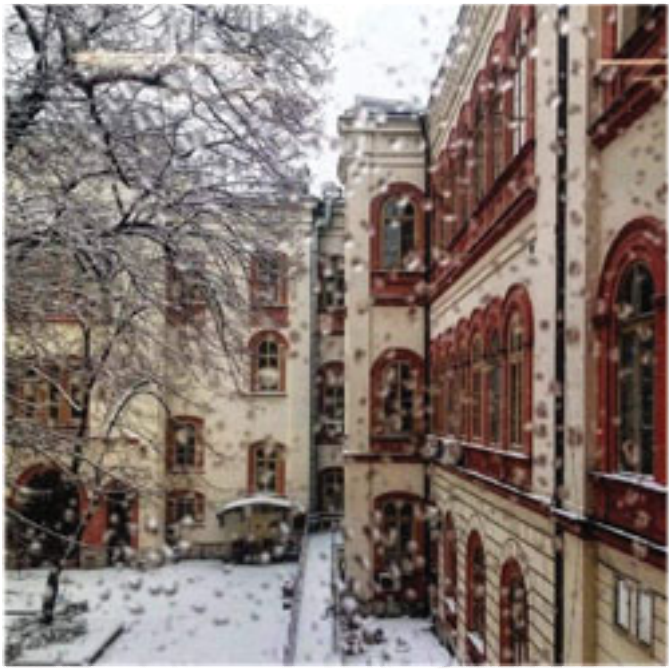
HUMANS AND SOCIETY IN THE TIMES OF CRISIS

Virtues and vices – between ethics and epistemology

Edited by Nenad Cekić



1838

UNIVERSITY OF BELGRADE
FACULTY OF PHILOSOPHY

Faculty of Philosophy, University of Belgrade | 2023

1838

Virtues and vices – between ethics and epistemology

Edited volume

Nenad Cekić (Editor)

Edition *Humans and Society in the Times of Crisis*

Virtues and vices – between ethics and epistemology
Edited volume
Prof. dr Nenad Cekić (editor)
Belgrade 2023

Publisher
University of Belgrade – Faculty of Philosophy
Čika Ljubina 18–20, Belgrade 11000, Serbia
www.f.bg.ac.rs

For the publisher
Prof. dr Danijel Sinani,
Dean of the Faculty of Philosophy

Referees
Prof. dr Snježana Prijić Samaržija,
University of Rijeka, Croatia
Dr Dejan Šimković,
The University of Notre Dame Australia, Sydney Campus,
School of Philosophy and Theology
Prof. dr Vojislav Božičković,
Faculty of Philosophy, University of Belgrade
Prof. dr Milorad Stupar
Faculty of Philosophy, University of Belgradey

Proofreader for Serbian
Irena Popović Grigorov

Cover art and design by
Ivana Zoranović

Set by
Dosije studio, Belgrade

Printed by
JP Službeni glasnik, Belgrade

Print run
200

ISBN 978-86-6427-257-5

This collection of papers was created as part of the scientific research project
Humans and Society in Times of Crisis, which was financed
by the Faculty of Philosophy – University of Belgrade.

CONTENTS

INTRODUCTION

The statement *everyone wants to live a fulfilled and happy life* may seem simple, self-evident, and even trivial at first glance. However, upon closer philosophical analysis, can we unequivocally assert that people are truly focused on well-being? Assuming they are, the question becomes: what guidelines should be followed and how should one behave in order to achieve true well-being and attain their goals? One popular viewpoint is that cultivating moral virtues and personal qualities is essential for a life of "true" well-being, rather than mere pleasure. This perspective is not particularly original, as it was advocated quite clearly by Aristotle. However, we must ask: what guarantees that virtue will lead to well-being? Can we really ignore the common-sense doubt that virtue (and moral life), as Kant points out, may not be rewarded with any form of well-being or "happiness" in the broadest sense during our earthly existence? How might we address this doubt?

If questions about the relationship between happiness and virtue have been asked since ancient times, why don't we have definitive answers yet? Perhaps there is something wrong with the questions themselves, as is often the case in philosophy. Views on this relationship have influenced the modification of theories that philosophers have put forth in order to provide satisfactory answers. One such theory is the renewed and modernized ethics of virtue, which places the greatest emphasis on building an individual's moral character, assuming that Aristotelian *character* virtues such as courage, honesty, generosity, prudence, self-control, and compassion will enable us to lead a happy and fulfilling life.

"Virtues" refer to a wide range of "excellences" or activities that individuals can use to perfect themselves. For example, an ordinary person might accept that courage can make their life more fulfilling, but it's important to recognize that virtues cannot be acquired overnight. Furthermore, many people believe that virtues are innate qualities that one *either possesses or lacks*. In contrast, ethics of virtue provides guidance and instructions for developing virtues over extended periods of time, as well as for continuously "practicing" the development of moral characteristics that will make us better individuals. The central and most significant concept in ethics of virtue is the *golden mean*, which can be interpreted in various ways. Essentially, ethics of virtue offers a path to a life filled with

pleasure, harmony, and virtuous action. However, this collection of essays, *Virtues and Vices: Between Ethics and Epistemology*, is not exclusively concerned with virtues or ethics. Knowledge, as well as the processes of belief formation and justification of those beliefs, are also essential to ethics of virtue. Going back to the systematic Greek philosopher and his teacher, we can emphasize that virtues and knowledge are inseparable. Without the epistemology of virtues, even a thematic collection like this one would be incomplete.

The Epistemology of Virtue is a relatively new branch of philosophy that has emerged as a response to questions about the value-neutrality of science. For a long time, the separation of values and scientific inquiry has been viewed as the correct methodological approach. However, it is clear that every scientist is also a human being with their own virtues and flaws. It is naive to assume that a scientist can leave their values and moral qualities outside the laboratory and treat their colleagues in the team differently than they would treat other people in their everyday life. Taking this into account, we can notice that Epistemology of Virtue is a theory that examines the influence of virtues on the processes of acquiring knowledge.

Epistemic virtues refer to qualities that enable individuals to acquire knowledge and make reliable judgments. These virtues include traits such as openness to diverse testimonies and arguments, critical thinking, a tendency to question one's own assumptions, and the ability to identify errors in one's own thought processes. Just like in virtue ethics, the Epistemology of Virtue raises the question of how we can cultivate the necessary virtues and how these virtues can have a positive impact on both individuals and society. The field also explores how epistemic virtues can be put into practice, such as in educational settings, scientific research, and various social activities. Finally, we turn to a topic that is often overlooked in philosophy – human flaws. While ethics typically views flaws as the opposite of virtues, requiring correction both individually and socially, the Epistemology of Virtue questions what happens when flaws occur in knowledge-related qualities. Do these flaws also need to be eliminated, or are they an inherent part of the process of acquiring knowledge and making judgments?

At first glance, prejudices, stereotypes, dogmatism, conformism, closed-mindedness towards arguments and evidence, as well as uncritical acceptance of authority, can all hinder our ability to acquire knowledge. Epistemologists view these qualities as vices, rather than flaws like ethicists do. The Epistemology of Vices is the newest of the three branches of philosophy covered in this thematic collection, and it examines the ways in which false and untrue beliefs develop and persist in society. This area of study seeks to answer questions about how people form delusions and whether it is possible to overcome them.

On the other hand, the Epistemology of Virtue is like the "older brother" that supervises these vices. It contributes to the investigation and overcoming of vices by developing virtues that enable individuals to more accurately assess the truthfulness of their beliefs. However, even though vices can have a destructive impact on knowledge acquisition, the question remains: are vices always detrimental in every context? Conversely, can virtues be detrimental? For example, does a virtue like solidarity sometimes hinder the efficient formation of true beliefs in time-sensitive situations such as finding a vaccine to combat a virus pandemic? Can vices be useful? Can stubbornness and uncooperativeness sometimes help a research team by allowing two sides to reach the same goal through different paths? Within this collection, some of these questions will receive answers, some of which may confirm existing impressions, while others may surprise readers.

The thematic collection, *Virtues and Vices: Between Ethics and Epistemology*, consists of four interesting chapters and 21 papers that cover the selected topic through the prism of philosophical disciplines such as ethics, epistemology, the history of philosophy, philosophy of mind, political philosophy, and philosophy of science. In this introduction, we will not discuss the contents of the articles so as not to disrupt the intended integral approach to reading the collection. A significant feature of this collection is the continuity that extends from one paper to another, from topic to topic, and from problem to problem.

Lastly, we would like to express our gratitude to the Faculty of Philosophy at the University of Belgrade for their support in realizing this collection, as well as to our colleagues from the Department and Institute of Philosophy who have made this collection a relevant and significant achievement with their contributions. We are also thankful to the reviewers of the collection as a whole, as well as to the referees of individual articles, who have provided us with detailed comments and remarks that have made this collection clearer and more comprehensive. We owe a special thanks to our colleagues from Zagreb, Rijeka, and Maribor, whose exceptional papers have contributed to making this collection an unabashed international contribution to debates about the ethics and epistemology of virtues and vices.

Editor

UVODNA REČ

Svi žele da žive ispunjenim i srećnim životom! Ova rečenica deluje jednostavno, samorazumljivo, pa čak i trivijalno. Međutim, šta se krije iza nje? Da li se, posle filozofske analize, zaista može bez rezerve tvrditi da su ljudi zaista jasno usredsređeni na blagostanje? Pretpostavimo, rasprave radi, da jesu. Ako ljudi teže blagostanju, kakva to uputstva treba pratiti i kako se ponašati da bismo do njega stigli da bismo te ciljeve i ostvarili? Jedan od trenutno pomodnih stavova glasi da za život „pravog" blagostanja (a ne tek pukog zadovoljstva) treba da razvijemo moralne vrline i posebne karakteristike. U takvom pristupu nema mnogo originalnosti jer ga je zastupao, i to vrlo jasno, još Aristotel. Međutim, zašto bi vrlina garantovala blagostanje? Zašto preskočiti zdravorazumsku sumnju u to da vrlina (i moralni život), na šta i Kant ukazuje, u ovozemaljskom životu ne mora da bude nagrađena nikakvim blagostanjem ili „srećom" u najširem smislu. Šta bismo na tu sumnju mogli da odgovorimo?

Ako se pitanja o sreći i vrlini postavljaju još od antike, kako to da na pitanja o njihovom odnosu još nemamo konačne odgovore? Možda nešto nije u redu sa samim pitanjima? To je, kao i što je u filozofiji uobičajeno, donekle i tačno. Naime, gledišta o odnosu sreće i vrlina uticala su i na izmenu teorija kojima su filozofi nastojali da ponude zadovoljavajuće odgovore. Jedna takva teorija je i obnovljena, osavremenjena – etika vrlina. U etici vrlina najviše pažnje se poklanja izgradnji moralnog karaktera pojedinca, sa pretpostavkom da će nam aristotelovske *karakterne* vrline poput hrabrosti, poštenja, velikodušnosti, razboritosti, samokontrole i saosećanja omogućiti srećan i ispunjen život.

