belief. One of those axioms (usually called Normality) assigns a value of 1 to every logical truth. In Bayesian epistemology this entails something like a rational requirement that agents assign certainty to all logical truths. (2015: 254)

He also discussed limitation needed, for instance:

To respond to the Cognitive Reach concern, we can restrict logical omniscience so that it requires certainty only in logical truths that are sufficiently obvious or accessible to the agent. (2015: 225).

So, here are, then, five degrees or stages of rationality of reasoning:

- 1 strongly crippled - link to epistemic vice
- 2 crippled to some extent, imperfect,
- 3 minimally positive
- 4 full (reflexive) - epistemically virtuous
- 5 idealized

Where do we put ordinary reasoning? I think the majority of people reason in the imperfect way that belongs to Stage Two. Here is a sketch of my argument for the claim that people in general do not reason rationally in the

positive, complimentary sense of the term, and are not virtuous reasoners.

The first premise. I claim that most people are like Doctor Watson, or worse, and reason on the level of Second Stage, in a manner characterized by imperfection, even by rationality being crippled to some extent. It embodies a lot of miserly strategies that the subject cannot control. Minimally positive, non-crippled reasoning is a privilege of a minority.

The second premise, that Second Stage reasoning is not rational in the ordinary sense of the word. Too many miserly, unacceptable strategies govern such a reasoning. Even if we accept the most prominent attempts at explanation and excuse, most of them do not allow us to ascribe to subjects rationality in the ordinary sense. Therefore, we conclude the following:

People in general do not reason rationally in the positive, ordinary sense of the word, their “rationality” is rather crippled and not virtuous at all (the fashionable term these days is H. Simon’s term “bounded rationality” (see Viale ed. 2021)).

Finally, is someone who is barely rational, in a somewhat crippled way thereby epistemically virtuous? Hardly. I would say that the epistemic quality of such a person’s mind is, at best, only a disposition for virtue.

The situation is similar with arithmetical capacities of ordinary, uneducated people. Is John epistemically virtuous as regards math? Hardly. He can become such, with some years of training, that’s all. Or take someone who walks normally, never runs and never exercises, is healthy and could become a sportsperson. Such a person does not have “sporting virtue” merely a disposition for it. I would claim that ordinary human being is in the same situation as far as the virtue of rationality is in question.

Following the line of least resistance, or epistemic miserliness, is constitutive for human ordinary thinking and reasoning. Virtue epistemology of reasoning should take into account this fundamental flaw. Note that requirements for virtue are relatively high: it is not just a mechanical adaptation to context; it requires effort and training, as well as polishing of capacities. Like in sport! Intellectual-cognitive gyms are needed!

#### 4 Questions for Smokrović

We are now ready to raise questions for our author. The main question concerns the evaluation of human ordinary reasoning.

Since Smokrović talks about development of logical capacities, I asked him whether he is ascribing his weakened picture to children instead of adults. He (kindly) answered in the negative: his proposal concerns rationality of grown-ups. Most importantly, Smokrović thinks that this weakened conception is enough for a moderate rationalist judgment: people do reason rationally in the sense sketched.



I think that this move is too undemanding. Suppose it holds, and suppose that together with such cases many miserly moves cognitive scientists see as all pervasive happen all the time; our Jane is prone to very miserly reasoning. Then we have minimal pieces of correct reasoning in the sea of non-logical heuristics and biases. It is a clear case of our Stage Two scenario. Why call such a reasoning rational, in the complimentary sense of the term? Smokrović himself proposes not to count Watson as rational; but *Watson is, in the novel and in general, a prototype of ordinary capacities; Holmes is an exception.*

At the end of the previous section I have claimed that most people are like Doctor Watson, or worse, and reason on the level of Second Stage, in a manner characterized by imperfection, even by rationality being crippled to some extent. It embodies a lot of miserly strategies that the subject cannot control. Minimally positive, non-crippled reasoning is a privilege of a minority.

Smokrović's favorite idea about local, unsystematic cases of logical micro-reasoning, like in the example with the bus platform, seems to me too undemanding. Suppose it holds, and suppose the miserly moves cognitive scientists see as all pervasive holds. Then we have minimal pieces of correct reasoning in the sea of non-logical heuristics and biases. It is a clear case of our Stage Two scenario. Why call such a reasoning rational, in the complimentary sense of the term? Smokrović himself proposes not to count Watson as rational; but *Watson is, in the novel and in general, a prototype of ordinary capacities; Holmes is an exception.*

This has led me to my second premise, that Second Stage reasoning is not rational in the ordinary, not crippled sense of the word. Too many miserly, normatively unacceptable strategies govern such a reasoning. Even if we accept the most prominent attempts at explanation and excuse, most of them do not allow us to ascribe to subjects rationality in the ordinary sense. Therefore, we concluded that people in general do not reason rationally.

The first line my (or our) opponents might take is to question the first premise: What if people are not as bad in reasoning as cognitive psychologists assume they are? Note first that this line is not opened for Smokrović; in his writings he is wisely on the side of science. The whole first part of his main (2004) book is a sympathetic exposition of the main negative examples, with Linda and Wason's selection task as main examples, discussed in detail.

### ***Discussion***

Smokrović can problematize my judgment, and some colleagues and friends, above all Edi Pavlović have done this in discussions (thanks go to them); I want to answer briefly.

My critic could re-iterate the standard judgment of defenders of bounded rationality as genuine kind of rationality, by claiming that what the subject in question, our Doctor Watson or Jane, lacks, are just resources, mostly informational, and not the reasoning capacities. So, why would our Watson commit the mistake in Selection Task? Not because he is limited in his rational capacities, but because he solves the problem under pressure, and cannot take into account all the theoretical possibilities that would be available to a subject not pressed by time. Under pressure, he searches for the most relevant options, so, first, he offers a re-reading of the task that narrows it down to alternatives relevant to him at the moment, second, searches for a reasonable answer, and finally, finds the answer. There is nothing non-rational about the strategy Watson deployed.

Here is my answer. First, bear in mind that Watson's answer is clearly wrong, in terms of the explicitly given task. Second, Watson is clearly not reflective about his effort; he is not even conscious of the work he has done, but only of its result. Now, if this is the case, how reasonable is his reasoning? It has ended up with the wrong answer, keep this in mind. Watson has no idea that he has unconsciously changed the shape of the task, nor would he be able to say why; we have to wait for ingenious intellectuals like Sperber, to tell us why the task has been surreptitiously changed. And like a pupil that systematically misrepresent his teacher's questions, Watson is doing this all the time. Is this merely a lack of informational resources, or is it rather a deeply misleading strategy of answering?

I think it is such a strategy. So, even if we re-interpret Watson's answer charitably, in the spirit of, say, ecological rationalists, the fact remains that it is a hasty answer to a misinterpreted question. No amount of charity can nullify the crippled component of the rationality involved.

Let me translate the debate into the vocabulary of overriding the suggestions derived from miserly processing, vocabulary due to Maggie Toplak, Keith Stanovich and their collaborators (2016). So, assume that Watson's cognitive system has delivered to him the wrong answer, produced in a miserly fashion. Watson has no means to override the delivered answer. What is he lacking? Is it just the information about the world, or are these skills of understanding? I don't see how mere lack of information about the world would be relevant in this case; it is rather the lack of reasoning skill that prevents overriding. And this makes the process somewhat crippled. Of course, besides being unavailable to Smokrović, this line is unjustifiably anti-naturalist.

The other line our opponents might take is to question the second premise. Authors like Gigerenzer argue that the limited, "ecological" rationality is an acceptable form of rationality. However, when we look at examples, it seems that they do not fit the general, commonsensical notion

of rationality. Patterns of thought that are valid only in very specific “ecological” circumstances are not rational in the ordinary sense of the term. Gigerenzer’s acceptance of quite heavy biases also speaks against counting this very limited form as rational in a positive sense. In short, such “ecological rationality” is not an epistemic virtue, which rationality is ordinarily taken to be.

## 5 Conclusion

Our guiding question has been: do people reason rationally? It divides into two: First, is their thinking rational in ordinary, positive, not crippled way? Second, is their thinking epistemically virtuous? To my mind the questions go together, and their answers should be either jointly positive - people think and infer in ordinary, positive, not crippled way, or jointly negative, their reasoning is not epistemically virtuous, but rather vicious.

In order to answer our question more systematically, we have distinguished five degrees or stages of rationality of reasoning. First, the strongly crippled one with a clear link to epistemic vice, corresponding to what is in ordinary language described as irrationality. Second, crippled to some extent, still shockingly imperfect; this stage is characterized by most of the mistakes associated with miserly processing and the use of heuristics and biases. It is the classical bad reasoning where crippled rationality impedes inquiry and prevents the thinker to reach truth. The third stage is the minimally positive one, the rationality in everyday life, far from perfect, but already rationality, It exhibits the predominance of justifiable methods, some valid universally, some in narrow circumstances (ecological, as cognitive scientists would call it). Also, the stage is the beginnings of correct reflective awareness of what one is doing when inferring. Stage four is the one of full rationality; it is normatively positive and fully reflexive, exemplifying the full epistemic virtue. The fifth stage is the one of idealized, fully rational individual endowed with extreme logical and computational powers.

So, at the top we have both ideal and real fully rational individuals. And at the bottom epistemically more or less crippled ones. There is a realistic option in between, our stage three.

What do we get from the structured picture of five stages? First, we avoid doubts and criticisms that restriction on too high options, or too generous extension of calling completely crippled reasoning “rational” normally encounter. The variety of options also helps with the unity, of cognitive science approaches, commonsense, and armchair philosophy. And also with unity across areas; I tried it in analyzing the (ir)rationality of nationalism and populism (to appear). So it is, hopefully, relevant for epistemology of rationality.

Back to my dialogue with Smokrović! Let me agree with that people are minimally rational if at all: people in general do not reason rationally in any complimentary sense. The next question is whether such rationality is an epistemic virtue; I would say, that it is not, or at, best that it is quite minimally virtuous. Smokrović sometimes sounds more optimistic, but I think we basically agree. Now, is someone who is barely rational, or rational in a somewhat crippled way thereby epistemically virtuous? Hardly. I would say that the epistemic quality of such a person's mind is, at best, only a disposition for virtue. Similarly with arithmetical capacities of ordinary uneducated people. Is John epistemically virtuous as regards math? Hardly. He can become such, with some years of training, that's all. Similarly with arithmetical abilities of uneducated people. Or take someone who walk normally, never runs and never exercises, is healthy and could become a sportsperson. Such a person does not have "sporting virtue", but merely a disposition for it. I would claim that ordinary human being is in the same situation as far as the virtue of rationality goes.

Let me say a few words about the road ahead that I find most interesting in relation to the present topic. I have already briefly alluded to the possibility of taking positive rationality as an epistemic virtue and crippled rationality (or simply irrationality) as an epistemic vice. Aristotle has anticipated this kind of approach in the book Z of his *Nicomachean Ethics*, where he talks about intellectual virtues, and characterizes each of them as being "meta logou" (e.g. 1140a), "involving reason" as the standard translation has it.

The present day commentators talk freely in these context of virtue considerations of rationality; let me mention as a typical example Gavin Lawrence's chapter from Richard Kraut's (2006) collection on *Nicomachean Ethics*; similar stance is taken by several authors in Ronald Polansky (ed.) (2014) *The Cambridge companion to Aristotle's Nicomachean Ethics*.

An Aristotelian supporting the view that virtue is normally the middle between two bad extremes, might claim that the same holds for virtue in reasoning. Here, she would say, we have positive reasoning capacities, as virtues-abilities, and on the side of character virtues inquisitiveness-curiosity and epistemic vigilance. On the side of vices, at one extreme one would have intellectual incapacities and miserliness in reasoning as intellectual vices, and sloth as a character vice. On the opposite extreme one would have intellectual hastiness and adventurism; on the ability side, it would be hard to find the right contrasting vice.

The line has not been clearly continued in the present-day virtue epistemology, except for a few rare occasions, like, for instance, Abrol Fairweather's and Carlos Montemayor's (2014) and in Ralph Wedgwood (2017). I would suggest that the issues of rationality should be clearly situated within mainstream epistemology, and my preferred approach would be the virtue-epistemology.

Given the present-day division usual in virtue epistemology, of epistemic qualities into epistemic virtues-abilities and virtues-character traits, and their vicious counterparts, one can divide epistemology of rationality into four parts: first, the study of rationality virtues, second of character traits, third rational disabilities-vices and irrational character traits-vices. To illustrate, the calculating competencies prominent in decision-theoretical debates on rationality would be a good example of epistemic virtues-abilities. The problematic features of our System-1 production of heuristics would belong to the group of vices-disabilities. Again, some positive features manifested in choice of goals would belong to the virtues-character traits, whereas the negative features, like some biases, would be manifestation of negative character-traits, epistemic vices. (For subtle divisions of epistemic virtues see the chapter "Rationality as a Virtue" from the (2017) book by Ralph Wedgwood).

So, Smokrović defends the thesis that human beings think and infer in an ordinary, positive, not crippled way and are to a significant extent rational, being, in my terms, epistemically virtuous (this is not the terminology he employs). I would argue in the opposite direction that epistemic miserliness and the line of the minimal resistance is constitutive for human ordinary reasoning and inference. Virtue epistemology, in particular the epistemology of vices, should deal with these fundamental flaws. And all this puts some high demand on human virtue of rationality: it is not a mechanical adaptation to context, but demands effort and learning, strengthening of motivation and honing of abilities. Like it is in sports! We urgently need to have more cognitive gyms!

So much for the present occasion; and thanks go to Nenad, Majda Trobok and David Grčki!

## Bibliography - Nenad Smokrović

- Nenad Smokrović (1997) Ali ljudi je sklepa racionalno?, *Analiza* – 3 pp. 13 - 29.
- Nenad Smokrović (2002) Semantic Naturalism and Normativity, *Croatian journal of philosophy*.
- Nenad Smokrović (2004) *Priroda prirodnog zaključivanja*, Biblioteka Filozofskih istraživanja, Zagreb.
- Nenad Smokrović (2012a), Logical Consequence and Rationality in M. Trobok et al. (eds.), *Between Logic and Reality, Logic, Epistemology, and the Unity of Science*, Springer.
- Nenad Smokrović (2012b) Logičko znanje i apriorno opravdanje,” in Prijić Samaržija Snježana; Bojanić Petar (eds) *Nenad Mišćević - sva lica filozofije*, Univerzitet u Beogradu, Institut za filozofiju i društvenu teoriju, Beograd, str. 159-172.
- Nenad Smokrović (2017) *Znanost i metoda: od opće metodologije do metodologije ekonomije*, KruZak, Zagreb.
- Nenad Smokrović (2017), “Curiosity and Argumentative Process”, in Bojan Borstner and Smiljana Gartner (eds.) *Thought Experiments between Nature and Society: A Festschrift for Nenad Mišćević*, Cambridge Scholars Publishing, 83-96.
- Nenad Smokrović (2018), Informal reasoning and formal logic: normativity of natural language reasoning, *Croatian Journal of Philosophy* Vol. XVIII, No. 54, 455-469.

## References

- Aristotle *Nicomachean Ethics*, translated and edited by Roger Crisp, Cambridge University press 2004.
- Robert Audi (2019), “Intellectual Virtue, Knowledge, and Justification”, in Heather Batataly (ed.), *Routledge handbook of virtue epistemology*, Routledge, 352-362,
- Berit Brogaard (2016), “Linda the Bank Teller’ Case Revisited Our useful reliance on contextual and non-literal information.”Posted Nov 22, at <https://www.psychology-today.com/intl/blog/the-superhuman-mind/201611/linda-the-bank-teller-case-revisited>
- Christopher Cherniak (1990) *Minimal Rationality*, The MIT Press Bradford Books.
- Dennis Chong (2013), “Degrees of rationality in politics” in Leonie Huddy, David O. Sears, Jack S. Levy (eds.) *The Oxford Handbook of Political Psychology*, Oxford University Press. 96-129.
- Simon J. Evnine (2008), *Epistemic dimensions of personhood*, Oxford University Press.
- Abrol Fairweather and Carlos Montemayor (2014) “Inferential Abilities and Common Epistemic Goods” in Abrol Fairweather (ed) (2014) *Virtue Epistemology Naturalized, Bridges Between Virtue Epistemology and Philosophy of Science*, Springer.
- Gerd Gigerenzer (2008) “Moral Intuition = Fast and Frugal Heuristics?” in Walter Sinnott-Armstrong (ed.) *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, MIT, 1-26.
- Gerd Gigerenzer (2021), “What is bounded rationality” in R. Viale (ed) *Routledge Handbook of Bounded Rationality* Routledge, 55-69.

- Dan Sperber, Francesco Cara, Vittorio Girotto (1995) Relevance theory explains the selection task, *Cognition* 57, 31-95.
- Clyde Hardin (2002) The Crippled Epistemology of Extremism. In Albert Breton, Gianluigi Galeotti, Pierre Salmon, Ronald Wintrobe Eds., *Political Extremism and Rationality*, ed. 3–22. Cambridge: Cambridge University Press.
- Clyde Hardin, (2009) "Extremism", ch. 9, in *How Do You Know? the economics of ordinary knowledge*, Princeton University Press.
- Allan Hazlett (2019) "Skepticism", in Heather Battaly (ed.), *Routledge handbook of virtue epistemology*, Routledge, 221-231.
- Daniel Kahneman: 2011. *Thinking, Fast and Slow*, Farrar Straus Giroux, New York.
- Gavin Lawrence 2006 Human Good and Human Function, in Richard Kraut (ed.) *The Blackwell guide to Aristotle's Nicomachean ethics*, Blackwell, 37 – 75.
- Stephen Nathanson (1999) *Ideal of Rationality: A Defense, Within Reason*, Open Court.
- Nisbett, Richard E. (2015), *Mindware: tools for smart thinking*, Farrar, Straus and Giroux.
- Robert Nozick (1993) *The nature of rationality*, Princeton University Press.
- Robert Weston Siscoe (2021), "Being Rational Enough: Maximizing, Satisficing, and Degrees of Rationality", *Australasian Journal of Philosophy*
- Julia Staffel (2019) *Unsettled thoughts*, Oxford University Press.
- Stewart Shapiro, (2014), *Varieties of logic*, Oxford University Press.
- Stalnaker, Robert (1999) *Context and Content*, Oxford: Oxford University Press.
- Keith E. Stanovich, Richard F. West, and Maggie E. Toplak 2013 "Myside Bias, Rational Thinking, and Intelligence", *Current Directions in Psychological Science*, 22(4) 259–264 .
- Julia Staffel *Unsettled thoughts*, Oxford University Press, 2019.
- Stanovich, Keith E., Toplak, Maggie E., West, Richard, 2016 *The rationality quotient: toward a test of rational thinking*, MIT Press.
- Michael G. Titelbaum (2015) Rationality's Fixed Point (or: In Defense of Right Reason) in Tamar Szabó Gendler and John Hawthorne eds. *Oxford studies in epistemology* Volume 5, Oxford University Press. 253-294.
- Toplak, Maggie E., 2018 The development of rational thinking: insights from the heuristics and biases literature and dual process models IN Linden J. Ball and Valerie A. Thompson (eds.), *The Routledge international handbook of thinking and reasoning*, 542-558 .
- Riccardo Viale (ed) *Routledge Handbook of Bounded Rationality* Routledge.
- Ralph Wedgwood (2017) *The Value of Rationality*, Oxford University Press.





INES SKELAC

## Uloga logike u ljudskom zaključivanju

**Sažetak:** Rasprava o normativnosti logike za svakodnevno zaključivanje, odnosno o odnosu deduktivnoga zaključivanja i svakodnevnoga zaključivanja postala je značajnom krajem devetnaestoga i početkom dvadesetoga stoljeća i vezuje se za pojavu psihologizma u logici i filozofiji. Između zaključivanja u logici i svakodnevnog zaključivanja postoje brojne razlike, ali riječ je o dvama procesima koji se mogu povezati i uspoređivati.

U najznačajnije suvremene zastupnike ideje o normativnosti logike za zaključivanje ubrajaju se Hartry Field i John MacFarlane, a protivnik je takve ideje Gilbert Harman. Od hrvatskih autora značajan je doprinos ovoj temi dao Nenad Smokrović koji također zastupa tezu o normativnosti logike, ali u vidu logičkoga pluralizma. U radu se analiziraju različiti pristupi normativnosti te navode prigovori koje je moguće uputiti, te se ukazuje na probleme koje je potrebno riješiti. Posebno se propituje tzv. načelo premošćivanja i pluralistički pristup normativnosti te se ukazuje na probleme koje je pri takvom pristupu potrebno riješiti.

**Ključne riječi:** deduktivno zaključivanje, logički pluralizam, ljudsko zaključivanje, načelo premošćivanja, normativnost logike.

### 1. Uvod

Temom odnosa ljudskog zaključivanja i deduktivnog zaključivanja u logici bave se mnogi logičari, filozofi logike i psiholozi. Brojna istraživanja ljudskoga zaključivanja (počevši od Wasona (1966) pa do danas) pokazuju kako ljudi uvelike griješe u zaključivanju ako se kao standard zaključivanja uzimaju logička pravila. Tu se javlja pitanje normativnosti, odnosno je li logika na neki način normativna za ljudsko zaključivanje.

U raspravi o normativnosti često se znaju pojaviti i argumenti koji govore da ljudi sve što rade, rade po nekim pravilima. To uglavnom i jest točno: ljudi žive u zajednici koja ima određena pravila, ponašaju se na određeni način, igraju igre na određeni način (po odgovarajućim pravilima), koriste jezik na određeni način. Ipak, postoji jedna značajna razlika. Kada se govori o slijeđenju pravila u bilo kojem aspektu života, osim u zaključivanju, ljudi u nekom trenutku mogu odlučiti da to pravilo više neće koristiti. Primjerice, žive u zajednici prema određenim pravilima, ali u svakom trenutku mogu odlučiti da više neće pripadati toj zajednici. S druge strane, ako se pretpostavi da je logika normativna za zaključivanje, onda je

ona normativna uvijek te se ta igra ne može napustiti. Stoga je rasprava o normativnosti vrlo značajna, a oba suprotstavljena stajališta, i ono prema kojem je logika normativna za ljudsko zaključivanje, i ono prema kojem nije, imaju vrlo dalekosežne implikacije.

Ako se pretpostavlja da postoji standard prema kojem ljudi zaključuju, često se kao standard pretpostavlja klasična logika, iako neki autori navode kako bi druge logike mogle biti povoljniji izbor ili da je čak moguće da se ne radi o jednoj, nego o više logika (Field 2009, Smokrović 2018). S druge strane, čini se intuitivnim da ljudi zaključuju konzistentno i vodeći se nekim pravilima jer kad ne bi bilo tako, teško da bi se ikad oko ičega mogli usuglasiti.

S obzirom na navedeno, rasprava o normativnosti, ovisno o tome je li zauzet stav o normativnosti ili stav protiv normativnosti, uključuje sljedeća temeljna pitanja: (1) Ako je logika normativna, na koji način jedan formalni deduktivni sustav može činiti standard ljudskoga zaključivanja?; (2) Ako logika nije normativna, čime je određeno ljudsko zaključivanje, odnosno kako ljudi zaključuju?

U ovom će se radu ukratko predstaviti različiti pristupi pitanju normativnosti u logici te će se istaknuti određene kritike koje se mogu uputiti normativistima. Naglasak je na iznošenju kritika tomu pristupu jer je na normativistima teret dokaza da je logika normativna. Nadalje, analizirat će se gledište koje sugerira Smokrović (2018), a koje uključuje mogućnost pluralizma u raspravi o normativnosti.

## 2. Ljudsko zaključivanje i deduktivno zaključivanje<sup>121</sup>

S obzirom na to da je u ovome radu fokus na analizi normativnosti logike, kao formalnoga sustava za ljudsko zaključivanje, prije samoga prelaska na temu potrebno je osvrnuti se na odnos ljudskoga zaključivanja i deduktivnoga zaključivanja u logici. Kada se u logici govori o deduktivnom zaključivanju, radi se o obliku zaključivanja pri kojem se na temelju danoga skupa premisa koji se smatra istinitim izvodi istinita konkluzija te konkluzija sadrži samo ono što je već sadržano u premisama.

Prva je razlika u odnosu na ljudsko zaključivanje u samim pravilima koja se koriste pri izvođenju takve konkluzije. U zaključivanju u logici koriste se pravila određenoga logičkog sustava u sklopu kojega se izvodi zaključak, dok je u ljudskome zaključivanju teško jasno odrediti o kakvim se pravilima radi. O načinima na koje ljudi zaključuju postoji velika povijest

---

<sup>121</sup> Dio ovoga rada dijelom se temelji na tekstu neobjavljenoga dijela doktorske disertacije pod nazivom *Logika i ljudsko zaključivanje*, koja je napisana pod mentorstvom prof. dr. sc. Nenada Smokrovića i obranjena na Filozofskom fakultetu Sveučilišta u Rijeci 2018. godine.

istraživanja, ponajprije u psihologiji zaključivanja, a općeniti je zaključak koji proizlazi iz mnogobrojnih empirijskih istraživanja da ljudi kod zaključivanja ne koriste skup logičkih pravila koji bi u potpunosti odgovarao bilo kojem logičkom sustavu<sup>122</sup>.

Druga se razlika sastoji u jeziku koji se koristi za izvođenje zaključka. Naime, u slučaju logike riječ je o formalnome, jednoznačnom jeziku koji pojedina logička teorija koristi, dok se ljudsko zaključivanje odvija posredstvom prirodnoga jezika, čiji su izrazi često i višeznačni te se prirodni jezici uvelike razlikuju po svojim sintaktičkim svojstvima. Za deduktivno je zaključivanje u logici ključno i to da se ono smije vršiti samo na temelju informacija sadržanih u premisama, dok se kod ljudskoga zaključivanja često uključuju pozadinska znanja i prešutne premise.

Oaksford i Chater (2009) ističu kako u svakodnevnome zaključivanju gotovo da i nema zaključaka koji se mogu smatrati apsolutno istinitima. Sve su istine do kojih možemo doći kontingentne, pa čak i one koje nam se čine trivijalnim, na primjer *Sva su živa bića smrtna*. Logički se zaključci temelje prvenstveno na svojoj formi, a ne ovise o sadržaju. Slijedom toga, logika govori o tome koje forme zaključaka čuvaju istinitost. Ako su premise istinite, samim je time istinita i konkluzija jer je ona već sadržana u premisama.

Konačno, dok za deduktivno zaključivanje u logici vrijedi monotoničnost, suprotno tomu, u svakodnevnom zaključivanju svojstvo monotoničnosti često nije očuvano te uvođenje dodatne premise može utjecati na mogućnost izvođenja konkluzije.

S obzirom na navedeno, vidljivo je da između zaključivanja u logici i svakodnevnog zaključivanja postoje brojne razlike. Ipak, riječ je o dva procesa koji se očito mogu povezati i uspoređivati, stoga će se u nastavku rada raspravljati o mogućnosti da je logika normativna za ljudsko zaključivanje. Međutim, navedene razlike treba stalno imati u vidu.

### 3. Kratak pregled dosadašnjih istraživanja

Kada se govori o normativnosti, Steinberger (2016) ističe barem tri moguća tumačenja pojma normi.

1. Norme kao smjernice koje vode prema odabiru djelovanja ili vjerovanja.
2. Norme kao vrednovanja u smislu postavljanja standarda prema kojima se pojedinačna djelovanja ili vjerovanja procjenjuju kao dobra ili loša, ispravna ili neispravna.

<sup>122</sup> Za rasprave o ovoj temi vidjeti radove L. J. Ripsa, Ph. Johnson-Lairda, J. St. B. T. Evansa, M. D. S. Brainea, C. Reverberija, A. Tverskyja, D. Kahnemana, i drugih.

3. Norme kao procjene koje se nalaze u osnovi pripisivanja pohvale ili kritike sebi ili drugima.

Za normativnost logike mogu biti značajna prva dva shvaćanja normativnosti, dok je treće više vezano za etiku. Prvo je pravilo normativno, dok je drugo deskriptivno. Za razliku od normativne teorije zaključivanja koja može ili govoriti o tome kako ljudi trebaju zaključivati ili može procjenjivati njihovo zaključivanje s obzirom na određeni standard, deskriptivna teorija govori o tome kako zapravo zaključuju te takvo zaključivanje usporuđuje s određenim standardom. Ipak, nije lako razlikovati deskriptivnu i normativnu teoriju zaključivanja jer normativna teorija mora početi od toga kako ljudi zapravo zaključuju kako bi pokazala koliko to odstupa od standarda, dok svaka deskriptivna teorija mora koristiti idealizaciju.

Iako se stavovi o normativnosti logike za zaključivanje mogu iščitati i iz radova Aristotela, stoika i Descartesa, ta je rasprava postala značajnom krajem devetnaestoga i početkom dvadesetoga stoljeća i vjerojatno je usko povezana s pojavom psihologizma u logici i filozofiji, pri čemu je dio filozofa i logičara inzistirao na tome da se logika udalji od psihologije. U najznačajnije suvremene zastupnike ideje o normativnosti logike za zaključivanje ubrajaju se Hartry Field (2009) i John MacFarlane (2004), a protivnik je takve ideje Gilbert Harman (1984, 1986, 2009), dok je značajna i Katarina Dutilh Novaes (2015), koja zastupa stajalište prema kojem logika može biti normativna za zaključivanje samo u društvenoj sferi. Od hrvatskih autora značajan je doprinos ovoj temi dao Nenad Smokrović (2017, 2018) koji, na tragu Fielda i MacFarlanea, također zastupa tezu o normativnosti logike.

Ako se tvrdi da je logika normativna za zaključivanje, potrebno je pokazati na koji način neka teorija ili pak neki logički zakoni mogu biti normativni za misaone procese koji se odvijaju u ljudskome umu. Zagovornici teze o normativnosti logike pretpostavljaju postojanje načelâ premošćivanja koja bi trebala riješiti taj problem. Smokrović (2018) detaljno razmatra MacFarlaneovu (2004) i Fieldovu (2009) varijantu načela premošćivanja te smatra da više od jednog formalizma može biti normativno prikladno za deduktivno zaključivanje, s obzirom na to da zaključivanje može pretpostaviti različite oblike koji su vođeni različitim ciljevima (str. 455).

### **3.1. Harman protiv normativnosti logike**

Gilbert Harman (1984; 1986; 2009) smatra da tradicionalno gledište prema kojem logika ima normativnu ulogu u zaključivanju počiva na *kategorijalnoj pogrešci* miješanja načela klasične logike s onim što on naziva teorijom zaključivanja, a klasična logika i normativna teorija zaključivanja suštinski se razlikuju po svom obliku. Dok teorija zaključivanja govori o tome kako

osobe koje zaključuju trebaju zaključivati i kako upravljati svojim vjerovanjima (ako je normativna) ili kako zaključuju i upravljaju svojim vjerovanjima (ako je deskriptivna), klasična se logika bavi odnosima među sudovima. Prema tome, Harman zaključuje da logička načela ne mogu izravno biti pravila prema kojima će se ispravljati vjerovanja jer se ne odnose na vjerovanja (Harman 1984: 107). Deduktivna pravila u logici vrijede bez iznimke, što ne može biti slučaj za teoriju zaključivanja nastalu na temelju praktičnih i epistemičkih razmatranja (*isto*, str. 109). Zbog same razlike u teoriji logike i teoriji zaključivanja logika se i ne može primijeniti na ljudsko zaključivanje u smislu bilo kakve norme jer apstraktna teorija ne može biti primjenjiva na procese koji se odvijaju u ljudskom umu.

U skladu s tim, Harman (1986) tvrdi da ne samo da logika sama po sebi nije teorija zaključivanja, nego i nema posebnu važnost za teorije zaključivanja jer logička načela nisu izravno pravila revidiranja vjerovanja ni pravila vjerovanja uopće. Na primjer, modus ponens ne govori da ako netko vjeruje  $p$  i ako  $p$ , onda  $q$ , može vjerovati  $q$ , niti postoje načela revidiranja vjerovanja koja izravno odgovaraju logičkim načelima kao što je modus ponens. Logička načela vrijede univerzalno, bez iznimki, dok odgovarajuća pravila revidiranja vjerovanja ne vrijede uvijek.

Kada bi logika imala posebnu važnost za zaključivanje, čini se da bi se oni koji zaključuju morali pridržavati barem sljedećih dvaju načela:

- 1) Načelo logičke implikacije: Činjenica da nečiji stavovi logički impliciraju  $p$  može biti razlog da on prihvati  $p$ .
- 2) Načelo logičke nekonzistentnosti: logička nekonzistentnost treba biti izbjegnuta.

Harman (1986) pojašnjava kako je za oba načela moguće navesti mnogo slučajeva u kojima ne vrijede. Primjerice, ako netko vjeruje  $p$  i ako  $p$ , onda  $q$ , prema načelu 1) to može dati razlog za vjerovanje  $q$ , ali ne govori da bi to trebalo spriječiti onoga koji zaključuje da istovremeno vjeruje i  $ne q$ . Vjerovanje  $ne q$  u slučaju kada osoba koja zaključuje vjeruje i  $p$  i ako  $p$ , onda  $q$  u suprotnosti je s načelom 2), ali ne i s načelom 1). Nadalje, prema načelu 1), ako osoba koja zaključuje vjeruje  $p$  i ako  $p$ , onda  $q$ , ima razloga vjerovati  $q$ , ali nekad to nije slučaj. Primjerice, ako *zaključivatelj* vjeruje da ako otvori kuhinjski ormarić, u njemu će pronaći zobenu kašu, te otvori kuhinjski ormarić, može se dogoditi da u njemu ne pronalazi zobenu kašu (možda je zaboravio da ju je pojeo ili ju je pojeo netko drugi).

Harman je stoga ponudio i oslabljene verzije navedenih načela:

- 1)' Zaključivatelj ima razloga vjerovati  $p$  ako prepoznaje da je  $p$  implicirano njegovim vjerovanjima.

2)' Zaključivatelj ima razloga izbjeći vjerovanja koja prepoznaje kao ne-konzistentna.

Ta su načela i dalje primjenjiva samo na zaključivatelje koji imaju uobičajene pojmove implikacije i nekonzistentnosti, ali to ne znači da ih svi imaju niti da ih se pridržavaju svi koji ih imaju, a usto su vrlo slaba. Iz navedenih razloga Harman zaključuje da nema očitog razloga zbog kojega je logika (u smislu logičke implikacije ili logičke nekonzistentnosti) posebno značajna za zaključivanje.

### 3.2. MacFarlane i načelo premošćivanja

Ako se želi zastupati tezu o normativnosti logike za zaključivanje, potrebno je pokazati na koji se način relacija logičke posljedice normativno pripisuje mentalnim stanjima. Sami zakoni logike ne mogu biti normativni, na što upućuje i Frege navodeći da na temelju njih ne slijede norme o tvrdnji, mišljenju, suđenju i zaključivanju (Frege 1918, cf. Steinberger 2017), već oni mogu implicirati normu kako bi se procijenilo misli li netko na odgovarajući način (u skladu s tom normom) (Steinberger 2017).

U cilju pokazivanja da je logika normativna za zaključivanje MacFarlane (2004) nastojao je pronaći poveznicu između logičkih pravila i zakona mišljenja u pogledu toga kako prva mogu biti obvezujuća za druge te kako je uopće moguće povezati jedne s drugima s obzirom na to da se radi o bitno drugačijim entitetima. Načela koja je pritom uveo nazivaju se načelima premošćivanja (engl. *bridge principles*). Za razliku od pravila dokazivanja, tvrdnje o logičkoj valjanosti nisu eksplicitno normativne na temelju svoga sadržaja. Ako postaju norme, to može biti jedino putem kakva načela premošćivanja koje povezuje tvrdnje o logičkoj valjanosti s normama vjerovanja. Radi se o implikacijama koje imaju odnos logičke posljedice među sudovima s jedne strane i tvrdnju o normativnim odnosima među tim sudovima s druge strane.

MacFarlane kreće od sljedeće sheme:

(NP) Ako  $A, B \Rightarrow C$ , onda (normativna tvrdnja o vjerovanju  $A, B$  i  $C$ ).

Oprimjerenja *normativne tvrdnje o vjerovanju*  $A, B$  i  $C$  sva imaju oblik implikacije gdje antecedens sadrži vjerovanja o  $A$  i  $B$ , a konsekvens vjerovanje o  $C$ . Na temelju opće sheme MacFarlane je izveo 36 različitih načela premošćivanja koja se razlikuju u pogledu deontičkoga operatora (obveza  $o$ , dopuštenje  $p$  i razlozi podložni opovrgavanju  $r$ ), polariteta vjerovanja (vjerovanje  $+$ , ne nevjerovanje  $-$ ) i dosega deontičkoga operatora (odnosi se samo na konsekvens  $C$ , odnosi se na antecedens i na konsekvens  $B$  ili se odnosi na cijeli kondicional  $W$ ).



Na primjer:

- (Co+) Ako se vjeruje  $A$  i  $B$ , onda se treba vjerovati  $C$ .
- (Bo+) Ako se treba vjerovati  $A$  i  $B$ , onda se treba vjerovati  $C$ .
- (Cp+) Ako se vjeruje  $A$  i vjeruje se  $B$ , onda se može vjerovati  $C$ .
- (Cr-) Ako se vjeruje  $A$  i vjeruje se  $B$ , onda postoji razlog da se ne odbacuje  $C$ .

Sam MacFarlane ističe kako su načela  $B$  preslaba. Prema  $Bo+$  ako se treba vjerovati u premise valjanoga argumenta, treba se vjerovati i u njegovu konkluziju. Ali što ako netko vjeruje u premise iako ne bi trebao? Tada  $Bo+$  ne govori ništa o tome što bi trebao ili ne bi trebao vjerovati. MacFarlane također smatra kako treba odbaciti načela  $C$ . Primjerice, vezano za  $Co+$ , gotovo će svaka logika sadržavati  $A \models A$  kao teorem, a u tom slučaju  $Co+$  implicira da za svaki  $A$ , ako se vjeruje  $A$ , mora se vjerovati  $A$ , ali vjerovanje nije samoopravdavajuće na taj način i često se vjeruje nešto što se ne bi trebalo vjerovati (str. 10).

Od načela  $W$  MacFarlane smatra da su sljedeća dva načela održiva:

- Wo- Treba se pobrinuti da ako se vjeruje  $A$  i vjeruje  $B$ , onda se ne može ne vjerovati  $C$ .<sup>123</sup>
- Wr+ Ima razloga za pobrinuti se da ako se vjeruje  $A$  i vjeruje  $B$ , onda se vjeruje  $C$ .

Ipak, ovako formuliranim načelima premoščivanja mogu se uputiti određeni prigovori. Ono što se događa u svakodnevnome životu jest da ako se vjeruje  $A$  i vjeruje  $B$  te je zaključivatelj svjestan toga da  $A, B \Rightarrow C$ , on se isto tako može pobrinuti za to da odbaci jedno od vjerovanja  $A$  ili  $B$  jer ima razloga zbog kojih ne želi vjerovati  $C$ . Načela premoščivanja vjerojatno su zamišljena kao neki univerzalni mehanizam, stoga ih takvi primjeri osporavaju, ali problem je naznačen i u doksastičkoj logici gdje je potrebno definirati o kakvom se tipu vjerovatelja radi pa nešto što vrijedi za idealnoga vjerovatelja ne vrijedi za standardnoga vjerovatelja, a teško je zamislivo da su svakodnevni zaključivatelji idealni vjerovatelji. Na to se nadovezuje problem s tvrdnjom u antecedensu načela premoščivanja. Što znači *Ako*  $A, B \Rightarrow C$ ? Ako zaključivatelj zna da  $A, B \Rightarrow C$ ? Ako je  $A, B \Rightarrow C$  tautologija? Što znači da netko zna da  $A, B \Rightarrow C$ ? Ovdje se dolazi do puno šireg proble-

<sup>123</sup> Prijevod s dvostrukom negacijom pomalo je nespretan jer se može činiti kao da se dvostruka negacija može pokratiti, ali u engleskome se jeziku koristi nešto jasniji izraz *not to disbelieve*. S druge strane, ovaj slučaj samo pokazuje da nema razloga smatrati da se dvostruka negacija u prirodnome jeziku može pokratiti ako je to svojstvo dvostruke negacije u jeziku logike. Naš jezik, koji dopušta dvostruku negaciju, dobar je primjer za to.

ma i pitanja o tome što je znanje, što prelazi opseg ovoga rada, ali čini se da se radi o važnom problemu koje načelo premošćivanja opet nije otklonilo jer se i dalje javlja u antecedensu pa, prema tome, načelo premošćivanja nije ništa riješilo.

Nevezano za same formulacije načela premošćivanja, prvo što se može dovesti u pitanje jest sam oblik takvih načela. Naime, kako je normativnost nekoga logičkog odnosa moguće prikazivati odnosom koji je također logički, i to također odnos logičke posljedice, iako se kod MacFarlanea govori o implikaciji? Kako i (NP) pripada metajeziku, i tu je riječ o logičkoj posljedici, pa dolazimo do toga da pri objašnjavanju odnosa jedne logičke posljedice i nečega u ljudskom umu ponovno koristimo logičku posljedicu. Za objašnjenje toga odnosa moglo bi se koristiti još jednu logičku posljedicu i tako u nedogled.

Sljedeći problem za (NP) jest taj što *A* ili *B* mogu biti neistiniti. U tom slučaju vrijedi da  $A, B \Rightarrow C$ , a također vrijedi da zaključivatelj vjeruje *A* i vjeruje *B*, te se prema (NP) treba pobrinuti da na temelju toga izvede *C*, čime zaključivatelj dobiva još jedno nevjerodostojno vjerovanje upitne istinitosne vrijednosti. Fitelson (2008) također ističe kako u kontekstu u kojem je skup vjerovanja nekonzistentan nema opravdanoga razloga za tvrdnju da ijedno načelo premošćivanja može poslužiti svrsi. Svako će načelo premošćivanja biti ili neistinito ili preslabo (u pogledu osiguravanja valjanosti zaključka) (str. 6. fn. 10).

Najveći i glavni prigovor načelu premošćivanja jest da, čak i kada bi se uspjelo izvesti načelo premošćivanja koje bi uspješno uspostavljalo normativnost logike u svakome pojedinom slučaju, to i dalje ne bi bilo dovoljno opravdanje za tvrdnju da je logika posebno značajna za zaključivanje, primjerice više od fizike ili neke druge prirodne znanosti jer sve one upućuju na slijeđenje istinitih normi. Kako bi logika bila posebno značajna za zaključivanje, logičke bi norme trebale utjecati na to da slijeđenje istinitih normi stvara više istinitih vjerovanja nego njihovo ne slijeđenje, a načelo premošćivanja to ne garantira. Štoviše, primjer s neistinitim tvrdnjama *A* ili *B* može činiti protuprimjer takvoj težnji.

### 3.3. Fieldov odgovor na kritike

Jedan je od glavnih zagovornika ideje o normativnosti logike Hartry Field. Svjestan je problema takve koncepcije koje navodi primjerice Harman (1986) te daje i odgovore na iznesene prigovore (2009), pri čemu predlaže i način na koji se može revidirati načelo premošćivanja koje je predložio MacFarlane. Jedan je od glavnih prigovora koje iznosi Harman već ranije spomenuta nekonzistentnost skupa vjerovanja. Field se s tim slaže, ali tvrdi

da ona nije sporna jer se radi o probabilistički nekonzistentnim vjerovanjima te je potrebno pokušati unijeti prilagodbe koje će ih ukloniti. Stoga uvodi stupnjeve vjerovanja pa navodi da ako  $A_1, \dots, A_n$  zajedno povlači  $B$ , onda stupanj vjerovanja u  $B$  mora biti najmanje toliko visok koliko je to vjerovanje u konjunkciju  $A_1, \dots, A_n$ .

Pokušavajući odgovoriti na sve prigovore koje iznosi Harman, uključujući slabljenje formulacije tako da više ne sadrži logičku posljedicu, Field normativnost logike izražava na ovaj način:

- (1) Način određenja kako logika može biti normativna jest u tome da norme koje zaključivatelj slijedi i norme koje upravljaju njegovim stupnjem vjerovanja jesu u skladu s pravilima te logike.
- (2) Pri vrednovanju nečijih vjerovanja i zaključaka ne vodimo se samo normama koje ta osoba slijedi, nego i normama koje smatramo dobrima (Field 2009: 262-263).

Field ističe kako (2) ne povezuje stvarne dužnosti (*trebanja*) sa stvarno ispravnom logikom (klasičnom logikom), ali povezuje prosudbe *trebanja* s onim što uzimamo kao *dobru logiku*, pri čemu je dobra logika određena ulogom koju igra u dobrim normama za vjerovanja. Što su dobre norme Field ne objašnjava, ali nastavlja kako je naš pojam dobrog argumenta bitno normativan pojam, te je logika normativna u tom smislu, ali nije zahvativ u smislu očuvanja istinitosti jer zbog nekonzistentnih vjerovanja moramo odbaciti tvrdnju da svi logički valjani zaključci čuvaju istinitost. Pritom se poziva i na Gödelov teorem nekompletnosti prema kojemu nijedna matematička teorija ne može dokazati vlastitu konzistentnost pa čak ni netrivialnost. Ipak, možemo zaključiti da je odustajanje od očuvanja istinitosti visoka cijena koju je potrebno platiti kako bi se sačuvala normativnost.

Iako Field rješava dio problema s kojima se suočava MacFarlane (2004), i dalje ostaje problem postupanja sa skupovima nekonzistentnih vjerovanja (bilo kod pojedinca, bilo kod zajednice). Drugim riječima, ako se ostaje pri tvrdnji da za pojedinca ili zajednicu može biti racionalno imati nekonzistentna vjerovanja, kao norma se mora uzeti logika koja ne sadrži načelo *Ex falso quodlibet* (EFQ) koje tvrdi da se iz kontradikcije može izvesti bilo što. Parakonzistentne logike odbacuju to načelo. Ali parakonzistentna logika odbacuje još neka načela koja želimo zadržati, primjerice disjunktivni silogizam, stoga ni ona nije dobar kandidat.

Zaključak do kojega možemo doći jest da kada bismo željeli uzeti bilo koju logiku kao normu za zaključivanje, morali bismo pripaziti na to da odabrana logika barem u većoj mjeri odgovara stvarnomu stanju jer je inače ta norma samo puki ideal za koji je odmah jasno da je nedostižan. Takvu

logiku vrlo teško možemo pronaći u formalnim sustavima jer ako i odaberemo neka svojstva koja nam odgovaraju, to će povući i neka druga svojstva koja ne želimo. Budući da svijet funkcionira prema logičkim načelima (onima klasične logike), tu nema nekonzistentnosti pa time ni potrebe za odbacivanjem EFQ-a.

### 3.4. Smokrović o normativnosti

Smokrović (2018) navodi kako ga zanima pitanje može li formalna logika, ili neka vrsta formalne logike, još uvijek, nakon svih prigovora koji su izneseni, imati normativne implikacije za zaključivanje, a to je pitanje važno jer nas drastično odbijanje normativnog utjecaja logike ostavlja bez sigurnog kriterija normativnosti. Nadalje ističe kako nam u tom slučaju preostaje samo pozivanje na intuicije koje bi trebale biti arbitrar ispravnosti zaključivanja. S druge strane, ako postoji vjerojatna teorijska veza između logike i zaključivanja, čak i ako se radi o neklasičnoj, nemonotoničkoj logici, naše razumijevanje zaključivanja bit će na mnogo čvršćem tlu (Smokrović 2018: 456). Ističe kako su ljudi skloni prepoznati lanac zaključivanja kao *dobar*, a argument kao *ispravan* iz perspektive prvog lica, povezujući stupanj povjerenja u ispravnost svojih prosudbi i druge ishode procesa zaključivanja. To se također javlja i iz perspektive trećeg lica, kada ljudi ocjenjuju zaključivanje drugih. Bez neke vrste norme, ne bismo mogli vršiti procjenu zaključaka drugih ljudi, kao ni slijediti pravilo po kojem sami zaključujemo.

U svom članku (2018) Smokrović diskutira o MacFarlaneovu (2004) i Fieldovu (2009) načelu premošćivanja, pri čemu prednost daje Fieldovu rješenju jer umjesto klasične logičke valjanosti uzima u obzir različite oblike valjanosti te logičku posljedicu zamjenjuje slabijom varijantom, odnosno varijantom za koju nije nužno da čuva istinitost. Ističe kako je takva varijanta liberalnija jer je bliža svakodnevnom zaključivanju, koje nastoji očuvati istinitost, ali samo onda kada je to važno (str. 463).

Na tragu MacFarlanea, Smokrović smatra da razumijevanje zaključka znači prepoznavanje toga da on ima određenu logičku strukturu, ne u smislu da zaključivatelj može točno odrediti tu strukturu, ali za što može preuzeti odgovornost i uputiti ili primiti kritiku. No, MacFarlaneovoj formulaciji, koja ne govori o valjanosti sheme zaključivanja, dodaje i taj dio jer je jedno pojedinačno zaključivanje oprimgjenjenje neke opće sheme, koja sama po sebi treba biti valjana. Ipak, ističe kako ovakav pristup ne može izbjeći prigovor koji se odnosi na slučaj revizije vjerovanja, a koji je naveden ranije. Stoga, za razliku od MacFarlanea i Fielda, predlaže pogled na deduktivno zaključivanje koje se očituje u različitim oblicima, a koji su vođeni različitim ciljevima, od kojih se svaki može modelirati drugačijom logikom.

Kao jednu mogućnost neklasične logike koja ovdje bolje pristaje od klasične logike Smokrović navodi probabilističku logiku jer smo uvijek u nekoj mjeri sigurni, na temelju istinitih premisa, da je konkluzija istinita te je takvo zaključivanje deduktivno. U toj se logici tradicionalno pitanje o istinitosti premisa zamjenjuje pitanjem o njihovoj vjerojatnosti. Takva je logika zasigurno deduktivna, iako nije monotonička (u početku dodjeljeni stupanj vjerojatnosti zaključku može se kasnije povući pred novim dokazom).

Govoreći o probabilističkome pristupu zaključivanju Harman (1984) ističe da konačna bića poput ljudi ne mogu djelovati probabilistički u većini slučajeva, osim u nekim vrlo specifičnim situacijama. Jedan od problema korištenja vjerojatnosti uključuje enormne memorijske i računске kapacitete koje bi takvo postupanje zahtijevalo. Stoga se čini da, s obzirom na svoje kapacitete i načelo ekonomičnosti koje primjenjuju, osobe što zaključuju ipak ne mogu ili nemaju razloga koristiti tako kompleksan sustav.

Druga je mogućnost zadana logika (engl. *default logic*), nemonotonička logika koja formalizira zaključivanje sa zadanim pretpostavkama. Dok klasična logika može izraziti samo je li nešto istinito ili neistinito, zadana logika može izraziti da je nešto istinito po zadanim postavkama. Stoga, obični se kondicional *Ako p, onda q* može zamijeniti kondicionalom *Ako p i ništa neuobičajeno, onda q* (Smokrović 2018: 467). Ova logika intuitivno djeluje bliže svakodnevnom zaključivanju, ali rješava samo slučajeve u kojima se zaključivanje prema načelima klasične logike narušava zbog neuobičajenog događaja.

#### 4. Pluralistički pristup normativnosti

O problemima koji proizlaze iz načela premošćivanja već je bilo riječi u prethodnim poglavljima, a Smokrović (2018) također ističe kako se u potpunosti ne mogu obraniti od Harmanova prigovora o mogućim revizijama vjerovanja. Ovdje ćemo se usmjeriti na razmatranje rješenja koje spominje Smokrović, a koji uključuje mogućnost različitih logika - logički pluralizam. Prema logičkom pluralizmu ne postoji jedna istinita logika, nego više njih. Također, nekad postoji više od jednog odgovora na pitanje *Je li ovaj zaključak valjan?* (Beall & Restall 2006).

Kao primjere logika koje bi se mogle koristiti navodi zadanu logiku i probabilističku logiku. Kako ne razlaže detaljnije ideju pluralizma ni korištenje svake od navedenih logika kao mogući standard zaključivanja, u ovom će dijelu rada fokus biti na razmatranju ideje logičkog pluralizma u kontekstu normativnosti, odnosno nastojat će se propitati je li pluralizam dobar smjer ako se želi pretpostaviti normativnost.

Prema pluralističkom gledištu, ako je logika normativna, onda postoji više mogućih normi, odnosno više logika od kojih je svaka normativna u nekoj domeni ili za neke potrebe. Tu se javlja pitanje što ako takve logike u nekom konkretnom slučaju daju nespojive preporuke o tome kako trebamo zaključivati? Prema kojoj od njih se treba ravnati?

Referirajući se na pluralizam koji predlaže Field (2009), a na čijem je tragu i Smokrović (2018), Steinberger (2019) ističe da on omogućava pojavu normativnog sukoba, s obzirom na to da se prema Fieldu može raditi o više kompetitivnih standarda vrednovanja te da se ne može za jedan skup normi reći da je jedinstveno ispravan. Stoga takav pluralist zapravo ne mora imati kao cilj izbjegavanje logičkih sporova, pogotovo ako se pretpostavi da ne postoje objektivni standardi za procjenu ispravnosti korištenja određenog standarda u određenoj situaciji. Ipak, to što ne postoji jedan skup logičkih normi, ne znači da se za neki skup normi ne može reći da je bolji od drugih. To je moguće jer se logičke norme koriste u skladu s epistemičkim ciljevima, te se mogu ocjenjivati s obzirom na to koliko učinkovito dovode do ostvarenja tih ciljeva (Steinberger 2019: 10). Međutim, i dalje ostaje problem kako zaključivatelj može znati koji je skup normi bolji te prema kojem bi se trebao ravnati u određenom slučaju.

Steinberger (2019) tvrdi da iako je pluralizam motiviran izbjegavanjem rasprava među zagovornicima različitih logika, ne može ponuditi valjani odgovor te na kraju ponovno dolazi do istih problema kao i monizam. Razlika u monizmu i pluralizmu, prema Steingergeru, svodi se na to da iako i monist i pluralist primjenjuju ista načela u istim kontekstima, logički je status tih načela različit. Prema pluralistu, ona su logički valjana u odnosu na određenu domenu, dok prema monistu nisu jer ovise o domeni. Ovdje se podrazumijeva da je logika prema kojoj se ravna monist uključena i u sustav pluralista. Problem koji se javlja kod monista sastoji se u tome što su jedini vjerodostojni zakoni logike oni koji vrijede u svim domenama, ali teško da se može pronaći logičko načelo za koje bi to vrijedilo. Stoga, ako za dovoljno mnogo područja najbolja teorija zahtijeva slabljenje logike, monist riskira da ostane na neograničeno slabom ili čak praznom skupu pravila. Taj prigovor Steinberger naziva prijetnjom logičkog nihilizma (str. 16). Međutim, ni pluralist ovdje nije u puno boljoj situaciji, jer se mora pridržavati načela koja vrijede u svim relevantnim domenama, što će opet dovesti do iste prijetnje kao i kod monizma.

Osim toga, problem s pluralizmima bilo koje vrste, na što se osvrće i Steinberger, jest u tome što ne dopuštaju objektivno načelo premošćivanja. Ako bi svaki pluralizam zahtijevao vlastito načelo premošćivanja koje će biti u skladu sa standardima svih logika koje su uključene, tada ne možemo govoriti o jednom načelu premošćivanja, a može se dogoditi i da su načela premošćivanja međusobno nekompatibilna.

S obzirom na navedeno, iako se čini da se rješenje nekih problema s kojima se susreću normativisti krije u pluralizmu, potrebno je vidjeti na koji bi se način moglo pristupiti pluralizmu kako bi se izbjeglo njegovo svođenje na iste probleme koji se javljaju u monizmu.

## 5. Zaključak

Rasprava o normativnosti i dalje ostavlja mnogo otvorenih pitanja. Čini se da nijedna dosad ustanovljena formulacija načela premošćivanja ne može proći bez zamjerki, a iako se ideja pluralističkog pristupa normativnosti čini dobrim putem, i tu postoje određeni izazovi te takav pristup zapravo dovodi u pitanje opstojnost bilo kojeg načela premošćivanja, kao i načela premošćivanja uopće. S druge strane, za zagovornike teze o nenormativnosti logike i dalje ostaje pitanje kako ljudi onda donose zaključke i kako procjenjuju zaključke drugih.

Bez obzira na raspravu o normativnosti, ljudsko zaključivanje, iako je puno grešaka, nedvojbeno daje rezultate i ima svoje metode kojih često i nije posve svjesno. Ipak, uloga logike nikako ne treba biti minorizirana, s obzirom na to da znanje logike daje vrlo korisnu metodu oblikovanja i usmjeravanja naših misli za izvođenje valjanih zaključaka, što je osobito važno u znanstvenome radu.

Točno je da se uvijek može postaviti neka norma koju će se imati u vidu dok se promatraju odstupanja, ali pitanje je koliko je to potrebno ako je razvidno da je ta norma nedostižna ili čak irelevantna. Analiziranje odstupanja s obzirom na takvu normu nije informativno kao što nisu informativni ni brojni ponovljeni eksperimenti koji uvijek iznova pokazuju da ljudi ne zaključuju u skladu s logičkim pravilima. U prvome je dijelu rada pokazano i da je puno toga potrebno riješiti kako bi se uopće moglo tvrditi da je logika normativna za zaključivanje, a to je prije svega vezano uz načela premošćivanja od kojih je svakomu moguće uputiti prigovor, kao i cijelomu pristupu.

## Literatura

- Beall, J. C., i Restall G. 2006. *Logical pluralism*. Oxford University Press, Oxford
- Dutilh Novaes, K. „A Dialogical, Multi-Agent Account of the Normativity of Logic“. *Dialectica*. 69 (4/2015), str. 587 – 609.
- Field, H. „Pluralism in Logic“. *The Review of Symbolic Logic*. 2 (2/2009), str. 342 – 359.
- Fitelson, B. „Goodman’s *New Riddle*“. *Journal of Philosophical Logic*. 37 (6/2008), str. 613 – 643.
- Harman, G. 1986. *Change in View: Principles of Reasoning*. MIT Press, Cambridge



- Harman, G. „Field on the Normative Role of Logic“. *Proceedings of the Aristotelian Society*. 109 (1/2009), str. 333 – 35.
- Harman, G. „Logic and Reasoning“. *Synthese*. 60 (1/1984), str. 107 – 127.
- Oaksford, M. i Chater, N. „Précis of Bayesian Rationality: The Probabilistic Approach to Human Reasoning“. *Behavioral and Brain Sciences*. 32 (2009), str. 69 – 84.
- MacFarlane, J. „In What Sense (If Any) Is Logic Normative for Reasoning“. (neobjavljeno) 2004.
- Smokrović, N. „Normativnost logike in vsakdanja sklepanja“. *Analiza*. 21 (2/2017), str. 45 – 61.
- Smokrović, N. „Informal Reasoning and Formal Logic: Normativity of Natural Language Reasoning“. *Croatian Journal of Philosophy*. 54 (1/2018), str. 455 – 469.
- Steinberger, F. „Explosion and the normativity of logic“. *Mind*. 125 (2016), str. 385 – 419.
- Steinberger, F. „Frege and Carnap on the Normativity of Logic“. *Synthese*. 194 (2017), str. 143 – 162. doi: 10.1007/s11229-015-0880-4.
- Steinberger, F. „Logic Pluralism and Logic Normativity“. *Philosophers' Imprint*. 19 (2019), str. 1 – 19.
- Wason, P. C. „Reasoning“. u: B. M. Foss. 1966. *New Horizons in Psychology*. Harmondsworth, Penguin

MATEJ SUŠNIK

## Priroda praktičnog zaključivanja

**Sažetak:** U tekstu razmatram odnos između teorijskog i praktičnog zaključivanja. Polazeći od elemenata teorijskog zaključivanja na koje ukazuje Nenad Smokrović, cilj rada jest utvrditi u kojoj su mjeri neki od tih elemenata primjenjivi u praktičnoj sferi. U prvom dijelu raspravljam o procesu, a u drugom o ishodu zaključivanja. U zaključku ukratko prikazujem poziciju kojoj se priklanam, a čiju je najutjecajniju verziju formulisao Bernard Williams.

**Ključne riječi:** praktično zaključivanje, proces zaključivanja, racionalnost, razlozi, teorijsko zaključivanje.

### 1. Uvod

Knjiga *Priroda prirodnog zaključivanja* (2004) Nenada Smokrovića bavi se teorijskim zaključivanjem. Teorijsko zaključivanje, on tvrdi, jest inferencijalni proces u čijoj se osnovi nalaze određena logička pravila. Premda ljudi nerijetko pogrešno zaključuju (kao što to uostalom i pokazuju različiti eksperimenti), to nije samim time i pokazatelj toga da model zaključivanja koji se temelji na logičkim pravilima treba odbaciti. Sasvim je lako moguće, tvrdi Smokrović, da izvor problema treba tražiti u samoj primjeni naših logičkih kompetencija i njihovoj podložnosti različitim ne-kognitivnim utjecajima. Smokrović se stoga priklanja stajalištu da “logika na određeni način jest standard racionalnosti” (Smokrović 2004: 23).

Imajući u vidu ovu sliku teorijskog zaključivanja, mogli bismo se upitati kako stoje stvari s praktičnim zaključivanjem. Je li praktično zaključivanje isto tako inferencijalno? Vrijede li logička pravila podjednako i u praktičnoj domeni? Je li logika ujedno i standard *praktične* racionalnosti? S obzirom na to da navedena pitanja nisu u središtu Smokrovićeve rasprave, on ih, razumljivo, ostavlja po strani. U ovom ću tekstu pokušati popuniti tu prazninu i razmotriti u kojoj su mjeri neki elementi teorijskog zaključivanja primjenjivi u praktičnoj sferi.

Teorijsko i praktično zaključivanje, barem nam tako izgleda, imaju različite ciljeve. Sasvim općenito govoreći, dok je cilj teorijskog zaključivanja utvrditi što vjerovati, praktičnim zaključivanjem nastojimo utvrditi što činiti. Relevantan dio današnjih rasprava o zaključivanju svodi se na pitanje odnosa između teorijske i praktične domene. Je li uopće opravdano povući

distinkciju između teorijskog i praktičnog zaključivanja? Ako jest, u čemu se sastoje njihove razlike? Ako nije, znači li to da postoji samo jedna vrsta zaključivanja?

Unatoč tome što Smokrović ne raspravlja o praktičnom zaključivanju, on ukazuje na distinkciju između teorijske i praktične racionalnosti. Ipak, on praktičnu racionalnost ponajviše spominje u kontekstu teorije odlučivanja te pritom ima u vidu ono što bismo mogli nazvati ekonomskom koncepcijom racionalnosti, koja se prvenstveno odnosi na racionalnost radnji ili izbora. Prema ovoj koncepciji, radnja je racionalna ako je uspješna, a uspješnost radnje ovisi o tome uvećava li očekivanu korist osobe koja radnju izvršava. Kada je riječ o teorijskoj racionalnosti, smatra Smokrović, stvari stoje drugačije. Ovdje je govor o racionalnosti prvenstveno vezan uz proces kojim dolazimo do određenog zaključka. Smokrovićevim riječima:

Kod praktičke racionalnosti čini se da je opravdanije racionalnost pripisati činu s obzirom na to da njime maksimiziramo korist. Kod teorijske se racionalnosti, pak, nekako čini da je logičnost procesa ta koja određuje racionalnost zaključka. (ibid., 41)

Nadovezujući se na ovu distinkciju, Smokrović također razlikuje “vanjsku” i “unutarnju” radnju. Dok je vanjska radnja – koja obično “izaziva neku promjenu u okolini” – ishod procesa odlučivanja, unutarnja radnja – shvaćena kao čin formiranja nekog vjerovanja – ishod je prethodnog procesa teorijskog zaključivanja (ibid., 40-1). Drugim riječima, ishod procesa zaključivanja (u teorijskoj domeni) jest određeni mentalni čin, a ishod procesa odlučivanja (u praktičnoj domeni) jest empirijski opažljiva radnja. Tablica 1 prikazuje Smokrovićevo razlikovanje između teorijske i praktične domene.

	TEORIJSKA DOMENA	PRAKTIČNA DOMENA
PROCES	Zaključivanje	Odlučivanje
ISHOD	Unutarnji čin zaključivanja (vjerovanje)	Vanjski čin (uključuje promjenu u okolini)

Tablica 1

Međutim, u suvremenim filozofskim raspravama o zaključivanju i racionalnosti, teoriji odlučivanja često se pristupa kao odvojenom području istraživanja. Jedno moguće objašnjenje jest u tome što teorija odlučivanja o zaključivanju zapravo ne govori mnogo. Premda ona govori o uvjetima koje naše preferencije moraju ispuniti da bi proces odlučivanja uopće mogao započeti, ona ne kaže ništa o tome kako zaključivati o našim preferencijama (usp. Mele i Rawling 2004: 4; Resnik 1987: 22).

Imajući to na umu, ako bismo željeli utvrditi sličnosti i razlike između teorijske i praktične domene, naše polazište u obama slučajevima trebalo bi biti jednako. U ovom tekstu stoga polazim od pretpostavke da je o zaključivanju moguće govoriti kako u teorijskoj tako i u praktičnoj domeni. Kada je riječ o ishodu zaključivanja, u oba slučaja govorit ću o činu zaključivanja, ostavljajući za sada otvorenim pitanje o kakvom je činu zapravo riječ. Ovaj pristup prikazan je u Tablici 2.

	TEORIJSKA DOMENA	PRAKTIČNA DOMENA
PROCES	Zaključivanje	Zaključivanje
ISHOD	Čin zaključivanja	Čin zaključivanja

Tablica 2

U nastavku ću razmotriti odnos između teorijskog i praktičnog zaključivanja usmjerujući se na navedene komponente. Dok ću u sljedećem odjeljku raspravljati o procesu zaključivanja, u trećem ću se odjeljku usredotočiti na ishod zaključivanja. Kao polazište moje rasprave uzet ću elemente teorijskog zaključivanja na koje ukazuje Smokrović. Pokušat ću utvrditi jesu li (i ako jesu, u kojoj mjeri) neki od tih elemenata prisutni i u praktičnoj domeni.

## 2. Proces

Većina autora o zaključivanju govori kao o nekoj vrsti procesa. Tako Bart Streumer, primjerice, smatra da je zaključivanje “proces modificiranja mentalnih stanja na racionalan način” (Streumer, 2010: 244), a John Broome zaključivanje vidi kao “proces koji vodi osobu od nekih njezinih postojećih mentalnih stanja do novog mentalnog stanja” (Broome 2000: 195). Uočimo da su navedene definicije neutralne s obzirom na teorijsku i praktičnu domenu. One nam ne govore što određeni proces zaključivanja čini *teorijskim* ili *praktičnim* procesom zaključivanja. Sasvim općenito, mogli bismo kazati da se ta dva procesa razlikuju po vrstama mentalnih stanja koja se u njima pojavljuju. Dok se u središtu teorijskog zaključivanja nalaze vjerovanja, proces praktičnog zaključivanja uključuje i druga mentalna stanja poput želja ili namjera. U tom smislu i Smokrović pod teorijskim zaključivanjem podrazumijeva “psihološke procese kojim iz postojećih mišljenja i vjerovanja kao rezultat dobivamo novo vjerovanje” (Smokrović 2004: 74). Slično određenje pruža i na drugom mjestu gdje govori o “procesu manipulacije vjerovanjima” koji osobu navodi ili na odbacivanje nekih već ranije usvojenih vjerovanja ili na formiranje nekih novih, prethodno nepostojećih vjerovanja (ibid., 42).

**Racionalnost**

Proces zaključivanja može biti racionalan ili iracionalan, a Smokrović o racionalnosti govori u kontekstu *pravila* i *razloga*. Tako on, s jedne strane, ističe da logička pravila predstavljaju standard racionalnosti odnosno da se “racionalnim smatra onaj proces zaključivanja koji se podudara s tako određenim *pravilima*” (ibid., 74, moj kurziv). No Smokrović također navodi da se priroda zaključivanja ne može u potpunosti objasniti naturalističkim putem, tj. kauzalnim odnosima između mentalnih stanja, nego da je za razumijevanje zaključivanja “potrebno uočavanje *razloga* na osnovi kojih je subjekt izveo određeni zaključak” (ibid., 58, moj kurziv). Razlozi koje on pritom ima u vidu su razlozi koji su obuhvaćeni logičkim pravilima.

Pokušajmo ovo ilustrirati primjerom. Uzmimo da, promatrajući neku šahovsku partiju, formiram vjerovanje da je (1) bijeli odigrao skakača na polje c6 i vjerovanje da (2) ako bijeli odigra skakača na polje c6, crnom prijeti mat u sljedećem potezu. Na osnovi navedenih vjerovanja, zaključujem da (3) crnom prijeti mat u sljedećem potezu. Navedeni proces zaključivanja prikazan je u Tablici 3.

Vjerujem da	...bijeli je odigrao skakača na polje c6.
Vjerujem da	...ako bijeli odigra skakača na polje c6, crnom prijeti mat u sljedećem potezu.
Prema tome, vjerujem da	...crnom prijeti mat u sljedećem potezu.

Tablica 3

Dok propozicije na desnoj strani označavaju sadržaj mojih mentalnih stanja, lijevi dio tablice prikazuje mentalni stav koji zauzمام prema navedenim propozicijama (propozicijski stav). Budući da tablica prikazuje primjer teorijskog zaključivanja, stav koji zauzمام prema gornjim propozicijama jest stav vjerovanja.

Međutim, ključno je naglasiti da je odnos između propozicija moguće analizirati sasvim neovisno o propozicijskim stavovima. Uočimo stoga da se propozicije u desnom dijelu tablice nalaze u odnosu koji je zahvaćen logičkim pravilom *modus ponens*. Označimo li propoziciju u prvom redu slovom  $p$ , a propoziciju u posljednjem redu slovom  $q$ , njihov međusobni odnos je sljedeći:  $p$ , ako  $p$  onda  $q$ , dakle  $q$ . Prema pravilu *modus ponens*, ako su prve dvije premise zaključka istinite (ako je  $p$  istinito i ako je  $p$  onda  $q$  istinito), onda je nužno da je i  $q$  istinito. U tom je smislu relaciju između mog (zaključnog) vjerovanja da  $q$  te mojih (početnih) vjerovanja da  $p$  i  $p$  onda  $q$  moguće objasniti upravo pozivanjem na logičku relaciju

koja postoji između navedenih propozicija. Uzimajući u obzir moje vjerovanje da  $p$  i vjerovanje da *ako  $p$  onda  $q$* , ta mi logička relacija daje *razlog* za vjerovanje da  $q$ .

Pretpostavimo li da se teorijsko zaključivanje odvija na ovakav način, moguće je upitati se nalazimo li slične elemente i u praktičnoj domeni. S obzirom na to da većina današnjih autora praktično zaključivanje najčešće prikazuje u instrumentalnom obliku, razmotrimo najprije jedan takav primjer zaključivanja. Recimo da danas namjeravam posjetiti obližnji otok te da vjerujem da jedini način na koji tu namjeru mogu ostvariti jest ukrcavanjem na trajekt koji tamo polazi u poslijepodnevnim satima. Polazeći od navedenih mentalnih stanja, procesom zaključivanja formiram namjeru ukrcati se na trajekt. Plauzibilno je pretpostaviti da je moje zaključivanje u ovom slučaju racionalno. Jer što reći o nekome tko namjerava ostvariti određeni cilj  $A$  te vjeruje da  $A$  može ostvariti samo ako učini  $B$ , ali pritom nema namjeru učiniti  $B$ ? Zar to ne bi bilo zbunjujuće?

Ali što je to što moje zaključivanje u ovom slučaju čini racionalnim? U teorijskoj sferi, vidjeli smo, racionalnost procesa zaključivanja moguće je objasniti pozivanjem na razloge koji se mogu izraziti logičkim relacijama između propozicija. No čini se da za takve razloge nema mjesta u praktičnoj sferi (usp. Smith 2004: 80). Slijedeći Humea, mogli bismo kazati da je to zato što mentalna stanja koja figuriraju u procesu praktičnog zaključivanja (namjere ili želje) nisu reprezentacijska i, prema tome, nemaju istinosnu vrijednost. Budući da nemaju istinosnu vrijednost, kako onda objasniti da moja namjera (posjet otoku) u kombinaciji s odgovarajućim vjerovanjem racionalno povlači drugu namjeru (ukrcavanje na trajekt)?

U nedostatku adekvatnog objašnjenja, nameće se zaključak da postoji bitna razlika između teorijske i praktične sfere. Ako je racionalno (teorijsko) zaključivanje ono koje pretpostavlja postojanje neovisnih razloga (koji se mogu izraziti logičkim pravilima), onda nam—imamo li u vidu da takvih razloga u praktičnoj sferi nema—ne preostaje ništa drugo nego objasniti racionalnost praktičnog zaključivanja na neki drugi način.

Generalno promatrajući, u suvremenim raspravama izdvajaju se dva različita odgovora (Kauppinen 2018). Dok neki smatraju da je praktično zaključivanje racionalno ako se odvija u skladu s odgovarajućim pravilima, drugi vjeruju da je proces praktičnog zaključivanja racionalan ako se odvija u skladu s prividnim ili stvarnim razlozima.

### ***Zaključivanje zasnovano na pravilima (ZP)***

Zastupnici ovog pristupa (npr. Broome, 2010) vjeruju da racionalnost postavlja određene zahtjeve i da te zahtjeve nerijetko možemo ispuniti upravo zaključivanjem. U određenim situacijama zaključivanje neće biti potrebno te ćemo zahtjeve racionalnosti ispunjavati nesvjesno. Recimo, osoba koja

vjeruje da je vani sunčan dan automatski će odbaciti to vjerovanje kao ne-utemeljeno nakon što primijeti da je vrijeme tmurno i oblačno. Time će ona nesvjesno primijeniti pravilo koje nalaže da osoba, ako je racionalna, nema kontradiktorna vjerovanja. No ponekad nesvjesni mehanizmi neće funkcionirati te će biti potrebno upustiti se u svjestan proces zaključivanja kako bi zahtjevi racionalnosti mogli biti ispunjeni (Broome 2010: 290-91).

U procesu zaključivanja, prema ovom shvaćanju, osoba ispunjava određeni zahtjev racionalnosti primjenom nekog unaprijed zadanog pravila. Primjerice, primijenim li pravilo *modus ponens*, ispunit ću zahtjev racionalnosti koji nalaže da ako vjerujem da  $p$  i ako vjerujem da *ako  $p$  onda  $q$* , onda vjerujem da  $q$ . Broome naglašava da mi ne moramo uvijek znati koje pravilo primjenjujemo u procesu zaključivanja. Kao što nismo uvijek u stanju identificirati gramatička pravila na koja se oslanjamo u govoru i pismu, tako ne moramo znati koja pravila reguliraju naše zaključivanje. Osim toga, pravila koja primjenjujemo ne moraju biti ispravna. Kao što možemo govoriti i pisati gramatički neispravno, tako možemo i neispravno zaključivati (Broome 2013: 288).

No što neki zahtjev čini zahtjevom racionalnosti? Premda je teško odrediti precizne kriterije koji moraju biti zadovoljeni da bi se nešto moglo nazvati zahtjevom racionalnosti, ono što je zajedničko svim takvim zahtjevima—bilo da je riječ o teorijskoj ili o praktičnoj domeni—jest ideja da se racionalnost prvenstveno tiče odnosa između mentalnih stanja. Racionalnost zahtijeva da naša mentalna stanja budu međusobno usklađena te da se međusobno podržavaju. Najjednostavnije rečeno, racionalna osoba jest ona osoba koja ima “red u glavi” odnosno osoba čija su mentalna stanja koherentna. Takav red ne nalazimo samo kod nekoga tko zaključuje u skladu s pravilom *modus ponens*, nego i nekoga tko, primjerice, u zaključivanju primjenjuje instrumentalno pravilo te na temelju postojećih mentalnih stanja (namjere da ostvari neki cilj te vjerovanja o tome kako to može učiniti) formira odgovarajuću namjeru.

Pitanju o tome je li neko zaključivanje racionalno, prema ZP-u, treba pristupiti odvojeno od pitanja o tome što imamo razloga vjerovati ili činiti. Zahtjevi koje postavlja racionalnost u tom su smislu neovisni o teorijskim i praktičnim razlozima. Pod pretpostavkom da  $p$  povlači  $q$ , za mene je racionalno vjerovati da  $q$  ako vjerujem da  $p$ . No to ujedno ne znači da imam razlog vjerovati da  $q$ . Moguće je da nemam takav razlog zato što u prvom redu nemam razlog vjerovati da  $p$ . U slučaju da doista nemam razlog vjerovati da  $p$ , taj razlog ne mogu stvoriti *ex nihilo*, naime samim time što ću formirati vjerovanje da  $p$ . Što je, prema tome, u navedenoj situaciji za mene racionalno vjerovati? S obzirom na to da racionalnost zahtijeva koherenciju između mentalnih stanja, zahtjevu racionalnosti ovdje mogu udovoljiti ili napuštanjem vjerovanja da  $p$  ili formiranjem vjerovanja da  $q$ .



Isto vrijedi i u praktičnoj sferi. Ako imam namjeru ostvariti neki cilj, a pritom nemam namjeru učiniti ono za što vjerujem da je jedini način na koji taj cilj mogu ostvariti, moja mentalna stanja neće biti koherentna. No formirati odgovarajuću namjeru nije jedini izlaz iz stanja nekoherencije. Ono što također mogu učiniti jest odustati od namjere za ostvarenjem cilja. Jedino što, promatrajući iz aspekta racionalnosti, ne smijem učiniti jest zadržati namjeru za ostvarenjem cilja, a istovremeno ne formirati namjeru poduzeti ono za što vjerujem da je jedino sredstvo putem kojeg cilj mogu realizirati.

Zastupnici ZP-a, prema tome, zaključivanje razumiju kao proces kojim možemo ispuniti zahtjeve racionalnosti. Premda je iluzorno očekivati da bi netko mogao sastaviti neku konačnu listu svih zahtjeva racionalnosti, u ovom je kontekstu dovoljno uočiti da je racionalnost, unutar ovog pristupa, shvaćena *strukturno* (usp. Scanlon 2007: 84; Kauppinen 2018: 7; Kieseletter 2017: 14). Najjednostavnije rečeno, ono što racionalnost od nas zahtijeva nije usvajanje nekog pojedinačnog mentalnog stava, nego formiranje odnosno izbjegavanje određene kombinacije mentalnih stavova. Pritom valja uočiti da mentalni stavovi o kojima je ovdje riječ ne uključuju samo vjerovanja. Primjerice, racionalnost zahtijeva da osoba istovremeno ne vjeruje da  $p$  i  $ne-p$ , ali isto tako zahtijeva da osoba istovremeno ne namjerava  $p$  i  $ne-p$ . Shvati li se racionalnost na taj način, pojavljuju se određene sličnosti između teorijske i praktične domene. Naime, ono što povezuje zahtjeve teorijske i praktične racionalnosti jest zahtjev za koherencijom među mentalnim stanjima. A s obzirom na to da strukturalna racionalnost nije utemeljena ni u čemu izvanjskom, taj zahtjev može biti ispunjen čak i ako se odnosi na mentalna stanja koja ne mogu biti istinita ili neistinita (npr. namjere ili želje).

### ***Zaključivanje zasnovano na razlozima (ZR)***

Prema ranije prikazanom pristupu teorijskom zaključivanju, racionalno teorijsko zaključivanje zasnovano je na pretpostavci o postojanju neovisnih normativnih razloga. No neki autori smatraju da takvi razlozi postoje i u praktičnoj sferi. Sasvim općenito govoreći, oni vjeruju da postoje istine o tome što imamo razloga činiti te da se uloga praktičnog razuma sastoji u otkrivanju tih istina.

Racionalna će osoba, prema njihovu shvaćanju, ponekad 'reagirati' na prividne, ali ne i stvarne razloge. Međutim, to ne predstavlja poteškoću. Jer, strogo promatrajući, mi zaključujemo na osnovi razloga za koje vjerujemo da postoje, a naša su vjerovanja ponekad neistinita. Tako racionalnost mog procesa teorijskog zaključivanja neće biti dovedena u pitanje čak i ako se pokaže da su moja početna vjerovanja neistinita (naime, čak i ako

pogrešno vjerujem da  $p$  ili ako pogrešno vjerujem da *ako  $p$  onda  $q$* ). Jer *ako* bi bilo istinito da  $p$  i ako bi bilo istinito da *ako  $p$  onda  $q$* , onda bi  $p$  i *ako  $p$  onda  $q$*  bili razlozi zašto  $q$ . Doduše, iz toga proizlazi da nemam stvarni razlog vjerovati da  $q$  u situaciji u kojoj su moja početna vjerovanja neistinita. Ali ako bi moja vjerovanja bila istinita, onda bih imao takav razlog. To je moguće objašnjenje za to zašto je, unatoč mojim pogrešnim početnim vjerovanjima, za mene racionalno formirati vjerovanje da  $q$ .