„Vrline" su raznovrsne „izvrsnosti", aktivnosti u kojima čovek može da se usavršava. Običan čovek bi možda bez primedbi mogao da prihvati da će, recimo, hrabrost učiniti njegov život ispunjenijim. Ali hrabrost se ne stiče preko noći. Štaviše, većina ljudi smatra da su vrline nešto što se *ima-ili-nema*, pa čak i da su one urođene. Etika vrlina, za razliku od nekih drugih etičkih teorija, daje savete i uputstva za razvijanje vrlina tokom (ponekad mukotrpno dugog) vremena, kao i za kontinuirano „vežbanje" razvijanja moralnih karakteristika koje će nas učiniti „dobrim" osobama. Središnji i najznačajniji koncept etike vrlina je *zlatna sredina*, koja se, to je istina, interpretira na različite načine. Ukratko, etika vrlina navodno nudi put ka životu ispunjenom zadovoljstvom, harmonijom i valjanim dela-

njem. Zbornik *Vrline i poroci: između etike i epistemologije* očigledno se ne bavi samo vrlinama, a ni isključivo etikom. Saznanje, kao i procesi formiranja verovanja i pružanja opravdanja tih verovanja imaju veliki značaj za etiku vrlina. Vrativši se na sistematičnog Grka, i njegovog učitelja, možemo da istaknemo da su vrline i znanje čak neodvojivi. Bez epistemologije vrlina ni ovaj tematski zbornik ne bi bio potpuna zbirka tekstova.

Epistemologija vrlina je novija grana filozofije koja se pojavila kao odgovor na pitanja o vrednosnoj neutralnosti nauke. Razdvojenost vrednosti i naučnog istraživanja dugo je preporučivana kao ispravan metodološki pristup. Ipak, potpuno je jasno da je svaki naučnik takođe i sasvim obična osoba sa sebi svojstvenim vrlinama i manama. Zamisao da će naučnik vrednosti do kojih drži i sopstvene moralne osobine ostaviti ispred laboratorije, a da će se, tokom svog istraživanja, prema kolegama u timu odnositi sasvim drugačije nego prema drugim osobama u svakodnevnom životu – naivna je. Kad to uzmemo u obzir, videćemo i šta je epistemologija vrlina. Ona je teorija o uticaju vrlina na procese sticanja znanja.

Epistemičke vrline su osobine koje pojedince čine sposobnim za sticanje znanja i donošenje pouzdanih sudova. Takve osobine podrazumevaju (na primer) otvorenost prema različitim svedočanstvima i argumentima, kritičko mišljenje, sklonost preispitivanju sopstvenih pretpostavki i sposobnost uočavanja grešaka u vlastitom mišljenju.

Poput etike vrlina, i u ovoj grani filozofije postavljamo pitanje kako možemo da razvijemo potrebne vrline i na koji način one mogu pozitivno da utiču i na pojedinaca i na društvo. Takođe, epistemolozi vrlina ispituju kako epistemičke vrline možemo da primenimo u praksi, uključujući kontekste obrazovanja, naučnog istraživanja i širokog spektra društvenih aktivnosti.

Konačno, dolazimo do teme o kojoj se, čak i u filozofiji, najnevoljnije govori – do ljudskih mana. Etika je uglavnom jasno postavljena prema manama, posmatra ih kao suprotnost vrlinama i osobine koje bi trebalo ispraviti i na pojedinačnom i na društvenom planu. Međutim, šta se dešava sa manama u epistemologiji? Da li su mane u pogledu saznanja osobine koje se moraju eliminisati, kao što je to slučaj u etici?

Na prvi pogled, predrasude, stereotipi, dogmatizam, konformizam, zatvorenost prema argumentima i evidenciji i nekritičko prihvatanje autoriteta predstavljaju razorne osobine po sticanje saznanja. Epistemolozi ne posmatraju mane kao što to etičari čine, zato ih drugačije i nazivaju – porocima. Epistemologija poroka je najnovija od sve tri grane filozofije kojima se bavimo u ovom tematskom zborniku. Ona se bavi ispitivanjem načina na koje se pogrešna i neistinita uverenja razvijaju i zadržavaju u

društvu. Epistemologija poroka odgovara na pitanja o načinima na koje ljudi formiraju zablude i da li ih je i kako moguće uspešno prevazići.

Epistemologija vrlina je „stariji brat" koji nadzire razuzdane poroke. Ona doprinosi ispitivanju i prevladavanju poroka razvijanjem vrlina koje omogućavaju pojedincima da ispravnije procenjuju istinitost svojih verovanja. Međutim, pitanje koje smo postavili je, uprkos očiglednosti razornog uticaja poroka na sticanje saznanja, i dalje relevantno. Da li su poroci u svakom kontekstu pogubni? S druge strane, da li vrline mogu biti pogubne? Da li vrlina poput solidarnosti nekada usporava efikasno formiranje istinitih verovanja u vremenskim osetljivim okolnostima kao što je hitnost pronalaženja vakcine zarad suzbijanja pandemije virusa? Da li poroci mogu biti korisni? Da li tvrdoglavost i nesaradljivost ponekad mogu da pomognu istraživačkom timu jer omogućavaju da dve strane različitim putevima stignu do istog cilja? U ovom zborniku neka od tih pitanja dobiće poneke odgovore. Neki od njih će potvrditi već postojeće utiske, dok će vas neki verovatno sasvim iznenaditi.

Tematski zbornik *Vrline i poroci: između etike i epistemologije* sastoji se od četiri interesantna poglavlja i 21 rada koji pokrivaju odabranu temu kroz prizmu filozofskih disciplina poput etike, epistemologije, istorije filozofije, filozofije duha, filozofije politike i filozofije nauke. O sadržaju samih radova nećemo govoriti u Uvodnoj reči da ne bismo narušili predviđeni integralni pristup njihovom čitanju. Bitna odlika ovog zbornika je kontinuitet koji se nastavlja iz rada u rad, iz teme u temu i iz problema u problem.

Na kraju, želimo da se zahvalimo Filozofskom fakultetu Univerziteta u Beogradu na podršci u realizaciji ovog zbornika i kolegama sa Odeljenja i Instituta za filozofiju koji su svojim doprinosima učinili ovaj zbornik relevantnim i značajnim postignućem. Zahvalni smo i recenzentima celokupnog zbornika i recenzentima pojedinačnih radova, koji su svojim detaljnim komentarima i primedbama ovaj zbornik učinili jasnijim i obuhvatnijim. Posebnu zahvalnost dugujemo kolegama iz Zagreba, Rijeke i Maribora, koji su svojim izuzetnim radovima doprineli tome da ovaj zbornik bez ustručavanja možemo da smatramo međunarodnim doprinosom debatama o etici i epistemologiji vrlina i poroka.

Priređivač

I. TO KNOW OR NOT TO KNOW? EXPLORATION OF VIRTUES AND VICES IN EPISTEMOLOGY

Živan Lazović*

LUIS „ZNA" I „SVAKA" ILI KAKO IZBEĆI „STENU FALIBILIZMA" I „VRTLOG SKEPTICIZMA"

Apstrakt: U svom poznatom članku „Elusive Knowledge" (1996) Dejvid Luis je izložio verziju epistemičkog kontekstualizma koja se oslanja na ideju o relevantnim alternativama. Osnovna kontekstualistička teza je da je semantički sadržaj predikata „zna" i rečenica oblika „*S* zna da *p*" osetljiv na konverzacioni kontekst govornih lica koja ih koriste za pripisivanje znanja. Da bi razjasnili semantiku glagola „znati", kontekstualisti su se obično pozivali na analogije sa naizgled nespornim kontekstualno osetljivim izrazima kao što su kaplanovski indeksikali, komparativni pridevi i reči koje se u običnom jeziku upotrebljavaju za kvantifikaciju. Luisu se obično pripisuje poslednji od ta tri semantička modela. Iako je ukazivao na vezu između upotrebe „zna" i univerzalnog kvantifikatora „svi", sadržanog u infalibilističkom uslovu otklanjanja svake mogućnosti pogreške, u ovom članku se zastupa stav da je Luis kontekstualnu zavisnost „zna" izveo iz semantičke osetljivosti prideva „relevantna", koja se ispoljava u kontekstualnom variranju skupa relevantnih alternativa čije je isključenje neophodno za znanje.

Ključne reči: „znati", „svaka", relevantne alternative, konverzacioni kontekst, semantička osetljivost

U svom poznatom članku „Nepostojano znanje" (1996) Dejvid Luis je izložio verziju epistemičkog kontekstualizma, čija je osnovna postavka da je pojam znanja, izražen glagolom „znati" u rečenicama oblika „*S* zna da *p*", semantički osetljiv na promenu konverzacionih faktora kao što su namere, potrebe ili interesi govornih lica. Većina komentatora smatra da je Luis kontekstualnu zavisnost znanja objašnjavao time što je u standardnoj infalibilističkoj definiciji tog pojma, prema kojoj znanje nekog iskaza

* Odeljenje za filozofiju, Filozofski fakultet Univerziteta u Beogradu, *zlazovic@f.bg.ac.rs.*

iziskuje isključenje *svake* mogućnosti pogreške, sadržana opšta pridevska zamenica „svaka", koja se, kao i njen pandan „svi", u svakodnevnom jeziku koristi za univerzalnu kvantifikaciju.[1] Iako je sam Luis isticanjem nekih analogija pružio osnov za takvo tumačenje, on nije nameravao da kontekstualnu zavisnost glagola „znati" izvodi iz semantičke osetljivosti kvantifikatora „svi" već, u duhu teorije relevantnih alternativa koju u ovom članku zastupa, iz prideva „relevantna". U najsvedenijem obliku prikazan, njegov stav je da prilikom procene da li subjekt *S* u datim okolnostima zna iskaz *p*, infalibilistički zahtev ograničavamo na alternative koje su u tim okolnostima relevantne, dok su kontekstualne varijacije u značenju predikata „zna" i rečenica „*S* zna da *p*" posledica promene skupa relevantnih alternativa, na koju delom mogu da utiču subjektivni konverzacioni faktori kao što su obraćanje pažnje na pojedine mogućnosti pogreške ili isticanje sličnosti između njih.

Pre nego što se upustimo u nešto detaljnije razmatranje tog Luisovog gledišta, izdvojićemo nekoliko opštih mesta.

1.