Sličan primjer, ali u praktičnoj domeni, pruža Derek Parfit. Pretpostavimo da imam neistinito vjerovanje o tome da je hotel u kojem se nalazim zahvatio požar. Premda u tom slučaju nemam stvaran razlog skočiti u obližnju rijeku, ipak je racionalno da to učinim (Parfit 2001: 17). Jer ako bi moje vjerovanje bilo istinito, onda bih imao razlog skočiti u obližnju rijeku. Drugim riječima, zastupnik ZR-a smatra da su razlozi neovisni o nama te da proizlaze iz određenih činjenica ili stanja stvari u svijetu. I kao što navedeni primjer ilustrira, mi ponekad imamo pogrešna vjerovanja o tim činjenicama ili stanjima stvari. Skočio sam u rijeku zato što sam pogrešno vjerovao da je požar. Ali da je doista bio požar, ta bi mi činjenica pružala razlog da skočim. No moje ponašanje svejedno predstavlja reakciju na razlog—u ovom slučaju, prividan razlog—te je zbog toga racionalno.

Na sličan način, ponekad imamo istinita vjerovanja o činjenicama, ali ne vjerujemo da nam one daju razloge. Možda istinito vjerujem da ću pušenjem uništiti svoje zdravlje, ali ne vjerujem da mi to daje razlog da prestanem. Ako mi ta činjenica doista daje takav razlog, onda se ne ponašam racionalno ako nastavljam pušiti. U tom slučaju pokazujem nesenzitivnost prema razlogu koji postoji neovisno o meni.

Postoji nekoliko načina na koje možemo utvrditi objektivno postojeće razloge. Za prepoznavanje razloga ponekad je sasvim dovoljno jednostavno biti upoznat s činjenicama odnosno određenim stanjem stvari. Na primjer, ako vidimo da se netko nalazi u životnoj opasnosti, većina ljudi to će automatski prepoznati kao razlog da nešto učine. Ali u mnogim situacijama to neće biti slučaj. Ponekad razlog neće biti evidentan te će biti potrebno dodatno reflektirati i upustiti se u proces zaključivanja (usp. Robertson 2009: 15-6).

Uočimo da ovaj pristup prihvaća drugačiju koncepciju racionalnosti od one ranije navedene. Za razliku od strukturalista—koji racionalnost promatraju kao nešto što se u prvom redu tiče relacija između mentalnih stanja—*supstantivisti* racionalnost usko povezuju s neovisno postojećim razlozima. Supstantivisti smatraju da je racionalna osoba ona čija su mentalna stanja (npr. namjere, želje) i djelovanje usklađeni s takvim razlozima, a oni kod kojih to nije slučaj podložni su racionalnoj kritici (usp. Hooker i Streumer 2004: 57-74; Parfit 2011: 62).

Meni na ovom mjestu nije namjera opredijeliti se za neku od navedenih koncepcija praktičnog zaključivanja. Oba pristupa danas su predmetom brojnih rasprava i kritika, a detaljno razmatranje njihovih osnovnih značajki nadilazi okvire ovog teksta. No potrebno je napomenuti sljedeće. Krenemo li od pretpostavke da se racionalno zaključivanje u teorijskoj domeni—kao što to, čini se, sugerira Smokrović—doista odvija u skladu s logičkim pravilima i neovisno postojećim razlozima, nameće se zaključak da između teorijske i praktične domene ipak postoje neke nepremostive razlike. Ako bismo željeli pokazati da su procesi zaključivanja u objema domenama relativno slični, trebali bismo ili napustiti Smokrovićevo razumijevanje teorijske racionalnosti ili pokazati da razlike između ZP-a i ZR-a u praktičnoj sferi nisu toliko velike kao što to možda izgleda na prvi pogled.

### 3. Ishod

Raspravljajući o racionalnosti, Smokrović također navodi da “[p]rocesom [teorijskog] zaključivanja produciramo vjerovanja o svijetu” (15). U teorijskoj domeni dakle svrha zaključivanja jest formiranje novih *vjerovanja*. Ali kako stoje stvari u praktičnoj domeni? Je li ishod praktičnog zaključivanja također određeno mentalno stanje ili je ovdje pak riječ nečem drugom? Čini se da postoje tri mogućnosti: (a) praktično zaključivanje rezultira formiranjem *namjere*; (b) praktično zaključivanje rezultira *radnjom*; i (c) praktično zaključivanje rezultira formiranjem *vjerovanja*. Tablica 4 prikazuje navedene mogućnosti.

PROCES	Teorijsko zaključivanje	Praktično zaključivanje		
ISHOD	Vjerovanje	Namjera?	Radnja?	Vjerovanje?

Tablica 4

Postoji nekoliko razloga koji idu u prilog stavu da je ishod praktičnog zaključivanja namjera. Jedan od njih proizlazi iz distinkcije između formiranja namjere i djelovanja na osnovi namjere. Polazeći od ove distinkcije, neki autori navode da je formiranje namjere krajnja točka do koje proces zaključivanja uopće može doći (usp. Raz 1978; Broome 2002). Kada bi ishod procesa zaključivanja bila radnja, to bi značilo da onaj kojeg taj proces ne dovodi do izvršenja radnje zapravo ima deficitarnu sposobnost zaključivanja. Ali to, naravno, ne mora biti slučaj.

Recimo da imam namjeru koncem tjedna otići na nogometnu utakmicu. Pod pretpostavkom da će sve ostalo biti jednako, tu ću namjeru koncem tjedna doista i ostvariti. Moj čin odlaska na utakmicu tada će biti

rezultat moje prethodne namjere. No sve ostalo možda neće biti jednako (netko bi mogao reći da stvari obično i nisu). Na utakmicu neću otići izvršim li u međuvremenu na hitnom kirurškom zahvatu. Odviju li se stvari na takav način, doista bi bilo neobično smatrati da je ovdje zakazala moja sposobnost zaključivanja. Slično tome, netko tko je paraliziran ili zavezan možda nije u stanju djelovati na odgovarajući način, ali to ne znači da mu je samim time oduzeta i sposobnost zaključivanja.

Ovaj argument ima snagu u onoj mjeri u kojoj se pod namjerom podrazumijeva *prethodna* namjera odnosno namjera koja prethodi činu. U tom slučaju distinkcija između namjere i radnje očito postoji. Ipak, potencijalni problem jest u tome što namjera ne mora uvijek prethoditi radnji. Postojanje prethodne namjere zapravo uopće nije nužan uvjet djelovanja, a tome svjedoči i činjenica da ponekad namjerno činimo stvari koje prethodno nismo namjeravali učiniti. Tako za vrijeme telefonskog razgovora mogu šetati gore-dolje po sobi, pogledati kroz prozor ili promijeniti kanal na televizoru, ali nijedna od navedenih radnji ne mora biti rezultat neke prethodne namjere. No to što spomenute radnje nisu rezultat prethodne namjere ne znači da ih u danom trenutku ne izvršavam namjerno: unatoč tome što nisam imao prethodnu namjeru pogledati kroz prozor, ipak sam namjerno pogledao kroz prozor. Ponekad se namjere prvi put pojavljuju u samom djelovanju (usp. Searle 1983: 84).

Budući da je pojam namjere moguće upotrebljavati na više različitih načina, postavlja se pitanje zašto bi ishod praktičnog zaključivanja bila prethodna namjera, a ne, recimo, namjera u djelovanju? Ako bi proces praktičnog zaključivanja mogao rezultirati i potonjim, onda se, kao što napominje Jonathan Dancy (2018: 168), čini da više ne postoji temelj za isključivanje radnje kao jednog od mogućih ishoda tog procesa. To je stoga što je namjera u djelovanju neodvojivi dio samog djelovanja i svaki pokušaj njihova razdvajanja unaprijed je osuđen na neuspjeh. Naravno, još je uvijek moguće razlikovati namjeru u djelovanju od određenog tjelesnog pokreta, no, kao što također upozorava Dancy, tjelesni pokret nije isto što i radnja.

Netko bi ovdje mogao odgovoriti da namjera u djelovanju ne može biti ishod praktičnog zaključivanja *upravo zbog toga* što je ona neodvojivi dio djelovanja. U tom kontekstu, kazati da je namjera u djelovanju ishod praktičnog zaključivanja u osnovi znači kazati upravo ono što protivnici tog stajališta dovode u pitanje (doduše pod nešto drugačijim opisom)—naime, da je djelovanje ishod praktičnog zaključivanja.

Čak i ako se pokaže da je gornji odgovor neodrživ, tvrdnja da proces praktičnog zaključivanja mora rezultirati izvršenjem radnje suočava se s daljnjom poteškoćom (Streumer 2010: 248). Pretpostavimo li da neku radnju možemo izvršiti tek u (daljoj) budućnosti, znači li to da će proces

praktičnog zaključivanja trajati sve dok radnja o kojoj je riječ ne bude izvršena? To zvuči apsurdno. Tako u ovom trenutku mogu razmišljati o tome hoću li za godinu dana upisati tečaj francuskog jezika, ali ako praktično zaključivanje mora rezultirati radnjom, onda bi iz toga slijedilo da će moj sadašnji proces zaključivanja o tome što ću učiniti za godinu dana trajati sve dok tečaj francuskog doista ne upišem. Ali jasno je da u međuvremenu mogu sasvim zaboraviti da sam ikada razmišljao o upisu tečaja francuskog. U tom je smislu krajnje neuvjerljivo tvrditi da moj proces zaključivanja još uvijek traje.

Je li ovaj prigovor moguće izbjeći priklonimo li se stajalištu prema kojemu praktično zaključivanje rezultira, primjerice, formiranjem prethodne namjere? Čini se da ne. Premda u ovom trenutku mogu intenzivno razmišljati o upisu tečaja francuskog, ne moram formirati namjeru da to učinim. Naprotiv, možda ću namjeru da upišem tečaj formirati tek za godinu dana. Znači li to da će moj proces praktičnog zaključivanja trajati sve do tog trenutka? Ponovno se suočavamo s istim problemom.

Mogući izlaz iz ove poteškoće jest prihvatiti (c), tj. tvrdnju da praktično zaključivanje rezultira formiranjem vjerovanja. S obzirom na to da proces deliberacije, kako se čini, može završiti prije formiranja namjere ili izvršenja radnje, netko bi mogao reći, jedini preostali kandidat za ishod tog procesa jest vjerovanje. Ovdje se pod vjerovanjem podrazumijeva normativno vjerovanje—vjerovanje o tome što imamo razloga učiniti ili što trebamo učiniti. Moje zaključivanje o tome hoću li sljedeće godine upisati tečaj francuskog jezika tako će, prema ovom pristupu, rezultirati formiranjem vjerovanja da to imam razloga učiniti, a formiranje namjere da to učinim kao i sam čin upisivanja tečaja nastupit će naknadno (ako uopće nastupi).

Međutim, ovom se rješenju često upućuje zamjerka da ostaje nejasno u kojem je smislu ovdje uopće riječ o *praktičnom* zaključivanju (usp. Brunero 2020; Raz 2015). Krenemo li od pretpostavke da namjera i čin mogu sasvim izostati nakon formiranja odgovarajućeg vjerovanja, otvara se pitanje što takvo zaključivanje čini praktičnim? Osim toga, izvorni problem i dalje ostaje. To što formiranje vjerovanja prethodi formiranju namjere i izvršenju radnje ne znači da do formiranja odgovarajućeg vjerovanja doista mora i doći. Možda sam se upustio u proces deliberacije o mogućem pohađanju tečaja iz francuskog, ali iz određenih razloga taj proces nisam priveo kraju te stoga nisam formirao odgovarajuće vjerovanje o tome imam li razloga to učiniti.

Rasprava o ishodu praktičnog zaključivanja sugerira sljedeće. Ako bi odgovor (c) bio prihvaćen, onda bi postojala velika sličnost između teorijske i praktične domene. Tada bi i teorijsko i praktično zaključivanje rezultiralo vjerovanjem. Određena sličnost također bi postojala ako bi odgovor

(a) bio ispravan. Naravno, ta bi sličnost bila nešto manja nego u slučaju (c), ali još uvijek bi se moglo reći da je ishod zaključivanja određeno mentalno stanje. Sličnost bi bila najmanja ako bi odgovor (b) bio ispravan. Dok se neki autori jasno opredjeljuju za neku od spomenutih opcija, drugi smatraju da postoje različiti oblici praktičnog zaključivanja te da nijedna od navedenih opcija ne isključuje drugu. Oni vjeruju da bismo jednostavno trebali napustiti ideju o tome da praktično zaključivanje mora imati samo jedan ishod. Ako su oni u pravu, morat ćemo se pomiriti s činjenicom da između teorijske i praktične domene—barem kada je riječ o ishodu zaključivanja—ipak postoje značajne razlike.

#### 4. Završna razmatranja

Premda nije upitno da između teorijskog i praktičnog zaključivanja postoje i stanovite sličnosti, ta se dva procesa u određenim aspektima ipak bitno razlikuju. Jedan takav aspekt jest činjenica da se praktično zaključivanje fundamentalno odvija iz perspektive prvog lica. Kada zaključujemo o tome što učiniti, mi ne razmišljamo isključivo o vanjskom stanju stvari, nego i o nama samima te specifičnim okolnostima u kojima se nalazimo. Na ovo distinktivno obilježje praktične domene prvi je ukazao Bernard Williams: “Kada razmišljam o svijetu i pokušavam utvrditi istinu o njemu, ja mislim o svijetu te izričem tvrdnje ili postavljam pitanja koja su o njemu, a ne o meni” (Williams 1985: 67). Mogu razmišljati o tome je li Washington glavni grad Sjedinjenih Američkih Država ili se pitati nastaju li doista plima i oseka zbog gravitacijskih sila Sunca i Mjeseca, ali takvo razmišljanje nije o meni te ne uključuje moje osobne projekte i ciljeve. Proces praktičnog zaključivanja esencijalno je o meni i uključuje razmišljanje o tome hoću li ja djelovati na određeni način. Radnja koju ću izvršiti nakon upuštanja u proces praktične deliberacije u tom smislu odražava moje specifične želje i preferencije.

Ovdje je potrebno naglasiti da govor o izvršenju radnje ne pretpostavlja ništa o ishodu praktičnog zaključivanja. Neovisno o tome je li izvršenje radnje dio procesa zaključivanja ili ono nastupa tek nakon što je taj proces završio, djelovanje u konačnici jest ono zbog čega se u praktično zaključivanje uopće upuštamo. Pokaže li se da je krajnja instanca praktičnog zaključivanja formiranje vjerovanja da imamo razlog nešto učiniti ili namjere da nešto učinimo, čini se razumnim pretpostaviti da bi nas ta mentalna stanja, pod jednakim uvjetima, trebala pokrenuti na djelovanje. Ako praktični zaključci imaju oblik mentalnih stanja, onda nije neobično smatrati da mora postojati određena relacija između praktičnih zaključaka i samog djelovanja.

Oslanjajući se na gornje uvide—naime, da se praktična deliberacija odvija iz perspektive prvog lica i da mora postojati određena relacija između djelovanja i konkluzije takve deliberacije—moguće je izvesti neke daljnje tvrdnje o prirodi ovog procesa. Želimo li sačuvati ideju da praktično zaključivanje sadrži elemente koji su prepoznatljivo *moji* te ideju da sam *ja* autor radnje koja u konačnici proizlazi iz takvog zaključivanja, plauzibilno je postaviti hipotezu da cijeli proces zaključivanja mora započeti iz nečega što se tiče mene osobno i što istovremeno ima motivacijsku snagu u smislu da me može pokrenuti na djelovanje. Williams to ‘nešto’ naziva “subjektivnim motivacijskim skupom”, a koji, kako on navodi, ne sadrži samo želje u užem smislu riječi, nego i različite dispozicije, emocije, projekte i obveze (Williams 1981: 105).

Prednost ovog pristupa jest u tome što njegov zagovornik sasvim lako može prihvatiti da se praktično zaključivanje odvija u skladu s određenim pravilima, a istovremeno sačuvati relaciju između praktičnog zaključivanja i razloga za djelovanje. Najjednostavnije rečeno, proces praktičnog zaključivanja, prema ovom gledištu, uvijek započinje iz elemenata subjektivnog motivacijskog skupa te nas—odvijajući se u skladu s unaprijed definiranim pravilima i procedurama—dovodi do zaključka o tome što imamo razloga učiniti u određenoj situaciji. To znači da rezultat zaključivanja nije samo određen subjektivnim motivacijskim elementima, nego bitno ovisi i o prirodi tih pravila i procedura.

U svrhu ilustracije, zamislimo da se motivacijski skup osobe *A* sastoji od samo jednog elementa, primjerice, želje da učini *x*. Ima li *A* razlog učiniti *x*? Odgovor na to pitanje ovisi o tome bi li *A*, nakon racionalnog zaključivanja, i dalje željela učiniti *x*. Ako bi osoba *A* to i tada željela, onda je istinito da ona ima razlog učiniti *x*, ali ako njezina želja ne bi ‘preživjela’ proces racionalnog zaključivanja, onda ona takav razlog nema. Međutim, čak i ako se složimo s time da proces zaključivanja uvijek započinje iz subjektivnog motivacijskog skupa, ono što imamo razlog učiniti u nekim specifičnim okolnostima i dalje će velikim dijelom biti uvjetovano prirodom tog procesa. Recimo, ako se racionalno praktično zaključivanje isključivo odvija u skladu s instrumentalnim pravilom, razlozi za djelovanje uvijek će imati hipotetički oblik u smislu da će u potpunosti ovisiti o tome što se nalazi u nečijem subjektivnom motivacijskom skupu. S druge strane, ako, uz ono instrumentalno, postoje i neka druga pravila zaključivanja (poput recimo kategoričkih), ono što imamo razloga učiniti u pojedinoj situaciji neće uvijek biti relativno u odnosu na ono iz čega je zaključivanje započelo (usp. Sušnik 2009).

Ostavljajući ovu raspravu po strani, dovoljno je uočiti da se praktično zaključivanje—u čemu god da se taj proces sastoji—ovdje pokušava odre-



diti neovisno o razlozima za djelovanje. To pritom ne znači da ti razlozi predstavljaju neko odvojeno područje istraživanja koje nije ni u kakvoj vezi s procesom zaključivanja. Naprotiv, jedino je riječ o tome da je odnos između razloga i racionalnog zaključivanja u teorijskoj i praktičnoj sferi dijametralno suprotan. Za razliku od razloga u teorijskoj domeni, praktični razlozi nisu neovisni, nego u potpunosti ovise o rezultatu racionalne praktične deliberacije. Ako je u teorijskoj sferi racionalnost definirana u kategorijama razloga, u praktičnoj su razlozi definirani u kategorijama racionalnosti. Dok je teorijski racionalna osoba ona koja zaključuje u skladu s neovisno postojećim razlozima, praktični razlozi su oni koji su izvedeni iz racionalnog praktičnog zaključivanja.

U ovom sam tekstu krenuo od pretpostavke da između teorijskog i praktičnog zaključivanja ne postoji oštra distinkcija te sam pokušao ispitati mogu li se neki elementi teorijskog zaključivanja također primijeniti i u praktičnoj sferi. Premda tu mogućnost nisam ni približno odbacio, nastojao sam ukazati na neke poteškoće koji se neminovno javljaju na putu formuliranja jedinstvene teorije zaključivanja, naime teorije koja bi jednako dobro funkcionirala kako u teorijskoj tako i praktičnoj domeni. U posljednjem sam dijelu nastojao ukratko skicirati poziciju kojoj se priklanjam te čiju je najutjecajnijiu verziju formulirao Williams.

Prilikom jednog od brojnih razgovora o razlozima i racionalnosti koje smo vodili na Filozofskom fakultetu u Rijeci, Nenad Smokrović je i sam spomenuo da bi bilo zgodno razmotriti je li moguće povući paralelu između teorijskog i praktičnog zaključivanja. Ovaj tekst predstavlja skroman pokušaj u tom pravcu.

## Literatura

- Broome, J. 2000. "Instrumental Reasoning" u: J. Nida-Rümelin i W. Spohn (ur.), *Rationality, Rules and Structure*, Kluwer: Springer, 195-207.
- Broome, J. 2002. "Practical Reasoning", u: J. Bermúdez i A. Millar (ur.), *Reason and Nature: Essays in the Theory of Rationality*, Oxford: Oxford University Press, 85-111.
- Broome, J. 2010. "Rationality" u: T. O'Connor i Constantine Sandis (ur.), *A Companion to the Philosophy of Action*, Wiley-Blackwell, 285-292.
- Broome, J. 2013. "Practical Reasoning and Inference" u: D. Bakhurst, B. Hooker i M. Little (ur.), *Thinking About Reasons: Essays in Honour of Jonathan Dancy*, Oxford: Oxford University Press, 286-309.
- Brunero, J. 2020. "The Conclusion of Practical Reasoning", *The Journal of Ethics* 25, 13-37.
- Dancy, J. 2018. *Practical Shape: A Theory of Practical Reasoning*. Oxford: Oxford University Press.

- Hooker, B. i Streumer, B. 2004. "Procedural and Substantive Practical Rationality", u: A. R. Mele i P. Rawling (ur.), *The Oxford Handbook of Rationality*, Oxford: Oxford University Press, 57-74.
- Kauppinen, A. 2018. "Practical Reasoning" u: D. Star (ur.), *The Oxford Handbook of Reasons and Normativity*, Oxford: Oxford University Press, 395-422.
- Kiesewetter, B. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press.
- Mele, A. R. and Rawling, P. 2004. "Introduction: Aspects of Rationality" u: A. R. Mele i P. Rawling (ur.), *The Oxford Handbook of Rationality*, Oxford: Oxford University Press, 3-13.
- Parfit, D. 2001. "Rationality and Reasons", u: D. Egonsson et. al. (ur.) *Exploring Practical Philosophy: From Action to Values*, London: Routledge, 17-39.
- Parfit, D. 2011. *On What Matters* Vol I. Oxford: Oxford University Press.
- Raz, J. 1978. *Practical Reasoning*. Oxford: Oxford University Press.
- Raz, J. 2015. "Normativity: The Place of Reasoning", *Philosophical Issues* 25, 144-164.
- Resnik, M. D. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis: University of Minnesota Press.
- Robertson, S. 2009. "Introduction: Normativity, Reasons, Rationality" u: S. Robertson (ur.), *Spheres of Reason: New Essays in the Philosophy of Normativity*, Oxford: Oxford University Press, 1-28.
- Scanlon, T. 2007. "Structural Irrationality", u: G. Brennan, R. Goodin, F. Jackson i M. Smith (ur.), *Common Minds: Themes from the Philosophy of Philip Pettit*, Oxford: Oxford University Press, 84-103.
- Searle, J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. New York: Cambridge University Press.
- Smith, M. 2004. "Humean Rationality" u: A. R. Mele i P. Rawling (ur.), *The Oxford Handbook of Rationality*, Oxford: Oxford University Press, 75-92.
- Smokrović, N. 2004. *Priroda prirodnog zaključivanja*. Zagreb: Hrvatsko filozofsko društvo.
- Streumer, B. 2010. "Practical Reasoning" u: T. O'Connor i Constantine Sandis (ur.), *A Companion to the Philosophy of Action*, Wiley-Blackwell, 244-251.
- Sušnik, M. 2009. "Internalizam razloga: dvije interpretacije", *Filozofska istraživanja* 114, 349-363.
- Williams, B. 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, B. 1985. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.



DANILO ŠUSTER

## A Mid-Blue Logic

As a rule, brighter colors are more casual than somber ones. A mid-blue suit, for example, is less formal than one in navy blue or charcoal gray. <https://www.gentlemansgazette.com/the-formality-scale-how/>, accessed October 4<sup>th</sup>, 2021.

Philosophy is that discipline which takes the abnormal to be the norm, e. g. hallucinations in perception, logically valid arguments in logic, neurotics in freedom of the will, and heroic self-sacrifice in ethics. [*Philosophy: A Commonplace Book*. Society for Philosophy & Culture: Wellington, NZ].

**Abstract:** I discuss Smokrović's work on the normativity of logic (Smokrović 2017, Smokrović 2018). I agree that the classical formal logic is not an adequate model for real-life reasoning. But I present some doubts about his notion of deductive logic and his proposal to model such reasoning in non-monotonic logic. No branch of formal logic by itself is likely to capture real-life inferential links (reasoned-inference). I use the logic of relevance as my case study and extend the pessimistic morals to modern systems of non-classical logic. Finally, I propose a more lax conception of normativity: there is a connection between logical assessment in the broad sense (as sanctioned by the notion of cogency) and the evaluation and criticism of reasoning.

**Key words:** normativity, non-monotonic logic, inference, relevance, consequence having, consequence drawing, informal logic.

### 1

For three decades or so I follow the work of Nenad Smokrović on the nature of human reasoning, with more intensive discussions in the last couple of years when we cooperated in the research project where he was the principal investigator. The title was characteristic for his work and dilemmas in this area: "Rationality: Between Logically Ideal and Commonsensical in Everyday Reasoning." We share interest in "real-life" reasoning and are both concerned about the usefulness and scope of classical logic in providing tools for the analysis and assessment of real-world reasoning. He insists on formal logic (or "deductive in a broad and weak sense"), I am

more open to the toolbox of logical assessment developed by the so called *Informal Logic Initiative*<sup>124</sup> (ILI). I think that the approaches are not or need not be exclusive.

I will discuss his latest and important work in this area (Smokrović 2017, Smokrović 2018) and present some critical remarks. One ought to agree with his view that every-day reasoning can assume different forms that are guided by different goals but the mainstream mathematical logic is not an adequate model for human argument and inference. Still, I have some doubts about the proposal to model such reasoning in logic which is non-monotonic and not strictly truth-functional but nevertheless *deductive* (Smokrović 2018: 459). This reminds one of the “chauvinism” of classical logic, sometimes ascribed to MacIntyre and sometimes to Sellars: “All inference is either deductive or defective.” The core of Smokrović’s views can be expressed by the following argument:

1. We should accept *logicism*, the claim that there is some connection between logical validity and the evaluation and criticism of reasoning.
2. But real-world reasoning in natural language is often probabilistic, context-dependent and content-sensitive.
3. Therefore, the appropriate way to model human reasoning is via deductive, although not classical logical systems (probabilistic, defeasible, non-monotonic ...).

Let me quickly summarize my views and express some hesitations. I basically agree with Smokrović – there has to be a connection between logic and the evaluation and criticism of reasoning. But I have doubts about the adequacy of *formal* validity. No branch of formal logic by itself is likely to capture real-life inferential links. I use the logic of relevance as my case study. Relevance logic was a project to reform (classical) entailment and offer a *realistic* theory of deductive reasoning. I think that the project failed, and the moral of this failure generalizes to contemporary defeasible, non-monotonic, default, auto-epistemic, ... formal systems discussed by Smokrović. Consequently, I present some considerations against (3). They are not decisive, but I propose a more lax conception of normativity: there is a connection between logical assessment in the *broad* sense (as understood by ILI) and the evaluation and criticism of reasoning.

---

<sup>124</sup> To use the description and abbreviation proposed by Koszowy and Johnson (2018).

## 2

Let me call the *Harman challenge* the view that logic is irrelevant for reasoning. Logic does not issue explicit prescriptions about what one ought to believe or how one ought to revise one's beliefs. As a vivid example take the classic Marx brothers movie *A Day at the Races* (1937). Dr. Hackenbush (Groucho, actually a veterinarian, a horse doctor) makes a famous observation: "Either he's dead or my watch has stopped." He is taking the pulse of Stuffy (Harpo Marx) who is jumping up and down on the chair. It is funny, because Stuffy is obviously alive, but also bewildering. If he really is dead, his pulse rate should be zero in, say, one minute. Not likely, given the circumstances. If the watch has stopped, his pulse rate should be, say, hundred and twenty in zero minutes – still not an indication that the person is dead. So let us assume that Groucho holds a pulsimeter and refers to this device as "a watch." As spectators we are led to believe that given the readings of the pulsimeter the person on the chair is dead (no beats). But the liveliness of this person contradicts the reading. The doctor's logical moves might then plausibly be:

- (1) The pulsimeter reads zero. (2) If the pulsimeter is reliable and it reads zero, then the person is dead. (3) This person looks alive. Therefore (4) either the pulsimeter is not reliable, or this person is dead.

The humorous effect is based on the fact that although the conclusion is *entailed* by the premises dr. Hackenbush is just deducing but not *inferring*. His set of beliefs *has* (4) as a consequence but (4) is not the conclusion that he should *draw*. MacFarlane (2020: 182-83) uses a very similar example. Once on a train he heard a young boy exclaim: "I have no pulse!" The boy may have also believed the conditional "If I have no pulse, I am dead." But he should obviously not apply *Modus Ponens* in order to come to believe "I am dead." He should reexamine his initial beliefs and interpret his initial observation as false. In a similar vein, dr. Hackenbush should dismiss the death of his patient as a serious doxastic option. According to Harman logic only tells you what a set of statements entails, it is not a theory of reasoning, a theory of "reasoned change in view." One often infers B from A because B provides the best explanation of A. In the case of dr. Hackenbush: From the alertness of this person I *infer* that the pulsimeter is not reliable as the best explanation of the evidence although I *deduce* the disjunction that the pulsimeter is not reliable or this person is dead.

An instructive way to map Harman's difference between deducing and reasoning is a contrast between consequence *having* and consequence *drawing*. Consequence-having occurs in logical space and consequence-drawing occurs in a reasoner's mind (Woods 2103a: 24):

Whether something is a consequence of a set of premises is wholly a matter of satisfying the requisite semantic conditions. Whether it is a consequence which it would be right (or necessary) to draw is partly a matter of semantics but also in large part tied to psychological factors.

Dr. Hackenbush is extracting the consequences but not drawing them. When a proposition  $Q$  is a consequence of something we believe,  $P$  (in the standard semantic sense, impossible for  $P$  to be true and  $Q$  false) this proposition is not always a candidate for drawing. Sometimes the right consequence to draw is that  $P$  is false (and thereby revise one's beliefs). According to Harman reasoning doesn't follow deductive principles (consequence having) and deductive reasoning (in the sense of deductive principles of consequence *drawing*) doesn't exist at all.

Smokrović disagrees with Harman, he argues that logic has a decisive *normative* role for reasoning. Reasoning can assume different forms that are guided by different goals. And each of these reasoning forms can be captured by a suitable logic: the normative standard for deductive reasoning is not (just) classical predicate logic. Deductive reasoning also includes reasoning in the conditions of uncertainty (modelled by probabilistic logic) and defeasible reasoning (modelled by default logic). This proposal immediately faces two questions: (1) What does *deductive* mean here? (2) Does this proposal offer an adequate reply to the Harman challenge?

### 3

First of all I am puzzled by the scope of *deductiveness*. The real-world reasoning is still described as deductive reasoning, to be evaluated by deductive logic according to the standards of logical validity. But deductive understood in “a weak and broad sense”, modelled by non-monotonic, probabilistic and default logic.

Smokrović rightly argues against a certain textbook picture of deductive logic as the norm of deductive reasoning. According to this view deductive logic equals classical predicate logic (CPL) combined with the soundness criterion of normativity. True premises and a valid form – the conclusion follows necessarily from the premises – is necessary and sufficient for the “goodness” of reasoning (argument). But this is a distorted picture of our everyday argumentative practice – we rarely infer with absolute certainty what follows necessarily from the available evidence. To quote from a recent book on deductive reasoning (Dutilh Novaes 2020: 19):

Indeed, in most practical real-life situations, the high degree of certainty afforded by deductive reasoning is not needed; in these situations, what we need to know is what is *likely* to follow from the available infor-



mation, given some background assumptions, such as that nothing abnormal is going on (the basic principle of some well-known non-monotonic logics [Stenning & van Lambalgen 2008] and default reasoning).

Dutilh Novaes is aware that human reasoning has a very strong component of defeasibility, *therefore* deductive logic is inadequate for modelling defeasible reasoning. She takes necessary truth preservation as a defining feature of deduction: the conclusion follows necessarily from the premises or is *entailed* by the premises, contrasted with inductive and abductive inferences, where the truth of the premises should make the conclusion more likely to be true (but it is still possibly false). Non-classical logical systems (probabilistic, default logics) that better suit defeasible real-life reasoning are standardly *not* classified as deductive logical systems. So, what could Smokrović (2018: 459) mean with: “Such logic is certainly deductive, although non-monotonic (initially assigned degree of probability to the conclusion may later be retracted in the face of new evidence) and not strictly truth-functional”?

I conjecture that he equates the domain of *deductive* with the domain of *formal*, an attitude well expressed by Burgess (2009: 2)

Logic, whether classical or extra- or anti-classical, is concerned with form. (On this traditional view of the subject, the phrase “formal logic” is pleonasm and “informal logic” oxymoron.) An argument is *logically valid*, its conclusion is a *logical consequence* of its premises, its premises *logically imply* its conclusions—three ways of saying the same thing—if and only if the argument is an instance of a logically valid form of argument.

Smokrović speaks about different kinds of validity grounded in different logics, but still validity remains a matter of *form*; one proposition is a consequence of others only if there is a valid pattern which the propositions together match. Non-classical logical systems are then formal deductive systems, where I will understand formal as formal<sub>2</sub> according to Barth and Krabbe (1982, 18). A system S is formal<sub>2</sub> when: (i) the (syntax of the) language to which S belongs is precisely formulated (the language is “formalized”); (ii) the validity concept in S is defined in terms of the forms of the sentences involved – it is a function of the definitions of the meanings of the logical constants concerned, and of the form of these sentences.

I propose to interpret “deductive in a weak and broad sense” in a very general and purely formal<sub>2</sub> sense, say: “a deduction is any sequence of statements each of which is derived from some initial set of statements (the

premises) or from a prior statement in the sequence.<sup>125</sup> The syntax of the language to which a statement belongs is precisely formulated or formalized. I think that the opposite of deductive in this sense is best understood as *informal* as in *informal* logic.

How successful is deductive in this broad sense as a *realistic* theory of deductive reasoning? I will address this question by discussing one of the earlier attempts to reformulate deductive logic as a *realistic* theory of deductive reasoning, the project of *relevance* logics. I think that the moral also applies to modern attempts to model real-life reasoning in formal systems, such as formal logics of dialogues and games, dynamic epistemic logic, defeasible logic, systems of default logic, auto-epistemic logic ... “FS\*” for short.

#### 4

Smokrović often stresses the fact that real-life reasoning in natural language is dynamic and non-monotonic (classical logic is monotonic and necessary truth preserving). An early model of a *deductive* (formal<sub>2</sub>) system that lacks the property of monotonicity was relevance logics (‘relevant logics’ in Britain and Australasia). These systems were initially developed as attempts to avoid the paradoxes of material and strict implication:

... the claim that ‘ $q \ \& \ \sim q$ ’ entails ‘ $p$ ’, in general, signals a breakdown in our intuitions not different in kind, though different perhaps in severity, from the kind of breakdown whose result is outright inconsistency, and similarly for the other paradoxes of implication, material or strict (Meyer 1971: 812).

But the ambitions were much higher, the aim of the project was to capture a true and correct formal counterpart of the *intuitive* notion of entailment. They often promoted a more encompassing “Aquarian” agenda of making and living in a better “world of reason.” What looked like purely technical logical results, one of the many specimen in the newly established zoo of non-classical logics, was promoted with the zeal of avantgarde manifestos against the tyranny of standard logic (Quine was a favourite target). According to Mares (2004: 3):

I suggest that what is wrong is that the standard notion of validity is too weak to provide a vertebrate distinction between good and bad arguments. It allows too many non-sequiturs to be classified as good arguments.

---

<sup>125</sup> Baker, Alan, “Non-Deductive Methods in Mathematics”, The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/mathematics-nondeductive/>, accessed October 4<sup>th</sup>, 2021.

Relevant logic was developed in part to avoid the so-called paradoxes of material and strict implication. The original sin, the paradigm “fallacy of relevance” is a classically valid inference that a contradiction entails everything and its twin, that a tautology is entailed by everything. In each case it is logically impossible for the premises to be true and conclusion false:

$$\begin{aligned} \textit{Ex falso quodlibet} \text{ (EFQ)} & (p \ \& \ \sim p) \supset q \\ \textit{Verum ex quodlibet} \text{ (VEQ)} & q \supset (p \vee \sim p) \end{aligned}$$

The objectionable thing about these paradoxes is that the antecedent and the consequent appear to be irrelevant to each other. There needs to be more of a connection between the content of the antecedent and the consequent in an implication if we are to say that the former implies the latter. Notice the close connection between “entailment” as characterizing the relationship between the premises and the conclusion in a valid argument and “implication.” Usually the connection is established via the deduction theorem (if a formula  $B$  can be derived from a set of formulas  $\Gamma$  together with a single formula  $A$ , then the implication  $A \rightarrow B$  can be derived from  $\Gamma$ ). For more orthodox tastes the connection defended by the relevantists is almost “incestuous.” Anderson and Belnap (1975: 473) retort provocatively : “the principle aim of this piece is to convince the reader that it is philosophically respectable to ‘confuse’ implication or entailment with the conditional, and indeed philosophically suspect to harp on the dangers of such a ‘confusion.’ (The suspicion is that such harpists are plucking a metaphysical tune on merely grammatical strings.)”

Relevance logicians have attempted to construct logics that reject theses and arguments that commit “fallacies of relevance”. The problem with EFQ and VEQ is that some of the premises of the inferences appear to have nothing to do with the conclusion. Intuitively, we require some kind of topic overlap between the premises and the conclusion, they should not have an entirely different subject matter. We also expect the premises to really *do* some “work” in establishing the conclusion of the inference. The relevantists propose two conditions as a formal capture of relevance between antecedent and consequent or premises and conclusion. First, the premises and the conclusion have to share some nonlogical content (a propositional variable)—this variable-sharing is a necessary condition for relevance. Secondly, they propose modifications of rules of inference for natural-deduction systems, all of the premises in an argument must really be *used* in the derivation of the conclusion—this condition is both necessary and sufficient.

Andersen and Belnap developed the logical system *E* as the system of entailment (relevant strict implication) and also formulated the nowadays standard logic *R* of relevant implication. Details are not important for our purposes, both systems are non-monotonic, since the premises really have to be *used* in a given derivation, one *cannot* move from:

- (1)  $A \vdash A$   
to  
(2)  $A, B \vdash A$

In the case of (2) *B* is doing no work. Still, *reflexivity* of the consequence relation expressed by (1) is declared to be unproblematic. Both systems incorporate the *The Law of Identity*: “ $A \rightarrow A$ .” According to Andersen and Belnap (1975: 8):

We take the law of identity to be a truth about entailment; it represents the archetypal form of inference, the trivial foundation of all reasoning, in spite of those who would call it “merely a case of stuttering.”

Yet “ $A$ , therefore  $A$ ” really looks like a poor candidate for “real” inference – the authors quote Strawson (1952: 15):

... a man who repeats himself does not reason. But it is inconsistent to assert and deny the same thing. So a logician will say that a statement has to itself the relationship [entailment] he is interested in.

And they offer a reply (Andersen and Belnap 1975: 8):

Strawson has got the cart before the horse: the reason that  $A$  and  $\sim A$  are inconsistent is precisely because  $A$  follows from itself, rather than conversely.

In a system *E* formula: “ $(A \rightarrow A)_{[\text{horse}]} \rightarrow \sim(A \ \& \ \sim A)_{[\text{cart}]}$ ” is a thesis but not vice versa (“ $\sim(A \ \& \ \sim A) \rightarrow (A \rightarrow A)$ ”) supposedly defended by Strawson. But let us have a look at what Strawson has to say about the cases of “follows from” (Strawson 1952: 14):

What is common to all the cases I refer to is the claim, signaled by the linking expressions, that it would be inconsistent to assert what precedes those expressions and to deny what follows them. The logician interests himself in cases in which this relationship holds between statements, irrespective of whether or not the transition from one statement to another so related to it is a transition which we should dignify by the name ‘step in reasoning’; irrespective even of whether it is something we should acknowledge as a transition.

A real “step reasoning” looks very similar to the core motivation for relevant logic, but immediately the differences appear (Strawson 1952: 14):

This explains why ‘study of the principles of valid deductive reasoning’ is too narrow a description of logic. A man who repeats himself does not reason. But it is inconsistent to assert and deny the same thing. So a logician will say that a statement has to itself the relationship he is interested in.