Osnovna kontekstualistička teza je metalingvistička, tiče se značenja glagola „znati" (upotrebljenog u predikatskoj funkciji) i semantičkog sadržaja (istinosnih uslova) rečenica oblika „*S* zna da *p*" koje se koriste za pripisivanje ili odricanje znanja. Kontekstualisti kao što su Koen (Cohen), Dirouz (DeRose), Luis i drugi tvrde: (i) da je predikat „zna" semantički osetljiv na kontekst upotrebe u tom smislu što se promenom konteksta mogu promeniti i njegovo značenje i, posledično, istinosni uslovi rečenice „*S* zna da *p*"; (ii) da na izmenu semantičkog sadržaja predikata „zna" bar delom utiču neki konverzacioni faktori kao što su namere, potrebe ili interesi govornih lica (pripisivača znanja). Ako je (i) tačno, to znači da jedna ista saznajna tvrdnja „*S* zna da *p*" u jednom kontekstu može biti istinita, a u nekom drugom lažna; za to je, prema mišljenju kontekstualista, moguće naći lingvističku potvrdu u svakodnevnom govoru o znanju. Takođe, kao što (ii) pokazuje, kontekstualisti značenje pojma znanja dovode u vezu sa konverzacionim kontekstom pripisivača znanja.[2]

1 Tako Šejfer piše da Luis tvrdi da „kontekstualna zavisnost ['zna'] proističe iz kontekstualne zavisnosti domene povezanog sa 'svaki' (...)" (Schaffer 2015: 475), a Ičikava da je „Dejvid Luis (1996) artikulisao kontekstualistički pristup znanju po modelu kontekstualno osetljivih ograničenja domena kvantifikatora" (Ichikawa 2011: 383). U knjizi *Jezik, kontekst i znanje*, koja bi trebalo uskoro da bude objavljena u elektronskom izdanju, provizorno sam prihvatio to tumačenje kako bih izložio i razmotrio kvantifikacioni model za semantiku glagola „znati".

2 *Cf*. Cohen 1986, DeRose 1992.

Šta se uopšteno uzev podrazumeva pod semantičkom osetljivošću jezičkih izraza na kontekst njihove upotrebe? Nesporan primer su indeksičke reči,[3] koje svoj semantički sadržaj (objekt referencije) menjaju sa promenom konteksta: zamenice „ja", „on", „ovo" i slične, prilozi „sada", „ovde" i drugi, te iz njih izvedeni pridevi kao što su „sadašnji", „ovdašnji" i sl.[4] U tu grupu obično se svrstavaju još neke kategorije reči kao što su komparativni pridevi „visok", „bogat" i drugi i reči „svi", „neki", „nijedan" i slične, koje u svakodnevnom jeziku koristimo za kvantifikaciju. Bez upuštanja u detaljniju analizu, njihovu zajedničku karakteristiku možemo izdvojiti tako što ćemo reći da je, posmatrano izvan konteksta upotrebe, njihov semantički sadržaj nepotpun i u govoru biva kompletiran izvesnim kontekstualnim obeležjima. Shematski, to njihovu semantičko obeležje možemo da prikažemo tako što ćemo uz takav izraz stavljati prazno mesto „[...]" rezervisano za odgovarajuću dopunu; ono može da se interpretira kao pozicija argumenta funkcije koju zauzima slobodna promenljiva čiju vrednost dodeljuje kontekst. Kada se kontekstualno uslovljen izraz javlja u rečenici, to prazno mesto čini fonološki neartikulisan deo njene sintaksičke strukture (sintaksičku elipsu). Posebna pitanja za razne leksičke kategorije indeksičkih izraza jesu gde je to mesto u sintaksičkoj strukturi rečenice, koje semantičke vrednosti im kontekst dodeljuje, kako to čini itd.[5] Ta pitanja možemo da razvrstamo u dve grupe. Na deskriptivnom planu, glavno pitanje je šta se dešava sa pomenutom sintaksičkom elipsom i koja obeležja konteksta upotpunjuju značenje određenog semantički osetljivog izraza; na eksplanatornom planu, pitanje je kako i zašto se semantička vrednost pojedinih indeksičkih izraza menja od konteksta do konteksta.

Gledano iz ugla *konverzacionog* kontekstualizma koji i Luis zastupa, bitno je istaći da za neke kontekstualno zavisne reči (na primer, kaplanovske čiste indeksikale kao što su „ja", „ti", „sada" ili „ovde") lingvistička pravila unapred diktiraju dopunu odgovarajućim nekonverzacionim kon-

3 Atribut „indeksički" ti izrazi su dobili u kaplanovskoj analizi konteksta kao indeksa, pod kojim se podrazumeva stanje stvari u odnosu na koje procenjujemo istinitost nekog iskaza. Okolnosti procene minimalno uključuju vreme i mogući svet odnosno par [*t*, *w*], ali kada su indeksički izrazi u pitanju, pošto oni svoj sadržaj kompletiraju u kontekstu izricanja, indeks uključuje bar još jednog subjekta i lokaciju [*t*, *l*, *w*, *s*]. Kaplan (Kaplan 1989) je lingvističko značenje indeksičkih izraza nazvao njihovim *karakterom*, zamišljajući ga kao funkciju od konteksta do njihovog *sadržaja* (vrednosti koju im kontekst dodeljuje).

4 Pojam semantičke osetljivosti je širi: indeksičnost je osobina pojedinačnih reči („ja", „sada" i dr.), dok se semantička osetljivost odnosi i na složenije izraze čiji su one deo (npr. „Ovo je knjiga koju sam želeo da pročitam").

5 Razni autori na njih nude različite odgovore. Za stupnjevite prideve vid. Ludlow 2005, Kennedy and McNally 2005; za kvantifikatore Gauker 2010, Heim 1982, Stanley and Szabó 2005.

tekstualnim faktorima (*ko* izgovara „ja", *kome* se govorno lice obraća sa „ti", *vreme* za „sada" i *mesto* izricanja za „ovde"). Za većinu indeksičkih izraza lingvistička pravila tu ulogu dodeljuju i konverzacionim faktorima, pre svega jezičkim namerama govornih lica: „on [osoba na koju govorno lice ukazuje sa *namerom* da na nju referira]"; „ovo [objekat na koji govorno lice pokazuje sa *namerom* da na njega referira]; „svi [objekti koji imaju neko zajedničko svojstvo po kojem ih govorno lice izdvaja]", „bogat [u odnosu na neku klasu poređenja ili standard bogatstva]" i sl.[6]

Pošto je predmet naše pažnje Luisovo dovođenje u vezu kontekstualne zavisnosti „zna" i „svaki", pogledajmo šta se dešava sa rečima koje u svakodnevnom jeziku koristimo za kvantifikaciju. Prema uobičajenoj interpretaciji,[7] one spadaju u semantički osetljive izraze čija se kontekstualna dopuna sastoji u preciziranju (ograničavanju) domena primene. Tu svoju kontekstualnu zavisnost kvantifikatori prenose na složenije jezičke sklopove u kojima se javljaju. Tako, na primer, da bismo mogli da procenimo istinosnu vrednost rečenica koje sadrže jedan ili više kvantifikacionih izraza, moramo prvo da utvrdimo na koji se domen objekata oni u datom kontekstu izricanja odnose.[8] Ponekad se taj domen eksplicitno specifikuje odgovarajućim imeničnim ili deskriptivnim izrazom, ali se češće izostavlja, prećutno podrazumeva i kontekstualno određuje. U takvim slučajevima, sintaksički gledano, kvantifikujući izraz u rečenicu unosi pomenuto prazno mesto rezervisano za tzv. *prikriveni indeksikal* (*hidden indexical*)[9], neizgovoren deskriptivni uslov izražen jednomesnim predikatom koji ima ulogu da ograniči domen kvantifikacije. Tako rečenica „Svi studenti su došli na čas", kojom profesor započinje čas, sadrži kvantifikujuću imeničnu frazu iza koje je sintaksička elipsa „svi studenti [...]" namenjena jednomesnom predikatu koji u datom kontekstu ograničava domen kvantifikacije, recimo „Svi studenti *koji su F* su došli na čas". Kada je izostavljen imenični dodatak – „Svi su došli" – kontekst bi trebalo i njega da odredi. U oba slučaja, profesor u datom kontekstu upotrebljava kvantifikator „svi" tako da ne govori o svim studentima u univerzumu ili na univerzitetu već

6 Zavisno od toga da li jezičke namere učestvuju u dodeljivanju semantičke vrednosti, Kaplan razlikuje čiste indeksikale i prave demonstrative (Kaplan 1989). Tu razliku je preciznije formulisao Peri, nazivajući prve automatskim a druge diskrecionim indeksikalima (Perry 1997).

7 *Cf*. Stanley and Williamson 1995, Stanley and Szabó 2000.

8 Stanley and Williamson 1995: 291.

9 Posebno je pitanje gde tačno smestiti tu prazninu u sintaksičkoj strukturi rečenice. Stenli i Sabo (Stanley and Szabó 2000) dovode je u vezu sa imeničnim dodatkom koji prati kvanitifikujući izraz („Svi [studenti...]"). Neki drugi autori, na primer Westerståhl (1985), tvrde da sintaksička elipsa ide uz kvantifikator („Svi [...]"). Gauker (Gauker 2010) nudi rešenje bez praznog mesta i prikrivenih indeksikala. Ovde usvajamo rešenje koje nude Stenli i Sabo.

samo o studentima *koji pohađaju njegov kurs*. Promenom konteksta može se promeniti i domen – rečenicom „Svi su došli, možemo da počnemo" predsedavajući može da otvori sednicu odbora, ali će tada na prazno mesto iza „svi" doći neartikulisani deskriptivni uslov *osobe koje su članovi odbora*.[10]

Pomenućemo još dva momenta koja se tiču semantičke osetljivosti kvantifikacionih izraza. Prvo, u određenim segmentima razgovora ili rečeničnim sklopovima promenljiva na poziciji argumenta u sintaksičkoj elipsi može da se javlja vezano; to je, recimo, slučaj sa tzv. kvantifikovanim kontekstima, kao u rečenici „Na svakoj polici u mom kabinetu svaka knjiga je na svom mestu"[11], u kojoj prva kvantifikaciona fraza („svaka polica u mom kabinetu") vezuje promenljivu na mestu iza fraze „svaka knjiga". Drugi, za naša zaključna razmatranja još važniji momenat, ogleda se u tome što se isti semantički efekat upotpunjavanja sadržaja rečenice tipa „Svi su došli" najčešće može postići različitim deskriptivnim uslovima – „studenti filozofije koji pohađaju kurs iz epistemologije", „studenti koji su upisali drugu godinu studija filozofije" itd.; štaviše, mogući su i konteksti u kojima se rečenica dopunjuje nabrajanjem (prozivka) ili nekim neverbalnim sredstvom kao što je gest pokazivanja na objekte ili individue koje ulaze u domen.

2.

Kada tvrde da predikat „zna" spada u indeksičke izraze,[12] kontekstualisti nude različite odgovore i na deskriptivno i na eksplanatorno pitanje. Na deskriptivnom planu oni se slažu da „zna" iziskuje kontekstualnu *semantičku* dopunu, ali se razlikuju u tumačenju prirode te uslovljenosti, odnosno u opisu *šta* se u kontekstu upotrebe „zna" dešava sa praznim mestom rezervisanim za neartikulisanu indeksičku komponentu. Dok se većina kontekstualista poziva na analogiju sa komparativnim pridevima, Luis se naizgled oslanja na model kvantifikujućih izraza. Na eksplanatornom planu, u objašnjenju *kako* semantički sadržaj „zna" varira od konteksta do konteksta, razlike su još upadljivije,[13] ali se u njih ovde nećemo

10 Ičikava (Ichikawa 2017: 252) smatra da bi to ograničenje trebalo da se uzme kao intenzionalno, zato što skup objekata ili individua koje u njega ulaze može da se menja od konteksta do konteksta (u našem primeru skup studenata razlikuje se od generacije do generacije).

11 Stanley and Szabó 2000: 242.

12 Upor. Cohen 1988: 97.

13 Među kontekstualistima nema pune saglasnosti u objašnjenju kontekstualnih varijacija. Dirouz se poziva na Nozikov uslov *osetljivosti* kao mehanizam pomoću kojeg

upuštati već ćemo se ograničiti samo na Luisov odgovor. Njegovu verziju kontekstualizma (iz 1996) prikazaćemo samo u osnovnim crtama.

Luis usvaja infalibilističku pretpostavku da je znanje nespojivo sa mogućnošću pogreške i formuliše sledeću početnu definiciju znanja:[14]

> (D1) Subjekt *S* zna propoziciju *p* akko *S*-ova evidencija otklanja svaku mogućnost u kojoj je ne-*p*.

Nevolja sa infalibilističkim zahtevom je u tome što on direktno vodi u skepticizam.[15] Luis ne bi želeo da od njega potpuno odustane ali smatra da se, kako to slikovito opisuje, može provući između „stene falibilizma" i „vrtloga skepticizma" (1996: 566) zahvaljujući indeksičnosti kvantifikujuće sintagme „svaka ne-*p* mogućnost". S obzirom na taj manevar, stiče se utisak da je ključni korak u njegovoj odbrani kontekstualne zavisnosti sadržaja predikata „zna" i istinosnih uslova rečenica oblika „*S* zna da *p*" u indeksičnosti kvantifikatora „svaka". Takav utisak naizgled potkrepljuju sledeća njegova zapažanja:

> „Šta znači reći da je svaka mogućnost u kojoj je ne-P eliminisana? Govor koji uključuje kvantifikaciju poput 'svaka' po pravilu je ograničen na neki omeđen domen. Ako kažem da je svaka čaša prazna i da je zato trenutak za sledeću turu, moji sagovornici i ja besumnje ignorišemo većinu svih čaša u celokupnom svetu i svim vremenima. One su izvan domena. One su irelevantne za istinitost rečenice koju sam izrekao.
>
> Slično tome, ako kažem da je svaka neotklonjena mogućnost mogućnost u kojoj P jeste slučaj, ili nešto što ima to značenje, besumnje ignorišem neke od svih postojećih neotklonjenih alternativnih mogućnosti. One su izvan domena, one su irelevantne za ono što sam rekao." (1996: 553)

Luis se očigledno drži uobičajenog tumačenja kvantifikatora kao izraza koji iziskuju kontekstualnu semantičku dopunu. Kao i u drugim indeksičkim izrazima,[16] lingvističko značenje reči „svaka" („svi") predstavlja funkciju od konteksta upotrebe do njihovog semantičkog sadržaja (objekti na koje se referira). Osim možda u nekim posebnim okolnostima – recimo, kada želimo da navedemo primer gramatički ispravne rečenice – tvrdnju „Sve čaše su prazne" nećemo izreći sa namerom da se odnosi na sve čaše u univerzumu već samo na (prećutno) omeđen domen objekata. Tako u Lu-

pripisivači znanja variraju zahtev u pogledu snage subjektovog saznajnog položaja, Koen smatra da je glavni mehanizam *isticanje* mogućnosti pogreške, dok Luis navodi ukupno sedam *pravila relevantnosti* alternativa.

14 Vid. Lewis 1996: 549, 551. Luisova definicija sadrži dve formulacije definiensa; ovde navodimo samo drugu.

15 Vid. Unger 1971; Cohen 1988.

16 To je Kaplan (Kaplan 1989) pokazao na primeru pomenutih čistih indeksikala („ja", „sada", „ovde" i sl.) i pravih demonstrativa („ovo", „onaj" i dr.).

isovom primeru rečenica „Sve čaše su prazne" sadrži sintaksičku elipsu iza kvantifikujuće sintagme „sve čaše" rezervisanu za objekte koji ispunjavaju neartikulisanu indeksičku komponentu preciziranu deskriptivnim uslovom *čaše koje su na našem stolu* ili *čaše iz kojih pijemo* ili nekim sličnim.

Uopšteno gledano, dakle, odgovor na deskriptivno pitanje o kontekstualnoj osetljivosti kvantifikacionih izraza bio bi da lingvistička pravila njihove upotrebe iziskuju odgovarajuće, obično fonološki neartikulisane jednomesne predikate pomoću kojih govorna lica ograničavaju domen kvantifikujuće fraze i na taj način upotpunjuju semantički sadržaj izgovorenih rečenica.[17] U skladu sa tim odgovorom Luis interpretira i upotrebu kvantifikatora „svaka" u definiensu (D1): u svakodnevnom govoru, kada za neku osobu tvrdimo da zna propoziciju *p*, ne podrazumevamo da ona raspolaže evidencijom koja otklanja apsolutno sve ne-*p* mogućnosti već samo one koje u datom kontekstu uključujemo u domen kvantifikatora „svaka"; ostale ne-*p* mogućnosti ignorišemo kao *irelevantne* za tačnost naše saznajne tvrdnje. Uzimajući to u obzir, Luis nudi novu formulaciju definicije znanja u kojoj uvodi *sotto voce* ograničenje i nagoveštava pretpostavljenu kontekstualnu zavisnost tog pojma:

> (D2) Subjekt *S* zna propoziciju *p* akko *S*-ova evidencija otklanja svaku mogućnost u kojoj je ne-*p* – Psst! – osim onih mogućnosti koje s pravom ignorišemo.[18]

Definicija (D2) u stvari predstavlja ključni korak ka primeni ideje o relevantnim alternativama.[19] Sintagma „one ne-*p* mogućnosti koje imamo pravo da ignorišemo" upotrebljena je u značenju u kojem se te mogućnosti mogu okvalifikovati kao irelevantne alternative, dok je domen onih ne-*p* mogućnosti koje nemamo pravo da ignorišemo kontekstualno omeđen imeničnom frazom „relevantne alternative". Pridev „relevantno" očigledno služi za kontekstualno ograničavanje domena „svaki" kojim dopunjuje semantički sadržaj predikata „zna", odnosno istinosne uslove rečenice „*S* zna da *p*".

Takav Luisov odgovor nam otkriva deskriptivni uslov pod kojim alternative ulaze u kontekstualno omeđen domen fraze „svaka alternativa",

---

17 Dopuna može da se interpretira i pragmatički: uzeta u doslovnom značenju, rečenica se zaista odnosi na sve čaše u univerzumu, ali se u kontekstu upotrebe domen kvantifikacije ograničava nekim grajsovskim mehanizmom (speaker's meaning). Na nedostatke pragmatičke i prednosti semantičke interpretacije ukazali su Stenli i Sabo (Stanley and Szabó 2000).

18 Vid. Lewis 1996: 553.

19 Iako se već na sledećoj strani (554) poziva na Angera, ideja je u stvari Dreckeova (ranu verziju nalazimo kod Ostina), s tim što je Drecke uticaj kontekstualnih činilaca ograničio na pragmatičku dimenziju znanja.

ali nam još uvek nedostaje odgovor na eksplanatorno pitanje *zašto* su to baš te-i-te a ne neke druge alternative. Specifikovanje domena kvantifikacionih izraza u situacijama kao što je kafanski govor o „svim čašama" po svoj prilici je potpuno arbitrarno i zavisno samo od namera govornih lica koja izriču rečenicu „Sve čaše su prazne" – u razgovoru za kafanskim stolom ništa me ne sprečava da govorim i o čašama na drugim stolovima, čašama na pultu i policama, čašama u drugim prostorijama, čašama koje su mi ostale kod kuće na trpezarijskom stolu i sl. Epistemički kontekst i govor o znanju ipak nameću izvesna ograničenja. Ako ništa drugo, nije nam dopušteno da po svojoj volji biramo koje ćemo ne-*p* mogućnosti ignorisati a koje ne – u protivnom, uvek bismo sa lakoćom mogli da ispunimo (D2) zahtev za znanje!

Upravo to Luis ima u vidu kada u definiensu (D2) dodaje „s pravom ignorišemo". On je, naime, svestan toga da su nam u epistemičkim kontekstima neophodni neki objektivni ili bar intersubjektivni kriterijumi relevantnosti alternativa. Prema njegovom mišljenju, da li je neka alternativa u datom kontekstu relevantna ili ne, regulisano je određenim pravilima o čijoj primeni sagovornici moraju da vode računa kada jedni drugima pripisuju ili odriču znanje – od njih zavisi da li ćemo u datom kontekstu imati pravo da ignorišemo neku ne-*p* mogućnost. Luis navodi ukupno sedam takvih pravila, od kojih su četiri prohibitivna (nalažu nam koje alternative nemamo pravo da ignorišemo), dok su tri permisivna (daju nam za pravo da ignorišemo neke alternative). Iako se većina tih pravila odnosi na objektivne kontekstualne činioce (aktualne okolnosti i saznajni položaj subjekta znanja), bar neka od njih otvaraju prostor za uticaj konverzacionih faktora na semantički sadržaj predikata „zna" i rečenica oblika „*S* zna da *p*".[20] Konkretno, to čine dva Luisova pravila: pravilo pažnje i pravilo (upadljive) sličnosti. Prvo pravilo nam nalaže da ne ignorišemo nijednu ne-*p* mogućnost koja je postala predmet naše pažnje,[21] dok nam drugo kaže da, kada već postoji neka ne-*p* mogućnost koju shodno drugim pravilima nemamo pravo da zanemarimo, ne bi trebalo da ignorišemo ni bilo koju drugu ne-*p* mogućnost koja joj je upadljivo slična.

---

20 Neki autori su prednost davali objektivnim kriterijumima relevantnosti kao što su, na primer, realna mogućnost i objektivna verovatnoća da je alternativa ostvarena u datim okolnostima (Dretske 1981; Goldman 1976). Stenli ih je, pak, doveo u vezu sa praktičnim interesima subjekta znanja (Stanley 2005).