One could interpret Strawson as saying that “A, therefore A” does not correspond to *real* reasoning (merely repeating). Andersen and Belnap object that this inference does not *really* follow via EFQ from “A & ~A” according to their reformulation of “follows from.” But they still accept “A, therefore A” as the trivial foundation of all reasoning. It is no longer based on the notion of classical consequence (impossible for the premise to be true and conclusion false), but we have an exact topic overlap between the premise and the conclusion and the premise (repeated as a conclusion) is really *used* in the derivation of the conclusion. Nevertheless, we should ask with Strawson, does such relevance really deliver the “goods” promised? For in instance, “I argue that relevant logic is useful. It provides us with a theory of inference” (Mares: 2004, viii)?

It will be useful to understand relevant entailment “ $A \rightarrow B$ ” as “If A then B for that *reason*” (Burgess 2009: 114). And let us further assume that this locution encapsulates “a correct description of the basis of inference” (Read, 1988: 2). There are then at least two ways of understanding “for that *reason*” as an inference claim. In the first sense B is deducible in virtue of A, it follows logically from A in a deductive, rule-governed system. Such a derivation is an inference in a technical, or perhaps *thin* sense, an abstract codification of an inference conforming to a certain collection of syntactic rules. *The Law of Identity* is relevantly impeccable in this sense: A is deducible from A for *that* reason, derivable from A using A only. But there is also a different, more substantial, *thick* notion of inference as an episode of reasoning, inference understood in an epistemically serious way. According to Adler “reasoning is a transition in thought, where some beliefs (or thoughts) provide the ground or reason for coming to another” (Adler and Rips, 2008: 1). To infer B from A is then to take up, to accept, B as a result of reflecting on A (Rumfitt 2015: 35). *The Law of Identity* is *not* an inference in this sense: A is not grounded in A, there is no rational transition from A as a premise to A as a conclusion.

We thus have two notions of inference, inference in the derivational sense (“d-inference”) and inference in the reason-giving sense (“r-inference”). Corresponding to these two notions are then two notions of rele-

vance: *derivational* relevance (in this sense  $A$  is relevant for  $A$ ). And, secondly, *reasoned* relevance – in this sense  $A$  is not relevant for  $A$ : we do not accept  $A$  (conclusion) as a result of reflecting on  $A$  as the premise in the argument “ $A$ , therefore  $A$ ”. The premise is not giving the right kind of reason for the conclusion. According to Strawson “ $A$ , therefore  $A$ ” is not an inference at all. Woods (2004: 34) writes in the same spirit: “For “ $p$ , so  $p$ ” is always a fallacious inference (not to be confused with the correct and unexceptional entailment, “ $p$  entails  $p$ ”).” And the same is true (for Woods) of EFQ: a failure for inference, not for entailment.

The logic of relevance offers an improved interpretation of consequence having (eliminates EFQ etc.) as a model of real inference. But this is still *d*-inference only. Contemporary formal systems championed by Smokrović follow the lead: they aim to improve consequence having in the derivational sense in their attempts to adequately map the real-life reasoning. Harman thought that all such projects were doomed to failure. He identified argument with proof that is governed by the (deductive) rules of implication and contrasted this with reasoned change in view which is governed by rules of revision which he called “rules of inference”. He was *equally* dismissive of what he called AI logics as improved systems of implication (nowadays listed as non-monotonic, dynamic, (auto)epistemic, etc.):

But, although this terminology emphasizes the noncumulative character of reasoned revision, it is also potentially misleading in calling the ordinary sort of proof or argument “monotonic reasoning,” because proof or argument is not of the same category as reasoned revision (Harman 1986: 4).

Even dynamic or epistemic logics cannot model human reasoning:

It may be a mistake to expect principles of reasoning to take the form of a logic. In short, distinguishing reasoning from argument can make one suspicious of certain arguments for inductive logic, practical syllogisms, a logic of entailment, and so on. It is unclear how work on such “logics” might contribute to the study of reasoned revision (Harman 1986: 6).

Reasoning escapes the structure of logical implication, even non-monotonic (etc.) logics do not really model how human agents infer. It is easy to see why Harman is complaining: all logical systems remain the systems of consequence-having, they are not norms of real-life consequence-drawing. But I think that it is impossible to map (any kind of) consequence-having as real-life consequence-drawing, that aim was unrealistic from the very beginning. Even an enriched logical toolbox cannot solve the Harman

problem. Smokrović proposes a default logic incorporating the closed-world assumption (CWA), which is a variant of default logic. In closed world reasoning, it is assumed that any information that would weaken or cancel the premisses' connection to the conclusion is already contained in the premisses:

$$(p \ \& \ \sim ab) \rightarrow q$$

This is to be read as “If  $p$  and nothing abnormal is the case, then  $q$ ” (Smokrović 2018: 467). Take the case of the young boy on a train: “I have no pulse! If I have no pulse, I am dead. So, I am dead.” In this case the problem is not the abnormality of his deductions but the abnormality of his *reasoning* (were he to conclude that he is dead via *modus ponens*). The boy should infer that the antecedent is false – something *abnormal* is the case. But what conclusion, exactly, should the boy *draw*? Probably there was something wrong with the instrument, perhaps with his measuring technique, or even his ability to reliably detect the results. What consequence is the best thing to *draw* from the data and how to modify one's beliefs is not really or not just a question of logic but a broader question of epistemic rationality or even practical rationality in general.

Well, the boy should not draw the “I am dead” conclusion because such a conclusion is *not* to be accepted as a result of reflecting on the premises. The verdict should be that the premises are unacceptable. And here is my worry: can any kind of *formal* logic (or deduction in the weak and broad sense) model *r*-inference? Even a reformed formal logic maps derivational inference only, this is clear from the fact that the principle of *reflexivity* is a necessary feature of *any* respectable *formal* consequence relation, non-monotonic (FS\* systems) included. Informally, a sentence is a consequence of any set of sentences of which it itself is a member (Woods, 2013a: 228). This structural principle of consequence immediately yields the problematic *Law of Identity* ( $A$  is a consequence of  $A$ ), which is not a principle of real-life reasoning.

To summarize the historical lesson: Strawson is “degrading” entailment (explained in classical terms) – classical logic does not capture real reasoning, but, we could say, derivational consequence-having only. Andersen and Belnap are trying to confer dignity back to entailment: only if the conclusion “really”, i. e. relevantly, follows from the premises (or the consequent from the antecedent) do we have a case of entailment. The non-classical systems further improved the notion of consequence, but the *worry* remains: formal logic (FS\* systems) does not, as such, constitute a theory of *r*-inference; it supplies a theory of derivational inference only.



## 5

Reasoned-inference *should* guide us in our rational consequence-drawing, but there are different formal and informal methods of reasoning involved as the principles of belief-modification (Bayesian updating, inference to the best explanation, etc.). Smokrović (2018: 468) is right to stress the *normative* role of logic: "... for any form and goal of deductive reasoning there is an adequate normative system that can direct this reasoning toward the "rational" achievement of the goal." I think that the "directedness" introduced by the goals of reasoning is a right way to go as a reply to the Harman challenge (to bridge the gap between consequence *having* and consequence *drawing*) and the proper way to capture real-life reasoning. But I am doubtful whether any branch of formal logic is likely to capture inferential links in the reason-giving sense. Smokrović (2018: 458) is aware "that reasoning performed in natural language is not syntactically or extensionally valid but at best intentional" and he thinks that FS\* systems and in particular a default logic with CWA offer an adequate model of real-life reasoning.

Let me mention some hesitations which are not decisive (I am not an expert on CWA systems). First of all, consider the core schema: "If  $p$  and nothing abnormal is the case, then  $q$ ." According to Woods (2013a: 285) the problem is to determine whether the assumption of normality is ever actually available to the reasoning agent, to beings like us, the world is never closed. If so, what are the conditions under which we are justified to invoke the closed world-assumption? This is hardly a matter of a *formal* system or deduction in a weak and broad sense.

And secondly, Smokrović (2018: 456) states: "I'm embracing the view that norms can be applicable to those who apprehend them." This seems plausible: a form reasoning is normatively justified if it can be connected with a type of validity that the thinker can apprehend or recognize as valid, where "validity" is to be understood in the weak and broad sense. The contemporary *formal* proposal is to improve consequence-having as a model of real-life logic. The project is implemented by *enhancing* the formal power and reach of logic, by developing a "heavy equipment" logic (cf. Woods 2013b). Just have a look at the *Appendix* of the paper on the logic of human reasoning: "Kleene 3-valued procedural semantics for logic programs in non-monotonic reasoning, based on models for definite programs, as fixed points of a three-valued consequence operator" (Varga, A., Stenning, K., and Martignon 2015). The formal way to make logic right for real-life reasoning seems to be by *complexifying* the logic's mathematical structure. But is such a structure really apprehensible for an everyday thinker? And

this is my second worry: when you insist on mathematically precise formulation and implementation of reasoning you risk the loss of apprehension required by the requirement of normativity.

Of course, the structure of apprehension is more complicated, Smokrović quotes approvingly MacFarlane (2004: 22):

My own view is that apprehension should not be intellectualized to the extent that it requires a completely explicit understanding of what an inference schema is, the kind one would get from an encyclopedia article on the subject. It is something more basic than that. But it is important that apprehension be something for which one can take responsibility and give or receive criticism.

The understanding of the *Appendix* on 3-valued procedural semantics is apparently not required to follow the requisite norm in your reasoning. Still, one apprehends an inference as an instance of inference schema if one is responsible in the sense that one intends to infer according to this scheme (Smokrović 2018: 466). If (full?) apprehension of an inference schema is required I remain doubtful about the normative legitimacy of heavy-equipment models of human reasoning.

These are sketchy remarks, they are not decisive, but they all point into a certain direction, beyond the walls of formality. Hartry Field, whose writings are quoted approvingly by Smokrović points into the same direction. According to Field an adequate normative system is not a system based on the rules of necessary truth preservation. He proposes to redefine validity, not as (necessarily) preserving truth in general but as (necessarily) doing so “when it matters” (Field 2009: 266). And a rule “preserves truth when it matters” if it preserves truth when applied to premisses that can be established or are rationally believable. This characterization will (presumably?) exclude *The law of Identity* as a degenerate instance of truth preservation since the premise (equivalent to the conclusion) will no longer count as acceptable. Field’s *formal* criterion (acceptable premisses and rules which preserve truth when it matters) comes surprisingly close to *cogency*, the basic normative notion of *informal* logic, cf. Govier (2018: 287-88): “If the premisses of an argument are rationally acceptable and are ordered so as to provide rational support for the conclusion, the argument is cogent.”

## 6

Roughly at the same time as relevance logic the *Informal Logic Initiative* emerged when many philosophers and logicians turned their attention to the analysis, evaluation and improvement of real life argument. Clas-

sic formal logic turned out to be an inappropriate educational tool for the analysis of real arguments. According to Scriven (1980: 147) it should be removed to the monsters part of the “academic zoo”:

It’s not good for children to see too much of the monsters part of the zoo; ... They grow up into poor little perverts who ... mutter things like “ $p$  is true if and only if  $p$ ,” then smile beatifically. Or they go around chanting “A false proposition implies any proposition; yes it does, yes it does ...”

ILI was motivated by the dissatisfaction with ‘the soundness doctrine’ or the normativity of classical logic. Classically valid form (necessary truth preservation) plus true premises is neither necessary nor sufficient as a criterion of a good argument. There are good arguments that are not sound (usually classified as *inductive*) and there are sound arguments that are not good arguments: any circular argument with a true premise. Johnson and Blair (1977) identified standards for the evaluation of an argument based on the leading question: what standard is violated in this or that traditional fallacy? They proposed a so called “RSA” criterion of *cogency*: in a good argument acceptable (A), individually relevant (R) and jointly sufficient (S) premises rational support the conclusion. In contrast to classical soundness, requiring valid arguments with true premises, the RSA criterion emerged as the central normative notion in the approaches that remain closer to the practice of argumentation.

It is clear that “A, therefore A” is a bad argument and no one would take it seriously. Sometimes the verdict is that it does not meet the RSA criteria, since its premise would not count as *acceptable* (Blair, 2012: 88). This is in line with a classical objection to *petitio*: a statement (premise) is made that presupposes or depends upon the point at issue (conclusion) and such a premise is judged to be unacceptable. But I find the Strawsonian objection more principled: “A, therefore A” is a failure of inference. This diagnose is based on the thick, reasoned conception of inference (A is not a reason to accept A). The relevance of premises and their sufficiency pertain to the adequacy of the inferential link: the reasons offered must be probatively relevant to the conclusion and they have to be sufficient for accepting it. The relevance criterion requires that the proposition  $P$  actually plays a supporting role for  $C$ , the premise  $P$  counts in favor of the truth of  $C$ . In *The Law of Identity* the premise is then not relevant, it offers no support or no grounds for the conclusion. Something of epistemic value (knowledge, justification, conviction ...) fails to be transmitted from premises to conclusion. An inference proper is “directed,” so to speak, there is a certain initial epistemic asymmetry between premises and conclusion. McKeon

(2015: 319) for instance proposes a very thick conception of inference: “if I infer  $q$  from  $p$ , then my belief that  $p$  explains, at least in part, why I believe  $q$ .” I would claim, perhaps more cautiously, that a proposition which entails itself is not relevant to establishing itself, so *The Law of Identity* fails as an inference.

Cogency as a central term of logical assessment covers different forms of reasoning (deductive, inductive and also a third type of connection between premises and conclusion, sometimes conceptualized as *conductive*). It incorporates broader epistemic norms, but the appraisal of arguments as based on the structure of reasoned-inferences is perhaps really closer to epistemic appraisal than to formal-logical evaluations. This was already the lesson from the Harman challenge. Informal logic includes directedness of reasoning guided by goals (emphasized by Smokrović) but also a dialectical dimension of reasoning. A defining feature of informal logic is its emphasis on arguments, not as abstract objects (ordered pairs premises - conclusion) but as arguments in *use*, arguments in the sense of argumentation as an interactive social process. Smokrović rightly observes that “reasoning can assume different forms that are guided by different goals.” He mentions goals such as proving the theorem, showing that an accused is guilty beyond any reasonable doubt, or coming to the conclusion about the whereabouts of a certain bus (Smokrović 2018: 457). He does *not* mention, however, the clearest case of goal-directedness: a piece of reasoning embedded in an *argument*. A type of discourse in which the arguer expresses a point of view (the conclusion) and offers one or more reasons (the premise(s)) in support of the conclusion. The premises are advanced as reasons in support of the conclusion. Arguments in this sense defy d-inferences and also introduce another, dialectical and pragmatic dimension, not easily captured by formal tools. Even Dutilh Novaes (2020), who takes necessary truth-preservation as a defining feature of deduction, explains the sources of deduction in dialectical exchanges and practices of debating which then evolved into deductive argumentative practices (cf. also Smokrović 2017).

Finally, consider again the argument we started with:

1. We should accept *logicism*, the claim that there is some connection between logical validity and the evaluation and criticism of reasoning.
2. But real-world reasoning in natural language is often probabilistic, context-dependent and content-sensitive.
3. Therefore, the appropriate way to model human reasoning are deductive, although not classical logical systems (probabilistic, defeasible, non-monotonic ...).

Normativity is a hot topic of theoretical discussions and (1) was famously questioned by Harman. I think that something like: “If *B* ‘really’ follows from *A*, then, if we accept *A*, we should accept *B* as a result of reflecting on *A* (as grounded in *A*)” should be a guiding principle in this area. Is this still *logicism*? Perhaps logicism in the weak and broad sense which includes *cogency*:

(1)\* We should accept the view that there is some connection between logical assessment in the broad sense and the evaluation and criticism of reasoning.

I have also raised some considerations against the formal conception of deductiveness operative in (3), so I would suggest instead: “we therefore have to model human reasoning with formal but also non-formal approaches.”

I understand the lesson from relevant logic to be that the project of incorporating relevance into the logical notion of validity in a formal way (d-inference) fails. We do get a better model of consequence-having (implication, entailment) but this is not enough to map the real-life reasoning based on reasoned-inferences. Even Andersen and Belnap realized that the project of relevance was too ambitious. Later they go more low-profile, they argue that relevance logic is preferable to classical logic for extracting information from a database that might contain inconsistent information (MacFarlane 2020: 187). I suspect that the moral generalizes to contemporary non-classical formal systems. I might be wrong however, prediction is very difficult, especially about the future, allegedly said Niels Bohr. It is safe to say that the logic for AI systems, dynamic logic, auto-epistemic logic, etc. will further improve the notion of consequence having and may even approximate real-life reasoning. Perhaps advanced formal systems which include goal-directedness might constitute a viable theory of inference.

But in general I have doubts about the attempt to make logic right for *argument* by complexifying logic’s mathematical structure. Admittedly, as a normative notion cogency turns out to be a loose evaluative concept. Smokrović stresses that we are engaged in various forms of reasoning, accomplishing different goals, we economize with our cognitive resources, “computational efficiency is an opportunity cost of expressive power” (Smokrović 2018: 457). But the same is true of logicians in their meta-logical theorizing about normative systems of reasoning. The rebellion of a new generation of logicians against the cliches and worn out admonitions of classical logic, in particular its treatment of implication and logical consequence was described as logic of “the Age of Aquarius” (Meyer 1971: 808). Astrologers apparently do not agree on when the Aquarian age will start or even if it has already started. Most published materials on the sub-

ject state that the Age of Aquarius arrived in the 20<sup>th</sup> century (29 claims), with the 24<sup>th</sup> century in second place with 12 claimants.<sup>126</sup> Well, from the perspective of the 21<sup>st</sup> century it looks that logic in the Age of Aquarius is and will continue to be pluralistic, including more *casual*, less precise and more fluid but consequently more flexible and more “real-life” approaches of informal logic.

## Literature

- Adler, J. E., & Rips, L. J. (2008). *Reasoning: Studies of Human Inference and Its Foundations*. Cambridge: Cambridge University Press.
- Anderson, A. R., Belnap, N. D. (1975). *Entailment: The Logic of Relevance and Necessity, Volume I*. Princeton: Princeton University Press.
- Barth, E. M., Krabbe, E. C. W. (1982). *From Axiom to Dialogue*. Berlin/New York: De Gruyter.
- Blair, J. A. (2012). *Groundwork in the Theory of Argumentation: Selected Papers of J. Anthony Blair*. Dordrecht: Springer.
- Burgess, J. P. (2009). *Philosophical Logic*. Princeton and Oxford: Princeton University Press.
- Dutilh Novaes, C. (2020). *The Dialogical Roots of Deduction: Historical, Cognitive, and Philosophical Perspectives on Reasoning*. Cambridge: Cambridge University Press.
- Field, H. (2009). “What is the Normative Role of Logic?” *Aristotelian Society Supplementary Volume 83* (1), 251–268.
- Govier, T. 2018. *Problems in Argument Analysis and Evaluation*. Windsor Studies in Argumentation Volume 6. Windsor Ontario Canada.
- Harman, G. (1986). *Change in View: Principles of Reasoning*. Cambridge, Mass.: The MIT Press.
- Johnson, R. H., Blair, J. A. 1977. *Logical Self-Defense*. Toronto: McGraw Hill-Ryerson.
- Johnson, R. H., Koszowy, M. (2018). “Logical Culture as a Common Ground for the Lvov-Warsaw School and the Informal Logic Initiative.” *Studies in Logic, Grammar and Rhetoric*, 55 (1), 187–229.
- MacFarlane, J. 2004. “In What Sense (If Any) Is Logic Normative for Thought?” Unpublished manuscript. Presented at the Conference on the Normativity of Logic organized by Central Division APA 2004.
- MacFarlane, J. (2020). *Philosophical Logic: A Contemporary Introduction*. London: Routledge.
- Mares, E. D. (2004). *Relevant Logic: A Philosophical Interpretation*. Cambridge: Cambridge University Press.
- McKeon, M. W. (2015). “Inference, Circularity, and Begging the Question.” *Informal Logic* 35 (3), 312–341.

<sup>126</sup> [https://en.wikipedia.org/wiki/Age\\_of\\_Aquarius](https://en.wikipedia.org/wiki/Age_of_Aquarius), accessed October 4<sup>th</sup>, 2021.

- Meyer R. K. 1971. "Entailment." *The Journal of Philosophy* 68, 808-818.
- Read, S. (1988). *Relevant Logic: A Philosophical Examination of Inference*. Oxford: Blackwell.
- Rumfitt, I. (2015). *The boundary stones of thought: An essay in the philosophy of logic*. Oxford: Clarendon Press/Oxford University Press.
- Scriven, M. (1980). "The Philosophical and Pragmatic Significance of Informal Logic." In J. A. Blair & R. H. Johnson (Eds.), *Informal Logic, The First International Symposium* (pp. 147–160), Inverness, CA: Edgepress.
- Smokrović, N. (2017). "Normativnost logike in vsakdanja sklepanja" (*Normativity of logic and everyday reasoning*). *Analiza* 21 (2017), 45-61.
- Smokrović, N. (2018). "Informal Reasoning and Formal Logic: Normativity of Natural Language Reasoning." *Croatian Journal of Philosophy* 18 (54), 455-469.
- Strawson, P. F. (1952). *Introduction to Logical Theory*. London: Methuen.
- Varga, A., Stenning, K., and Martignon, L. (2015). "There Is No One Logic to Model Human Reasoning: the Case from Interpretation." In U. Furbach and C. Schon (eds.). *Proceedings of the first workshop on bridging the gap between human and automated reasoning*. Berlin: 32–46.
- Woods, J. (2004). *The Death of Argument: Fallacies in Agent Based Reasoning*. New York: Springer.
- Woods, J. (2013a). *Errors of Reasoning: Naturalizing the Logic of Inference*. London: College Publications.
- Woods, J. (2013b). "Advice on the Logic of Argument". *Revista de Humanidades de Valparaíso*, 28.



MAJDA TROBOK

## The Role of Argumentation<sup>127</sup> In Honour of Nenad Smokrović<sup>128</sup>

**Abstract:** In this paper I will concentrate on the notion of argumentation and discuss the approach that Smokrović takes with respect to the properties, the role and the importance of argumentation.

**Key words:** argumentation, group reasoning thesis, truth.

In his papers (Smokrović 2015; 2017), the endorsed basic view concerning argumentation amounts to the idea that argumentation is a specially good means for achieving the truth and for the extension of knowledge. Smokrović takes argumentation to be "the form of communication consisting of the claim, reason(s) (or evidence) supporting it and the relation between the claim and the reasons, where the addresser produces a claim and reasons while the addressee evaluates it" (Smokrović 2015: 226). It is a social, two-sided (usually informationally asymmetric) temporal process in which two sides<sup>129</sup> enter, each with some initial stock of beliefs,<sup>130</sup> and

---

<sup>127</sup> Research for this paper was carried out under the project "Critical thinking and Society: Education, Science, Politics and Religion", funded by the University of Rijeka. uniri-human-18-254.

<sup>128</sup> Nenad Smokrović and I have been collaborating and discussing through the years on many topics both in the philosophy of logic and mathematics. He has also been either a formal or an informal reviewer of virtually everything I have written so far. I am hence both glad and honoured to write a paper-contribution in honour of his work.

<sup>129</sup> In an argumentative process we have, according to Smokrović, basically two parts - the addresser and the addressee. We might wonder whether there can be an argumentative process in which a person is both the addresser and the addressee? (in the same way in which, e.g. we might play chess against ourselves.) Another possibility is the case in which we argue against someone's theory/idea/view written in a paper/book - could such a situation be taken as an example of a, non-standard, argumentative process? Smokrović does not take such cases into consideration, we could however take them as possible example of argumentative processes since they might fulfil the posed conditions for a process to be argumentative.

<sup>130</sup> Before entering a discussion focused on analysing the truth value of a given statement S, we might have several different epistemic options. First of all, we might either be convinced that the statement S we defend is true or not. In the former case, we might even be convinced of having the proof of S (i.e. an argument that should show that S is true) or we might just be convinced S being true without (for the moment) good evidence (this could be the case in many disparate scenarios. Let us just remember Ramanujan and his conviction of the trueness of certain mathematical statements be-

through the process consisting of producing and evaluating arguments, each of the, eventually reach the point where they know more than they did before. (Smokrović 2015: 225)

The AP (argumentation process) is structured in a way that supports participants 'epistemic curiosity to acknowledge whether a certain proposition  $p$  is true.

AP hence being a specific form of communication, a social activity and a temporal process in which each of the (usually two) involved sides "reach the point where they know more than they did before" (Smokrović 2015: 225). One problem that often arise during the argumentation process is the confirmation bias, i.e. the situation in which someone affirms and beliefs that  $P$  and hence has no motivation to find out wether  $P$  is false. Smokrović, interestingly, offers a way to avoid the confirmation bias when it appears in the argumentative process. The remedy is what Williamson describes in the following lines: "...sometimes we step back of our beliefs and regard them as psychological phenomena on a par with the belief of others, in equal need of both psychological explanation and epistemological criticism" (Williamson 2007: 247). Smokrović, then, goes one step further in aiming to show that the argumentation case is precisely the one in which we are incited to "step back" in the sense described by Williamson. In proposing the adequate epistemic theory for the idea of extension of knowledge, Smokrović opts for the Williamson's knowledge founded on the safety principle, i.e. on the concept of safe avoidance of error. He then applies elements of Williamson's theory (Williamson 2000: 9) to the argumentation process and formulates the following: "S safely believes  $P$  in situation  $\alpha$ , if, using method  $M$  in time  $T$ , S truly believes  $P$  and could not easily have been wrong in similar situations  $\beta$ ." (Smokrović 2015: 231).

Smokrović takes as the basic statements of his theory Sperber and Mercier's advantage of group reasoning thesis (AGRT) and their evolutionary thesis (ET) (Mercier and Sperber 2011; 2012). The former (AGRT) claiming that "groups do better at reasoning tasks than individuals, and, *in some cases* (italics by Trobok), do even better than any of their individual members (occasionally even better than theory best member)". The latter (ET) claiming that "reflective reasoning has been designed by evolution as a communicative competence (rather than aiming at enhancing individual inference)" (Smokrović 2015: 224).

---

cause Goodness Namakiri told him so in his dreams). In the latter case, we might not be convinced the statement  $S$  being true but presuppose (or have an intuition about) the statement  $S$  being true and in need of (further) evidence/proof. Let us furthermore suppose that we enter an argumentative process with an interlocutor. The interlocutor (or addressee) might, as it is the case with us, be in any of the above mentioned epistemic states once we introduce the statement  $P$  in the discussion.

I find Smokrović's theory most compelling and interesting. And I agree with most of the statements he defends. I would like, however, to provoke a further discussion on the topic by attempting to point to some potential difficulties with his theory. We shall, hence, look into the theory in more details.

Let us start by focusing on the AGRT (advantage of group reasoning thesis). It might appear to be a variant of the famous Tuscan proverb: A wise man and a fool together, know more than a wise man alone (*Sanno più un savio e un matto, che un savio solo*). The Tuscan proverb being, however, more straightforward than the AGRT, since it is truth in virtue of set-theoretical/mathematical relations only. If we take what someone/a wise man knows to be a set  $W$  of statements, than what a wise man and a fool know is the union of the set  $W$  of statements that a wise man knows and the set  $F$  of statements that a fool knows. In this case it holds that:  $k(W) \leq k(W)+k(F)$ , where  $k$  stands for the cardinality of set(s) of statements and hence the number of statements that someone knows. As for the AGRT, it deals with the reasoning tasks that is supposed to be such that groups do better than individuals. It appears that determining if the AGRT is true is much more difficult than determining the truth if the Tuscan proverb. If anything, it is not clear at first what should be regarded as "doing good/bad at reasoning tasks," one might assume it means making fewer inference errors or choosing the applied deduction rules more elegantly or avoiding more often the informal fallacies or being able to determine more logical consequences (or the more appropriate ones) from the premisses or a certain combination of all mentioned. If, however, we look at the way Sperber and Marcier (2011: 65) defines reasoning, a different picture appears:

The mental action of working out a convincing argument, the public action of verbally producing this argument so that others will be convinced by it, and the mental action of evaluating and accepting the conclusion of an argument produced by others correspond to what is commonly and traditionally meant by reasoning (a term that can refer to either a mental or a verbal activity).

The keyword in this definition is: *convincing*. Reasoning is about determining *convincing* arguments, i.e. arguments that will *convince* others. Doing good at reasoning tasks could hence mean offering convincing arguments. If so, than determining whether someone is more or less good at reasoning tasks should involve determining if the proposed arguments are convincing to the addresser(s). Such a process is certainly a function (at least) of time: what might have been taken to be a convincing argument a hundred

years ago, might not be convincing today.<sup>131</sup> So far so good. However, being good at reasoning tasks appears not to have clear meaning outside the temporal determination and possibly social context. Why is that a problem? Because such delineation of reasoning and argumentation appears to be distant from Smokrović's .

The keyword "convincing" is missing from Smokrović characterisation of the argumentation process, he rather concentrates on developing arguments aiming to determining whether a certain proportion  $p$  is true ( $p$  being the object of discussion). I wonder whether the discrepancy in the core thesis between Smokrović's theory on one hand and Sperber and Mercier's on the other might produce a much more severe discrepancy along the way. Namely, Smokrović is jointly asserting the AGRT (advantage of group reasoning thesis) and the view that in the argumentation process two sides driven by the propositional curiosity are motivated to do their best to find out whether a proposition  $p$  is true (Smokrović 2017: 85). The problem might arise given that in the domain of "convincing others" and producing "convincing arguments" we might get lost in the plethora of possible situations in which someone who has both produced outstanding arguments and has proved a statement  $p$  to be true may not be convincing at all.<sup>132</sup> The statement that group does better than individuals, given the above characterisation of reasoning, might be construed in a way that is (too) distant from what Smokrović is focusing on and gives credit to. His insisting on reaching the truth might not be on the line of what Sperber and Mercier are concentrated on. Namely, to be convincing need not to coincide with reaching the truth - we already mentioned two examples showing the dis-

---

<sup>131</sup> An example might be Aristotle's theses concerning physics or zoology that were convincing for a time, but not today. For more details see (Ramachandran V. S.; Rogers-Ramachandran, D. 2010) Another example might be the one concerning the arguments in defence of the Euclidean geometry as the only possible one. Such arguments were convincing for more than 20 centuries, until the beginning of the 19th century. For more details see (Bonola 1911/2012)

<sup>132</sup> We might mention the well known case of the German mathematician Georg Cantor who was unconvincing to (most of) his colleagues when presenting the diagonal procedure, i.e. the proof that there are more real than rational numbers. For more details see (Boolos 1997). Or, to give a medical example related loosely to covid, we might mention the case of Ignaz Semmelweis. He was a Hungarian doctor of the mid 19th century who was practicing in Vienna, and who was unable to convince his colleagues to wash their hands when switching from doing a most mortem in the morning to assisting at child-birth in the afternoon, notwithstanding his good arguments and empirical proofs. For more details see (Potter 2001: 368).

These are just two, unfortunately not isolated, cases in which the addressers not just were unconvinced by good arguments but did actually try to virtually destroy (socially and psychologically) the addressees themselves.

crepancy between "being convincing" and "speaking truthfully." The ulterior problem lying in the possibility that the AGRT might not be correct, if we replaced Mercier and Sperber's focus on convincing the audience (i.e. the addresser(s)) with Smokrović's one on epistemic curiosity and truth determining. If this is to be the case, Smokrović might not have good reasons for adopting the AGRT in the first place. The rejection of the AGRT might, on the other hand, lead to the demand for draconian emendations in any theory that, like Smokrović's, takes AGRT to be one of its tenets.

Apart from the characterisation of the argumentative process given above (Smokrović 2015: 226), in (Smokrović 2017) he goes on in determining more details connected with the notion of argumentation. He narrows the notion of argumentation process by defining it exclusively as a process that can occur only if specific condition is fulfilled, i.e. when "the intrinsic curiosity<sup>133</sup> whether *p* is true takes the leading motivational role in the process". It means that "each of the participants is primarily interested (curious) in the truth of the claim. The addressee wants to avoid false beliefs by hypothesis, while the addresser, sincerely believing that *P* is committed to the belief that belief in *P* is true (because *P* implies that *P* is true" (Smokrović 2015: 227). There are, on the other hand, situations that we would call or find it natural to call argumentation process but would not fit in the conditions Smokrović insists upon. The first case that might pop up on our mind is the political debate. Two sides arguing about a statement/situation, having contrasting views and exchanging reasons for one or the other option. Not intending to extend knowledge, but rather to increase the number of votes in the forthcoming election. And once we insert of the riddle different possible intentions, there are many more argumentative options that standardly do not necessarily aim to increase knowledge. Even if we are not politicians and have the initial idea to sincerely make an effort to find out the truth, engaging in argumentation process open-mindedly is far easier said than done. Yet, we might still want and find it natural to say that we were engaged in argumentation process.

Smokrović himself is very well aware of those possible situations that in the literature are often labelled as argumentative processes. In his posing the question should the argumentative process be metaphorically comparable more to a boxing match or rather to dancing in pairs, he undoubtedly opts for the latter and makes it clear that:

---

<sup>133</sup> Smokrović accepts curiosity to be "the disposition to have desires of having true answers and, accordingly, to motivate the pursuit of their fulfilment." (Smokrović 2017: 83)

The AP (*argumentative process - note by M. Trobok*) in the generic sense covers an area of a huge variety of motives and goals that drive participants to engage in the AP, and accordingly, the types of action participants can perform. ...participants might be driven by the plethora of other private motives. Many of them can be purely practical or pragmatic: to win in the discussion, to impress the audience, to check's one's rhetoric skill, to outwit the audience and the like. ...Which motives and goals will prevail and, accordingly, which form the AP may assume is, of course, an empirical question. ...it is plausible to suppose that, depending on the motives and goals of the participants, the AP can take art in one the the poles in the continuum, traditionally divided to two parts, "rhetoric" and "dialectic" part of the continuum. (Smokrović 2017: 94)

In footnote 2 (Smokrović 2015) he makes it clear that the situation in which the addresser does not believe P even though he's claiming it (the addresser believes that P is false and just want to deceive the addressee) is not counted as being a situation within an argumentative process. Further more, within this initial condition, the process itself is confined to very specific situations. The situations being those in which the addresser both sincerely believes that P and that he has good grounds for it, while the addressee sincerely make an effort to find out whether P. They are both genuinely interested in finding out whether P is true.

By confining the notion of argumentation and imposing very specific constraints, Smokrović makes the way clear for implying that argumentation is a means for extending knowledge. Once Smokrović narrows the notion to a certain extend, the aspect of knowledge extension follows from the characterisation of argumentation he starts from. Namely, to say that argumentation is "a social, two-sided (usually informationally asymmetric) temporal process in which two sides enter, each with some initial stock of beliefs, and through the process consisting of producing and evaluating arguments, each of them eventually reach the point where they know more than they did before" (Smokrović 2015: 225) is, in other words, to say that the result is extending knowledge. If so, that the argument presented is certainly valid, but a *petitio principii*: the statement that argumentation process leads to extending knowledge is already presented in the premiss, i.e. in the definition of AP that Smokrović presents.

Once we accept Smokrović notion of argumentation, it follows rather smoothly that in such situations one is incited to avoid the confirmation bias. Given that the situations in which e.g. the addressee is, by definition, such that when confronted with the addresser's evidence, even though that contradicts his intuitions or beliefs, his goal is still determining the truth.

He then, by definition, does not try to discredit the interlocutor and the proposed arguments.

In the literature, different are the ways of defining/characterising argumentation. What the authors mostly agree on is the argumentation's layout, taking it to be a specific sort of dialogue. Walton defining it as "a chain of arguments, where the conclusion of one inference is a premise in the next one", the argumentation approach being dialogical (or dialectical) in the sense that "it looks at two sides of an argument, the pro and the contra. According to this approach, the method of evaluation is to examine how the strongest arguments for and against a particular proposition at issue interact with each other, and in particular how each argument is subjects to probing critical questioning that reveals doubts about it. By this dialog process of pitting the one argument against the other, the weakness in each argument are revealed, and it is shown which of the two arguments is the stronger." (Walton 2009: 2) the interaction of different arguments for and against some conclusion,

Hintikka promotes the idea of considering "...argumentation as a question-answer sequence, interspersed by logical (deductive) inferences" (Hintikka 1996: 307-308). Argumentation only occurs when, upon making a claim, someone receives a request for further support for the claim in the form of reasons, or estimates herself that further justification is required (Jackson & Jacobs 1980; Jackson, 2019). In such cases, dialogues of "giving and asking for reasons" ensue (Brandom, 1994; Bermejo Luque 2011). What is common in these characterisation is the absence of interest in the underlying motivation of those involved in the argumentation process. What has been analysed is the outcome of the process.

At this point the concern is that Smokrović's characterisation of argumentation process is too narrow and comprises just a very specific and limited argumentative process 'situations while the justification or motivation for opting for such a narrow definition might be in need for (additional) clarification and motivation.

So, two are the factors that we are going to put into relation and see if and when their conjunction is or might be problematic.

Firstly, according to Smokrović, determining the truth (of a statement) is the central point of interest when talking about the AP (argumentative process). His ideal situation is the one in which "each party is doing whatever is in their intellectual power to find out whether that which is claimed is true" (Smokrović 2015: 229).<sup>134</sup>

---

<sup>134</sup> The underlying truth theory consistent with such a view is, at least, realism in truth-value, possibly in ontology too. Realism in truth value being the view according to which statements have objective truth value independently of our mind, language



On the other hand, a process can be called argumentative if and only if the participants are driven by curiosity and truly interest in determining the truth, e.g. if P is true or not (and it holds for any P in question) Truth-seeking is one of the most relevant features and "behind [...] practical goals, there is a more fundamental one. It is the epistemic goal of acquiring knowledge" (Smokrović, *ibidem*)

It might seem natural to combine these two traits in a theory: the focus on determining the truth and, in addition, the interest in having those involved in a AP curious and genuinely and exclusively interested in determining the truth.

This certainly would be the optimal option. The potential problem consists in Smokrović's request that it is to be the only acceptable when determining what an AP is and what it consists of. Namely, at this point someone might ask what justifies Smokrović's insistence on demanding for both the addresser and the addressee (i.e. everyone involved in a AP) to be driven by and only by curiosity as the central motivating epistemic virtue.

If I find a counterexample to someone's argument, how and why is my motivation in the context of the argumentation and truth determining process relevant? My seeking for counterexamples could be done out of boredom, or ego-feeding desire/need or could be based on my intrinsic truth-seeking motivation: whatever might be the case, it is not clear how any of the cases can influence the importance or role of my counterexample within the discussion or my interlocutor's argument.<sup>135</sup> Of course, if interested in the psychological aspect or the epistemic *modi* within an argumentation process, it is clearly relevant to investigate on the possible motivations and/or mental states of the interlocutors. But, if the core aspect is the search for the truth (of a proposition or set of propositions/theory) than what motivates a single interlocutor might be irrelevant. In the same way in which, e.g. we generally have no interest in the intrinsic motivation of our doctor (could it be the money he gets, the fame he wishes to obtain or his genuine interest in our health) as long as the doctor's diagnosis is correct and the suggested cure help us recover: After all, why should or would his intrinsic motivation be of any importance in the context of his

---

or conventions. They are true or false independently of our capacity to determine their truth value. For more details about the differentiation between the two see (Trobok 2006: 17-47).

<sup>135</sup> I confine myself to the cases in which the interlocutor is, as a matter of fact, interested in determining if *p* is true (without any intrinsic curiosity being the motivational force). It is, as Smokrović himself mentions (Smokrović, 2015, p. 225), possible that the interlocutor has no genuine interest in knowing if *p* is true while, e.g. trying to convince the interlocutor or the audience that *p* is true and hence win the argument - I don't think such cases are of some relevance in this discussion though.

successfully determining the diagnosis and helping us to recover? Obviously, the ideal situation is one in which our doctor is both successful and intrinsically interested in our health, but to say that just such doctors are the one included in the definition of the term might be too limited. Should we demand not to consider as ("true") doctors or "real cure" all those practitioners not intrinsically motivated in helping people, independently of their related great success in doing so? That is hardly the case. Doing otherwise might get the argument to start smelling of the "no true Scotsman" logical fallacy: once we demand that our doctors are intrinsically curious and sincerely interested in the health of their patients in order to say they are doctors, doctors by definition give the correct diagnosis and their cure is helpful; one might reply that a person can be greedy and still give the true diagnosis and right cure, in which case the answer would be that such a person cannot be seen as a *true* doctor.