21 *Sotto voce* ograničenje u (D2) posredno ukazuje na značaj koji Luis daje pravilu pažnje. Ono igra glavnu ulogu u Luisovom odgovoru skeptiku jer objašnjava kako i najbizarnije skeptičke alternative tipa „mozgovi u posudi" mogu postati relevantne. Ipak, kao glavni „krivac" za eluzivnost znanja, ono je jedna od naslabijih tačaka Luisove verzije kontekstualizma i s razlogom je bilo predmet kritike (vid. Williams 2004).

Ovaj kratak prikaz Luisove verzije kontekstualizma sasvim je dovoljan za potrebe naše glavne teme. Mislim da je Ičikava načelno u pravu kada primećuje da su pozivanje na kontekstualnu zavisnost „svaka" i artikulacija pravila relevantnosti dva nezavisna projekta.[22] U svakom slučaju, oni se uzajamno nadopunjuju kao odgovori na deskriptivno i eksplanatorno pitanje o pretpostavljenoj indeksičnosti glagola „znati". Štaviše, moglo bi se reći da je za gledište *konverzacionog* kontekstualizma koje Luis zastupa važniji odgovor na drugo, eksplanatorno pitanje. Čak i ako su neka od pravila koje Luis nudi sporna – najčešće su predmet kritike upravo pravila u kojima se uzima u obzir konverzacioni kontekst pripisivača znanja, pravilo pažnje i pravilo sličnosti – za zastupnike tog gledišta neophodno je da pokažu da neki konverzacioni faktori utiču na kompletiranje semantičkog sadržaja predikata „zna" i rečenica „*S* zna da *p*".

Ovde ćemo ipak ostaviti po strani probleme koji se mogu javiti u vezi sa pojedinim pravilima relevantnosti. Bavićemo se pre svega deskriptivnim pitanjem i odgovorom koji se obično pripisuje Luisu: da „zna" svoju pretpostavljenu kontekstualnu zavisnost duguje kvantifikatoru „svaka" i to tako što na naznačeno prazno mesto iza „zna" u rečenicama oblika „*S* zna da *p*" dolazi uslov isključenja svih onih alternativa koje su *relevantne* u kontekstu njihovog izricanja.

3.

Jednostavnosti radi, prvo ćemo preformulisati Luisove definicije. Pošto je pojam alternative jednoznačno i akontekstualno određen – neki iskaz *q* je alternativa iskazu *p* akko *q* i *p* ne mogu da budu istovremeno istiniti – umesto „ne-*p* mogućnost" pisaćemo „alternativa", a pridev „relevantna" dodaćemo kao odrednicu za one alternative koje, izraženo Luisovom terminologijom, u datom kontekstu nemamo pravo da ignorišemo:[23]

> (D1') Subjekt *S* zna propoziciju *p* akko *S*-ova evidencija otklanja svaku alternativu.
>
> (D2") Subjekt *S* zna propoziciju *p* akko *S*-ova evidencija otklanja svaku alternativu – Psst! – osim onih alternativa koje su irelevantne.

Na prvi pogled, čini se kao da je u (D2') napravljen prilično grub previd. Naime, pošto je *sotto voce* ograničenjem skup svih alternativa sveden na podskup onih koje su relevantne, izgleda kao da definiens izlaže samo

22 Vid. Ichikawa 2011. Za Ičikavu problematičan je drugi, a prihvatljiv prvi projekat koji razvija u: Ichikawa 2017.

23 I sam Luis kasnije u tekstu često termin „alternativa" koristi kao zamenu za „ne-*p* mogućnost".

nužan uslov istinitosti rečenice „*S* zna da *p*" i da bi između njega i definienduma umesto „akko", kao veznika ekvivalencije, trebalo da stoji „ako". Prema jednom od pravila relevantnosti na koje se Luis poziva, formulacija ipak nije pogrešna. U pitanju je pravilo aktualnosti, koje nalaže da ni u jednom kontekstu nemamo pravo da ignorišemo onu alternativu koja je aktualizovana. U svetlu tog pravila, za tačnost „*S* zna da *p*" irelevantna je i s pravom ignorisana *svaka* alternativa *q* koja *de facto* nije aktualizovana (da jeste, *p* ne bi bilo tačno). Samim tim, to što *q* ne ulazi u domen relevantnih alternativa ujedno znači da u datom kontekstu ona nije aktualizovana; u protivnom, pošto Luisovo pravilo aktualnosti nalaže da aktualnost ne sme biti ignorisana, ona bi morala da bude relevantna.[24]

Uvođenje pojma relevantne alternative omogućuje još jednostavniju formulaciju definicije (D2') i ekspliciranje glavne postavke teorije relevantnih alternativa:

> (D3) *S* zna *p* akko *S*-ova evidencija otklanja svaku relevantnu alternativu *q*.

Kao i prethodne, i ova definicija je formulisana neformalno, u objekt jeziku. Uostalom, Luis je celokupan svoj tekst napisao u tom obliku, zaključivši ga sledećom porukom:

> „Mogao sam veoma pažljivo da razlikujem (1) jezik koji koristim za govor o znanju, ili bilo čemu sličnom, i (2) drugi jezik koji koristim za govor o semantičkim i pragmatičkim aspektima prvog jezika. Ako želite da čujete moju priču ispričanu na taj način, onda ste verovatno već dovoljno upućeni da taj posao i sami obavite." (1996: 567)

Poslušaćemo ovu Luisovu poruku pa ćemo posao ekspliciranja kontekstualne zavisnosti definienduma i definiensa u (D3) obaviti sami tako što ćemo je formulisati metalingvistički:

> (D4) „*S* zna da *p*" je istinito u *K* akko je „*S*-ova evidencija isključuje svaku *relevantnu* alternativu" istinito u *K*.[25]

Kao što vidimo, analogija sa ponašanjem univerzalnog kvantifikatora „svi" Luisu je poslužila da se provuče između „stene falibilizma" i „vrtloga skepticizma", pri čemu je uvođenjem *sotto voce* ograničenja početnu infalibili-

24 Utisak je da pojedini komentatori to gube iz vida. Šejfer na primer Luisovu definiciju formuliše metalingvistički „Rečenica oblika '*S* zna da *p*' istinita je u kontekstu *K* akko *S*-ova evidencija otklanja svaku ne-*p* mogućnost koja je relevantna u *K*" (2015: 475), ali za „akko" tu nema mesta bez naznake da se iz nekog razloga istinitost *p* već podrazumeva.

25 Ova metalingvistička formulacija ima dodatnu prednost što je neutralna u pogledu tradicionalnog (invarijantističkog) i kontekstualističkog (varijantističkog) tumačenja znanja (upor. Schaffer 2015: 475).

stičku definiciju pojma znanja *de facto* pretvorio u falibilističku. Drugim rečima, premda od nas traži da iz pojmovnih razloga načelno prihvatimo infalibilizam, Luis priznaje da smo praktično, to jest u primeni pojma znanja falibilisti.

4.

Pretpostavimo da je teza konverzacionog kontekstualizma tačna, da Luis tu tezu povezuje sa idejom o relevantnim alternativama i da, ukazujući na kontekstualnu zavisnost izraza „svaka alternativa" koji figurira u definiensu (D1), objašnjava zašto je i glagol „znati" semantički osetljiv na kontekst upotrebe. Da li je to dovoljno da zaključimo da je Luis kontekstualnu zavisnost „znati" objašnjavao prema modelu univerzalnog kvantifikatora „svi"? Da li tako shvaćena kontekstualizacija pojma znanja proizlazi iz sintagme „svaka alternativa" sadržane u definiensima (D1') i (D2')?

Prvo, ako bi to bio slučaj, glagol „znati" bi morao da preuzme bar neka važnija semantička i sintaksička obeležja karakteristična za kvantifikacione reči kao što je „svi". To, međutim, nije slučaj. Na neke značajne razlike ukazao je Džejson Stenli u svojoj lingvističkoj kritici kontekstualizma. Kao što smo na početku naveli, on spada u autore koji smatraju da je Luis kontekstualnu zavisnost predikata „zna" izvodio iz univerzalne kvantifikacije sadržane u definiciji pojma znanja:

> „Prema Luisu, semantika reči 'zna' uključuje univerzalnu kvantifikaciju nad mogućnostima. (...) Kontekstualizam u pogledu 'zna' Luis, dakle, izvodi iz, prvo, tvrdnje da 'zna' uključuje univerzalnu kvantifikaciju nad mogućnostima i, drugo, iz činjenice da je kvantifikacija u prirodnom jeziku po pravilu ograničena." (Stanley 2005: 61)

Ipak, jedna od karakteristika većine tipičnih indeksičkih izraza, uključujući i kvantifikatore kao što je „svi", jeste fleksibilnost njihove upotrebe. Naime, konverzacioni kontekst u kojem ih koristimo obično ne nameće neka posebna ograničenja u sadržaju (referencijama) koji im govorna lica dodeljuju, tako da se može desiti da u relativno kratkom segmentu razgovora, pa čak i u istoj rečenici jednu istu semantički osetljivu reč ili frazu smisleno i razumljivo upotrebimo dva ili više puta sa različitim semantičkim sadržajem.[26] Ilustrativan je sledeći Stenlijev primer kratkog dijaloga između osoba A i B:

26 To je obeležje i kvantifikatora kao što je „svi" i većine tipičnih indeksičkih izraza (cf. Stanley 2005: 58). Izuzetak su donekle kaplanovski čisti indeksikali kao „ja", „ti", „ovde" i slični, a njihov objekt referencije fiksiran je nekim vanjezičkim obeležjem konteksta u skladu sa odgovarajućim lingvističkim pravilom.

A: Svaka Van Gogova slika nalazi se u holandskom Nacionalnom muzeju.

B: Došlo je, znači, do promene. Kada sam poslednji put posetio Nacionalni muzej, video sam svaku Van Gogovu sliku, ali su neke definitivno nedostajale.[27]

I sagovornicima i nama kao posmatračima sa strane potpuno je razumljivo da je kvantifikaciona fraza „svaka Van Gogova slika" u prvoj rečenici upotrebljena tako da se odnosi na *sve* Van Gogove slike, dok osoba B govori o svim Van Gogovim slikama *koje su bile izložene* u trenutku kada je ona posetila muzej.

Sličnu situaciju imamo i u sledećem Stenlijevom primeru: „U Atlanti ima mnogo serijskih ubica, ali nema mnogo nezaposlenih ljudi", gde nam je jasno da se kvantifikaciono upotrebljena reč „mnogo" u dva javljanja odnosi na različite domene.[28] Možda bi još ilustrativniji primer bio izveštaj sportskog komentatora sa početka utakmice: „Igrači oba tima istrčali su na teren i svako se pozdravio sa svakim", u kojem nam je očigledno da se reč „svaki" u prvom javljanju odnosi na igrače jednog, a u drugom na igrače drugog tima.