When Smokrović is discussing the difference between believing and arguing, he also argues that the sufficient condition for being rationally warranted for begin arguing for  $p$  is holding  $p$  defensible (Smokrović 2017: 90):

What I want to stress is that holding  $p$  defensible is sufficient for being rationally warranted to start arguing for  $p$ . It is normative not in the sense that one should argue for  $p$  but that it is epistemically permissible to start arguing for  $p$ . (p.89)

Technically, many argumentations start out of pure hypothesising or by applying the *reductio od absurdum* rule of inference. Hence, we might start defending  $p$  while hoping to arrive at the contradiction and hence prove not  $p$ . Or, it is possible to start defending  $p$  while hoping to arrive at the contradiction (e.g. Saccheri<sup>136</sup> believing in and hoping to show that there could not be any non-euclidean geometry and to prove the Parallel Postulate of Euclid while presupposing and trying to argue for the opposite)

One of the tenets of Smokrović's theory is the idea that "the argumentative process naturally guides participants to extend their knowledge given that a) reasoning has a biological function to work optimally in argumentation, and b) that the very *argumentative process* is structured so that it optimally supports participants 'epistemic curiosity to acknowledge whether a proposition P (the object of discussion) is true" (Smokrović 2015: 224). Such a view hence implies that the argumentative process is to lead to the extension of knowledge. Such a definition of argumentative process implicitly invokes the absent of fallibilism. Just by adding the additional condition that no mistakes occurs with that process, it follows that the argu-

---

<sup>136</sup> In his Saccheri, Girolamo, 1733, *Euclides ab Omni Naevo Vindicatus* (*Euclid Cleared of Every Flaw*). Interestingly enough, Saccheri was actually wrong.

mentative process is or, better, might be knowledge-extending. It sounds like a common-sense addition, obviously that in case of errors, there cannot be any extension of knowledge. It has, however, to be stated explicitly otherwise many of processes that would otherwise fall under Smokrović's definition would fail to lead to knowledge extension. One can be honestly curious about knowing if  $p$  is true, while making several mistakes during the process: inferential mistakes along the way that neither the addresser nor the addressee are aware of. Such a process is not one of extended knowledge, even though all the conditions of the AP narrow definition are formally fulfilled: the willingness of the participants to know the truth, their genuine curiosity and intellectual camaraderie within such process. Nevertheless, none of those factors can, *per se*, prevent them from making mistakes. Inferential mistakes might not happen often, yet they are possible and they happen. Hence, even if we narrow the definition of AP as Smokrović proposes, it does generally not follow that AP is an extending knowledge-process.

In practice, the interlocutors' curiosity need not be neither a necessary nor a sufficient condition for an AP to be knowledge-extensive.

I hope the points underlined in this paper as potentially problematic or in need of further discussion will spur a new, interesting argumentation with Smokrović. I certainly look forward to finding out what his replies are.

## References

- Bermejo-Luque, Lilian, 2011, *Giving Reasons: A Linguistic-Pragmatic Approach to Argumentation Theory*, Springer.
- Bonola, Roberto, 1911/2012, *Non-Euclidean Geometry*, Dover publications.
- Boolos, George, 1997, Constructing Cantorian counterexamples, *Journal of Philosophical Logic* 26 (3), pp. 237-239.
- Brandom, Robert, 1994, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Harvard University Press.
- Hintikka, Jaakko, 1996, Strategic thinking in argumentation and argumentation theory, *Revue Internationale de Philosophie*, vol. 50, no. 196 (2), Major trends in argument theory today (May 1996), pp. 307-24.
- Jackson, Sally 2019, Reason-giving and the natural normativity of argumentation, *Topoi*, 38(4), 631-643.
- Jackson, Sally and Jacobs, Scott, 1980, Structure of conversational argument: Pragmatic bases for the enthymeme, *Quarterly Journal of Speech*, Volume 66, Issue 3, 251-265.
- Mercier, Hugo, and Sperber, Dan, 2011, Why do humans reason? Arguments for an argumentative theory, *Behavioral and Brain Sciences* 34, pp. 57-111. [https://www.researchgate.net/publication/50906835\\_Why\\_Do\\_Humans\\_Reason\\_Arguments\\_for\\_an\\_Argumentative\\_Theory](https://www.researchgate.net/publication/50906835_Why_Do_Humans_Reason_Arguments_for_an_Argumentative_Theory)

- Potter, Polyxeni, 2001, *Ignaz Philipp Semmelweis (1818-65), Emerging Infectious Diseases Journal*, Vol. 7, Number 2, <https://doi.org/10.3201/eid0702.ac0702>.
- Ramachandran, Vilayanur. S.; Rogers-Ramachandran, Diane, 2010, Illusions: Aristotle's Error, *Scientific American Mind* 21, 1, 20-22 (March 2010).
- Shapiro, Stewart, 2000, *Thinking about Mathematics*.
- Smokrović, Nenad, 2015, Arguments as a Means for Extending Knowledge, *Croatian Journal of Philosophy*, Vol. XV, No. 44, Kruzak, Zagreb.
- Smokrović, Nenad, 2017, Curiosity and the Argumentative Process, in Borstner B. and Gartner S. (eds) *Thought Experiments between Nature and Society: A Festschrift for Nenad Mišćević* Cambridge Scholars Publishing, pp. 83-96.
- Sperber, D. And Mercier, H., 2012, Reasoning as a Social Competence, in Landemore and Elster (eds.) *Collective Wisdom: Principles and Mechanisms*, Cambridge University Press, pp. 369-392.
- Trobok, Majda, 2006, *Platonism in the Philosophy of Mathematics*, Filozofski fakultet u Rijeci.
- Walton, Douglas, 2009, Argumentation theory: A very short introduction, in Rahwan, I. And Simari, G., *Argumentation in Artificial intelligence*, Springer, Berlin, pp. 1-24.
- Williamson, Timothy, 2007, *Philosophy of Philosophy*, Oxford: Blackwell.



ANDREJ ULE

## Implicit and Explicit Knowledge in Argumentation

**Summary:** I continue the conversation with Nenad Smokrović on the concept of argumentation, especially on the relationship between social and individual-psychological aspects of argumentation. Nenad believes that argumentation is a kind of synthesis between social and individual knowledge. The question remains as to how this integration occurs. I explain the thesis on the argumentative process as a continuous process of extending and externalization of the objective spirit of the reasoning situation. The cornerstone of this continuous process is learning and exhibition of appropriate argumentation forms. However, we cannot presuppose the existence of any strict norms or general rules of reasoning which would cover all forms of rational reasoning. Many norms and rules of reasoning necessarily remain latent, implicitly present in the everyday argumentation practice in different social situations.

**Key words:** argumentation, reasoning, implicit knowledge, common knowledge, objective spirit.

In my article, I continue the conversation with Nenad Smokrović on the concept of argumentation, especially on the relationship between social and individual-psychological aspects of argumentation. The conversation began with Nenad's contribution to the collection on my philosophical path (Smokrović 2019). Nenad there defends the middle way between the individual-cognitivist and the fundamental social, communicative nature of reasoning and argumentation. He criticizes my "Wittgenstein" notion of following the rules, which in his opinion rejects the possibility of "mental" following the rules (Smokrović 2019: 172-3). The rules are supposed to be "not-justified, but indubitable foundation of other more explicit rules" (Ule 2017: 230). As I stated in my answer, I emphasize that this only applies to basic rules, e.g., logical rules on the use of identity, on the avoidance of obvious contradictions, on the use of the ponens mode, on the derivation of the specific and the individual from the general. These rules seldom, if at all, become the subject of reflexive consciousness, only a trained logical mind can truly make them aware. However, "blind following" does not generally apply to all rules, especially not to more demanding tracking of complex rules, e.g., to maintain coherence between one's own views, to observe the unexpected consequences of one's own or other people's views, to

consider relevant facts or information about the subject of the discussion, etc. There is a great need for individuals' own mental activity, but even this activity is ultimately based on well-practiced "blind" following of certain basic rules of intelligible speech and reasonable argumentation.

At the same time, my attention is drawn to the basic social or the collective nature of knowledge circulating in argumentation. Nenad believes that argumentation is a kind of synthesis between social and individual knowledge. This is probably true, but the question remains as to how this integration occurs, how individual knowledge "socialize", or "social" knowledge individualizes and what it means, because "common epistemics" (common knowledge and common beliefs) cannot in principle be understood as a purely logical conjunction of individual epistemic bases. (more about this a bit later but see also Ule, 2008) I have always been interested in how we can properly formally describe this common epistemic and how it comes about at all., which means that it is social in principle and arises only in eminently social situations that give rise to common rules. In what follows, I will not need a distinction between common knowledge and common belief, so I will mostly talk about common knowledge.

The common knowledge of  $CGp$  of some group  $G$  about the fact  $p$  can be formally characterized as an infinite conjunction of "shared knowledge of the group" ( $SG$ ) of different levels of all members of the group about the fact  $p$ , namely as  $SGp \ \& \ SG(SGp) \ \& \ SG(SG(SGp)) \ \& \dots$ . Articles  $SG(SG(SGp))$  of any degree of iteration of shared knowledge appear in sequence, which seems impossible, since no epistemic subject can have an infinite number of epistemic states of all possible degrees of iteration at the same time. They are only implicitly present in the epistemic situation of the group; thus the common knowledge is in principle implicit. Even when the common knowledge is publicly manifested, e.g., in a gesture with all recognizable meaning, only finally many members of the chain of iterative epistemic states are explicitly taken by individuals, the rest remain implicit. However, *the shared knowledge there is a common knowledge of the meaning of a gesture in a group*, is possible, which means that all members of the group explicitly know that they have certain common knowledge, but they do not know how much explicitly is shared their knowledge. When I speak of the "explication of a piece of common knowledge," then I mean *explicitly shared knowledge* of the existence of common knowledge, not some common knowledge made *de facto* explicit.

I think this kind of partial explication of common knowledge typically comes in learning to follow rules. Once a student has mastered a given rule, then both he and his teachers (say this group "learning group for a given rule") are undoubtedly familiar with this fact, because otherwise someone



could object or question the student. All members of the study group then know that they are in a situation that also includes the shared knowledge of everyone that the student masters the rule.

The learning group therefore first has

- a shared knowledge of the student's mastery of the rule, but also
- a shared knowledge that the given situation implies (contains)
- a shared knowledge of the student's mastery of the rule.

From this, it logically follows that the learning group has

- a shared knowledge that all its members are in a situation that implies that
- all members have a shared state of knowledge about the student's mastery of the rule.

Of course, we get implications of increasingly extensive iterations of shared states. We can go with them to infinity, and this fact corresponds exactly to the (implicit) common knowledge of the learning group about the student's mastery of the rule. When a student "proves" his mastery of a rule, then the learning group comes to the *explicit shared knowledge that they have a common knowledge that the student has mastered the rule*.

A good example of such "learning situations" are multi-actor games. The learning these games is an eminently social situation where students must first master some basic rules of gaming that are often implicitly recognized. There, implicit common knowledge emerges, which is the basis for explicit forms of individual and common knowledge. (Osborne, Rubinstein, 1994) How this knowledge arises, where and how it is encoded in our brains is not yet known, it seems that this process can to some extent be modelled with groups of parallel distributed actors (e.g., in neural networks). They "learn" from each other in interaction with each other. (Halpern, Moses 1990; Geanakoplos 1992)

During argumentation, another important thing appears, namely the occurrence of "weak" conclusions or arguments, i.e., conclusions that are not entirely necessary and logically correct, but which we nevertheless often follow with advantage. Especially in rhetorical argumentation, such an occurrence is a quite regular occurrence. Consider the frequent reasoning by non-deductive reasoning models, e.g., inference by analogy, by abduction, inference based on probabilistic assumptions, practical inferences, and on various kinds of the so-called non-monotonic, default or defeasible reasoning.

Aristotle had already dealt with them in his *Topics*. It is interesting that people are so fond of and without hesitation resorting to such weak

inference patterns. The basic assumption offered to us here is that this happens because all actors in argumentation rely on a series of implicit assumptions that they accept at least as a common belief, if not as common knowledge. Strengthening the argumentative premises and the steps in reasoning with these assumptions would greatly strengthen the arguments made, perhaps even turning them into logically valid arguments. However, even this does not explain all the relevant examples of weak reasoning.

As a sample of such reasoning, I cite a well-known example of a non-monotonic or default inference:

The birds are flying  
Tweety is a bird, then  
Tweety flies.

This conclusion is usually fortunate, but not necessarily so, as it only applies to “typical” birds and not to atypical ones such as e.g., penguins, ostriches, etc. And Tweety may be some atypical bird. We say that this is “non-monotonous reasoning”, namely with the addition of a new premise “Tweety is a penguin”, the conclusion becomes false. This cannot happen with “monotonous” conclusions, they remain valid if we consider any additional premises. Logically (deductively) valid conclusions are monotonic, while the non-deductive inferences are generally non-monotonic.

Perhaps someone will say the aforementioned conclusion can be changed to deductive if we change it to

Typical birds fly  
Tweety is a typical bird, then  
Tweety flies.

However, this is a sophisticated circular conclusion, because for a typical bird, flying is simply an essential feature, so that the equality “typical bird = flying bird” applies. Therefore, the first premise is unnecessary because with the statement “Tweety is a typical bird” we say “Tweety is a bird”. Another option would be to supplement the premise “Birds fly” by listing all the exceptions that do not fly, e.g., “Birds fly except penguins, ostriches, etc.” In that case, you would get the so called “default conclusion”, but it is significantly incomplete, because we simply cannot list all possible exceptions, so we have a statement that substantially includes “etc.”, which means that it is not complete. Nevertheless, we like to use such and similar conclusions very much and constantly in everyday life, in science and in philosophy. These and similar non-monotonic inferences have been intensively dealt with by modern computer science, especially artificial intelligence, for the last forty years, but to date it has not been able to find satisfactory

computer models for all relevant forms of non-monotonic inference. more generally, common-sense reasoning (Bochman 2007).

For the most part, these forms of soft reasoning cannot be transformed into logically valid conclusions, even if we consider some additional, implicit assumptions, because they most often simply deviate in principle from the models of deductive reasoning. At the same time, we usually do not have any easily recognizable basic rules of reasoning, which would be learned “blindly” in common practice of following the rules, i.e., as undoubted components of common knowledge in representative cases of argumentation. However, one can lean on some other forms of sound, although defeasible guidance in reasoning, like cases of good reasoning in similar cases, looking for “strong” or “weak” points in arguments etc. I agree that such strategies are rather various forms of methodical reasoning and not some “real arguments” but they in “common-sense reasoning” often are surprisingly effective, but of course they are not generally valid.

However, implicit social epistemics (implicit common belief and common knowledge) play a key role in these forms of reasoning too. It is true, that silent premises and silent rules of this reasoning, in principle cannot be fully explicated, but still, some common guidance in weak forms of reasoning must be fully available to all participants in the argumentation processes. Weak forms of reasoning presuppose an implicit common epistemic knowledge base, which supports, but does not fully justify such reasoning in concrete situations. I say that there are various forms of “weak (defeasible) social rationality”.

We may ask to what extent this rationality is “proper” to the corresponding epistemic groups and to what extent it is peculiar to the mental structures of individuals. Nenad believes that at least in some deductive argumentation processes, individuals’ reflection on the course of argumentation plays an important role, namely by becoming aware that they are better off changing their pattern of reasoning by revising it to attain higher deductive results (Smokrović 2018: 173). In his essay from the *Analiza* “Normativnost logike in vsakdanja sklepanja”, he adds that with this correction and additional learning, the beginner can spontaneously progress in his deductive abilities and approach the level of an expert (ibid., 2017: 57). We could say that the individual develops his individual “deductive competence” in argumentation processes, and this is somehow “embodied” in the individual’s mental apparatus. Therefore, deductive rationality is not unique to argumentative or, more broadly, to epistemic communities, but also to individuals or their individual argumentative competencies.

I agree with this proposal, but I add that we can similarly assume that the participation of individuals in various forms of rational argumentation al-

lows the individual to progress in their argumentation skills and approach the levels of experts. Here, too, it is possible to assume that the individual develops his argumentative competence, and this is somehow embodied in the individual's mental apparatus. Here, too, the broader argumentative rationality is inherent in both epistemic communities and individuals.

However, the question is still how this kind of internalization of the deductive or argumentative rationality occurs in individuals. Argumentation process regularly occurs in the relevant argumentative communities (with at least two participants). They develop their own form of implicit common knowledge about permissible forms of argumentation in given epistemic situations. As a rule, the epistemic "power" of common knowledge is greater than the power of individual knowledge. An exception to this "rule" is represented by individual experts in their field, who anticipates in their individual knowledge the common knowledge of highly relevant epistemic groups, e.g., groups of researchers in their field of knowledge. With such experts, what they propose (basing on their individual knowledge) may become an integral part of the common knowledge of any group of competent experts. We can also assume the individual progresses in his argumentative abilities and approaches the levels of experts. Here, too, it is possible to assume the individual develops his argumentative competence. Such competence must be somehow embodied in the mental apparatus of the individual. A broader argumentative rationality is thus inherent in both epistemic communities and individuals.

I assume that the transition of the implicit acceptance of a common content into the explicit individual acceptance of the content presents mostly the transition of a part of (implicit) common knowledge into partly explicated and then individually accepted common knowledge, and not a change of (implicit) common knowledge into a purely individual knowledge.

Therefore, it is rational to assume that the appropriation of argumentative rationality generally takes place as a gradual change of implicit common argumentative guidance into partially explicit individual forms of this guidance and not as the formation of completely individual argumentative abilities in individuals. This is, in my opinion, even more present in the teaching of non-deductive forms of argumentation, for in the teaching of deductive forms we are blinded by the intuitive self-evidentness of certain basic logical rules, which give the appearance of an a priori preconceived notion. However, I do not see a fundamental gap between the appropriation of deductive and non-deductive inference forms in argumentation practice.

I understand thus the formation of argumentative rationality as the gradual growing certain argumentative competence. It is a special example of the formation of situation-specific potential of human actors for a meaningful and rational response to common situations, specifically to situations of discussion, persuasion, and proof between two or more persons. With the help of my reaffirmation of the Hegelian notion of the objective spirit, I conceive of this potential as the “objective spirit of a given situation.” (Ule 2011; 2015)

My thesis is that the objective spirit of a situation (at a given time) consists of all common acceptances of those contents and patterns of speech, reasoning and action which may be relevant for a competent mastering of the given situation. Because of its implicit nature, objective spirit transcends the experience and the conscious mind of any individuals (acting in the given situation). It “supervenes” on the common verbal and non-verbal interactions of individuals included in the situation. Because its contents and patterns may transcend all unconscious motives, objective spirit transcends the unconscious minds of individuals too. In this sense, the objective spirit of a situation transcends any individual mind-form or mind-structure.

Insofar as individuals rarely completely master their individual or common situations and they sometimes simply “blindly” follow the latent forms of guidance in mastering the situations and implications of their thoughts, speech, reasons, and actions, we may say that objective spirit “has a life of its own” and evolves according to its own hidden “logic”. However, this life and logic are not an autonomous “spiritual life” and autonomous “spiritual logic” but are our projection of some common traits of human thinking, speech and acting onto the level of common potentialities. However, if objective spirit is conceived of as a trans-personal potential of thought, speech, and action, it seems rational to speak of the “actualizations of objective spirit” when some individuals (one or more) explicitly use and express some contents of objective spirit in their thought, speech, or actions.

To be sure, we need a “medium” which preserves and transmits the trans-personal potential of objective spirit in individuals and in society without being a concrete actualization of this potential. I would like here to refer to the concept of *institutions of meaning* as the fundamental bearers of objective spirit. This idea was presented by Vincent Descombes in his books *La denrée mentale* (1995) (*The Mind's Provisions* (2001)) and *Les institutions des sens* (1996). When we think of objective spirit as the *constitutive background* of human culture and society then it appears as the *implicit order of content* (meaning, values, definitions of rules etc.) and *implicit or-*

*der of discourses*. We cannot and should not reduce objective (or any other form of) spirit to singular processes or to its structural aspects. Social institutions take on a life of their own and allow us to think, judge, decide what we are unable to do purely in our heads, or even in many heads.

Objective spirit, however, is not simply externalized in social institutions, etc. as its objective works; it works in its externalizations that call forth further cognitive activity. We can thus speak of *extension* and not only of *externalization* of spirit (Gallagher, Crisafi 2010: 125). I assume accordingly the argumentative context of the argumentative practices of two or more people may present a minimal institution of meaning with its implicit order of content and implicit order of discourses which enables some forms of extension and externalization of the objective argumentation spirit.

When we think of argumentative process as *something lived between two or more people* then it appears as a *continuous process of extending and externalization of the mind (spirit)* using different meaning and knowledge resources which are put inside or outside of human heads (memory, inner thought, language games, written notes, computer data, presentation devices etc.). The cornerstone of this continuous process is learning and exhibition of appropriate argumentation forms. It also includes the introduction of new rules of reasoning and new language games in common use. However, we cannot presuppose the existence of any strict norms or general rules of reasoning which would cover all forms of rational reasoning. Because of different “incompleteness proofs” in logic we do not have even any complete system of the norms or rules of deductive reasoning. Many norms and rules of reasoning necessarily remain latent, implicitly present in the everyday argumentation practice in different social situations. Sure, it is especially useful and valuable to find and to explicate some of them, but we must know that they are only a smaller part of the implicit common knowledge of the participants in the argumentation practices, the greater part present silent agreements about the validity of the actual reasoning. Sure, such agreements may be weak and defeasible, but they still bear grains of contextually adequate (even if not monotonous) forms of reasoning.

## Literature

- Bochman, Alexander (2007): "Nonmonotonic Reasoning." In Dov M. Gabbay, J. Woods, *Handbook of the History of Logic Volume 8. The Many Valued and Nonmonotonic Turn in Logic*. Amsterdam, Elsevier: 557-662.
- Descombes, Vincent (1995): *Le denrée mental*. Paris, Minuit. (Trans. *The Mind Provisions. A Critique of Cognitivism*, Princeton University Press, 2001)
- Descombes, Vincent (1996): *Les institutions de sens*. Paris, Minuit. (Trans. *The Institutions of Meaning: A Defense of Anthropological Holism*, Harvard University Press, 2014)
- Gallagher, Shaun; Crisafi Anthony (2010): "Hegel and the Extended Mind." *AI & SOCIETY* 25(1):123-129
- Geanakoplos, John (1992): "Common knowledge." In Moses, Y. (ed.), *Reasoning About Trans. The Institutions of Meaning: A Defense of Anthropological Holism*, Harvard University Press, 2014 *Knowledge. (TARK 1992)*. Monterey, Morgan Kaufmann: 254-315.
- Halpern, Joseph; Moses, Yoram (1990): "Knowledge and common knowledge in a distributed environment." *Journal of ACM*, 37, 3: 549-587.
- Osborne, Martin, J.; Rubinstein, Ariel (1994): *A Course in Game Theory*. Cambridge/M, MIT.
- Smokrović, Nenad (2017): »Normativnost logike in vsakdanja sklepanja.« *Analiza*, 2: 45-60.
- Smokrović, Nenad (2019): "From argument to argumentative dialogue." In Markič, O., Malec, M., *Filozofska pot Andreja Uleta*. Ljubljana, Znanstvena založba FF: 165-174.
- Ule, Andrej (2008): "Knowledge attributed to collectives." In Ule, A, *Circles of Analysis*. Wien, Berlin, LIT: 273-286.
- Ule, Andrej (2011): "The concept of spirit: A critical re-conceptualization of a metaphysical category." *Contemporary Science and Faith*. Mostar, Sveučilište: 85-112.
- Ule, Andrej (2015): "Consciousness, mind, and spirit: Three levels of human cognition." *Interdisciplinary Description of Complex Systems*, 13, 4: 488-500.
- Ule, Andrej (2017): "Mental models in scientific work." In Borstner, B., Gartner, S. (eds.), *Thought Experiments between Nature and Society: A Festschrift for Nenad Miščević*, Newcastle upon Tyne: Cambridge Scholars Publishing: 219-233.





LINO VELJAK

## O utemeljenju metodologije znanstvenog istraživanja

**Sažetak:** Redukcionizam utemeljen na ontologijskom monizmu i konceptu jedinstva znanosti očituje se u načelnom izjednačavanju metodologije znanstvenog istraživanja u svim područjima znanosti. Tako oblikovan znanstveni realizam pretendira na aproksimativno istinito spoznavanje. No, svođenjem zbilje na predmet koji se aproksimativno istinito može spoznati ukoliko se izrazi kvantitativnim pokazateljima, realizam ostaje nužno zarobljen nemogućnošću uvida u proizvedenost danosti koja se u redukcionistički zasnovanim metodologijama izjednačava sa zbiljom. Zaključak glasi: metodologiju znanstvenog istraživanja treba osloboditi od apsolutizacije kvantitativnoga.

**Ključne riječi:** metodologija, monizam, dualizam, metafizika, znanstveni realizam, redukcionizam, danost, kvantitativnost.

Jedno od bitnih pitanja u suvremenoj filozofiji, prvenstveno filozofiji znanosti, a svakako i središnje pitanje metodologije znanstvenog istraživanja sastoji se u dilemi: jesu li metode znanstvenog istraživanja u svim područjima i poljima znanosti jedinstvene (pa se stoga može govoriti o jednoj jedinoj paradigmi metodologije, što može implicirati redukcionizam, a općenito se imenuje kao metodologijski monizam) ili pak pojedina područja znanosti posjeduju različite metode i različite istraživačke standarde (to se prvenstveno odnosi na rascjep – ili navodni rascjep – između metodologije istraživanja u prirodnim znanostima i metodologije istraživanja u društvenim znanostima, a imenuje se kao metodologijski dualizam)<sup>137</sup>.

Zastupnici dualističke pozicije (ali i neki drugi kritičari koncepta jedinstva znanosti) ponekad imenuju metodologijski monizam kao *redukcionizam*.<sup>138</sup> Taj se pojam, međutim, ne mora odnositi isključivo na sferu metodologije. Naime, redukcionizam može, kako to izlaže britanski filozof

---

<sup>137</sup> Riječki filozof Nenad Smokrović formulira taj rascjep između monista i dualista na sljedeći način: “Jedni su smatrali da prirodne i društvene znanosti koriste i trebaju koristiti različite metode, dok su drugi smatrali da je u znanostima na djelu jedan i jedinstven standard znanstvenog postupanja i istraživanja”. (Smokrović 2017: 11).

<sup>138</sup> Za opsežnije izvode o pojmu redukcionizma i njegovim posljedicama za metodologiju znanstvenog istraživanja usp. Veljak 2020: 157 i dalje. Ovdje se daje sažetak argumentacije sadržane u citiranom prilogu relevantne za temu ove rasprave.

*kontinentalne* orijentacije Nicholas Joll<sup>139</sup>, označavati ili 1) filozofijsku poziciju, ili 2) metodologijsku strategiju ili pak 3) princip svodenja ranijih znanstvenih teorija na termine nove teorije odnosno novouspostavljene znanstvene paradigme.<sup>140</sup> U prvom se slučaju radi o ontologijskom (metafizičkom) stajalištu, a u drugom i trećem slučaju o epistemologijskoj poziciji i njezinoj primjeni na metodologiju znanstvenih istraživanja.

Britanski fizičar i teolog John Polkinghorne (2002a, 2002b) definira pak redukcionizam kao poziciju prema kojoj bilo koji složeni sistem nije ništa drugo doli zbroj svojih dijelova, uslijed čega se sistem može shvatiti tako što će se razmotriti svi njegovi pojedinačni dijelovi (čemu se suprotstavlja holistički antiredukcionizam koji drži cjelinu nečim višim od pukog zbroja dijelova te se stoga cjelina ne može adekvatno opisati pomoću opisivanja njezinih dijelova).<sup>141</sup> On razlikuje tri osnovna tipa redukcionizma: *strukturni ili konstitutivni redukcionizam* (koji tvrdi da su dijelovi nekoga kompleksnog sistema koji se u njima rastvara isključivo njegovi konstitutivni dijelovi; taj je tip redukcionizma blizak metodologijskom redukcionizmu koji dolazi do izražaja u oblikovanju strategija znanstvenog istraživanja), *pojmovni ili epistemologijski redukcionizam* (smatra da se pojmovi koji se koriste za opis ili objašnjenje neke cjeline mogu u potpunosti izraziti pojmovima kojima se opisuju dijelovi te cjeline) i *kauzalni redukcionizam* (smatra da uzroci koji djeluju na cjelinu nisu ništa drugo doli zbroj učinaka onih uzroka koji djeluju na pojedinačne dijelove te cjeline). Ta vrsta redukcionizma veoma je bliska ontologijskom redukcionizmu, što ga Polkinghorne definira kao stav prema kojemu cjelina *jest* zbroj dijelova.

Drugačiju definiciju redukcionizma daje američki fiziolog i psihijatar Eric Richard Kandel, dobitnik Nobelove nagrade za medicinu. U nastojanju da prevlada rascjep između kulture kakvu njeguju prirodni znanstvenici i kulture svojstvene zastupnicima humanističkih i društvenih znanosti, Kandel vidi most između dviju kultura u redukcionizmu koji je zajednički kako znanstvenicima svih vrsta tako i umjetnicima. Taj se redukcionizam sastoji u destiliranju *znanstvenih i umjetničkih zamisli u manje komponente kojima je onda lakše upravljati i koje su ono o čemu se zapravo radi* (Čačinić 2020: 128).

Ako neki autor ili neka škola prihvaća sva tri tipa redukcionizma, onda se može govoriti o *jakom redukcionizmu*, a on je, prema autorima poput Polkinghornea, utemeljen na takvoj ontologiji koja zastupa stav da zbiljska opstojnost pripada isključivo bazičnim sastavnicama. Stoga je fizika

---

<sup>139</sup> Usp. Joll. <https://www.iep.utm.edu/con-meta/> (posjećeno 2. 2. 2021.)

<sup>140</sup> Usp. Ney. <https://www.iep.utm.edu/red-ism/> (posjećeno 11. 5. 2020.)

<sup>141</sup> Polkinghorne, nekadašnji profesor matematičke fizike u Cambridgeu, kod nas je poznat kao popularizator pomirenja religije i znanosti (usp. Polkinghorne 2019.).

elementarnih čestica jedina fundamentalna disciplina: sve druge discipline, od ostalih grana fizike, preko biologije, pa do psihologije, antropologije i sociologije nisu ništa drugo doli učinci onoga što je temeljno. No, te bazične sastavnice ne moraju biti elementarne čestice, kao što je to slučaj u fizikalizmu,<sup>142</sup> nego mogu biti, primjerice, i geni, kao što je to slučaj u stanovitoj vrsti biologizma<sup>143</sup> koja je svoju široku primjenu našla u sociobiologiji. U svim ovim slučajevima na djelu je *metafizički redukcionizam* koji proizlazi iz apsolutizacije nekog entiteta ili područja (te dovodi do ontologizma, gnoseologizma, antropologizma ili antropocentrizma, ili pak do biocentrizma, geocentrizma, teocentrizma, scijentizma, naturalizma, fizikalizma, itd.), a događa se podjednako i u otvoreno metafizičkim, kao i u onim verbalno antimetafizičkim orijentacijama i filozofijskim školama. Bitak, zbiljnost, svijet ili kozmos svedeni su na jedan apsolutizirani entitet, bilo da je riječ o Engelsevoj i Lenjinovoj materiji, bilo o elementarnim fizikalnim česticama Otta Neuratha i Rudolfa Carnapa, bilo pak o Dawkinsovim genima (ili, na planu metafizičkih orijentacija u užem smislu riječi, o idejama, nepokrenutom pokretaču, itd.).<sup>144</sup>

Vrijedilo bi obratiti pozornost na kritiku naturalističke, pozitivističke i neopozitivističke metafizike koju je u svom nastojanju da naturaliziranu metafiziku poveže s jednom voluntarističkom epistemologijom dala britanska filozofkinja Sophie Allen, vrlo precizno uočivši kako se u tom kontinuitetu metafizike i empirijskih istraživanja uspostavlja pretenzija da *znamo objektivno postojeću ontologiju svijeta prirode koji je neovisan od uma* (Allen 2012: 221). Na temelju imanentne kritike ona će zaključiti kako bi ova ili ona ontologijska teorija mogla predstavljati istinu o objektivnom

<sup>142</sup> Usp. Poland 1994. ili Tye 2009.

<sup>143</sup> Usp. npr. Dawkins 1976. Za kritički pristup biologizmu (ali i njegovim praktičkim, kako epistemologijskim tako i socijalno-političkim učincima) vidi usp. Steven i Hilary Rose 2001.

<sup>144</sup> Metafizičko utemeljenje redukcionizma do svojega punog očitovanja dolazi na vidjelo među onim autorima koji otvoreno priznaju metafizički karakter svojih umotvorina. U tom smislu vrijedilo bi obratiti pozornost skupini britansko-južnoafričkih autora koji su 2007. objavili knjigu pod naslovom *Svaka stvar mora ići – naturalizirana metafizika* (usp. Ladyman et al. 2007., posebno str. 2, 17, 27. i 75.). Oni razumiju metafiziku kao interdisciplinarno istraživanje konzistentnosti znanosti i znanstvenih teorija. Znanost (preciznije, današnja fizika) nije, doduše, još dospjela do potpune spoznaje svojega predmeta, ali ona se stalno razvija i stječe nove spoznaje o fenomenima, otkrivajući nam neočekivane stvari o fenomenima što ih istražuje, a njezina je epistemička supremacija neupitna, te je utoliko opravdana njezina ambicija da zahvati i samu *stvar po sebi* (usp. Ladyman et al. 2007., str. 2, 17, 27. i 75.). Argumentaciju u pogledu teze prema kojoj se takva (otvorena) metafizika ne može pomiriti s Carnapovom (prešutnom) fizikalističkom metafizikom ponudio je dansko-norveški filozof Rasmus Jakslund (2016., str. 125. i dalje).

svijetu, ali *mi ne možemo reći koja je to teorija* (Isto, str. 232.). Uglavnom, naturalizirana metafizika ne može se osloboditi svojih precizno identificiranih teškoća i kontradikcija. No, i tu se može naići na jednu zamku, koju se može izbjeći tek odbacivanjem lažne alternative između naturalističkog redukcionizma i idealizma filozofije imanencije, kao što je to u svojoj rekonstrukciji geneze kritičke teorije Maxa Horkheimera uvjerljivo pokazao njemački filozof Olaf Asbach (1997: 61-73.).

Valja, međutim, naglasiti da – neovisno o svim mogućim teškoćama naturalističkih tipova metafizike (kako u njihovim fizikalističkim, tako i u biologističkim inačicama) – sve one predstavljaju tek varijante *metafizičkog monizma*. Moglo bi se reći da se redukcionizam svojstven tim metafizikama daje sa svoje strane reducirati na obuhvatan pojam monizma. Alternativu monizmu – i tu se srećemo s vječnim metafizičkim sporom, koji se proteže od samih početaka filozofije pa sve do današnjih rasprava u okviru po samorazumijevanju transmetafizičke filozofije uma – čini dualizam (kao i, u širem smislu, pluralizam),<sup>145</sup> čija se metafizička narav ne bi dala kvalitativno razlikovati od metafizičnosti monizma. I monizam i pluralizam predstavljaju *filozofijske pozicije* (a njihov metafizički karakter ne dovodi u pitanje ocjenu da se tu radi upravo o filozofijskim, prvenstveno ontologijskim pozicijama).

Već spomenuti Nenad Smokrović zastupa jednu sofisticiranu monističku poziciju, koju – u kontekstu razmatranja mogućnosti utemeljenja metodologije znanstvenog istraživanja – određuje na tragu *main-stream* pozicije u današnjoj analitičkoj filozofiji<sup>146</sup> kao *znanstveni realizam*:

Stav **znanstvenog realizma** (...) smatra da su teorije razvijenih znanosti **aproksimativno istinite** (...) *Što je stvarnost?* (...) Materijalna, fizička stvarnost koja nas okružuje sasvim je sigurno 'objektivna' utoliko što postoji nezavisno od nas (...) Teorije fizike ili astronomije, na primjer, odnose se na stvarnost nezavisnu od ljudskog djelovanja (...) Možemo li ljudski govor, mišljenje, mehanizme djelovanja također smatrati dijelom objektivno postojeće stvarnosti? Odgovor je, da! Sam govor, jezik kao društveni fenomen, kao i mišljenje, nastali su evolucijski i na taj način postoje objektivno. Tu spadaju i mehanizmi odlučivanja i motivacije (...) Istina je da različite društvene i ekonomske institucije ne bi postojale bez čovjeka i društva, međutim moramo prihvatiti da, primjerice, država kao institucija ili, pak, tržište postoje objektivno. One neumitno utječu na nas, bez obzira na to što mi o njima mislili (Smokrović 2017: 111).

---

<sup>145</sup> Usp. Robinson 2003: 85-101.

<sup>146</sup> Usp. npr. Leplin 1997.

Znanstveni realizam precizno je definiran pomoću tri postavke:

a) da objektivno postoji aspekt stvarnosti koji određena znanost istražuje, b) da možemo o tom aspektu stvarnosti imati znanje, c) da se to znanje izražava tvrdnjama koje su doslovno istinite ili lažne (Isto, str. 186.).

No, ovdje se ne može neposredno govoriti o metafizičkom redukcionizmu, već i utoliko što autor izbjegava dogmatske odgovore o zbiljskoj prirodi dotične objektivno opstojeće *stvarnosti*, s pravom se ograničavajući na konstataciju prema kojoj ti entiteti posjeduju objektivnu zbiljnost.

Taj opravdani oprez jasno dolazi do izražaja u pitanju o ontologijskom statusu matematičkih entiteta gdje Smokrović ne daje jednoznačan odgovor, već upućuje na to da neki matematičari smatraju kako brojevi i drugi entiteti postoje objektivno, nezavisno od nas. Drugi pak drže da tih entiteta nema nezavisno od nas, nego bi oni bili proizvod naše misli, matematičkog genija, te on na kraju zaključuje:

Koje je od ovih gledišta istinito? Postoje uvjerljivi argumenti za svako od njih, kao i karakteristični problemi koje svako od njih otvara. Ostavljamo čitatelju da sam odluči koju će poziciju odabrati (Isto, str. 79.).<sup>147</sup>

Moglo bi se reći da je u središtu Smokrovićeve misaone pozornosti epistemologijska dimenzija pitanja o istini, a da se dimenzija ontologije i metafizike uzima u obzir samo u mjeri nužnoj za izgradnju epistemologije; iz toga što je, kako on navodi, ekonomija *jednako egzaktna znanost kao što je to i fizika* (Isto, str. 11.) ne slijedi da su entiteti koje istražuje ekonomijska znanost nužno svodljivi na fizikalne entitete. Kriteriji *objektivnog postojanja* nekog entiteta su njegova nezavisnost od ljudskih stavova i mišljenja te kauzalna moć koju on iskazuje (usp. Isto, str. 198), a što se tiče ontologijskog aspekta problema objektivnosti, Smokrović se priklanja Searleovom stavu prema kojemu *institucionalne činjenice postoje ontološki subjektivno a epistemološki objektivno* (Isto, str. 201.). Ovdje nije riječ o entitetima kao takvima, niti o činjenicama općenito, već o ljudskim društvenim tvorbama. Izrazit primat epistemologijskog interesa nad ontologijsko-metafizičkim propitivanjima dolazi do izražaja u najznačajnijem dijelu radova Nenada Smokrovića (usp. Smokrović 2015). Stalo mu je do pouzdanog znanja, takvog znanja koje bi dovelo do osiguranja aproksimativne istinitosti teorijâ razvijenih znanosti, a te bi teorije trebale istinito predstavljati *stvarnost* (Smokrović 2017: 114, 175).

---

<sup>147</sup> Razjašnjenju te dvojbe koja se prepušta čitatelju ne doprinosi ni razmatranje pitanja na što se odnose matematičke formulacije logičkih konzekvencija, budući da se tu razrješava problem prevodenja običnih logičkih iskaza u matematičke, a da se pitanje njihova ontologijskog statusa ne tematizira (usp. Smokrović 2012).

No, je li Smokrovićev znanstveni realizam, suprotstavljen znanstvenom antirealizmu (u varijantama instrumentalizma, operacionalizma i pozitivizma) ontologijski neutralan? Sasvim sigurno – ne! Tako on tvrdi da je *materijalna, fizička stvarnost koja nas okružuje sasvim (...) sigurno 'objektivna' utoliko što postoji nezavisno od nas*, te da se to odnosi, primjerice, na teorije fizike ili astronomije, koje se odnose *na stvarnost nezavisnu od ljudskog djelovanja* (Isto, str. 111.).<sup>148</sup> Ovdje sadržana implicitna kritika antirealizma daleka je od zdravorazumskih karikiranja antirealizma kao stajališta prema kojemu ljudsko mišljenje i drugi kognitivni procesi doslovno stvaraju zvijezde i dinosaure u njihovoj konkretnoj egzistenciji<sup>149</sup>. No, anti-realizam (i to upravo ontologijski antirealizam) podvrgnut je i eksplicitnoj kritici, koja polazi od (posve korektne) definicije predmeta kritike:

Neprihvatanje ontološkog (realizma) označava najčešći oblik antirealizma (...) Prema ontološkom antirealizmu, čak su i u jakim znanstvenim teorijama djelomičan konstrukt koji proizlazi iz znanstvenog svjetonazora (...) Entiteti (nacija, porezni obveznici, rasa, rod, itd.) ovisni (su) o tome kako ih vide članovi društva i što sami članovi grupe o sebi vjeruju (...) Utoliko su oni u cijelosti društveni konstrukti (Smokrović 2017: 188).

Smokrović sažimlje kritiku antirealizma na osnovi koncepcije kauzalnog realizma:

Ovo što vrijedi za entitete jakih znanosti možemo primijeniti i na ekonomsku i, općenito, društvenu stvarnost (...) **Vjeronjanja** članova zajednice omogućuju da određeni komad metala bude novac. Novac je dijelom određen vjeronjanjima **pojedinaца** (podebljano L. V.)<sup>150</sup>. Ako je

---

<sup>148</sup> Međutim, tu se postavljaju najmanje dva pitanja: prvo, ima li smisla govoriti o objektivnoj opstojnosti bića ili entiteta koja nisu na bilo koji način dodirnuti ljudskim spoznavanjem (a spoznavanje je oblik ljudskog djelovanja, kao što su spoznaja, teorije, paradigme, itd., učinci takvog djelovanja) te, drugo, kad je već riječ o teorijama fizike, jesu li spoznati fizikalni entiteti identični sebi samima prije nego što su postali predmetom spoznavanja, kako je to pokazao još Heisenberg svojom teorijom neodređenosti, ustanovivši, osim svega, i to da u kvantnoj teoriji nema nikakva putokaza za povezivanje matematičkih simbola s pojmovima običnog jezika (usp. Heisenberg 1997: 143.).

<sup>149</sup> Kao primjer takve zdravorazumske kritike antirealizma usp. npr. Devitt i Sterelny 2002, posebno str. 287.

<sup>150</sup> Kritički valja primijetiti kako Smokrović ovdje (ali i sistematski, kako se može vidjeti na više drugih mjesta u ovoj knjizi, npr. na str. 197.) neupitno preuzima jednu ideologijsku floskulu neoliberalne/neoliberističke (upravo tačeriističke) provenijencije, prema kojoj postoje samo atomizirani pojedinci (i njihove obitelji, koje on, za razliku od britanske političarke, ne spominje), što označava pad ispod Aristotelova uvida u to da je cjelina više od zbroja njezinih dijelova (usp. *Metafizika*, VII 17, 1041b). Društvo kao skup Demokritovih atoma?



tako, postoji li on objektivno? U skladu s kriterijem kauzalnosti – da! Mi smo u interakciji s novcem (...) Javni dug sasvim sigurno djeluje na nas (Isto, str. 191.).

O ontologijskoj neutralnosti ne može se dakle govoriti ni u pogledu onih entiteta koji su proizvod ljudskog djelovanja (poput države i tržišta) ili se ne mogu javljati bez ljudi (mišljenje, govor, jezik, mehanizmi odlučivanja). Smokrovićev je odgovor na pitanje o objektivnoj opstojnosti takvih entiteta jednoznačno afirmativan:

Možemo li ljudski govor, mišljenje, mehanizme djelovanja također smatrati dijelom objektivno postojeće stvarnosti? Odgovor je, da! Sam govor, jezik kao društveni fenomen, kao i mišljenje, nastali su evolucijski i na taj način postoje objektivno. Tu spadaju i mehanizmi odlučivanja i motivacije (...) Istina je da različite društvene i ekonomske institucije ne bi postojale bez čovjeka i društva, međutim moramo prihvatiti da, primjerice, država kao institucija ili, pak, tržište postoje objektivno. One neumitno utječu na nas, bez obzira na to što mi o njima mislili (Isto, str. 111.).

Pri tom konstrukt *objektivno postojeća stvarnost* služi za utemeljenje epistemologije i metodologije istraživanja. Raspravljajući o spoznatljivosti neopažljivih termina, Smokrović se oslanja na Ernana McMullina<sup>151</sup> te na pitanje *Opisuje li teorija koja u svojim iskazima koristi neopažljive termine 'svijet na istinit način'?* Daje jednoznačan odgovor: *Za znanstvenog realistu, elektron (ili novac) je upravo ono što teorija o elektronu (ili teorija o novcu) kaže da jest* (Isto, str. 180.). A jednako važi i za društvene entitete:

Iskazi 'profit je glavni cilj poslovanja', ili, 'demokracija je ključno svojstvo suvremene civilizacije' vrlo (su) udaljeni od direktnog opažanja, premda su oba tipa iskaza tvrdnje o stvarnosti. Iskazi (hipoteze) sastoje se od termina koji se odnose na neko svojstvo stvarnosti. Ovakvi se termini u znanstvenom iskazu izražavaju kao varijable. Realizam, rekli smo, smatra da svojstva na koje se varijable odnose, postoje nezavisno od promatrača (...) S obzirom na nezamjetljiva svojstva, znanstveni realista obavezan je pokazati (...) da ona korespondiraju sa stvarnim entitetima u svijetu (Isto, str. 181.).

I dalje: *Entiteti prirodne stvarnosti postoje nezavisno od ljudskog uma i ljudske djelatnosti* (Isto, str. 197.). Daleko je to od ontologijske neutralnosti! Valja usput naglasiti kako bi se teško moglo tvrditi da bi takva neutralnost predstavljala nekakvu poželjnu opciju (osim ako se ne opredijelimo za radikalni skepticizam).

---

<sup>151</sup> O McMullinu usp. Godfrey-Smith 2003.

U funkciji garantiranja pouzdane i istinite spoznaje kao i valjanosti znanstvenih teorija također stoji onaj možda najproblematičniji element Smokrovićeve epistemologije: pribjegavanje teoriji odraza, što se najjasnije očituje u sljedećim iskazima:

Teorije čiji iskazi ne odražavaju stvarnost ili pružaju iskrivljenu sliku stvarnosti trebaju se korigirati ili zamijeniti novima (...) U empirijskim znanostima teorije uvijek nastoje biti odraz stvarnosti (Isto, str. 106.).

Iz ontologijskog realizma konzekventno proizlazi i adekvatna gnoseološkijska teorija, koja bi mogla biti posve legitimna u okviru klasičnih teorija istine svojstvenih Aristotelu ili Tomi Akvinskom, ali svakako dvojbeni u kontekstu utemeljenja metodologije društvenih znanosti. Ključan pojam koji upućuje na svu problematičnost jedne takve teorije jest *danost*.<sup>152</sup> Ignoriranje razlike između *bića* (koje prema postulatima svake realističke ontologije opstoji neovisno o subjektu, promatraču ili spoznavatelju) i *predmeta* (kojim to biće postaje predmetom promatranja, spoznavanja i modificiranja) svojstveno teoriji odraza rezultira ne tek nemogućnošću razumijevanja konstituiranja svakoga bića s kojim na bilo koji način imamo posla, nego i nemogućnošću modifikacije zatečenoga i danoga koja bi nadmašivala granice mehaničkoga ili evolutivnog reproduciranja vječno istoga.<sup>153</sup> Konzekvencije zanemarivanja spomenute razlike očituju se, primjerice, u kanonizaciji činjeničnog stanja obilježenoga time što je profit svrha ekonomijske djelatnosti u prirodnu i nepromjenjivu *nužnost*. Tako se činjenica uzdiže na rang jedne jedine i jedino moguće zbilje. Tu je na djelu jedan značajan tip redukcionizma: biće se (metafizički, koliko god zastupnici takvog redukcionizma s gnušanjem odbacivali opravdanost bilo kakvog povezivanja s metafizikom) svodi na danost, a napor spoznavanja na pokušaj što vjernijeg zrcaljenja te danosti, što se može definirati kao strukturni ili konstitutivni redukcionizam.

Kakve su konzekvencije utemeljenja metodologije znanstvenog istraživanja na naznačenom redukcionizmu? Ako strukturni ili konstitutivni redukcionizam predstavlja formu metafizičkog redukcionizma, to se ne bi moglo reći za njegovu metodologijsku primjenu u oblikovanju strategija znanstvenog istraživanja. Jednako vrijedi i za pojmovni ili epistemologijski, kao i za kauzalni redukcionizam. Model tih primijenjenih tipova redukcionizma dao je zapravo još Auguste Comte u svom nacrtu pozitivne znanosti (ako tu zaslugu ne bi trebalo pripisati skolastičaru Williamu Ockhamu i

---

<sup>152</sup> Za opsežnu argumentaciju u prilog postavke o naznačenoj problematičnosti usp. Veljak 1979.

<sup>153</sup> O problematici vezanoj uz redukcionistički karakter danosti usp. Marion 1989.

njegovoj poznatoj britvi odnosno principu ekonomičnosti mišljenja),<sup>154</sup> a danas on posebno intenzivno živi među zagovornicima ideala *jedinstvene znanosti*, ideala (ili programa) što su ga formulirali upravo utemeljitelji fizikalizma iz Bečkog kruga.<sup>155</sup> Zagovornici jedinstvene znanosti (ili jedinstva znanosti), kako to definira jedna Merriam-Websterova enciklopedija,<sup>156</sup> uvjereni su da se teorijski entiteti, stavovi i činjenice do kojih dolazi svaka znanost trebaju verificirati opserviranjem empirijskih podataka te da se teorijski entiteti određenih znanosti mogu definirati u terminima bazičnih znanosti, a da se zakonitosti koje su otkrila istraživanja u nekom znanstvenom polju mogu objasniti zakonitostima bazičnih znanosti (odnosno da ih valja svesti na te temeljnije zakonitosti).

Nenad Smokrović je upravo na tom tragu. Bazična je znanost i za njega, kao i za sve relevantne suvremene istraživače u sferi metodologije znanosti – matematika. U tom će smislu on naglasiti:

Pomoću relacija, jednadžbi i funkcija možemo opisati gotovo svako stanje, događaj ili proces u svijetu, bio on jednostavan ili složen, ticao se on fizikalnog, kemijskog, biološkog, ali i društvenog, interpersonalnog i ekonomskog aspekta stvarnosti. U ovom smislu jezik matematike je univerzalno primjenjiv na svaku znanost (Isto, str. 75.).<sup>157</sup>

Premda odatle ne slijedi da među pojedinim područjima i poljima znanosti nema bitnih razlika,<sup>158</sup> jedinstveni je jezik znanosti primjenjiv u svim domenama, uključujući, dakako, i ljudski svijet. Jer,

---

<sup>154</sup> Usp. Losee 1977.

<sup>155</sup> Koncept *jedinstvene znanosti* javlja se, čini se, tek 1933. godine, a oblikuju ga pripadnici Bečkog kruga. To čine matematičar Hans Hahn u spisima *Logika, matematika i spoznaja prirode te Suvršene bitnosti* (usp. Hahn 1975a i 1975b.) i Otto Neurath u djelu *Jedinstvena znanost i psihologija* (usp. Neurath 1933). No, možda je u tom sklopu najznačajniji programatski spis pod naslovom *Znanstveni svjetonazor – Bečki krug* koji potpisuju Hahn, Neurath i Carnap (1975).

<sup>156</sup> Usp. Doniger 1999: 911.

<sup>157</sup> Smokrović će upravo na temelju primjene matematike opovrgavati Keynesovo metodsko razlikovanje istraživanja fizikalnih entiteta od istraživanja socijalnih činjenica, koje su, prema Keynesu, obilježene otvorenošću, jer *ljudsko djelovanje u društvenom kontekstu karakterizira otvorenost i određeni stupanj neodređenosti* (Smokrović 2017: 208). No, suvremena istraživanja u psihologiji demantiraju Keynesa, naglašava Smokrović, a to se posebno odnosi na formalno matematički utemeljenu teoriju odlučivanja (počevši od 40-ih godina prošlog stoljeća) *koja daje odgovor na pitanje 'kako ljudi trebaju odlučivati' pod pretpostavkom da su racionalni* (Isto, str. 209.). Ostaje posve nejasno kako bi se pretpostavka o ljudskoj racionalnosti mogla matematički ili bilo kako drugačije demonstrirati.

<sup>158</sup> Naime, *ono što čini razliku je različiti vokabular koji je različit za svaku znanost. Vokabular određene znanosti određuje varijable koje ta znanost koristi* (Isto, str. 76.).

sam čovjek i njegove evolucijskim putem nastale sposobnosti također su objektivno postojeći i nezavisni 'predmeti'. Znanstvenici koji proučavaju čovjeka i njegove sposobnosti proučavaju ga kao objektivno postojeći entitet, poput drugih životinjskih vrsta (...) Ljudske temeljne sposobnosti zaključivanja dizajnirane su evolucijski i također ih treba smatrati objektivno i nezavisno postojećim entitetima (Isto, str. 196).

Stoga bi se i psihologija i antropologija, poput drugih društvenih i humanističkih znanosti, metodički morale osloniti na primjenu matematičkih metoda, kako bi aproksimativno dospjele do istinitog znanja o entitetima koje istražuju. A to one i čine, dakako, s velikim uspjehom.

Nema nikakve dvojbe da su rezultati primjene tih metoda kojima se dopijeva do kvantitativnih pokazatelja iznimno korisni (da se – bez pretjerivanja – ustvrdi, upravo dragocjeni) za uvid u karakter istraživanih fenomena i procesa u ljudskom svijetu, u društvu, u spoznaji individualnih i kolektivnih osobina i svojstava ljudskih bića i njihovih interakcija. No, kako će neko kvantitativno istraživanje moći razlikovati istovjetan stupanj osjećaja zadovoljstva postignut uživanjem u gledanju vulgarnog *realityja* od kvantitativno jednakog stupnja zadovoljstva stečenoga slušanjem (ili i gledanjem) vrhunskog koncerta. Današnja empirijska i eksperimentalna psihologija raspolaže (i to u sve većoj mjeri i sa sve višim stupnjem pouzdanosti) metodama mjerenja zadovoljstva (kao i frustracije i koječega drugoga). No, kako se matematički mogu razlikovati kvalitativno nesvodljive pojave? Dvije jabuke i tri kruške čine pet voćaka. Ako nisu trule – u redu je, imamo pet jestivih i ukusnih komada voća. No, što čine kilogram smeća i dva kilograma krumpira? Tri kilograma čega?

A stvari postaju još složenije kad se suočimo s rezultatima znanstvenog (dakle, kvantitativnog) istraživanja neke društveno-političke konstelacije koja je oblikovana kroz svrhovito djelovanje određenih aktera usmjereno na postizanje nekih ciljeva koji bi bili primjereni interesima tih aktera. Da se ilustrira jednim konkretnim primjerom! Neka zajedničkim interesima povezana grupa zaključi da će medijski posredovanom sistematskom aktivnošću (zasnovanom na poznavanju metoda primjena rezultata psihologijskih istraživanja) produbljivati etničku distancu u određenom multietničkom prostoru.<sup>159</sup> U slučaju uspješnosti kampanje širenja predrasuda,

---

<sup>159</sup> Kritički će čitatelj na ovom mjestu postaviti legitimno pitanje: Zašto bi netko uopće htio pokretati kampanje usmjerene na stvaranje i produbljivanje nepovjerenja i mržnje? Odgovor glasi: zato što je racionalna analiza pokazala da bi od uspjeha neke takve kampanje moglo biti (materijalne, financijske) koristi. O modelima i metodama racionalnog odlučivanja usp. npr. Sikavica et al. 2014. ili Milić 2014. Modeli odlučivanja na planu poslovanja neke kompanije primjenjivi su (dakako, uz kompleksniji sadržaj potrebnih varijabli) i na planu stvaranja javnog mnijenja, kao što je to veoma davno praktički po-

nepovjerenja i mržnje taj će prostor nakon nekog vremena biti obilježen visokim stupnjem zatrovanosti međuetničkih odnosa. Znanstvena (sociologijska, sociopsihologijska i politologijska) istraživanja vjerno će otkriti stupanj te zatrovanosti, identificirat će trendove, komparirati aktualne rezultate istraživanja s rezultatima ranijih istraživanja te – sasvim u skladu s time – zaključiti da je stupanj održivosti te multietničke zajednice nizak ili čak i nikakav. Neće, međutim, problematizirati generiranost evidentiranog stanja (posebno ukoliko ta istraživanja ovise o onoj interesnoj grupi koja je to stanje proizvela), već će zaključiti kako dotična zajednica nema nikakve perspektive. Tu se danost uzima kao neupitna zbiljnost, dok njezina generiranost izmiče kvantifikaciji te na taj način ostaje nevidljivom. Kako kvantificirati odluku one spomenute interesne grupe da krene u sistematsku kampanju širenja nepovjerenja mržnje? Stoga ta odluka i njezini učinci izmiču problematiziranju i istraživanju.

Ako se hoće generalizirati, valjalo bi postaviti pitanje: *Smije li se kompleksnost ljudskog bića te još veća kompleksnost društvenih, historijskih i povijesnih tendencija i procesa svesti na ono što se može iskazati brojevima i brojčanim omjerima? Zahvaća li takvo znanstveno istraživanje doista sve potencijale zatečenog stanja ili ono zahvaća samo neke njegove aspekte (a vodeći računa o nerijetkoj pojavi da je istraživanje vrijednosno usmjereno u skladu sa zahtjevima naručitelja i financijera istraživanja, koliko god znanstveno istraživanje proklamativno – pa i po samorazumijevanju – bilo objektivno i vrijednosno neutralno te stoga ono osvjetljava i afirmira samo one tendencije koje su poželjne sa stajališta interesa spomenutih financijera i naručitelja, najčešće one tendencije koje vode učvršćivanju *statusa quo*)?*

Kakav bi se zaključak iz naznačenoga mogao izvesti? On bi se mogao formulirati u obliku retoričkog pitanja: *Ne bi metodologiju znanstvenog istraživanja vrijedilo osloboditi od apsolutizacije kvantitativnoga?*

---

kazao Goebbels, primijenivši Bernaysove marketinške metode na širenje antisemitizma i općenito nacionalsocijalističke ideologije (usp. Veljak 2018), dok se danas primjenjuje u (barem naizgled) manje zloćudnim oblicima poželjnog oblikovanja javnog mnijenja. U slučaju uspješnosti takve kampanje, rezultati (forenzičko-financijskih) istraživanja pokazat će rast profita i bogatstva spomenute interesne skupine i njezinih pripadnika u periodu uspješnog provođenja kampanje. Činjenica da se takva istraživanja (kada se uopće i provode) ne dovode u vezu s politologijskim, sociologijskim i sociopsihologijskim istraživanjima etničke distance vjerojatno u prvom redu ovisi o povezanosti financijera znanstveno-istraživačkog pogona s naručiteljima, akterima i dobitnicima kampanja nepovjerenja i mržnje.

## Literatura

- Allen, Sophie R. 2012. „What Matters in (Naturalized) Metaphysics“. *Essays in Philosophy*, 13.
- Aristotel *Metafizika*. Sveučilišna naklada Liber, Zagreb, 1985.
- Asbach, Olaf. 1977. *Von der Erkenntniskritik zur Kritischen Theorie der Gesellschaft*, Leske+Budrich, Opladen.
- Čačinović, Nadežda. 2020. „Varijante redukcionizma“. u: Borislav Mikulić i Mislav Žitko (ur.), *Metafilozofija i pitanja znanstvene metodologije*, Filozofski fakultet Sveučilišta u Zagrebu, Zagreb.
- Dawkins, Richard. 1976. *The Selfish Gene*. Oxford University Press, Oxford.
- Devitt, Michael i Sterelny, Kim. 2002. *Jezik i stvarnost: Uvod u filozofiju jezika*. (preveo Boran Berčić) KruZak, Zagreb.
- Doniger, Wendy (ur.). 1999. *Merriam-Webster's Encyclopedia of World Religions*. Merriam-Webster, Springfield, Mass.
- Godfrey-Smith, Peter. 2003. *Theory and Reality. An Introduction to the Philosophy of Science*. Chicago University Press, Chicago.
- Hahn, Hans. 1975a. „Logik, Mathematik und Naturerkennen“. u: Schlechter, Hubert (ur.). *Logischer Empirismus - Der Wiener Kreis*. Fink, München.
- Hahn, Hans. 1975b. „Überflüssige Wesenheiten“. u: Schlechter, Hubert (ur.). *Logischer Empirismus - Der Wiener Kreis*. Fink, München.
- Hahn, Hans; Neurath, Otto i Carnap, Rudolf. 1975. „Wissenschaftliche Weltauffassung – Der Wiener Kreis“. u: Schlechter, Hubert (ur.). *Logischer Empirismus - Der Wiener Kreis*. Fink, München.
- Heisenberg, Werner. 1997. *Fizika i filozofija*. (preveo Stipe Kutleša) KruZak, Zagreb.
- Jakslund, Rasmus. 2016. *The Possibility of Naturalized Metaphysics*. Institute for Media, Cognition, and Communication, University of Copenhagen, Copenhagen.
- Joll, Nicholas. „Metaphilosophy“. *Internet Encyclopedia of Philosophy*, <https://www.iep.utm.edu/con-meta/> (posjećeno 2. 2. 2021.)
- Ladyman, James et al. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press, Oxford.
- Leplin, Jarrett. 1997. *A Novel Defense of Scientific Realism*. Oxford University Press, Oxford.
- Losee, John. 1977. *A Historical Introduction to Philosophy of Science*. Oxford University Press, Oxford.
- Marion, Jean-Luc. 1989. *Réduction et donation*. P.U. F., Paris.
- Mikulić, Borislav i Žitko, Mislav (ur.). 2020. *Metafilozofija i pitanja znanstvene metodologije*, Filozofski fakultet Sveučilišta u Zagrebu, Zagreb.
- Milić, Mladen. 2014. *Odlučivanje i poslovna inteligencija*. PIM, Banjaluka.
- Neurath, Otto. 1933. *Einheitswissenschaft und Psychologie*. Gerold & Company, Wien.
- Ney, Alyssa. „Reductionism“. *Internet Encyclopedia of Philosophy*. <https://www.iep.utm.edu/red-ism/> (posjećeno 11. 5. 2020.)
- Poland, Jeffrey. 1994. *Physicalism. The Philosophical Foundations*. Clarendon Press, Oxford.

- Polkinghorne, John C. 2002a. „Riduzionismo“. u: *Documentazione Interdisciplinare di Scienza e Fede (DISF)* <http://www.disf.org/riduzionismo> (posjećeno 1. 2. 2021.)
- Polkinghorne, John C. 2002b. „Riduzionismo“. u: Tanzella-Nitti, Giuseppe i Strumia, Alberto (ur.). 2002. *Dizionario interdisciplinare di scienza e fede*. Urbaniana University Press – Città Nuova Editrice, Roma.
- Polkinghorne, John C. 2019. *Vjera u Boga u doba znanosti*. (preveo Miroslav Furić) Verbum, Split.
- Robinson, Howard. 2003. „Dualism“. u : Stich, Stephen P. i Warfield, Ted A. (ur.). *The Blackwell Guide to Philosophy of Mind*. Blackwell, Malden, str. 85-101.
- Rose, Hilary i Rose, Steven (ur.). 2001. *Alas, Poor Darwin: Arguments Against Evolutionary Psychology*. Jonathan Cape, London.
- Schleicher, Hubert (ur.). 1975. *Logischer Empirismus - Der Wiener Kreis*. Fink, München
- Sikavica, Pere et al. 2014. *Poslovno odlučivanje*. Školska knjiga, Zagreb.
- Smokrović, Nenad. 2012. „Logical Consequence and Rationality“. u: Trobok, Majda; Mišević, Nenad i Žarnić, Berislav (ur.). *Between Logic and Reality*. Springer, Dordrecht, str. 123-133.
- Smokrović, Nenad. 2015. „Argumentation as a Means for Extending Knowledge“. *Croatian Journal of Philosophy*, 15, 2 (44), str. 223-232.
- Smokrović, Nenad. 2017. *Znanost i metoda, Od opće metodologije do metodologije ekonomije*, KruZak, Zagreb.
- Stich, Stephen P. i Warfield, Ted A. (ur.). 2003. *The Blackwell Guide to Philosophy of Mind*. Blackwell, Malden.
- Trobok, Majda; Mišević, Nenad i Žarnić, Berislav (ur.). 2012. *Between Logic and Reality*. Springer, Dordrecht.
- Tye, Michael. 2009. *Consciousness Revisited: Materialism Without Phenomenal Concept*. MIT Press, Cambridge, Mass.
- Veljak, Lino. 1979. *Marksizam i teorija odraza*. Naprijed, Zagreb.
- Veljak, Lino. 2018. „Zastarjelost istine“. *In Medias Res*, br. 12, str. 1799-1806.
- Veljak, Lino. 2020. „Metafilozofija i redukcionizam u metodologiji društvenih istraživanja“, u: Borislav Mikulić i Mislav Žitko (ur.) *Metafilozofija i pitanja znanstvene metodologije*, Filozofski fakultet Sveučilišta u Zagrebu, Zagreb.





MICHAEL WATKINS

## The Mastery of a Concept: Dispositions and Skills

**Abstract:** Nenad Smokrovic defends the view that understanding or grasping a concept is constitutively tied to being disposed to assent to certain sentences (henceforth, the “constitutive account”), and in particular to sentences that we might reasonably take to be analytically true. Most notably, he defends the constitutive account against purported counterexamples presented by Timothy Williamson. I argue that we can only understand Smokrovic’s response to Williamson if we have some way to tell the difference between a counterfactual situation in which you obtain a disposition, and a disposition that you currently have that would be manifested in a counterfactual situation. I think it is unlikely that that we can understand that distinction unless we treat dispositions non-reductively, and it is hard to see what is gained by Smokrovic’s account if that is the case.

**Key words:** Williamson, concepts, dispositions.

In a series of papers, Nenad Smokrovic defends the view that understanding or grasping a concept is constitutively tied to being disposed to assent to certain sentences (henceforth, the “constitutive account”), and in particular to sentences that we might reasonably take to be analytically true. Here I will focus on one such paper, “Are Dispositions to Believe Constitutive for Understanding,” in which Smokrovic defends the constitutive account against purported counterexamples presented by Timothy Williamson. My aim here is not to settle the debate between Smokrovic and Williamson. My suspicion is that the debate cannot be settled, at least not in the terms set by Smokrovic and Williamson. Indeed, I will argue that despite Smokrovic’s valiant and clever attempts to defend the constitutive account, what emerges finally from the debate is that the debate, as Smokrovic understands it, cannot be settled. I’ll conclude with some suggestions as to why.

### 1

To begin, we need to get clearer about the position which Smokrovic wishes to defend. He tells us that according to the constitutive account, there is a constitutive link between understanding a concept and assenting, or being disposed to assenting, to a sentence. As Williamson and Smokrovic note, there are stronger and weaker versions of this position. But mini-

mally, and as an example, Smokrovic tells us the constitutive account accepts that, necessarily, whoever understands the sentence “every vixen is a vixen” has a disposition to assent to it (p. 2). Now we need to do a bit of unpacking. The claim is not that to understand a sentence is to be disposed to assent to it; although Smokrovic at times suggests as much, this cannot be what he means. I perfectly well understand the sentence “ $2+2=5$ ”, but I never (or hardly ever) assent to it since I know it to be false. And I might perfectly well understand the sentence “no mammal lays eggs” whether I assent to it or not. So I assume that what is being claimed is that, necessarily, someone understands a sentence that is conceptually necessarily true only if that person is disposed to assent to it; and being disposed to assent to such sentences is constitutive of understanding the concepts expressed by that sentence. I assume, too, that the view also entails that, necessarily, someone understands a sentence that is conceptually necessarily false only if that person is disposed to dissent to it; and being disposed to dissent to such sentences is constitutive of understanding the concepts expressed by that sentence.

Now as I have put the constitutive position, the claim is that being disposed to assent to “every vixen is a female fox” is only partially constitutive since I treated the claim as necessary, but not sufficient, for understanding. Smokrovic is committed only to the necessary condition, although he seems tempted by both. In any case, let’s put that aside. To commit to the necessary condition is trouble enough.

The constitutive position is *prima facie* plausible since a neighboring claim seems trivially true. If I assent to “ $2+2=5$ ” and dissent to “ $2+2=4$ ”, that is surely strong evidence that I don’t understand addition. If I assent to “all mammals are feathered animals”, then that is surely strong evidence that I don’t understand what is meant by ‘mammal’, or ‘feathered’, or ‘animal.’ Of course, you may understand the concept *mammal* although you assent to “all mammals are feathered animals”; perhaps you understand Croatian, but very little English (the reverse of me). You not being disposed to dissent to “all mammals are feathered animals” perhaps shows that you do not understand English; your being disposed to dissent to “*svi su sisavci pernate životinje*” is evidence that you understand the concepts; that you are such that *if* you learn English, then you *would be* disposed to dissent to “all mammals are feathered animals” is evidence that you understand the relevant concepts. Or should the last be that you *are* disposed to dissent to “all mammals are feathered animals” *if* you learn English? There is a puzzle here, and it relates to the debate between Smokrovic and Williamson. How do we tell the difference between a counterfactual situation in which you obtain a disposition, and a disposition that you currently

have that would be manifested in a counterfactual situation? Likewise, how do we tell the difference between a disposition that is not being manifested because it is blocked, and a situation that eliminates that disposition? The importance of these distinctions will be obvious momentarily. For now, I put them aside.

For now I want only to make two observations. First, you might think it certainly true that for anyone to understand a concept, they have to know “how to go on” using that concept, as Wittgenstein puts it, but not accept the constitutive position. If you think that having mastered a concept is simply having mastered a certain skill, then you might well countenance that people often more or less master a concept, and so someone might understand a concept but not be disposed to assent to every analytically true statement (or dissent from every analytically false statement) employing that concept. Second, and independently of the first, you might think that whether someone who has mastered a concept is disposed to assent to every analytic sentence employing that concept depends upon what else she believes. This will be especially true for those skeptical of the analytic-synthetic distinction. Anyway, let the above serve as an introduction. Now to work.

## 2

Williamson presents various cases intended as counterexamples to the constitutive account. I will focus on one. According to Williamson, the defender of the constitutive account is committed to the view that, for anyone who understands the concept *vixen*, if that person understands English, then that person is disposed to assent to “every vixen is a female fox.” But now imagine Peter, Williamson’s fictional character. Peter is a fox-skeptic due to his having fallen prey to a conspiracy theory. Peter also believes, along with Aristotle, that universal quantification carries with it an existential commitment. So Peter believes that if “every vixen is a female fox” is true, then “there is at least one fox” is true. Since he denies the consequent, he denies the antecedent. Williamson concludes that although Peter understands the concept *vixen*, he is not disposed to assent to “every vixen is a female fox.”

Now you might be skeptical that the case of Peter can tell us anything about the constitutive position. After all, “every vixen is a female fox” is not, as Peter understands it, analytically true. Since, as Peter understands it, “every vixen is a female fox” has existential import, the sentence is empirically true, although Peter thinks it is empirically false. Peter’s denial, we might think, tells us nothing about his conceptual mastery. What Peter denies, we might think, is not that all vixen are foxes, but (as we would put

it) that all vixen are foxes *and* there is at least one fox. But what then constitutes his understanding of the concept *vixen*? What must he be disposed to assent to assuming both that Peter understands the concept *vixen* and that the constitutive account is true? After all, by Peter's lights, we have not properly understood his denial of "every vixen is a female fox". The suggestion was to think that what Peter actually denies is the conjunction "all vixen are foxes and there is at least one fox," and that is compatible with his believing that every vixen is a female fox. But according to Peter, "every vixen is a female fox and there are no foxes" is a contradiction.

Smokrovic's strategy is to argue that, *pace* Williamson, Peter is disposed to assent to "every vixen is a female fox." Peter's disposition to assent is simply "defeated" or "overridden" (to use Smokrovic's terms), but not eliminated. This takes us back to the worry that I raised towards the end of §1. Intuitively, there is a difference between having a disposition that is no longer manifested even under appropriate conditions, as opposed to something's once having had a disposition that it no longer has. To take an example of Saul Kripke's, we can imagine an object that is painted yellow, but the paint is such that immediately upon looking at it, and prior to having the appropriate experience, you are killed. Anything painted killer yellow might be thought to have the disposition to look yellow, even if it never would look yellow. Perhaps we might say that it is disposed to look yellow, but that disposition is defeated. The problem is that there are no conditions under which it would look yellow. So it would seem that any justification we might have for thinking that the thing is yellow is not its being disposed to look yellow (it never will look yellow to anyone ever!), but that it has some property that we've learned is, or is a determinate of, yellowness. The color now explains the disposition; and so the disposition must not be the color. This, of course, takes us into the rather thick weeds of the metaphysics of dispositions. I will not go there here, having said what I have to say elsewhere. I am skeptical that much explanatory work can be accomplished by appeal to dispositions alone, but let's put those thorny issues aside. For although there is much work to be done concerning dispositions that I think must be done before settling this debate between Williamson and Smokrovic, and, indeed, much work that much be done before we can get clear about what that debate concerns, the deeper disagreement between them does not concern the nature of dispositions.

3

So here is how Smokrovic thinks of Peter.

1. Peter acquires the concept *vixen* only when he acquires the disposition to assent to “all vixen are female foxes” and only for so long as he is so disposed (on the condition that he understands English).
2. Peter later acquires other dispositions due to his skepticism about foxes and his belief that universal statements carry existential import.
3. But (2) does not remove Peter’s disposition to assent to “all vixen are female foxes”; it merely prevents that disposition from being manifested.

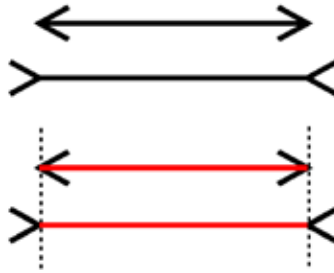
Williamson thinks that Peter’s skepticism and his commitments about universal statements result in Peter’s no longer being disposed to assent to “all vixen are female foxes.”

Let’s look at another example. Steve is disposed to assent to “if *p*, then  $\sim p$  is a logical contradiction”. But Steve knows how to construct truth-tables. He knows, more particularly, that the material conditional is true unless the antecedent is true and the consequent false. And he knows that a sentence is contradictory if and only if every line in that sentence’s column is false. After constructing the truth-table, he realizes that ‘if *p*, then  $\sim p$ ’ is not contradictory. So did Steve not understand what contradictions are until he completed the truth-table, or did he have the disposition to dissent to “if *p*, then  $\sim p$  is a logical contradiction” (if only his disposition had not been defeated)? I’m not sure that there is fact of the matter.

The more important point about Steve is that he seems to understand what a contradiction is, at least more or less. He seems somewhat competent. And this, I take it, is the deeper disagreement between Smokrovic and Williamson. Smokrovic is certainly correct that understanding a concept requires being able to make certain connections, and that brings with it a set of dispositions. But is there *a* disposition necessary for understanding a concept. Is there a linguistic test, a *shibboleth*, for whether someone understands any particular concept? With Williamson, I am skeptical.

Thus far I have done little but put forward competing intuitions. Again, the question is, in part, whether it is more plausible to believe that Peter’s disposition to assent to “all vixen are female foxes” is being blocked, or whether it is more plausible to believe that he is simply not so disposed. Since, as a matter of fact, he will not assent, there is at least a *prima facie* case for his not being disposed to assent. So Smokrovic owes us an argument, and he accepts the challenge. He presents an argument, relying in part on an analogy with perception.

Imagine looking at a standard Muller-Lyer illusion:



You are disposed, we might say, to see the top line as shorter than the bottom line and so to assent to “the top line is shorter than the bottom line.” Of course, most of us who are aware of this illusion will not assent to “the top line is shorter than the bottom line.” Being aware that it is an illusion, we have beliefs that block our disposition to assent to “the top line is shorter than the bottom line.”

Smokrovic claims that Peter’s case is analogous. According to Smokrovic, although Peter believes that *there are no foxes* (and so no vixens), he is still far from having the belief that *‘every vixen is a vixen’ is not true*. To dissent from “every vixen is a vixen” he must have a particular metasemantic theory. According to Smokrovic: “his dissenting is dissenting from the semantic form of the sentence, and only secondarily about vixen’s properties” (p. 9).

It seems to me that this cannot be quite right. What Peter dissents from, when he dissents from “every vixen is a vixen,” cannot be at a linguistic level above what he assents to, when he assents to “there are no foxes.” Peter’s mistakes do not include a use-mention error. But Smokrovic can make his point, I think, without insisting that Peter is even more confused than he actually is. What Smokrovic needs is only that Peter is disposed to assent to “all vixen are female foxes” just as we are disposed to assent to “the top line is shorter than the bottom line”; but just as our disposition is rightly blocked by good reasoning, Peter’s is blocked by bad reasoning. The question remains, however, is this true? In the Muller-Lyer case, the justification for thinking that we have a disposition that is blocked is that, as a matter of fact, the top line continues to look shorter than the bottom line even though we know it isn’t. And it is not obvious that we have anything like this kind of justification to treat Peter as Smokrovic insists that we should.

Think of another case. I do not believe that there are dragons, but I know that ‘drakaina’ means *female dragon*. I am disposed to assent to “a drakaina is a female dragon.” So far, so good. But I also believe not only that there are no dragons, but necessarily there are no dragons. Dragons,



being mythical, are necessarily mythical. At least, that is what I believe. So I will not assent to “if there were a drakaina, then there would be a female dragon.” I understand that sentence as claiming that in every possible world (or at least in the closest possible world) in which there is a drakaina, there is a female dragon. But there are no such worlds. Of course, given my views, it is simply not analytic (indeed, it is necessarily false) that “if there were a drakaina, then there would be a female dragon.” But the sentence “every vixen is a female fox” seems no more or less analytic than “if there is a vixen, then there is a fox.”

#### 4

In so far as dispositions are evidenced by their manifestations, as seems to be the case with the Muller-Lyer illusion, Steve would seem not to mean by ‘contradiction’ what we do. And that seems false. Steve makes a mistake; he doesn’t make a claim that is true using a concept different than ours. Moreover, we seem at a loss to say whether Peter retains a disposition that is blocked, or whether he simply does not have the disposition. If dispositions are instead thought of as the powers bestowed by properties, then to know what ‘vixen’ means is simply to know what ‘vixen’ means. This brings us back, once again, to worries about the metaphysics of dispositions. And I don’t see a way out of the dispute between Smokrovic and Williamson, at least as Smokrovic casts that debate, that avoids those worries.

But there is another way to think about this debate, and I think it favors Williamson. Imagine that we have two competing theories of what it is to master a concept. According to the first, to master a concept is to be disposed (whatever that comes to) to assent to any analytically true statement employing that concept. According to the second, mastering a concept is like mastering a jump shot in basketball; it is to have acquired a skill. The second theory predicts that, just as someone who has mastered a jump shot will sometimes miss, someone who masters a concept will sometimes apply the concept incorrectly. Mastery of a skill admits of degrees, and perfect mastery (or, at least, all of the mastery that is humanly possible) falls short of perfection. The second theory predicts that cases like Peter and Steven will sometimes occur. It predicts that mastery of a concept is consistent with misapplications and confusions in applying that concept. The second theory predicts, in other words, exactly what we find to be the case.



TIMOTHY WILLIAMSON

## Idealized Rationality in Models of Knowledge and Probability

**Abstract:** The chapter opens with a brief appreciation of Nenad Smokrović, his work, and the author's intellectual interactions with him. It continues with his preferred theme of rationality by discussing different levels at which rationality assumptions, for instance of 'logical omniscience', seem to be built into standard models of epistemic and doxastic logic: through their treatment of propositions as sets of (classical) worlds, through their semantic clauses for evaluating knowledge and justified belief ascriptions over models, through structural constraints on the doxastic or epistemic accessibility relation of a model, and through the reliance on 'toy' models. At stake are the relative contributions of idealization, simplification, stipulation, and error-theoretic accounts of attitude ascription. Consequences for the debate between intensionalist and hyperintensionalist theories of attitude ascription are assessed, and comparisons made with corresponding issues about models of probability.