Ako bi svoju kontekstualnu zavisnost dugovao kvantifikatoru „svi", i predikat „zna" bi trebalo da ispoljava sličnu fleksibilnost, što znači da bismo mogli u istom konverzacionom kontekstu da ga upotrebljavamo tako da prećutno referiramo na različite skupove relevantnih alternativa. Izgleda, ipak, da epistemički konverzacioni kontekst to ne dopušta. Naime, takva slobodna upotreba predikata „zna" sobom donosi opasnost od zapadanja u protivrečnost. Najjednostavniji primer su konjunktivne formulacije koje su već na prvi pogled neprihvatljive, kao što bi bilo sledeće zapažanje: „On zna da se svaka Van Gogova slika nalazi u holandskom Nacionalnom muzeju, ali ne zna da neke od njih nedostaju" ili implikacija Dreckeovog i Nozikovog rešenja problema skepticizma koje poriče deduktivnu zatvorenost znanja: „On zna da ima ruke, premda [on] ne zna da nije bestelesni mozak u posudi". U oba slučaja je očigledno da se drugim članom konjunkcije negira semantička implikacija koju sobom nosi prvi član konjunkcije.[29]

Pošto takve rečenične konstrukcije izgledaju nedopustive, Stenli zaključuje da je semantička analiza predikata „zna" po uzoru na kvantifikator „svi" neodrživa.[30] Taj zaključak možemo dodatno da potkrepimo

27 *Ibid.*: 65.

28 *Ibid.*: 67.

29 Iz tog razloga je Dirouz (DeRose 1995) takve konjunkcije okvalifikovao kao *nepodnošljive* (*abominable*).

30 Stenlijev zaključak naizgled važi samo za pokušaj da se semantika glagola „znati" tumači prema modelu kvantifikacionih reči koje se upotrebljavaju sa ograničenim

ukazivanjem na neke sintaksičke razlike između te dve reči. Istakli smo na početku da kvantifikatori, poput ostalih semantički nepotpunih indeksičkih izraza, u složenije jezičke sklopove unose sintaksičku elipsu rezervisanu za argument funkcije kojoj kontekst dodeljuje vrednost; u različitim kontekstima govornim licima su na raspolaganju brojni imenični izrazi i jednomesni predikati (čak i neverbalna sredstva) pomoću kojih mogu da ekspliciraju domen kvantifikacije. Sa „znati" to naizgled nije slučaj. Na pretpostavljeno prazno mesto iza „zna" u rečenicama oblika „*S* zna [...] da *p*" ne može da stupi bilo koji od neograničenog broja određujućih izraza. Naprotiv, kao što definicija (D3) pokazuje, ako je „zna" uopšte indeksički izraz, onda njime prilikom upotrebe, odnosno izricanja rečenice „*S* zna da *p*", govorna lica prećutno referiraju na skup koji je određen samo jednim jednomesnim predikatom – na alternative koje su shodno odgovarajućim pravilima (Luisovim ili nekim drugim) u datom kontekstu stekle svojstvo *relevantnosti*.

Ilustrujmo to na Luisovom primeru rečenice „Sve čaše su prazne" izrečene u kontekstu u kojem je njen upotpunjen sadržaj „Sve čaše *na našem stolu* su prazne". Sintaksička struktura te rečenice može da se prikaže u obliku „Za svako *o* u [čaša, *F*(*i*)], *o* je prazno", u kojem je sintaksička elipsa rezervisana za objekte *o* koji ulaze u domen kvantifikacije omeđen jednomesnim predikatom *F*. Taj predikat će, međutim, moći gotovo neograničeno da varira: za rečenicu „Sve čaše *koje su F* su prazne" u jednom kontekstu će to biti čaše *koje su na kafanskom stolu* za kojim sagovornici sede, u drugom će to biti čaše *koje su na šanku* uz koji oni stoje, u trećem čaše *na koje govorno lice pokazuje* itd. Paralelno sa tim, pretpostavimo da tvrdimo da je ispunjen uslov iz (D1) „Sve alternative su eliminisane". Odgovarajuća sintaksička struktura trebalo bi da bude „Za svako *o* u [alternativa, *F*(*i*)], *o* je eliminisano". Kada bi postojala analogija u semantičkoj osetljivosti između „zna" i „svi", deskriptivni uslov *F* bi i u toj konstrukciji mogao neograničeno da varira: u jednom kontekstu bi to mogle da budu alternative *na koje smo obratili pažnju*, u drugom alternative *koje su zabavne*, u trećem alternative *koje su napisane na tabli* itd. Teorija relevantnih alternativa, međutim, nameće jedinstven deskriptivni uslov ekspliciran u definiensu (D3) – neophodno je isključiti *relevantne* alternative. Jednomesni predikat „relevantna" ostaje isti u svakom kontekstu pripisivanja zna-

domenom (D-kvantifikatori). Ima autora koji tvrde da je taj zaključak prestrog i da „zna" u nekim kontekstima ispoljava izvestan stepen fleksibilnosti (Ludlow 2005: 36–37; Baumann 2016: 178–181; Ichikawa 2011). Takođe, Šejfer i Sabo su ponudili semantički model adverbijalnih kvantifikatora (A-kvantifikatori) kao što je „uvek", koji su sličniji kaplanovskim čistim indeksikalima i sa kojima, prema njihovom mišljenju, „zna" deli obeležje nefleksibilnosti (Schaffer and Szabó 2014). Neki noviji kontekstualisti opredeljuju se za to rešenje (Ichikawa 2017, Blome Tillman 2022).

nja. Promenom konteksta ne menja se *atribut relevantnosti* – kriterijumi relevantnosti su fiksirani pravilima – menja se jedino *skup* alternativa čije je isključenje nužan i dovoljan uslov istinitosti rečenice „*S* zna da *p*". Prema tome, ako je tačno da je „*S* zna da *p*" kontekstualno zavisno zbog toga što infalibilistički zahtev „*S*-ova evidencija eliminiše svaku alternativu" podvrgavamo *sotto voce* ograničenju i primenjujemo u obliku „*S*-ova evidencija eliminiše svaku relevantnu alternativu", izbegavajući tako „stenu falibilizma" i „vrtlog skepticizma", izvor semantičke osetljivosti predikata „zna" nije u indeksičnosti kvantifikatora „svi" već u kontekstualnoj zavisnosti prideva „relevantna" sadržanog u imeničnoj frazi „relevantna alternativa".

Uostalom, sadržinski gledano ne gubimo ništa ako u definiensu metalingvističke formulacije (D4) izostavimo kvantifikator „svaka" a zadržimo pridev „relevantna" jer je on dovoljan za specifikaciju domena alternativa čija je eliminacija uslov istinitosti rečenice „*S* zna da *p*":

> (D5): „*S* zna da *p*" je istinito u *K* akko je „*S*-ova evidencija isključuje *relevantne* alternative" istinito u *K*.

Naravno, dodavanje univerzalnog kvantifikatora „sve" moglo bi da se pravda čisto formalnim razlozima, ali je sadržinski (D5) i bez njega sasvim u redu s obzirom na to kako u neformalnoj prezentaciji na eksplanatornom planu funkcionišu Luisova pravila relevantnosti: kombinovanom primenom u *K* ona fiksiraju skup *relevantnih* alternativa, a nužan i dovoljan uslov istinitosti tvrdnje „*S* zna da *p*" izrečene u *K* jeste da *S*-ova evidencija *te* alternative isključuje. Ako je tačno da je predikat „zna" semantički osetljiv na kontekst upotrebe i da kao takav u sintaksičku strukturu rečenice „*S* zna da *p*" unosi prazno mesto rezervisano za slobodnu promenljivu kojoj kontekst dodeljuje odgovarajuću vrednost, tu prazninu uvek i jedino popunjava skup kontekstualno relevantnih alternativa: *S* zna u odnosu na alternative (*a*) koje su relevantne (*R*) u kontekstu *K* da *p*, odnosno „*S* zna [*Ra*, *K*] da *p*".[31] Drugim rečima, dok je kvantifikator „svi" indeksički izraz nezavisno od imeničnog dodatka sa kojim formira kvantifikujuću frazu, kada se uvede *sotto voce* ograničenje u (D2), predikat „zna" svoju pretpostavljenu kontekstualnu zavisnost dobija od atributa „relevantna" kao sastavnog dela imenične fraze „relevantna alternativa".[32]

---

31 Zahtev za isključenjem skupa alternativa može da se prikaže i kao epistemički standard koji pripisivači znanja primenjuju u datom kontekstu: „*S* zna [u odnosu na *K*-standard] da *p*". Vid. Ludlow 2005.

32 Ako su ta naša zapažanja na mestu, Luis izbegava Stenlijev prigovor i nedopustive konjunkcije. Naime, pošto kontekstualnu zavisnost predikata „zna" ne izvodi iz kontekstualne zavisnosti reči „svaka", on nije obavezan da dozvoli slobodno javljanje „zna" i formulacije tipa „*S* zna da ima ruke, ali *S* ne zna da nije bestelesni mozak u posudi".

Reference:

Baumann, Peter. 2016. *Epistemic Contextualism*. Oxford: Oxford University Press.

Blome-Tillmann, Michael. 2022. *The Semantics of Knowledge Attributions.* Oxford: Oxford University Press.

Cohen, Stewart. 1986. Knowledge and Context. *Journal of Philosophy* 83: 574–583.

Cohen, Stewart. 1988. How to be a Fallibilist. *Philosophical Perspectives* 2: 91–123.

DeRose, Keith. 1992. Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research*. 52 (4): 913–929.

DeRose, Keith. 1995. Solving the Skeptical Problem. *Philosophical Review* 104: 1–52.

Dretske, Fred. 1981. The Pragmatic Dimension of Knowledge. *Philosophical Studies* Vol. 40, No. 3: 363–378.

Gauker, Christopher. 2010. Global Domains versus Hidden Indexicals, *Journal of Semantics* 27: 243–270

Goldman, Alvin. 1976. Discrimination and Perceptual Knowledge. *Journal of Philosophy* 73: 771–91.

Heim, Irene. 1982. *The semantics of definite and indefinite noun-phrases*. PhD thesis, University of Massachusetts, Amherst.

Ichikawa, Jonathan Jenkins. 2011. Quantifiers and Epistemic Contextualism. *Philosophical Studies* 155: 383–98.

Ichikawa, Jonathan Jenkins. 2017. *Contextualising Knowledge*. Oxford: Oxford University Press.

Kaplan, David, 1989. Demonstratives. In: Almog, J., Perry, J. and Wettstein, H. (eds.). *Themes from Kaplan*. Oxford University Press: 481–563.

Kennedy, Christopher and L. McNally. 2005. Scale structure and the semantic typology of gradable predicates. *Language* 81: 345–81.

Lewis, D. 1996. Elusive Knowledge. *Australasian Journal of Philosophy* 74 (4): 549–567.

Ludlow, Peter. 2005. Contextualism and the New Linguistic Turn in Epistemology. In: Gerhard Preyer and Georg Peter (eds.) *Contextualism in Philosophy*. Oxford: Clarendon Press: 11–50.