**Key words:** intensionalism, hyperintensionalism, epistemic logic, doxastic logic, ideal rationality, logical omniscience, model-building.

I first encountered Nenad Smokrović in 2000, at the Austro-Slovene Philosophical Conference in Celje, where the clarity, organization, and good sense of his paper made a very positive impression on me—though not quite as strong an impression, it must be admitted, as that made on me by Ana Mladenović (Williamson), now my wife, whom I first met within minutes of leaving the conference. A few years later, Nenad and I got to know each other properly as co-teachers for a series of summer schools at the Inter-University Centre in Dubrovnik, and on my visit to the University of Rijeka. One of our Dubrovnik interactions led to an exchange in print on a theme from *The Philosophy of Philosophy* (Smokrović 2013; Williamson 2007, 2013). Nenad's professionalism and well-balanced reasonableness have continued to impress me, not to mention his excessive modesty. He has made a very important contribution to the development of analytic philosophy in Croatia and in the wider region. It is a pleasure to contribute to his much-deserved *festschrift*.

As well as exemplifying rationality, Nenad has also made it the focus of his research. In this paper, I will approach his favourite topic by considering models in epistemology which are often charged with giving a hope-

lessly over-idealized picture of rational agents. As so often, the issues turn out to be far more complex than they seem at first sight. My main focus will be on models of epistemic and doxastic logic, but I will sometimes mention analogous issues for models of epistemic and doxastic probability.

## Background

In 1962, Jaakko Hintikka published his seminal book *Knowledge and Belief*, which established epistemic and doxastic logic as a branch of applied modal logic. He explained very clearly its relevance to classic epistemological problems. It was later taken up by computer scientists and theoretical economists, in particular because it was so well-adapted to the analysis of questions about common knowledge and common belief. However, for several decades, Hintikka's work had hardly any impact on mainstream epistemology. The main reason was that it was widely taken to be discredited by its commitment to principles of logical omniscience, according to which agents' knowledge and belief is automatically closed under logical consequence. Understandably, though mistakenly, such principles were felt to be so far beyond the capacities of real agents as to be virtually irrelevant to the serious study of human knowledge and belief. Of course, the limitations of many epistemologists' logical training may have made the conclusion that there was no need to grapple with Hintikka's technicalities come as something of a relief.

Another factor in the reception of Hintikka's book was the absence of a model-building sensibility from most twentieth-century philosophy. Amongst analytic philosophers, the predominant ethos was falsificationist: a principle must be rejected once a counterexample has been found, though modifying the principle with epicycles of arbitrary complexity was permitted. Hardly on the horizon was a model-building methodology, of exploring the consequences of a precise but drastically over-simplified description to gain insight into the complex, messy reality (see Yap 2014, Williamson 2017).

Better known than epistemic or doxastic logic was probability theory. It was most familiar as a branch of mathematics going back to the seventeenth century, but the subjective Bayesian interpretation of probability in the tradition of Frank Ramsey and Bruno de Finetti made probability theory into a kind of graded doxastic logic, and raised similar issues about the contrast between the logical omniscience of ideally rational agents and the computational limitations of real human beings. For mainstream epistemologists, it was less esoteric than epistemic or doxastic logic, but still remained somewhat marginal. Its focus on operationally defined degrees

of belief (credences), instead of psychologically vivid outright belief, and its epistemologically impoverished conception of rationality as formal probabilistic coherence or refusal of Dutch books, did not help. Formal epistemology of the probabilistic sort tended to serve as the last refuge of logical positivism.

More recently, interaction has increased between mainstream epistemology and formal epistemology, comprising both epistemic and doxastic logic and probabilistic (mostly Bayesian) epistemology, and combinations of the two. That has been to the advantage of both sides. Formal epistemology has become more philosophically sophisticated, less driven by internalist dogma. Mainstream epistemology has become more mathematically sophisticated, more able and more willing to use a model-building methodology to test out ideas.

Since assumptions in the vicinity of logical omniscience have played a major role in the difficult relationship between formal epistemology and mainstream epistemology, they merit exploration in more depth. They turn out to arise at four different levels of modelling, in both models of epistemic logic and models of probability theory. At each level, we must ask whether to interpret them as prescribed ideals of rationality, or described structural features of the underlying reality, or mere artefacts of the need for simple, tractable models.

### **Level 1: the space of propositions**

Both in epistemic and doxastic logic and in probability theory, a model is based on an underlying set. Formally, it can be any nonempty set; only the number of its members makes a relevant mathematical difference. Informally, its members are understood as mutually exclusive, jointly exhaustive possibilities (in some loose sense), each maximally specific in all relevant respects. For a model of epistemic or doxastic logic, the underlying set is typically called ‘ $W$ ’, its members are typically called ‘worlds’, and its subsets are typically called ‘propositions’; a proposition  $p \subseteq W$  is *true* at a world  $w \in W$  if and only if  $w \in p$ ; otherwise  $p$  is false at  $w$ . Thus any proposition is either true or false at any world, and not both. For a model of probability theory, the underlying set is typically called ‘ $\Omega$ ’, its members are typically called ‘outcomes’, and its subsets are typically called ‘events’; an event  $e \subseteq \Omega$  *occurs* at an outcome  $o \in \Omega$  if and only if  $o \in e$ . The differences in terminology reflect the different origins of probability theory on the one hand and epistemic and doxastic logic on the other, but are otherwise inconsequential. For present purposes, we will prefer to use the terminology of epistemic and doxastic logic, with worlds, propositions, and truth.

Given the set-up so far, we can define truth-functional operations from propositions to propositions in a natural, easy way.

The negation  $\neg p$  of a proposition  $p$  is the set of worlds in  $W$  but not in  $p$ . Thus the proposition  $\neg p$  is true at a world  $w$  if and only if  $p$  is false at  $w$ .

The conjunction  $p \wedge q$  of propositions  $p$  and  $q$  is the set of worlds in both, in other words  $p \cap q$ . Thus the proposition  $p \wedge q$  is true at a world  $w$  if and only if both  $p$  is true at  $w$  and  $q$  is true at  $w$ .

The disjunction  $p \vee q$  of propositions  $p$  and  $q$  is the set of worlds in either, in other words  $p \cup q$ . Thus the proposition  $p \vee q$  is true at a world  $w$  if and only if either  $p$  is true at  $w$  or  $q$  is true at  $w$ .

The material implication  $p \rightarrow q$  of a proposition  $q$  by a proposition  $p$  is the set of worlds in  $q$  or not in  $p$ . Thus the proposition  $p \rightarrow q$  is true at a world  $w$  if and only if either  $p$  is false at  $w$  or  $q$  is true at  $w$ .

Given these definitions, negation, conjunction, and disjunction on propositions constitute a Boolean algebra, so all the equations of a Boolean algebra hold for all propositions. For example, De Morgan's laws hold:  $\neg(\neg p \wedge \neg q) = p \vee q$  and  $\neg(\neg p \vee \neg q) = p \wedge q$ . These are genuine equations: they assert the *identity* of the propositions, not their mere necessary equivalence. Slightly less familiar examples are the equations  $(p \wedge q) \vee p = p = (p \vee r) \wedge p$ .

Such identities make it imperative to distinguish *propositions* from *sentences*. We must not assume that differences in the *structure* of two sentences project onto differences in the structure of the propositions they express: as sentences, a disjunction and a negated conjunction differ in structure, but they may express the very same proposition. Similarly, we must not assume that differences in the *subject matter* of two sentences project onto differences in the subject matter of the propositions they express: as three linguistic expressions,  $(p \wedge q) \vee p$ ,  $p$ , and  $(p \vee r) \wedge p$  may differ in subject matter— $q$  occurs only in the first, and  $r$  only in the third—but they all express the very same proposition.

Since models of epistemic logic are interpreted as containing a domain of propositions, the natural way to model *knowledge* in them is with an operator  $K$  mapping each proposition  $p$  to a proposition  $Kp$ . Informally, we understand  $Kp$  as the proposition that one knows  $p$ , where 'one' refers to the agent of the model and the present tense of 'know' refers to the time of the model. Similarly, the natural way to model *belief* in models of doxastic logic is with an operator  $B$  mapping each proposition  $p$  to a proposition  $Bp$ , informally understood as the proposition that one believes  $p$ . If we want to model a situation with  $n$  agents, we can introduce  $n$  subscripts for  $K$  or  $B$ , one for each agent, but for present purposes we can ignore the extra complications of multi-agent models, since they make little difference to the issues of logical omniscience.

In models of Bayesian probability theory, we model *probability* with an operator Prob mapping each proposition  $p$  to a real number  $\text{Prob}(p)$ , informally understood as one's probability or credence in  $p$ . For combinatorial reasons, in a model with infinitely many outcomes,  $\text{Prob}(p)$  may not be defined for all propositions  $p$ .

So far we have said nothing specific about the K, B, and Prob operators. Nevertheless, a serious problem of logical omniscience already arises at this level. For operators are just functions, and so are subject to the elementary logical principle that if  $x = y$  then  $f(x) = f(y)$ . Thus if  $p = q$  then  $Kp = Kq$ ,  $Bp = Bq$ , and  $\text{Prob}(p) = \text{Prob}(q)$ . For example, since  $(p \wedge q) \vee p = p$ ,  $K((p \wedge q) \vee p) = Kp$ .

We can work the problem through informally, with respect to a fixed model. By the Boolean equations, sentences [1]-[3] express the same proposition as each other:

- [1] Either it is raining and the vaccine is safe or it is raining.
- [2] It is raining.
- [3] It is either raining or hot and it is raining.

Thus, if we apply to the sentences [1]-[3] the same sentential operator expressing the same propositional operator, the result will be three more sentences expressing the same proposition as each other. For example, if the sentential operator 'Mary knows that' expresses the knowledge operator K on propositions, then the resulting sentences [4]-[6] will express the same proposition as each other:

- [4] Mary knows that either it is raining and the vaccine is safe or it is raining.
- [5] Mary knows that it is raining.
- [6] Mary knows that it is either raining or hot and it is raining.

Similarly, if the sentential operator 'John believes that' expresses the belief operator B on propositions, and we apply it to the sentences [1]-[3], then the resulting sentences [7]-[9] will express the same proposition as each other:

- [7] John believes that either it is raining and the vaccine is safe or it is raining.
- [8] John believes that it is raining.
- [9] John believes that it is either raining or hot and it is raining.

Again, if the sentential operator 'It is 97% probable for Jo that' expresses an operator defined from the probability operator Prob, and we apply it to the sentences [1]-[3], then the resulting sentences [10]-[12] will express the same proposition as each other:



- [10] It is 97% probable for Jo that either it is raining and the vaccine is safe or it is raining.  
[11] It is 97% probable for Jo that it is raining.  
[12] It is 97% probable for Jo that it is either raining or hot and it is raining.

When a sentence expresses a proposition, the truth-value of the sentence should be simply the truth-value of the proposition it expresses. Thus, since the model predicts that the sentences [4]-[6] express the same proposition as each other, it also predicts that [4]-[6] are either all true or all false. Similarly, since the model predicts that the sentences [7]-[9] express the same proposition as each other, it also predicts that [7]-[9] are either all true or all false. Again, since the model predicts that the sentences [10]-[12] express the same proposition as each other, it also predicts that [10]-[12] are either all true or all false.

A very natural reaction is that these predictions can very easily turn out false. It is raining, and Mary is soaked through; she knows that it is raining. Thus [5] is obviously true at the given time. But Mary is in no mood to do pointless exercises in the propositional calculus. If you present her with sentence [1] or [3], she will not know straight away that it is logically equivalent to [2], and will not assent. Thus, one naturally supposes, she lacks the corresponding knowledge, so [4] and [6] are false at the given time. Consequently, [4]-[6] do not all have the same truth-value as each other, contrary to the model's prediction. Of course, there are similar apparent counterexamples to the model's prediction that [7]-[9] are uniform in truth-value, and to its prediction that [10]-[12] are uniform in truth-value.

The screw can be turned with examples where the initial logical equivalences are much harder to work out than they are for [1]-[3], but the point seems clear even without that.

We could avoid these counterexamples by reinterpreting the sentential operators in play. Instead of 'Mary knows that' we could have 'Mary knows something necessarily equivalent to the proposition that', and likewise for 'John believes that' and 'It is 97% probable for Jo that'. But, in themselves, such redefinitions just seem to change the subject. If they make a crucial difference, they do so by substituting for knowledge, belief, and probability other relations less helpful for understanding the agent's actions and reactions. We would find ourselves trying to work out what those other relations told us about the agent, and in effect trying to model the agent's knowledge, beliefs, and probabilities by those other relations. We might as well be honest and admit that what we are really trying to model *is* the

agent's knowledge, beliefs, and probabilities. This paper will be no further concerned with the redefinition strategy, though its possibility is worth noting.

A more fruitful line of thought concerns the individuation of propositions. A critical choice-point in the theory of propositions is between *intensionalism* and *hyperintensionalism*. Intensionalism is the view that necessarily equivalent propositions are identical (for present purposes, we need not take a stand on exactly which kind of necessity is at issue). The negation of intensionalism is hyperintensionalism, on which there are distinct but necessarily equivalent propositions. Examples like [4]-[12] pose an obvious challenge to intensionalism.

Hyperintensionalism comes in several flavours. *Fregean* hyperintensionalists individuate propositions cognitively. Propositions are thoughts, senses of declarative sentences, or modes of presentation (in Frege's case, of a truth-value, but it could also be of a worldly state of affairs). Since the sentences [1]-[3] present things differently (in both structure and subject-matter) to a thinker who understands them, they express different modes of presentation, in other words different propositions, despite their necessary equivalence, so the thinker can have a given attitude to one of them and not to another, so there is no problem in [4]-[6] or [7]-[9] or [10]-[12] being non-uniform in truth-value. By contrast, *Russellian* hyperintensionalists individuate propositions in terms of the worldly objects, properties, and relations they are about. Thus if 'Hesperus' and 'Phosphorus' are analysed as two names for the same astronomical object, the sentences 'Hesperus will explode' and 'Phosphorus will explode' express just one Russellian proposition, but may express two Fregean thoughts in virtue of some subtle cognitive difference in how they present the same astronomical object. Nevertheless, [1]-[3] still express different Russellian propositions, since they differ in structure and subject-matter. Thus a thinker can still have a given attitude to one of them and not to another, so there is still no problem in [4]-[6] or [7]-[9] or [10]-[12] being non-uniform in truth-value.

As the most fine-grained of the views, Fregean hyperintensionalism starts off with an obvious advantage over both Russellian hyperintensionalism and intensionalism in its capacity to treat charitably pre-theoretic assessments of propositional attitude ascriptions such as [4]-[12]. However, it has performed poorly in trying to take advantage of its headstart. On one side, its cognitive criterion of identity for thoughts has led it into individuating them ever more finely in terms of the thinker's state, with no end in sight, a direction of travel in some tension with Frege's insistence

that thoughts can form part of the common heritage of mankind. On the other side, some examples suggest that the focus on modes of presentation fundamentally misconceives how propositional attitude ascriptions work. For example, thinking of John as 'I' and thinking of him as 'you' involve two very different modes of presentation of John. Thus, when Mary says to John 'You think that I am richer than you', we might expect her remark, given Fregeanism, to have a reading on which it is false simply because John does not think of Mary as 'I' and of himself as 'you'; but no such reading is available. The recent revival of hyperintensionalism has mainly concerned the Russellian variety, despite its troubles with 'Hesperus' and 'Phosphorus'.

Both Fregean and Russellian versions of hyperintensionalism face another threat: inconsistency. They are susceptible to forms of the Russell-Myhill paradox. For example, in brief, for any property of propositions there is at least one propositional operator which yields a truth when applied to any proposition with the property and a falsehood when applied to any other proposition, so there are at least as many propositional operators as properties of propositions; but (by a Cantorian diagonal argument) there are more properties of proposition than propositions, so there are more propositional operators than propositions. On the other hand, for any propositional operator a proposition results from applying it to a fixed atomic proposition; moreover, by the structured conception of propositions, different propositions result from applying different propositional operators to the atomic proposition, so there are at least as many propositions as propositional operators. That is a contradiction. Attempts to block such paradoxes lead to implausibly *ad hoc* restrictions.

One more extreme version of hyperintensionalism is also worth mentioning. To capture very fine-grained differences, some theorists introduce 'impossible worlds' (for a recent example see Berto, French, Priest, and Ripley 2018, and for critical discussion Williamson 2020: 256-262). Typically, such worlds are simply arbitrary sets of sentences, which are identified with the sentences true at that world. Thus both a sentence and its negation, or neither, can be true at a given impossible world. In such a set-up, not even [1]-[3] need be true at the same worlds. As a result, the logical expressions in the language no longer correspond to the operations on sets of worlds described above: for example, exactly one of the propositions  $p$  and  $\neg p$ , understood as mutually complementary sets of worlds, is true at any given world, but that does not extend to the sentences ' $p$ ' and ' $\neg p$ '. From the perspective of a model-building methodology, this tradition of 'impossible world' models looks retrograde. It requires 'writing in by hand' all the phenomena one wants to obtain, rather than deriving them from a much

simpler basis. Notoriously, one can cheaply model whatever one likes by granting oneself enough degrees of freedom; in doing so, one learns nothing much. Model-building with such impossible worlds grants itself virtually unlimited degrees of freedom, and is correspondingly unexplanatory.

These problems suggest that the meagre explanatory benefits of a hyperintensional theory of propositions do not adequately repay its enormous additional complexity. By contrast, intensionalism is simple, elegant, and immune to Russell-Myhill paradoxes. Robert Stalnaker has defended just such an account of propositions for many years, on which propositions are in effect just sets (or classes) of possible worlds (Stalnaker 1984 and 1999). Thus [1]-[3] literally express the same proposition. The differences in structure and subject matter are features of the sentences, not of the one proposition they all express. Given that propositions are the objects of knowledge, belief, and probability (as they are for Stalnaker), the sentences in each of the trios [4]-[6], [7]-[9], and [10]-[12] also express the same proposition, and so have the same truth-value.

Of course, intensionalists must still explain why the sentences in such a trio can still *appear* to differ in truth-value. On pain of gross implausibility, they must do so while acknowledging that, by normal linguistic standards, the agent *understands* each of the sentences [1]-[3]. The problem is at its most extreme in mathematics, since all true sentences of pure mathematics express necessary truths, and so necessary equivalents of each other, while all false sentences of pure mathematics express necessary falsehoods, and so necessary equivalents of each other. Thus, given intensionalism, at the level of propositions there is just one truth in pure mathematics, and just one falsehood. To make sense of what is going on, the intensionalist can use the idea of knowing, believing, or assigning a probability to a proposition *under a guise*, where in such cases the guise is a sentence (or a sentence in a context, to handle demonstratives and other expressions whose reference depends on context). Thus Mary may know that it is raining under the guise of sentence [2], but not under the guise of sentence [1] or [3]. Then the knowledge-ascribing sentence [5] may seem true, while the knowledge-ascribing sentences [4] and [6] seem false, because in ascribing knowledge and other propositional attitudes we rely primarily on a heuristic which treats sentences (or sentences in contexts) as adequate proxies for propositions. Such a heuristic is quick and easy to use, but not wholly reliable. We are forced to modify it when we have to deal with speakers of another language, and in other tricky cases, but that does not mean that we have reflective access to another standard for attitude ascription which *is* wholly reliable. We may have no better option available than to *improvise* as best we can, more or less opportunistically. Despite all that, the seman-

tics of our attitude ascriptions may be intensional, because that is what makes the best overall sense of our practice with them.

This is not the place to come to a definitive verdict on the dispute between intensionalism and hyperintensionalism. What is of most interest here is the difference between such accounts of propositions in their implications for models of epistemic and doxastic logic and of probability theory.

For hyperintensionalists, models which treat propositions as sets of possible worlds are drastic over-simplifications. They are not *idealizations* in a normative sense, because hyperintensionalism does not make sets of possible worlds somehow ‘better’ than structured propositions. However, the *effect* of the simplification is much less for ideally rational agents, since they are immediately aware of the logical equivalence of [1]-[3], so in each of the other trios the sentences will in fact have the same truth-value when the designated agent is ideally rational, though without logical equivalence.

For intensionalists, by contrast, models which treat propositions as sets of possible worlds are not simplifications at all. They simply reflect the real nature of propositions.

Consequently, in such cases, one cannot determine which features of the model simplify reality without taking a controversial stand on the structure of reality itself.

## Level 2: the definitions of K and B

At the second level, we introduce a new component of epistemic and doxastic models, which will enable us formally to define the knowledge operator K and the belief operator B. For each primitive operator, the corresponding component is a binary ‘accessibility relation’ between worlds (members of  $W$ ). Thus for K a model has an epistemic accessibility relation, and for B a doxastic accessibility relation. These are not usually intended to analyse knowledge and belief in philosophically more basic terms. Rather, we can regard the epistemic accessibility relation  $R$  as *encoding* the epistemic structure of the model, and the definition of K in terms of  $R$  as *decoding* that structure. Similarly, we can regard the doxastic accessibility relation  $S$  as *encoding* the doxastic structure of the model, and the definition of B in terms of  $S$  as *decoding* that structure.

Informally, for any worlds  $w$  and  $x$ ,  $wRx$  if and only if, for all the agent knows in  $w$ , the agent is in  $x$ ; that is, whatever the agent knows in  $w$  is true in  $x$ : in  $w$ ,  $x$  is an epistemically open possibility for the agent.

Again informally, for any worlds  $w$  and  $x$ ,  $wSx$  if and only if, for all the agent believes in  $w$ , the agent is in  $x$ ; that is, whatever the agent believes in  $w$  is true in  $x$ : in  $w$ ,  $x$  is a doxastically open possibility for the agent.

Such informal accounts of epistemic and doxastic accessibility are useful for making philosophical sense of the models, but play no role in the formal development. What do play a formal role are the definitions of K and B in terms of R and S respectively. These definitions have exactly the same structure as each other.

For any proposition  $p$  and world  $w$ ,  $Kp$  is true at  $w$  if and only if  $p$  is true at every world epistemically accessible from  $w$ . More formally:  $Kp = \{w: \forall x(wRx \Rightarrow x \in p)\}$ . More informally: what you know is whatever is true in all possibilities epistemically open for you.

Similarly, for any proposition  $p$  and world  $w$ ,  $Bp$  is true at  $w$  if and only if  $p$  is true at every world doxastically accessible from  $w$ . More formally:  $Bp = \{w: \forall x(wSx \Rightarrow x \in p)\}$ . More informally: what you believe is whatever is true in all possibilities doxastically open for you.

Irrespective of any specific features of the accessibility relations R and S, these definitions immediately imply a further dose of logical omniscience. We will examine this in detail for the case of knowledge; the case of belief is exactly analogous.

First, we define some useful relations between propositions, all relative to a fixed model. A proposition  $p$  *entails* a proposition  $q$  if and only if  $p$  is a subset of  $q$ , in other words, at every world at which  $p$  is true,  $q$  is true too. A set  $\Pi$  of propositions entails a proposition  $q$  if and only if the conjunction of propositions in  $\Pi$  entails  $q$ , in other words, the intersection of  $\Pi$  is a subset of  $q$ , that is, at every world at which every proposition in  $\Pi$  is true,  $q$  is true too. We also define  $K\Pi = \{Kp: p \in \Pi\}$ , in other words, the propositions in the set  $K\Pi$  together say in effect that the agent knows every proposition in  $\Pi$ .

We can now define an *unqualified multi-premise closure principle for knowledge*. It says that for any set  $\Pi$  of propositions and proposition  $q$ , if  $\Pi$  entails  $p$ , then  $K\Pi$  entails  $Kp$ . In other words, if premises entail a conclusion, and the agent knows every premise, then she knows the conclusion too.

The proof of multi-premise closure is immediate from the definition of K. If every proposition in  $K\Pi$  is true at a world  $w$ , so  $Kp$  is true at  $w$  for every proposition  $p$  in  $\Pi$ , so  $p$  is true at any world  $x$  epistemically accessible from  $w$  for every proposition  $p$  in  $\Pi$  (by definition of K), so  $q$  is true at  $x$  if  $\Pi$  entails  $q$ , in which case  $q$  is true at any world  $x$  accessible from  $w$ , so  $Kq$  is true at  $w$  (by definition of K again). Thus  $K\Pi$  entails  $Kp$ , as required.

Notably, no additional assumption was made to the effect that the agent is aware of the entailment or has deduced  $q$  from premises in  $\Pi$ ; the connection is automatic. In knowing the premises, the agent *ipso facto* knows the conclusion. This is a much bolder principle than those usually contem-

plated in the epistemological literature on closure principles for knowledge (for a debate on which see Dretske 2014a and 2014b and Hawthorne 2014).

We can note some salient cases of unqualified multi-premise closure for knowledge. Each of them goes beyond the closure principle noted at level 1, which is in effect the closure of knowledge under necessary equivalence. These are in effect new forms of logical omniscience. We state them informally.

First, since the set  $\{p \rightarrow q, p\}$  entails  $q$ , whenever the agent knows  $p \rightarrow q$  and knows  $p$ , the agent also knows  $q$ . In other words, the agent's knowledge is closed under modus ponens.

Second, since  $\{p, q\}$  entails  $p \wedge q$ , and  $p \wedge q$  entails  $p$  and entails  $q$ , whenever the agent knows  $p$  and knows  $q$ , the agent knows  $p \wedge q$ , and vice versa. In other words, the agent's knowledge is closed under the introduction and elimination rules for conjunction. Closure under the introduction rule, like closure under modus ponens, shows that the agent's knowledge is not *compartmentalised*: the agent combines any two parts of their knowledge.

Third, since  $p$  entails  $p \vee q$ , whenever the agent knows  $p$ , the agent also knows  $p \vee q$ . Thus the agent's knowledge is closed under the introduction rule for disjunction. This is striking because the premise does not specify any relation between the agent and  $q$ , which could be any proposition whatsoever.

Fourth, since the empty set of propositions entails any tautology, the agent knows any tautology. Thus the model is *not* vacuously satisfied by an agent who knows nothing whatsoever. However ignorant agents are in ordinary respects, they must at least know some logic!

Since the definition of the belief operator B in terms of the doxastic accessibility relation S has exactly the same structure as the definition of the knowledge operator K in terms of the epistemic accessibility relation R, belief has all these new closure properties in doxastic models too. The agent satisfies the unqualified multi-premise closure principle for belief, so the agent's beliefs are closed under modus ponens, the introduction and elimination rules for conjunction, and the introduction rule for disjunction, and they include all tautologies.

Unsurprisingly, the probability operator 'the probability for one is at least  $x$  that' does not generally satisfy a corresponding unqualified multi-premise closure principle except when  $x = 0$  (trivially) and when  $x = 1$  (for finitely many premises), for the conjunction of two propositions may be less probable than either of them. However, such a probability operator does satisfy an unqualified principle of *single*-premise closure: if  $p$  entails  $q$ , then the probability of  $q$  is at least as high as the probability of  $p$ .



Hyperintensionalists about the individuation of propositions are likely to see these stronger closure principles for knowledge, belief, and probability as extreme idealizations, holding at best of logically perfect agents.

By contrast, although intensionalism about the individuation of propositions by itself does not mandate the new closure principles, intensionalists are already committed to the situation with propositional attitude ascriptions being so different from what it seems at first sight that they may be willing to bite the bullet and accept them. Stalnaker has even made a more positive case for unqualified multi-premise closure principles as corollaries of a functionalist account of mental states such as belief (Stalnaker 1984 and 1999).

However, if a non-Stalnakerian intensionalist holds that knowing or believing  $p$  involves knowing or believing  $p$  under some guise or other, then even unqualified single-premise closure may fail for knowledge and belief. A simple case is disjunction introduction. By single-premise premise closure, an agent who knows or believes  $p$  also knows  $p \vee q$ , for any proposition  $q$ . Given the proposed role for guises, an agent who knows or believes  $p$  under some guise also knows or believes  $p \vee q$  under some guise. A Neolithic agent knows and believes under some guise, perhaps a sentence in her language, that there is a bear in that cave. But under what guise does she know or believe that either there is a bear in that cave or in 2021 there is a Large Hadron Collider near Geneva? Similar issues arise for her distribution of probabilities. Thus, for an intensionalist who requires guises, the models' definitions of the knowledge, belief and perhaps probability operators may involve an idealization to agents who have guises for all propositions in the model. Even if an agent does have a guise for  $p \vee q$  in her repertoire, she is unlikely to deploy it instantaneously. Thus, at a given moment, she may count as *already* knowing or believing  $p$  under some guise, but as *not yet* knowing or believing  $p \vee q$  under any guise, thereby violating single-premise closure for knowledge or belief at that moment. Thus the models may embody an even stronger idealization to agents who have guises for all propositions and deploy them instantaneously.

On the simplest Stalnakerian view, knowledge, belief, and probability require no such guises, the violations of closure are merely apparent, and in these respects the models are not idealized. However, one may worry that the simple Stalnakerian view will in the end be forced to make correspondingly extreme idealizations elsewhere, perhaps in relating such cognitive states to action. In determining how to act, even quite simple creatures do some sort of reasoning from their knowledge, beliefs, or probabilities; it is unclear how that can happen if they have nothing like a guise or vehicle for a proposition at issue.



### Level 3: constraints on accessibility

Normally, theorists impose some constraints on the accessibility relations for knowledge and belief, prior to building particular models. Such constraints may be compared to the Kolmogorov axioms for probability, which are usually taken for granted in advance of selecting a specific probability distribution.

In epistemic logic, the most obvious example is the factiveness of knowledge: the truth of  $p$  is a precondition for knowing  $p$ . Thus, for any world  $w$ , whatever one knows at  $w$  is true at  $w$ : in other words, every world is epistemically accessible from itself. In consequence, the epistemic accessibility relation  $R$  is standardly required to be reflexive. Thus, by the definition of  $K$ ,  $Kp$  entails  $p$  for every proposition  $p$  in the model. This is not an idealization or simplification, just an obvious feature of knowledge.

Since false belief is all too possible, the doxastic accessibility relation  $S$  is not required to be reflexive. However, it is often required to be *serial*, in the sense that for any world  $w$ ,  $wSx$  for *some* world  $x$ : there is always at least one doxastically accessible world. Every reflexive relation is serial, but not every serial relation is reflexive. Seriality in effect requires the agent's beliefs to be consistent, since whatever the agent believes is true at any doxastically accessible world. Thus  $Bp$  entails  $\neg B\neg p$ : if you believe a proposition, you do not also believe its contradictory. This does look like an idealization, since inconsistency in one's beliefs seem to be quite common, even in a great logician such as Frege.

If we lift the restriction to consistent belief, we allow worlds where no world is doxastically accessible. The definition of  $B$  then implies that the agent vacuously believes all propositions whatsoever in such a world, since vacuously they are all true in all doxastically accessible worlds. But is it plausible that the price of ordinary inconsistency in belief is believing all propositions whatsoever? We seem able to keep our inconsistencies local. For hyperintensionalists, this is another gross simplification built into doxastic models' treatment of belief: it disallows the localization of inconsistency. Even intensionalists who require every believed proposition to be believed under some guise or other can allow local inconsistencies. If one believes each proposition in an inconsistent set under a guise, it does not follow that one believes *every* proposition under a guise; as already noted at level 2, one may have no guise for some propositions. By contrast, the simple Stalnakerian view, which does not require guises, cannot allow for inconsistency in belief, on pain of collapsing the agent's beliefs altogether. This opens an uncomfortable gap between what agents *say*, to themselves as well as to others, by normal standards sincerely, and what they *really* believe. Again, this is not the place to resolve such theoretical disputes.

In the tradition of epistemic logic, much stronger constraints are often imposed on the epistemic accessibility relation. In particular, it is often required to be an equivalence relation (reflexive, symmetric, and transitive), and so to partition the worlds into equivalence classes (see for example the classic text Fagin, Halpern, Moses, and Vardi 1995). Transitivity is equivalent to the positive introspection principle that if you know, you know you know ( $Kp$  entails  $KKp$ ). Together, the constraints also yield the negative introspection principle that if you don't know, you know you don't know ( $\neg Kp$  entails  $K\neg Kp$ ). In the presence of reflexivity, negative introspection entails positive introspection.

It is easiest to see the problem for negative introspection. Since we are not sceptics, we allow a *good case*, a world  $g$  in which you know for example that you have hands. But sceptical scenarios are not metaphysically impossible, so there is also a *bad case*, a world  $b$  indiscriminable from the inside from the good case but in which you falsely believe that you have hands. In both the good and the bad case, it seems to you that you are in the good case. Thus, in the bad case, you don't know that you have hands, but you also don't know that you don't know that you have hands, because it seems to you that you *do* know that you have hands. Thus negative introspection fails in the bad case.

We can capture all this in an epistemic model whose worlds include  $g$  and  $b$ . Let  $h$  be the proposition that you have hands. Thus  $Kh$  is true in  $g$  but  $h$  is false in  $b$ . Consequently, by the definition of  $K$ ,  $g$  does not have  $R$  to  $b$ : the bad case is epistemically inaccessible from the good case. However, in the bad case, for all you know, you are in the good case: the good case is epistemically accessible from the bad case;  $b$  has  $R$  to  $g$ . Thus epistemic accessibility is not symmetric. This failure of symmetry is typical of sceptical scenarios, whether mild or extreme. Correspondingly, negative introspection fails in the bad case. For  $\neg Kh$  is true in  $b$  (since  $\neg h$  is true in  $b$  and  $bRb$ ), but  $K\neg Kh$  is false in  $b$  (since  $Kh$  is true in  $g$  and  $bRg$ ). In the bad case, you don't know and you don't know that you don't know. Thus assuming negative introspection seems to prevent us from understanding what sceptical arguments are about. Hintikka already discussed such examples in his 1962 book.

Such a model of the failure of negative introspection shows that it is no failure of logical omniscience, for unqualified multi-premise closure holds in the model. The failure does not even show the agent to be less than perfectly rational, at least in their cognitive dispositions. For even an agent with perfectly rational dispositions can fall victim to a sceptical scenario: only an agent who refuses opportunities for knowledge in the good case is guaranteed not to be deceived in the bad case, and refusing opportunities

for knowledge does not seem perfectly rational. The agent in the bad case has less than perfect self-knowledge, but it is not for want of trying.

The non-symmetry of epistemic accessibility does not prove its non-transitivity; other considerations are needed to refute positive introspection. One warning sign of its strength is that positive introspection automatically iterates: it makes  $Kp$  entail  $K^n p$ , where  $K^n$  is a sequence of  $n$   $K$ 's, for any finite number  $n$ . If knowing  $p$  implies knowing  $p$  under some guise, then positive introspection is immediately problematic. For most ordinary propositions  $p$  can be true without being known, making  $Kp$  a distinct proposition from  $p$ , so an agent might know  $p$  under a guise without having or at least without deploying a guise for  $Kp$ , and so on the envisaged view without knowing  $Kp$ . For Stalnaker, by contrast, one can know  $p$  without having a guise for  $p$ , and he defends the literal truth of positive introspection, though not of negative introspection. Others defend more qualified versions of positive introspection: for example, that if one knows  $p$  and has the concept *know*, one is *in a position to* know that one knows  $p$ . Elsewhere, I have argued that the problems for positive introspection go much deeper than this, and come from the very nature of knowledge (see Williamson 2000: 114-131 and 2021a; the latter has references to recent defences of positive introspection). I will not repeat those arguments here.

A natural conclusion is that both positive and negative introspection are *modelling assumptions*, simplifying principles which are not universally true but which may usefully be assumed for some limited purposes. Indeed, there is a specific point to assuming both introspection principles in many applications of epistemic logic, especially in theoretical economics (including game theory) and computer science, where modelling is widespread. For those applications concern *multi-agent* epistemic logic, whose focus is on interpersonal effects, for example, where everyone knows  $p$ , but not everyone knows that everyone knows  $p$  (or some analogue of that in a distributed computing system). For such purposes, it makes sense to idealize the individual agents by assuming that they satisfy both introspection principles, for instance in order to separate interpersonal obstacles to common knowledge from intrapersonal obstacles. Such reasons for building the introspection principles into a model clearly provide no evidence that the principles are universally true.

Unfortunately, many epistemic logicians have not seen things that way. Instead, they have elevated the introspection principles to the status of dogma and gone into denial about the objections to them. This is particularly striking in the case of negative introspection, where the central objection is so well-known and so easy to understand. Part of the reason why philosophers have been so much more alert to the problem is perhaps

that long experience has made them sensitive to implicitly sceptical ways of thinking, and consequently to the danger of making the bad case epistemically accessible from the good case. By contrast, non-philosophers are more willing to make that move in a specific local setting, to keep accessibility symmetric, without seeing the disastrous long-term consequences of that strategy. This provides another example of the contentious issues which can easily be involved in separating modelling assumptions from exceptionless generalizations.

#### Level 4: specific models

The final level is that of idealizations or simplifying assumptions specific to particular models. If intensionalism is correct, this may turn out to be the level at which most such assumptions are made in epistemic and doxastic logic, and in probabilistic modelling.

Of its nature, this level is harder to generalize about. But we can easily illustrate it by considering the problem for negative introspection discussed at level 3, with the good case  $g$  and the bad case  $b$ . We can build a toy model with a domain of just two worlds,  $\{g, b\}$ . As already explained, the epistemic accessibility relation we need for this model is given by:  $R = \{ \langle g, g \rangle, \langle b, g \rangle, \langle b, b \rangle \}$ . With only two worlds, the model has only four propositions. We can easily calculate the effect of the knowledge operator on each proposition:

$$\begin{aligned} K\{g, b\} &= \{g, b\} \\ K\{g\} &= \{g\} \\ K\{b\} &= \{\} \\ K\{\} &= \{\} \end{aligned}$$

This childishly simple model contains within it the whole problem of scepticism, in a nutshell.

The most blatant departure from reality here is the restriction to just two worlds. There is nothing ideal about that; it is just a gross simplification. Obviously there are numerous good cases, numerous bad cases, and numerous cases which are neither good nor bad—for example, where one is in the dark and so knows that one does not know. Moreover, many cases are good in some respects and bad in others, and so on. The toy model ignores all that. Nevertheless, by eliminating all unnecessary complications, it presents the heart of the problem to us in the most easily tractable and manipulable form. It does just what a model should do.

## Bibliography

- Berto, Francesco, French, Rohan, Priest, Graham, and Ripley, Dave. 2018: 'Williamson on counterpossibles', *Journal of Philosophical Logic*, 47: 693-713.
- Dretske, Fred. 2014a: 'The case against closure', in Steup, Turri, and Sosa 2014: 27-40.
- Dretske, Fred. 2014b: 'Reply to Hawthorne', in Steup, Turri, and Sosa 2014: 56-59.
- Fagin, Ronald, Halpern, Joseph, Moses, Yoram, and Vardi, Moshe. 1995: *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.
- Hawthorne, John. 2014: 'The case for closure', in Steup, Turri, and Sosa 2014: 40-56.
- Hintikka, Jaakko. 1962: *Knowledge and Belief*. Ithaca, NY: Cornell University Press.
- Smokrović, Nenad. 2013: 'Are dispositions to believe constitutive for understanding?', *Croatian Journal of Philosophy*, 13: 37-47.
- Stalnaker, Robert. 1984: *Inquiry*. Cambridge, Mass.: MIT Press.
- Stalnaker, Robert. 1999: *Context and Content*. Oxford: Oxford University Press.
- Steup, Matthias, Turri, John, and Sosa, Ernest. 2014: *Contemporary Debates in Epistemology*. Oxford: Wiley-Blackwell.
- Williamson, Timothy. 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.
- Williamson, Timothy. 2007: *The Philosophy of Philosophy*. Oxford: Blackwell.
- Williamson, Timothy. 2013: 'Replies to Trobok, Smokrović, and Mišćević on the philosophy of philosophy', *Croatian Journal of Philosophy*, 13: 49-64. Reprinted in Williamson 2021b.
- Williamson, Timothy. 2017: 'Model-building in philosophy', in Russell Blackford and Damien Broderick (eds.), *Philosophy's Future: The Problem of Philosophical Progress*. Oxford: Wiley-Blackwell, 159-173. Reprinted in Williamson 2021b.
- Williamson, Timothy. 2020: *Suppose and Tell: The Semantics and Heuristics of Conditionals*. Oxford: Oxford University Press.
- Williamson, Timothy. 2021a: 'The KK principle and rotational symmetry', *Analytic Philosophy* 62: 107-124.
- Williamson, Timothy. 2021b. *The Philosophy of Philosophy*, 2<sup>nd</sup> ed. Oxford: Wiley-Blackwell.
- Yap, Audrey. 2014: 'Idealization, epistemic logic, and epistemology', *Synthese*, 191: 3351-3366.

NENAD SMOKROVIĆ

## Acknowledgments, comments and answers

It would be impossible to name all the colleagues, authors, and friends without whom my work would never have come to the stage; calling for such a reflection—be it deserved or undeserved. Thinking of those actively involved in putting together this volume and enriching it with their contributions, I must emphasise that I am aware that the texts in front of me make a body of work that is certainly greater than it deserves.

Therefore, my debt of gratitude is all the greater to each of the contributors: Miloš Arsenijević, Igor Bajšanski, Gabriela Bašić Hanžek, Hanoch Ben-Yami, Boran Berčić, Aleksandra Golubović, Marko Jurjako, Jelena Kopajtić, Paolo Labinaz, Nenad Miščević, Edi Pavlović, Ines Skelac, Matej Sušnik, Danilo Šuster, Majda Trobok, Andrej Ule, Lino Veljak, Michael Watkins, and Timothy Williamson. It is a great honour to be presented by papers from such remarkable theoreticians gathered in one volume.

By no means would this book be possible without the enormous effort of its editors, Boran Berčić, Aleksandra Golubović and Majda Trobok. They started the project and carried it out patiently and cleverly to its end, now available to readers in both an online and a book version.

Regardless of whether they engage with some of my views or their essays are unconnected with them, all the papers in this volume have given me a most valuable incentive to reflect more deeply on my ideas, to enrich my views and to situate my position with clarity. Nevertheless, due to the space restrains, my responses will mostly be directed to those authors that took interest in some of my particular papers, commenting on and criticising them.

There are, of course, some comments and criticisms I gladly agree with, giving just short remarks. There too are some collegial insights that provoke my wider and more careful elaboration of their arguments prompting my alternative views. I sincerely hope that our disagreements will advance the precision of our arguments as well as engender our knowledge of the phenomena we have been discussing for years.

Last but not least, I would also like to express my enjoyment in reading articles and writing the answers to challenges arisen. To some of the articles I will be remarking briefly, while the others I will furnish with more detailed comments. As readers would expect, I will be responding to the contributions in their order of appearance in this volume.

Let me start by expressing my warm thanks to all those authors whose works are not related to my views but who have contributed to this volume with their original, highly valued essays. Having a restricted space, I am in a position to mention them with gratitude only. In her essay *Gabriela Bašić Hanžek* continues to elaborate her long-standing interest in different aspects of pragma-dialectic. This is a fortunate occasion that *Hanoch Ben-Yami* and *Edi Pavlović* whose *Quantified Argument Calculus* (Quarc), being a recognised contribution to contemporary logic, have offered a paper that "after having been in circulation for the better part of the last decade", appears in print for the first time in this volume. *Boran Berčić*, as in most of his works, elaborated a very original and provocative idea. *Aleksandra Golubović* and *Jelena Kopajtić* scrutinise quite interesting educational aspects of critical thinking. *Marko Jurjako* problematises the relationship between the concept of rationality, reasons for action and moral requirements. *Ines Skelac* submits the essay called "The role of logic in human reasoning", furthering the project that, I can say proudly, we initiated together several years ago. *Matej Sušnik's* essay "The nature of practical reasoning" provides a proper insight into differences between theoretical and practical reasoning. *Lino Veljak* in "On foundation of scientific investigation" argues against the thesis of scientific monism. Finally, special thanks go to *Timothy Williamson* whose essay has grossly enriched this volume. Williamson kindly accepted to contribute to this volume offering the modelling approach to epistemology of beliefs and knowledge. Rather than scrutinise a specific problem in the field, he provides in his essay a wider and comprehensive view of the problem of ideal rationality and its consequence, logical omniscience on the one hand and hyperintensional approaches on the other. His model approach opens a vista to a plausible solution to the problem.

*Miloš Arsenijević* finds my proposal I called the logicist thesis, to be a general framework convenient for applying to his two great topics of interest, namely, the problem of future contingencies and the continuum problem (spatial and/or temporal). I owe a particular gratitude to Miša for his excellent review of my logicist thesis (I developed in: Smokrović, 2018), making my point of view much clearer than I would ever be able to do. Having done that, Arsenijević continues by exposing his two theories and concludes that "After all, we have found logical forms which are in congruence both with *everyday reasoning and mathematical analysis*", which is in fact the core of the logicist thesis (p. 16).

*Igor Bajšanski's* essay presents an excellent analysis of the recent theoretical stance in the cognitive science. Here, in an understandable and above all interesting way, but also showing a huge expert knowledge of the



scientific approach to rationality, he points to problems and possible solutions. In the Conclusion Bajšanski relates to my theory of argumentation characterising it as "One of situations in which reasoning can substantially contribute to the efficiency of thinking and this is a specific argumentative context in which interlocutors are sincerely motivated by curiosity, i.e. with the desire to know what's true" (p. 34). His question for me is, "To what extent is this situation typical, and to what extent is it just an idealisation?" (Ibid). I agree that it is an idealisation. My approach is not a description of the most common or the typical situations. It intends to be a kind of a model, a conditional situation of the form: if required conditions are met, the desired form of behaviour (argumentation that extends knowledge) will occur. The required conditions, socio-pragmatic and epistemic, are not typical, but when they are present they will guide the interlocutors to the specific goal.

*Paolo Labinaz* specifically discusses what I have to say about the argumentation process. Effectively and with sharp and clever insight into details he reviews my theory of argumentation, better, I would say, than I myself did. Labinaz proposes a reading of my theory that I indicated but have never fully developed. He points out the weak points in my position and immediately suggests solutions. Giving an account of my view on argumentation, he makes the connection between the socio-pragmatic and epistemic aspect (which is attaining safe knowledge) in an exceptionally visible and clear way. Paolo concludes his discussion putting succinctly forward what I myself hold to be the core point of the theory, "I have tried to show, only if we consider reasoning in its basic form, that is as a reason-giving device, can we understand why two or more people driven by their curiosity can get together in a collaborative effort to safely establish whether a certain proposition is true, without manipulating each other. Indeed, those involved in the argumentation process described by Smokrović are expected to benefit from the exchange by acquiring something of value, namely safe knowledge." (p. 153/154). I cannot but agree with him with gratitude.

*Nenad Mišćević* discusses and challenges my account of the epistemology of rationality of reasoning in this paper. The issue is, roughly speaking, whether everyday reasoning, which is by no means perfect, can be counted as rational or its imperfection incurs the accusation for irrationality. Mišćević advocates the latter while I am prone to defend the former approach. In answering Mišćević's objections, I am going to start with his nicely articulated explicit questions to me. As the objections are grounded in his positive proposal concerning degrees of rationality, my answer extends to this topic as well. Here are the questions:



”Most importantly, Smokrović thinks that this *weakened conception* (my italic) is enough for a moderate rationalist judgment (my italic): people do reason rationally in the sense sketched. I think that this move is too undemanding. Suppose it holds, and suppose that together with such cases many miserly moves cognitive scientists see as all pervasive happen all the time; our Jane is prone to a very miserly reasoning. Then we have minimal pieces of correct reasoning in the sea of non-logical heuristics and biases. It is a clear case of our Stage Two scenario. Why call such a reasoning rational, in the complimentary sense of the term? Smokrović himself proposes not to count Watson as rational; but Watson is, in the novel and in general, a prototype of ordinary capacities; Holmes is an exception.” p. 172/3

To be able to master the objections as precisely as possible, let me indicate their targets putting them in a broader context of salient trends in cognitive science and in formal epistemology. As I see it, Mišćević’s main objections target my view calling it a weakened conception that enables a moderate rationalist judgment. This view of mine is “too undemanding” because, as I understood his point, it evaluates instances of everyday reasoning as rational though they are not. I am not quite certain what his characterization of the “weakened conception” is precisely supposed to mean. Rationality, many would agree, is a property of reasoning. Ascribing the property of rationality to reasoning (either as a process or as a product of this process) the standard for judging rationality is what matters. Rationality is always assessed or measured according to a particular standard. Generally speaking, there are two ways as to how the rationality standard can be determined. It can be expressed either in logical or non-logical terms. In this short response, I will consider only the logical aspect. Employing logic as a standard against which cognitive performances are assessed, is traditionally entrenched in both relevant disciplines having reasoning as its subject, namely, in formal epistemology and in cognitive sciences (we are here focusing on psychology). Traditional or orthodox approach takes classical (propositional and/or predicate) logic to be the normative standard of reasoning. In the first part of the last century the enthusiasm concerning the role of logic in reasoning prevailed in both disciplines. In formal epistemology, Hintikka’s (1962) standard epistemic logic represents this view. In psychology it is particularly present in early researches conducted by Wason (1960), who unproblematically held classical propositional logic to be the standard, and, at least partly, by Kahneman, Slovic, Tversky (1982), in whose experiments probabilistic calculus figured as norms for measuring rationality. According to this standard, the average people’s reasoning abilities are judged as cognitively miserable (as Mišćević notes in the quoted passage). Confronting ideal norms (what people should do) and factual

behaviour (what they are able to do), the result does not surprise. "Cognitive miserly movies" are "all pervasive".

The difference between norms and behaviour results in the gap between normative and descriptive. On the epistemological side, Harman (1986) proclaims the divorce between logic and reasoning. The trend in both epistemology and psychology was to bridge the gap somehow. On the cognitivist side, Evans and a bit later Stanovich introduced the more structured cognitive architecture, distinguishing system 1 (automatic, fast and fallible) and system 2 (slow and reflective).

To sum up, two theoretical stances are distinguishable. There are, on the one hand, those that allow the ideal normative standard of rationality expressed in the form of logical principles. In the conducted experiments, as a rule, the responses of the majority of individuals grossly depart from the normative model. Inevitably, we get few experts assessed as rational and the great majority of individuals as irrational. Let me refer to this position as the elitist position. There are, on the other hand, those that want to make the gap more shallow. The most famous are Stanovich and West. They responded to this division referring to individual differences, continuing on Slovic and Tversky's (1974) line of thought that suggests that the gap is due "to an initial failure to fully process and/or understand the task" (Stanovic & West, p. 651). According to this, "The more reflective and engaged reasoners are more likely to affirm the appropriate normative model for a particular situation" (ibid). This insight is theoretically articulated as a distinction between the lazy system 1 and the reflexively engaged system 2. What is important for our purposes is to note that latter approach allows different stages of rationality. Some authors also try to indicate the reasons or obstacles that prevent agents to obtain a higher level in actual performances. Johnson-Laird & Oatly (1992) mention factors such as dispositions toward premature closure, cognitive confidence, dispositions toward confirmation bias, and so on (in Stanovich & West, 2000, 664). Let me, for short, call this theoretical stance the structured one. In addition, there are theoreticians called meliorists that hold that some of these obstacles might be removed under favourable conditions .

Coming back to our dispute, I would say that Miščević's positive proposal naturally belongs to the structured theory but strongly avoids the meliorist view. His proposal consists in introducing a closed rationality scale containing five stages. Before proceeding, let me mention constraints that, in my mind, any closed scale has to meet. Since the scale is closed, it is in need of determining the upper and lower bounds of the domain the scale is referring to. Lower bound determines cognitive actions that are still considered as rational. The usual way of doing this is to propose some,

at least general requirement for the rational performance. As far as I am aware, Miščević's theory is missing such a proposal. Also, all enumerated stages, from the lowest up to the highest, describe the space of rationality so that instances of inference falling in any of the stages should be counted as rational. It seems to me that Miščević's proposal does not meet this condition. However, let us scrutinise the proposal in more details.

Miščević's scale begins with stage One, which is the "stage of a strongly crippled rationality" (p. 167). This seems a bit perplexing. If rationality in question is crippled to such an extent that such performance actually ceases to be rational, then it does not belong in the scale. If this were the case, the former stage Two would become stage One. This is the stage of the "partly crippled rationality". Characterising this stage Miščević says, "Thinkers with partly crippled rationality (maybe even those with completely crippled one) would be capable of performing such simple logical inferences; unfortunately, they will do it only within a framework of extremely unreliable cognitive tendencies" (p. 1699). Then he continues: "The partially crippled rationality would go with the use of all basic constants (see S. Evinine (2008))" (ibid). However, it seems that Miščević holds the agents that fall into this stage irrational, too. Using my illustration where Holmes and doctor Watson discuss a criminal case, Miščević classifies Watson's reasoning as "a clear case of our Stage Two scenario" and asks, "Why call such a reasoning rational, in the complimentary sense of the term?". Since Watson is "a prototype of ordinary capacities", the majority of people falling onto stage Two are irrational. Interestingly enough stage Two does not belong to the scale containing stages of rationality.

Back to the higher stages. Here is a brief reference to the next, stage Three, "It would be positive, with some logical reasoning and with successful avoidance of the worst traps of miserly reasoning". Miščević emphasises that "Stage Three is the first one that corresponds to the ordinary notion of "being rational"(ibid). Note that whatever ordinary notion of "being rational" might mean, stages One and Two, as members of the scale of rationality, are not actually stages of rationality, because the "ordinary notion" of rationality in the given scale is marked only with stage Three. Stage Four is exemplified with genius individuals as Gödel, while stage Five depicted a completely ideal reasoner.

Let me discuss Miščević's proposal starting from the stage Three. As it is said, "It would be positive, with some logical reasoning". To make the expression "logical reasoning" more precise, it should be said that it is the reasoning obeying general logical requirements which usually supposes that inference has to be abstract, global, closed and synchronic. That it is abstract means that agent's inferences are valid regardless of the content

of the used propositions. That it is global means that the validity of inferences affects all agent's attitudes (in the case of deduction, beliefs), explicit and implicit ones. Importantly, inferences should be closed under logical constants and operators. Finally, synchronicity requires that whenever the premises are given, the conclusion is given at the same time. It is clear that the satisfaction of such requirements implies logical omniscience. No finite agent can satisfy them. On the contrary, any model of rationality that intends to be at least minimally realistic supposes that agent's reasoning is diachronic, concrete, local and if closed at all, only inside a restricted domain of attitudes. What does Miščević's claim exactly mean when he writes that at the stage Three the agent's cognitive behaviour should be endowed with "some local reasoning"? If "some" means that the space of attitudes is restricted in a way, I guess that it refers to a form of synchronicity, locality, and concreteness. But, if this is the case, I wonder how stage Three (with some logical reasoning) differs from stage Two (with less or perhaps without logical reasoning) on one hand and from stage Four (that, I guess, should be almost perfect) on the other? If stage Two lacks rational reasoning and stage Four is characterised by almost perfect reasoning, what remains for ordinary rational people is stage three but that is achievable only for such exceptional characters as Sherlock Holmes.

That is exactly where Miščević and I disagree. I agree with the graded concept of rationality, but, taking Miščević's structure, stage Two is the stage where the majority of agents are able to master basic logical constants and simple inferences. This stage should still be counted as rational. As a matter of fact, bounded rationality that is still counted as rationality, is usually defined as a process in which an agent is able to perform inferences correctly according to the rules for constants, and combine only few of them at the time. Any step in reasoning can be considered an ability to master basic simple inferences, while the stages of rationality differ according to how many steps an agent is able to perform in the process of explicit reasoning.

*Danilo Šuster*, in his excellent comment and criticism of my view (mainly in Smokrović, 2018), starts with the claim that we both "... share interest in "real-life" reasoning and are both concerned with the usefulness and scope of classical logic in providing tools for the analysis and assessment of real-world reasoning" (p. 211). However, we take different paths when providing the account for the role of logic in every-day reasoning. Prior to focusing on Šuster's criticism, let us be reminded of the crucial distinction in the domain we are dealing with. It is the distinction between the relation of entailment and the relation of transition in reasoning, as Harman famously stated. While entailment describes the relation between

propositions (or, better to say, formulas), the transition in reasoning has to do with the relation between attitudes of different types (beliefs, knowledge, intentions, and so on). Since our discussion is focusing on deductive reasoning, I will consider only beliefs. Attitude token, unlike attitude type, has a form: S believes p. The reasoning transition (or inference, as it is usually referred to) is an act of mind where, from a set of attitudes (beliefs), a subject derives a new belief, apprehending the former as a basis or reason for the latter. The question is whether the relation of entailment (described as a set of logical rules and principles) can be applied to the reasoning transition or inference, either as a normative standard that should guide rational reasoning (prescribing what to infer having a certain set of beliefs) or as a model that is not primarily concerned with normativity, but can help us understand reasoning better. The point where Šuster and I disagree, in a nutshell, is this. I hold that there is a particular non-classical logic (in my 2018 paper I argued that it is default logic) that can be used as a proper model for reasoning, while Šuster denies that any branch of formal logic by itself is likely to capture real-life reasoning.

I have named my own position a logicist thesis which, expressed in MacFarlane's words, is the claim that there is "some connection between logical validity and the evaluation and criticism of reasoning" (Smokrović, 2018). Šuster "basically agrees" with a part of the thesis, namely, that "There has to be a connection between logic and the evaluation and criticism of reasoning" (p. 212), but expresses doubts concerning the other part "about the adequacy of formal validity". Putting an emphasis on formal, he claims that no formally valid system can capture reasoning. If there is a connection between logic and evaluation of reasoning, this logic can only be informal logic.

Šuster's paper contains a provoking and exceptionally well-argued criticism of the logicist thesis, but he also offers a positive part where he argues that informal logic (instead of any branch of formal logic) has a decisive role in modelling real-life reasoning. In the rest of this paper I am going to provide answers to Šuster's criticism, but also question some assumptions of his positive proposal.

Having agreed with me that classical logic is not a plausible candidate for modelling human reasoning, Šuster opens his list of objections questioning my concept of deduction. Here I agree with him (and I am very grateful for his remarks) that the necessary condition for deduction is the necessary truth preservation, the condition that many of non-classical logics lack. Since I argue for the view that, as Šuster formulates it, "the appropriate way to model human reasoning is via deductive, although not classical logical systems", the way to save deducibility of logic I opt for,

Šuster suggests, is to "equate the domain of deductive with the domain of formal" (215). In this sense he proposes to interpret my use of deductive "in the weak and the broad sense" in which "the validity remains a matter of form; one proposition is a consequence of others only if there is a valid pattern which the propositions together match" (215). So far, so good, but here comes the substantial aspect of Šuster's criticism, namely, he strongly dismisses the idea that any "branch of formal logic" can model everyday reasoning.

Let us observe Šuster's reasons. To show that the whole family of non-classical logics fails to be a suitable model for reasoning, he examines the relevant (or relevance) logic as the representative for the whole class. The conclusion is "that the project failed, and the moral of this failure generalizes to the contemporary defeasible, non-monotonic, default, auto-epistemic, ... formal systems discussed by Smokrović" (p. 212). I am afraid that it is not quite clear whether the failure of relevant logic generalises only to default logics (especially because there is no any indication of the connection between these two kinds of logic) or to the wider family of all non-classical logics? As can be seen from the claim that "no branch of formal logic is by itself likely to capture real-life inferential links", the failure obviously extends to the whole class of non-classical logics. I would just note here that Šuster considers the family of non-classical logics perhaps too narrow. Many of non-classical logics are non-monotonic, but the scope of this class of logics includes many others, besides different forms of defeasible logics. Not to forget, it includes modal logics, (particularly non-normal modal logics), not to mention intuitionistic and many-valued logics, that are not affected by Šuster's criticism.

Coming back to the relevant logic, let us see in what sense the project failed and why its failure should be extended to all non-classical logics? Šuster succinctly and precisely expresses the "core motivation" (219) for relevantists. It was to make this logic suitable for reasoning that is considered a train of thought proceeding through episodes of inference or "steps of reasoning". In each following episode, the antecedent of the inference is supposed to be a ground or a reason for the consequence. Andersen and Belnap accept the principle of identity, " $A \rightarrow A$ ", read as "A, therefore A" as the trivial foundation of all reasoning" (p. 219). However, Šuster, relying on Strawson, claims that this principle is not a step in reasoning. In Strawson's wording, "a man who refers to himself does not reason". Why is it so? Let us see the explanation: "We thus have two notions of inference, inference in the derivational sense ("d-inference") and inference in the reason-giving sense ("r-inference"). Corresponding to these two notions, there are then two notions of relevance, firstly, the derivational relevance (in this sense

A is relevant for A). Secondly, the reasoned relevance – in this sense A is not relevant for A: we do not accept A (conclusion) as a result of reflecting on A as the premise in the argument “A, therefore A” (p. 220). Hence, it is claimed, “ $A \rightarrow A$ ” is not the step in reasoning in the sense of r-inference. In the reasoned relevance sense, A (as antecedent) is not the right kind of reason for A (as a conclusion). If this holds, “ $A \rightarrow A$ ” can’t be the trivial foundation for all reasoning and the system of relevant logic fails.

However, can we find any decisive argument supporting this claim? We can see that reasons against relevant logic, offered in this paper, rely on the inappropriateness of “A, therefore A” as a foundation for reasoning. It is not such a foundation because it is not reasoning at all. However, the view that the principle of identity is not a case of reasoning is a strong and widespread intuition, but still intuition. On the other hand, I am inclined to follow a different intuition. Let me offer two versions of this intuition. One version takes that one can (though not necessarily) explicitly take A as a reason for believing A. The other one is based on the strictly implicit or tacit grasping of A as a reason for A. The former version goes like this: since reasoning connects attitudes, steps in reasoning “in an epistemically serious way” should be expressed as a believed or known proposition, for instance “I believe that it is raining”. In this case we have: “I previously believed that A, therefore I believe that A”. The part of my evidence for believing that A is my prior belief that A. According to Williamson’s E = K principle in knowledge-first theory, knowledge is a total evidence one has. One’s believing that p is part of the total evidence for believing that p. This is Williamson’s formulation: “Call S’s belief in p explicitly evidence-based if it is influenced by prior beliefs about the evidence for p (p. 722, *Mind*, Vol. 106, 424). Hence, it seems reasonable to say A is the (part of) reason for accepting A, and “A, therefore A” is a reasoning step after all.

Let me just briefly indicate the latter variant of this intuition. I am reaching here for a variant in the spirit of Boghossian view. “ $A \rightarrow A$ ” is, everyone would agree, perfectly valid as d-inference but it is also a basic, axiomatic schema. Such a basic inference cannot and need not be justified, even for the naive reasoner, relying on some other inferences. The reasoner grasps it tacitly or “blindly” as Boghossian says. In Peacocke’s (Peacocke, 1987) terms, such an inference strikes one as “primitively compelling”. In this case, too, it can be said that “ $A \rightarrow A$ ” is a legitimate, although implicit, step in reasoning.

Of course, there is no decisive solution to the issue. I am just offering an intuition contrary to Šuster’s. Nevertheless, I hold that this move can, at least partly, impede the power of Šuster’s argumentation. If this were the case, I could move to the other part of the argument, namely, that the



failure of relevant logic to capture every-day reasoning can be extended to other non-classical logics, including defeasible logic. If the power of the first part of the argument is at least restricted, in the sense that it cannot be decisively shown that relevant logic failed, the other part of the argument can hardly be accepted.

I am going to conclude with very brief remarks concerning the alleged primacy of informal logic over a branch of formal logic in modelling reasoning. While the criteria for the evaluation of reasoning are well defined in formal logic (classical or non-classical, though it varies among logics), it is not so in informal logic. There are different proposals in contention. The majority of authors accept a cogency as a criterion, which is usually cashed out, originally due to Johnson & Blair (1977), as a triad containing: acceptability, relevance and sufficiency. My main objection for taking cogency expressed as a set containing acceptability, relevance and sufficiency is that each of them is all too much subject-relative. It depends on the subject to judge whether a premise is acceptable or relevant or if they are jointly sufficient. It is interesting to see how David Hitchcock, one of the most important informal logicians, evaluates the elements of the triad. Concerning acceptability, he remarks that it is "relative to the particular evaluator, or to the particular audience for whom the evaluator is judging the worth of the argument" (p. 443). The other two members of the triad that are supposed to give general instructions for the good reasoning (argument) fare just as badly. The criterion of relevance is particularly controversial. Hitchcock says: "It is doubtful whether relevance of each premise is a necessary condition for a good argument. To say so is to imply that a good argument can be turned into a bad one by adding an irrelevant premise" (p.443). For sufficiency it is said: "It is true that a good argument must be such that its premises, if true, provide enough support to the conclusion. But, how much is enough, in what circumstances?" (Ibid). In conclusion, I would say that I cannot see why a particular non-classical logic is a worse candidate for evaluating everyday reasoning than informal logic.

*Majda Trobok's* paper challenges my view "concerning argumentation. It amounts to the idea that argumentation is an especially good means for achieving the truth and for the extension of knowledge" (p. 229). The first point she finds suspicious is what she calls "AGRT (advantage of group reasoning thesis)" claiming that "It might appear to be a variant of the famous Tuscan proverb: A wise man and a fool together, know more than a wise man alone". However, it is a plain misunderstanding of my idea. The so-called "group reasoning thesis" does not mean a mere aggregation of knowledge. Far from that, what matters is the concept of argumentative process. The process can emerge only when particular conditions are met.

In this kind of argumentative process, both interlocutors should be curiosity-driven, while the adversarial nature of argumentation should gradually be transformed into cooperation. It is only in this case that the process might become the means for acquiring safe (extension of) knowledge in which (may I be forgiven for using Paolo Labinaz's wording) none of the interlocutors emerges as "loser". This is, of course, not the description of the situation fitting typical real-life arguing. This account of reasoning intends to be a model<sup>1</sup> that includes a degree of idealisation, but can nevertheless help to better understand the specific kind of argumentation. This kind of argumentation is not typical but it still emerges sometimes. Moreover, as a desirable form of argumentation, it is a role-model for less successful (in the epistemic, not pragmatic sense) forms.

In his article *Andrej Ule* continues our debate extending its domain from a simple argumentative situation to the "common epistemic" (p. 242). He introduces into discourse a very interesting and important concept of common knowledge in its implicit and explicit aspects. Not only that I completely agree with Andrej's ideas, but his reflections prompt me, hopefully, to further the investigation of these epistemic ideas.

Regarding *Michael Watkins'* criticism, I have to say that I accept it. My work was originally an attempt to challenge Timothy Williamson's denying of the inferentialists' claim that the assent to a sentence is constitutive for understanding it (in "Philosophy of Philosophy", 2013). However, Williamson's answer to my objection (in the Croatian Journal of Philosophy) convinced me that my attack fails. Michael's criticism testifies that Tim was right (and I was obviously wrong).

---

<sup>1</sup> For this account to be a model in the model-building sense, the formal logical description of the particular aspect of the domain is needed.

---

**Index of Names**

- Adler Jonathan, 219  
 Aharoni Eyal, 127  
 Aikin Scott, 39  
 Albert Hans, 40  
 Alfano Mark, 98  
 Allen Sophie, 253  
 Alston William, 105  
 Anderson Alan, 3, 217  
 Anderson Elisabeth, 79, 91  
 Aristotle, 1, 4, 5, 9, 10, 15, 84, 85, 87,  
     176, 232, 243, 267  
 Arsenijević Miloš, x, 291, 292  
 Asbach Olaf, 254  
 Austin John, 40  
 Bajšanski Igor, x, 291-293  
 Barcan Markus Ruth, 54  
 Barth Else, 40, 215  
 Bašić Hanžek Gabriela, xi, 43, 50,  
     291, 292  
 Bayes Thomas, 171, 222, 274-276  
 Beall Jeffrey, 191  
 Belnap Nuel, 3, 9, 54, 68, 217-226  
 Ben-Yami Hanoch, xi, 53-55, 58,  
     61, 62, 67, 68, 75, 291, 292  
 Benacerraf Paul, 10, 11  
 Berčić Boran, xi, 90, 115, 291, 292  
 Bermejo Luque Lilian, 235  
 Bernays Paul, 10, 16  
 Berto Francesco, 280  
 Biro John, 39, 46, 142  
 Black Max, 10  
 Blair Rachel, 127  
 Blair Jonathan, 224  
 Blečić Martina, 135  
 Bochman Alexander, 245  
 Bortolotti Lisa, 124  
 Brogaard Berit, 168  
 Broome John, 128, 197-203  
 Brzović Zdenka, 135  
 Burgess John, 215, 219  
 Cantor Georg, 232, 280  
 Cardella Valentina, 127  
 Carnap Rudolf, 171, 253, 259  
 Chater Nick, 127, 183  
 Chihara Charles, 10  
 Chisholm Roderick, 79  
 Colyvan Mark, 124  
 Comte Auguste, 258  
 Copp David, 128  
 Crisafi Anthony, 248  
 Čuljak Zvonimir, 96, 104  
 Cuneo Terence, 116  
 Cushman Fiery, 30  
 Dancy Jonathan, 204  
 Davidson Donald, 119  
 Dawkins Richard, 104, 107, 253  
 de Finetti Bruno, 274  
 De Neys Wim, 22, 23  
 Dennett Daniel, 104  
 Descombes Vincent, 247  
 Dewey John, 97, 100  
 Doniger Wendy, 259  
 Dove Ian, 47  
 Dretske Fred, 283  
 Dunn Jon, 54, 68  
 Dutilh Novaes Katarina, 116, 150,  
     184, 214, 215, 225  
 Edwards Paul, 90  
 Elder Linda, 98, 100, 102, 109  
 Elqayam Shira, 124  
 Ennis Robert, 100  
 Evans Jonathan, 21, 22, 26, 34, 124,  
     145, 183, 295  
 Evnine Simon, 169  
 Fagin Ronald, 287  
 Fairweather Abrol, 177  
 Feigl Herbert, 10  
 Feldman Richard, 39

## Index of Names

---

- Fernbach Philip, 29  
Festinger Leon, 26  
Field Hartry, 2, 181, 182, 184, 188-193, 223  
Fisher Alec, 100  
Fitch Frederic, 64  
Fitelson Branden, 188,  
Flew Anthony, 104, 107  
Frederick Shane, 22  
Frege Gottlob, 186, 279, 280, 286  
French Rohan, 280  
Gallagher Shaun, 248  
Garssen Bart, 39, 42-44  
Gaus Gerald, 127, 133, 134  
Gauthier David 127, 132  
Gavran Miloš Ana, 106  
Geanakoplos John, 243  
Geil Molly, 145  
Gibbard Allan, 79  
Gigerenzer Gerd, 167, 169, 175  
Gintis Herbert, 128  
Goldman Alvin, 39, 141-143, 151  
Golubović Aleksandra, xi, 291, 292  
Govier Trudy, 223  
Gratzl Norbert, 53, 54, 75  
Grčki David, 177  
Grice Paul, 40, 149  
Grootendorst Rob, 39-41, 50, 142, 153  
Grossmann Igor, 133  
Grünbaum Adolf, 1, 10-15  
Grüne-Yanoff Till, 80  
Guala Francesco 124, 127  
Habermas Jürgen, 142  
Hahn Hans, 259  
Hahn Ulrike, 144  
Haidt Jonathan, 26  
Halpern Joseph, 243, 287  
Hamblin Charles, 41, 42, 142  
Handley Simon, 145  
Hansson Sven Ove, 79-81  
Hardin Clyde, 167  
Hare Richard, 85, 90  
Harman Gilbert, 1, 2, 115-125, 135, 165, 181, 184-191, 213, 214, 220-226, 295  
Harris Sam, 104  
Hawthorne John, 283  
Henkin Leon, 58, 68-75  
Hilbert David, 10, 15, 16  
Hintikka Jaakko, 42, 235, 274, 287, 294  
Hitchens Christopher, 104  
Hitchcock David, 301  
Hooker Brad, 203  
Hopster Jeroen, 133  
Hornikx Jos, 144  
Hume David, 114, 125, 199  
Jackson Sally, 235  
Jacobs Scott, 235  
Jaksland Rasmus, 253  
Janis Allen, 12, 13, 15  
Jaśkowski Stanisław, 64  
Jefferson Anneli, 128  
Johnson Ralph, 212, 224  
Johnson-Laird Philip, 161-163, 183  
Joll Nicholas, 252  
Jurjako Marko, xi, 118, 127, 291, 292  
Kahneman Daniel, 22, 161-163, 167  
Kandel Eric, 252  
Kant Immanuel, 85, 114, 115, 121, 126, 127, 132  
Keynes John Maynard, 259  
Kiesewetter Benjamin, 201  
Kohlberg Lawrence, 126  
Kopajtić Jelena, xi, 291, 292  
Kölbel Max, 88, 89  
Korsgaard Christine, 114, 125  
Koszowy Marcin, 212  
Krabbe Erik, 39, 40, 42, 44, 215  
Kripke Saul, 54, 68, 268

- Labinaz Paolo, xi, 150, 291, 293, 302  
 Ladyman James, 253  
 Lanzet Ran, 53, 54, 62  
 Leblanc Hugues, 54, 68  
 Lehrer Keith, 96, 97, 101  
 Leiter Brian, 116, 125, 127  
 Lerner Jennifer, 153  
 Levy Neil, 115  
 Lewandowsky Stephan, 33  
 Lewis David, 9, 128  
 Łukasiewicz Jan, 5  
 Lumer Christopher, 39, 40, 45, 47-49  
 Macagno Fabrizio, 38  
 MacFarlane John, 2, 88, 89, 165, 168, 181, 184, 186-190, 213, 223, 226  
 MacIntyre Alasdair, 107, 212  
 Mackie John, 90  
 Maibom Heidi, 127  
 Malatesti Luca, 127  
 Mares Edwin, 216, 219  
 Martignon Laura, 222  
 Maxwell Grover, 10  
 McGraw Peter, 91  
 McKeon Matthew, 224  
 McMullin Ernan, 257  
 Mele Alfred, 196  
 Mercier Hugo, 19, 20, 26-28, 33, 98, 141, 142, 144, 145, 149, 150, 152, 154, 230, 232, 233  
 Meyer Robert, 216, 226  
 Milić Mladen, 260  
 Millgram Elijah, 125  
 Mišćević Nenad, xi, 114, 291, 293-297  
 Miyazono Kengo, 124  
 Montemayor Carlos, 177  
 Moore G.E, 79, 90  
 Moses Yoram, 243, 287  
 Moshman David, 145, 150  
 Muller-Lyer Franz, 270, 271  
 Nathanson Stephen, 160  
 Neurath Otto, 253, 259  
 Newel Allen, 29  
 Nichols Shaun, 119  
 Nosich Gerald, 99, 100  
 Nowell-Smith Patrick, 83  
 Nozick Robert, 160  
 Oaksford Mike, 127, 183  
 Ockham William, 258  
 Olbrechts-Tyteca Lucie, 142  
 Osborne Martin, 243  
 Papineau David, 122, 123  
 Parfit Derek, 202  
 Paul Richard, 100-102, 109  
 Pavlović Edi, xi, 53, 54, 75, 174, 291, 292  
 Perelman Charles 142  
 Plantinga Alvin, 105  
 Plato, 83  
 Polkinghorne John, 252  
 Polović Vlatko, 82  
 Pölzler Thomas, 133  
 Popper Karl, 40, 41  
 Priest Graham, 280  
 Prijic-Samaržija Snježana, 98, 100  
 Quine Willard, 57, 125, 216  
 Raab Jonas, 53, 54  
 Ramsey Frank, 274  
 Rawling Piers, 196  
 Raz Joseph, 203, 205  
 Read Stephen, 219  
 Reed Chris, 38  
 Resnik Michael, 121, 196  
 Restall Greg, 191  
 Rips Lance, 183, 219  
 Rose Hilary, 253  
 Rose Steven, 253  
 Rubinstein Ariel, 243  
 Russell Bertrand, 279-281  
 Saccheri Girolamo, 237

- Salmon Wesley, 10  
 Scanlon Thomas, 90, 118, 120, 201  
 Scheler Max, 81, 84  
 Schlick Moritz, 4  
 Schwarz Norbert, 33  
 Schwayder D. S, 10  
 Scriven Michael, 224  
 Searle John, 40, 204, 255  
 Sekulić Dragana, 37, 42, 118  
 Sellars Wilfrid, 212  
 Shapiro Stewart, 171  
 Siegel Harvey, 39, 46, 141, 142  
 Sikavica Pere, 260  
 Simon Herbert, 29, 124, 172  
 Siscoe Robert, 167  
 Skelac Ines, xii, 291, 292  
 Sloman Steven, 21, 29  
 Smith Michael, 115-124, 128-132, 135, 199  
 Sosa Ernest, 79, 145  
 Sperber Dan, 19, 20, 26-28, 98, 141-145, 149-154, 167, 174, 230-233  
 Staffel Julia, 166, 167  
 Stalnaker Robert, 281, 285, 286, 288  
 Stanisavljević Dean, 81  
 Stanovich Keith, 21, 24, 26, 161, 164, 169, 174, 295  
 Stein Edward, 124,  
 Steinberger Florian, 183, 186, 192  
 Stenning Keith, 2, 3, 215, 222  
 Strawson Peter, 218-221, 224  
 Street Sharon, 120  
 Streumer Bart, 197, 203, 204  
 Sušnik Matej, xii, 114, 118, 135, 291, 292  
 Šuster Danilo, xii, 291, 297-299  
 Taylor Richard, 10  
 TeHennepe Eugene, 10  
 Tetlock Philip, 153  
 Thaler Richard, 95  
 Thompson Valerie, 23, 25, 145  
 Thomson James, 10, 15  
 Titelbaum Michael, 171  
 Toplak Maggie, 168  
 Toulmin Steven, 41  
 Trobok Majda, xii, 83, 95, 177, 234, 235, 291, 301  
 Turiel Elliot, 126, 127  
 Tversky Amos, 22, 162  
 Ule Andrej, xii, 247, 291, 302  
 van Boven Leaf, 91  
 van Eemeren Frans, 39-41, 44, 50, 142, 153  
 Vardi Moshe, 287  
 Varga Alexandra, 2, 3, 222  
 Veljak Lino, xii, 251, 261, 291, 292  
 Velleman David, 83  
 Viale Riccardo, 172  
 von Wright Georg, 86, 88  
 Walton Douglas, 38, 39, 42, 151, 153, 235  
 Warren Caleb, 91  
 Wason Peter, 161, 167, 168, 170, 173, 181, 294  
 Watling John, 10  
 Watkins Michael, xii, 292, 302  
 Wedgwood Ralph, 121, 177  
 Weyl Hermann, 10  
 Williams Bernard, 119, 120, 206-208  
 Williamson Timothy, xii, xiii, 97, 100, 141, 145, 146, 230, 265, 267-269, 271, 273, 280, 288, 291, 292, 300, 302  
 Wilson Timothy, 26  
 Wisdom John, 10, 104  
 Wolff Jonathan, 92  
 Wolterstorff Nicholas, 104, 105  
 Woods John, 42, 49, 213, 220-222  
 Wright Jennifer, 133  
 Yap Audrey, 274  
 Zenker Frank, 39, 43



**NENAD SMOKROVIĆ** is professor emeritus at the University of Rijeka. He was teaching at the *Faculty of Economy* from 1980 to 2020, and at the *Department of Philosophy* of the *Faculty of Humanities and Social Sciences* from its beginning in 1998 to 2020. Currently he is teaching and mentoring at the PhD program *Philosophy and Contemporaneity* at the Department of Philosophy.

This collection of articles is a tribute to Nenad Smokrović, our friend and colleague from the Philosophy Department. He has been working on human rationality for over 25 years. Consequently, articles in this collection are either directly or indirectly related to this subject matter.

In this volume, 19 authors from Croatia, Slovenia, Serbia, Italy, Israel, USA, and GB write about human rationality. What human rationality consists in? What is the relationship between its normative aspect (logic) and its descriptive aspect (psychology)? Do people *in fact* reason according to the laws of logic? What is the nature and the function of reasoning, argumentation, and inference? ... We are publishing this volume in the hope that it will provide further contribution to the research in the field. (Editors)

**MILOŠ ARSENIJEVIĆ**  
**IGOR BAJŠANSKI**  
**GABRIELA BAŠIĆ HANŽEK**  
**HANOCH BEN-YAMI**  
**BORAN BERČIĆ**  
**ALEKSANDRA GOLUBOVIĆ**  
**MARKO JURJAKO**  
**JELENA KOPAJTIĆ**  
**PAOLO LABINAZ**  
**NENAD MIŠČEVIĆ**  
**EDI PAVLOVIĆ**  
**INES SKELAC**  
**MATEJ SUŠNIK**  
**DANILO ŠUSTER**  
**MAJDA TROBOK**  
**ANDREJ ULE**  
**LINO VELJAK**  
**MICHAEL WATKINS**  
**TIMOTHY WILLIAMSON**



9 789533 610641

Faculty of Humanities and Social Sciences  
University of Rijeka 2022  
[www.ffri.uniri.hr](http://www.ffri.uniri.hr)