Perry, John. 1979. The Problem of the Essential Indexical. *Noûs* 13: 3–21.

Perry, John. 2006. Using Indexicals. In: Michael Devitt and Richard Hanley (eds.). *The Blackwell Guide to the Philosophy of Language*. Blackwell: 314–334.

Schaffer, Johnatan 2015. Lewis on Knowledge Ascriptions. In: Barry Loewer and Jonathan Schaffer (eds.). *A Companion to David Lewis*. Chichester: Wiley-Blackwell: 471–90.

Schaffer, Jonathan and Zoltán G. Szabó. 2014. Epistemic Comparativism: A Contextualist Semantics for Knowledge Ascriptions. *Philosophical Studies* 168(2): 491–543.

Stanley, Jason & Timothy Williamson. 1995. Quantifiers and context dependence. *Analysis* 55: 291–95.

Stanley, Jason and Zoltán G. Szabó. 2000. On Quantifier Domain Restriction. *Mind and Language* 25: 219–61.

Unger, Peter. 1975. *Ignorance: A Case for Scepticism*. Oxford: Clarendon Press.

Westerståhl, Dag. 1985. Determiners and context sets. In: van Benthem, J. and ter Meulen, A. *Generalized Quantifiers in Natural Language*. Dordrecht: Foris, 45–71.

Williams, Michael. 2004. Knowledge, Reflection and Sceptical Hypotheses. *Erkenntnis* 61: 315–343.

Živan Lazović

LEWIS, „KNOWS" AND „EVERY", OR HOW TO AVOID „THE ROCK OF FALLIBILISM" AND „THE WHIRPOLL OF SCEPTICISM"

Summary: In his influential article „Elusive Knowledge" (1996), David Lewis presented a version of epistemic contextualism that relies on the idea of relevant alternatives. The basic contextualist thesis is that the semantic content of the predicate „knows" and the sentence of the form „*S* knows that *p*" is sensitive to the conversational context of speakers who use them to attribute knowledge. In order to clarify the semantics of the verb „to know", contextualists have usually appealed to analogies with seemingly indisputable context-sensitive expressions such as Kaplan's indexicals, comparative adjectives and words used in ordinary language for quantification. Lewis is usually credited with the last of these three semantic models. Although he pointed to the connection between the use of „knows" and the universal quantifier „every", contained in the infallible requirement of eliminating all possibility of error, this paper aims to show that Lewis derived the contextual dependence of „knows" from the semantic sensitivity of the adjective „relevant" which is manifested in the contextual variation of a set of relevant alternatives whose exclusion is necessary for knowledge.

Keywords: „knows", „every", relevant alternatives, conversational context, semantic sensitivity

Bojan Borstner*
Niko Šetar**

PROVIDING KNOWLEDGE AND VIRTUE TO OTHERS: THE THIRD RESPONSIBILITY

Abstract: As is well known among epistemologists and those interested in the field, one of the main approaches in virtue epistemology is virtue responsibilism, which in a nutshell claims that epistemic agents are responsible for acquiring virtuous character traits which allow them to obtain knowledge. The notion of responsibility widens once epistemic vices are introduced into the knowing game, and Cassam (2019) defines two types of responsibility that pertain to the latter, i.e., acquisition responsibility and revision responsibility. However, one may sometimes not be held responsible for acquisition, and revision may often, as we will show in this article, be next to impossible, or the opportunity for revision might arrive insanely late and take a lamentable moral toll. We argue that there is a need for preventing the need for vice revision, and that that can be only achieved if we are to somehow to improve our virtue acquisition. The latter seems like a gargantuan task to undertake but might not be so if we withdraw from the subpersonal approach to virtue and vice and instead look at the whole epistemological process, from obtaining virtue, to gathering knowledge, all the way to revising vice, as a social process. I will attempt to defend this view and introduce a new kind of responsibility, one that requires that people provide virtue to each other, and analyse several ways in which that may be achieved, as well as elaborate on how that reduces need for revision and what outstanding issues still remain.

Keywords: virtue epistemology, vice epistemology, responsibilism, acquisition, revision, provision

1. Introduction to Virtue and Vice

Epistemic virtue is a concept derived from Aristotelean conception of virtues, albeit with a number of modifications. Aristotle conceived of

* Faculty of Philosophy, University of Maribor, bojan.borstner@um.si
** Faculty of Philosophy, University of Maribor, niko.setar1@um.si

virtues as manners of conduct that are neither exaggerated nor lacking – in context of moral virtue, an example would be the virtues of generosity and frugality, which are both situated between two extremes; wastefulness (being the extreme of generosity) and stinginess (being the extreme of frugality). Even though his virtues are often considered fundamentally ethical, Aristotle did not restrain himself to morality and also accounted for virtues of the mind, which, differing from virtues of the soul that lead, when exercised, to moral good, pertain mostly to intellectual or cognitive conduct that when exercised leads to epistemic good (Nichomachean Ethics VII).

This is the strand of virtues that was picked up by epistemologists in the second half of the previous century, and thus contemporary virtue epistemology started filling the pages of journals. The first important definition of virtues in epistemology was authored by Ernest Sosa (1991), who took what is called the reliabilist (see also Greco, 1999) stance towards these virtues. He claimed that the latter are certain cognitive and sensory faculties that must be correctly developed and function reliably in order for their holder to obtain knowledge. Later, an approached named the AAA Model, which argues that epistemic performances must be accurate (achieve their goal), adroit (competent in using faculties required to achieve that goal) and apt (accurate because of being adroit), of reliabilist virtue was constructed (see Sosa, 2007). This model, we would argue, paves the way for a compatibilist view of reliabilism with the next approach we will describe, responsibilism (although that possible compatibility falls outside the scope of this article).

Responsibilism can be traced back to Montmarquet (1992), who pointed out that knowledge is obtained by intentional and voluntary pursuit of truth, which is done through virtuous intellectual conduct. The approach was named and refined by Linda Zagzebski (1996), who argued that virtues are knowledge-conductive character traits, and that an epistemic agent is responsible for developing such traits throughout her epistemic pursuits. Zagzebski claimed that reliabilist approach fails to account for the link between epistemic and moral virtue, and that regarding virtues as reliable faculties tends to discriminate against those who are, due to congenital or other medical conditions, unable to develop such faculties; it cannot be a failure on their part, they cannot be considered responsible.

This view allowed the conception of epistemic vices as virtues' negative counterparts, that is, initially, as character traits that inhibit one's ability to obtain knowledge. In what is perhaps the most important work in vice epistemology to date, Cassam (2019) defines epistemic vice as "intellectual defects that get in the way of knowledge" (ibid, ix), which is not

to say they should be defect in a reliabilist sense, but rather intellectual conducts that are defective even while subsisting on reliable "hardware". Further, he asserts that knowledge is true belief in which we have to be reasonably confident – we are vicious when we foster a belief (true or not) that we are confident in for the wrong reasons, or our confidence is supported by faulty evidence. The author splits these vices into three categories: character traits, attitudes and postures, and ways of thinking. Character traits are the most high-fidelity of the three, meaning that they are expressed in all or near-all epistemic processes the agent undertakes, with examples of them being intellectual/epistemic hubris, arrogance, close-mindedness, and such. Attitudes and postures usually apply to one particular context or field that the agent is biased towards in a certain way – it is worth noting that attitudes are considered involuntary, while postures are voluntary. An example of an attitude might be epistemic insouciance, where one ignores truth-value of their belief entirely because doing so benefits them, while an example of a posture is epistemic malevolence, where one intentionally hides or twists the truth. Ways of thinking as the third type of vice according to Cassam have the lowest fidelity; that is to say they might appear only in one instance in a person who is otherwise epistemically virtuous – such are certain conspiratorial thoughts and false presuppositions.

2. Responsibilities

We will derive our further account and problematisation of virtue and vice related responsibilities from Cassam's work as well, starting with his notions of reprehensibility and blameworthiness. Cassam claims that all vices are reprehensible because of their nature and epistemic consequence, but he maintains that not all vices are blameworthy, i.e., that not all vicious agents are to blame for their vices. It is easy to see why that is; if someone is not aware of their own viciousness – say that a person was raised dogmatically and is therefore unaware of her own dogmatism because she lacks the notion of something being dogmatic – they cannot be blamed for possessing that vice. The vice itself, however, still is reprehensible and should be done away with if at all possible. This is where responsibility first comes in: when an agent becomes aware of their viciousness, as well as aware that this viciousness is epistemically bad for her, she is obligated to stop exercising the vice in question. Cassam calls this revision responsibility – the responsibility to revise a vice once aware of its presence. This goes for all vices that are exercised involuntarily and without

agent's awareness. However, for voluntary vices that the agent is aware of, and that are blameworthy, the agent is also acquisition responsible – responsible for having them in the first place – which is the condition for their being blameworthy.

However, revision responsibility, according to Cassam, seemingly arises only when one becomes aware of her vicious conduct, and that may well never happen – if it does happen, it may come too late. When we say it may come too late, we are referring to Cassam's solutions for self-improvement, which demand problematic circumstances such as cataleptic experiences, exposure to intellectual authority, etc., and we will delve into that further in subsection 2.1. Furthermore, if one is not acquisition responsible and does not become aware of her viciousness then she may well die a vicious person, and if one is acquisition responsible, then she may never have any sort of need or want to revise her epistemic behaviour. Where then, does any kind of responsibility lie, and by what means can we stop or prevent epistemically vicious conduct?

In the rest of this section, we will overview why and how acquisition fails, and why relying on revision when that happens is overly optimistic. In section 3 we will go on to argue there is an important aspect of epistemic process that is vasty undervalued or even entirely ignored by many philosophers dealing with virtue and vice.

2.1. Issues of Acquisition and Revision

2.1.1. Acquisition-blocking Factors

Foremostly, the main reason acquisition of virtue fails and vice is acquired instead is some form of epistemic bad luck. Battaly (2017) defines bad luck as "constitutive or environmental factors that are beyond her [agent's] control." She speaks of the notion in context of testimonial injustice, drawing on Fricker (2007) to point out two cases, that of the jury in *To Kill a Mockingbird*, and that of Greenleaf in *The Talented Mr. Ripley*. In the former case, the all-white jury fails to consider evidence in favour of African-American defendant Tom Robinson, or in fact consider Mr. Robinson himself a reliable witness or, ultimately, a plausibly innocent man, because of racial prejudice. Similarly, in the latter case, Herbert Greenleaf fails to acknowledge Marge Sherwood as a potential source of knowledge because of gender prejudice. While Battaly and Fricker both argue in their respective works that Greenleaf did not have an opportunity to know better because of historical social context and the jury somehow did (possibly because the white lawyer, Atticus, did, etc.), that point of view is difficult to defend. *To*

Kill a Mockingbird is set in the early 1930s when discrimination against racial minorities was just as rampant and engrained in society as discrimination against women was in the early 1950s when *The Talented Mr. Ripley* takes place. It can be argued that both Herbert Greenleaf and the members of the jury had the possibility to look beyond socially-embedded vice and act virtuously, but failed to do so, or argued that none of them had that possibility precisely because how deeply embedded those prejudicial perceptions were in their respective historical environments; however, it is difficult to defend the position that one had the opportunity to do so but failed while the other had no such opportunity at all. The actions of both Greenleaf and the jury are still reprehensible, but none of them can be considered blameworthy or responsible for having acquired their vicious attitude, while both can be regarded as having been revision responsible.

In fact, it is possible to pinpoint numerous cases when epistemic bad luck is responsible for the formation of a vice. For example, if a child grows up in a strictly religious community, she is more than likely to develop the vicious character trait of dogmatism, provided there are no or very few strong external influences that could prevent her from conforming to epistemic conduct prevalent in her social context. Similarly, a child who is being taught school materials in an outdated manner involving memorising facts (or "facts") without being provided an explanation or evidence will later expect that she can disseminate those or other similarly obtained facts without being required to provide explanation or evidence as well, which leads to intellectual arrogance (Paul, 2000).

2.1.2. Catastrophic Moral Consequences

Even though Cassam classifies stealthy vices as a separate category, but it can be argued that all vices that are not voluntary are also stealthy as those who exercise them are usually unaware that their conduct is vicious and require some sort of a trigger that causes the epistemic agent to become aware of her viciousness. Some of these triggers are problematic in and of themselves, such as traumatic experiences, cataleptic experiences and epiphanies, while some such as testimony of others seem to present no issue. Nevertheless, even such testimony requires that a vice is made readily apparent to others who subsequently provide the testimony, or in the case of cognitive therapy and retrospective analysis, vice must be expressed in a way that allows the epistemic agent and her peers to perceive it.

Let us briefly review several of Cassam's main examples. One is the Iraq fiasco example, where Donald Rumsfeld, due to vicious traits of ar-

rogance and closed-mindedness, chose to disregard military analysts' estimates about how many troops will be required for a maximally peaceful takeover of Iraq, which resulted in nearly nine years of war, with the middle eastern country still suffering consequences of the conflict. Similar example is that of the Yom Kippur invasion when Eli Zeira ignored intelligence reports of Egyptian activity near the Israeli border, in turn failing to prepare his country for an invasion that took thousands of lives. Another example is that of the trial of Birmingham Six, when the judges failed to acknowledge evidence showing that confessions of the defendants were obtained through police violence, because of a momentary disposition that led them to fail at accepting the notion that police could be so corrupt – this resulted in the six defendants serving sentences despite being innocent.

In all of the above examples we can note that the behaviour of epistemic agents could not have been construed as vicious before the practical outcome of their decisions came about. Rumsfeld would have been considered a wise senior if the Iraq invasion had gone well and was probably considered as such by many who supported his decision to occupy Iraq with such-and-such number of troops. Likewise, had the Yom Kippur invasion of 1973 never happened, Zeira would have been considered either as having "known better" than the intelligence officers, or his decisions would not have been discussed at all. Should the Birmingham Six turn out to be indeed guilty and allegations of police violence refuted, the judges on the case would have been congratulated on not taking such wild accusations seriously.

However, this was not the case. What we are trying to point out here is that in all of these (and many more possible) examples, the attention to the responsible epistemic agents' viciousness was brought about by the consequences of their intellectual conduct, and not by that conduct itself. These consequences, in and of themselves, hold no epistemic or intellectual value, but a moral one – the outcomes of all these agents' conducts was morally disastrous, leading to lives lost and lives ruined. While the agents may have become aware of their vicious conduct after the consequences of their decisions came to light, and while they may or may not have revised such conduct in the future, mass casualties or wrongful imprisonments are hardly a trigger that could be considered morally acceptable for vice revision. In such cases, to avoid such consequences, the vice should not have been present in the first place; therefore, something should have been done to prevent its acquisition in the first place.

2.1.3. Transformative Experiences

Another way that Cassam suggests can bring about awareness of one's own viciousness and serve as a trigger for revision are several types of transformative experiences that apply mostly to epistemic postures. The first of those is traumatic experience, which is precisely what its name suggests: in order to be motivated to revise her vice, the agent must experience a traumatic consequence of her vicious agency. Cassam himself states that one such example is the experience of the Yom Kippur invasion. We have already argued why this is problematic – traumatic experience demands severe moral consequences that are not acceptable as a regularly occurring scenario.

The next one is called simply 'transformative experience' and is defined as a radical new acquisition, such as the example of Mary in the Black-and-White Room in the famous thought experiment. The issue here is not any moral toll this kind of experience would demand, but rather that such experiences seem to be phenomenally rare. In fact, it is difficult to even imagine a practical example in which one would acquire some new knowledge or belief so radical it would cause a revision of her belief-acquiring faculties themselves. Typical transformative experiences in a Joe Average's life are, for example, childbirth, religious conversion, or major medical procedures (Paul, 2014), but they seem to not be quite radical enough to cause vice revision. Say one converts to Buddhism – such conversion does cause her system of beliefs to change, but it does not necessarily change the intellectual faculties or character traits by which she arrives to those beliefs. In fact, it is more than likely that the agent converted to Buddhism precisely because her existing means of acquiring beliefs (may they be virtuous or vicious) led her to acquire beliefs that are better suited to Buddhism than whatever religion she subscribed to beforehand. She might also convert, say from the Orthodox church to Protestantism, in order to be accepted by a community or to enter wedlock with a member of Protestant church, but in that case such conversion may be purely pragmatic and not alter any belief at all. In undergoing serious medical procedures, one may certainly change some beliefs – say someone undergoes heart surgery due to cardiovascular disease caused by eating unhealthily; it seems plausible that she knew, even before the surgery, that her life-style was unhealthy and now decided to change it, but that means that neither her belief nor her manner of acquiring beliefs had changed, it was merely a shift in her everyday habits. Similarly, we may observe during the Covid-19 pandemic, that oftentimes when conspiracy theorists about the pandemic become ill with the virus, their vicious conspiratorial thinking does not change due to this often-serious illness. Instead, they

tend to shift their belief from say "the virus does not exist" to "the virus was made on purpose to hurt people like me," while retaining the same vicious conduct that led them to the former as well as the latter belief. In short, regular transformative experiences tend to cause changes in belief, but not in belief-acquiring mechanisms, while irregular transformative experiences that are radical enough to be able to cause a shift in agents' intellectual conducts are rare enough to be an exceptional circumstance indeed, and not something that could be considered a common or reliable revision trigger.

Similar can be argued about quantum change, a transformative process featuring breakthrough or epiphany. Cassam offers the example of Ebenezer Scrooge, which again brings forward the question of how likely such an event is outside of fictional scenarios. That is, while epiphany seems to be a sure way of changing one's core beliefs to the point of changing conducts that lead to those beliefs, the problem lies in whether epiphany in traditional or Joycean sense of a sudden recognition of an important truth that shakes the foundations of the experiencer's beliefs, indeed happens or is it simply a literary figure derived from a concept of mystical or religious epiphany. Another type of epiphany, however, is making a breakthrough, usually a scientific one. In this sense one may claim that Archimedes' Eureka moment might be a moment of some kind of epiphany, or that Newton's observation of a falling apple was one as well. Yet again we are faced with the same issue as above: when Archimedes stepped into his tub and realised the volume of the water displaced is equal to the volume of object immersed, he already had in place notions of volume and displacement, he already had, as a scientist, virtuous qualities such as open-mindedness, curiosity, conceivably a fair level of intellectual humility, etc. Same goes for Newton who probably possessed most of these qualities, quite certainly already knew that things fall when dropped from a height, and simply at that point realised that this must be a universal law. While these are both important scientific breakthroughs, both of them simply added a new element (equality of relevant volumes; universality of falling) to their respective agents' systems of belief but did not change in any way the qualities by which these agents acquire belief. It is difficult to explain how an epiphany would trigger the revision of those same mechanisms that led to the epiphany, unless we are introducing a new kind of epiphany with negative consequences, in which case we can further argue that we have regressed back to traumatic experience. This is also the case with cataleptic experiences, or emotional self-realisations, which Cassam offers along with other transformations that could lead to acknowledgement of own vice and revision. The problem is, again, that something emotional

that would lead us to need to revise something about ourselves cannot be positive, it has to be emotional in a negative sense, sense of regret, grief, etc. That is to say it has to be, in a sense, traumatic. I will not venture further into this particular argument on this occasion, but suffice it to say that most triggers for revision that fall into this bundle present an issue either because they a) require seriously problematic moral consequences that bring about awareness of epistemic vice, b) change beliefs but not intellectual or epistemic conducts that lead to the formation of beliefs, or c) are extremely uncommon or unlikely to be radical enough to trigger vice revision.

2.1.4. Invulnerability to Epistemic Influence

Lastly, we will glance at what is likely the most optimistic self-improvement mechanism that Cassam suggests, and that is exposure to epistemic authority. He argues that vicious character traits can be done away with by exposing the vicious agent to epistemic authority. It is a mighty strange notion, seen as the most common vicious traits seem to be close-mindedness, epistemic arrogance, epistemic hubris, and such. It is an integral quality of those intellectually or epistemically arrogant that they consistently fail to recognise others as intellectual or epistemic authority on one particular matter, or oftentimes in all matters they consider themselves qualified in.

Let us take the example of the former president of the United States, Donald Trump Jr. To illustrate our point, we will again invoke a case related to the Covid-19 pandemic – the case of how at the time president Trump handled the pandemic. There was outright denial of the existence of virus, there was blaming China of releasing the virus purposefully, and then there was denying that restrictions like masks and social distancing work. President Trump was exposed to all epistemic authorities possible, from regular doctors to emissaries of the World Health Organisation, to, perhaps most notably, NIAID director, immunologist Anthony Fauci, all of whom failed to make the slightest of progresses even with Trump's purported beliefs, much less with his vicious character traits.

This is partially because Trump is likely to have been epistemically malevolent in addition to those vices and furthered his agenda despite knowing better than what he was pushing in public, and partially because if one's vice involves not recognising epistemic authorities, then it is very unlikely that it can be revised by exposing them to epistemic authorities.

Donald Trump is not an isolated case by far, Boris Johnson's insouciant attitude towards all objective information regarding the consequences

of Brexit was also quite obviously impervious to any voice of epistemic authority. Conspiracy theories about Flat Earth and about the moon landing being fake also cannot be disproved by epistemic authorities as those holding conspiratorial thinking as a high-fidelity trait do not consider those authorities actual authorities but presuppose that those authorities are being epistemically malevolent.

Numerous vicious traits possess this common property; agents who exercise these vicious traits tend to disregard epistemic authority as authority, and therefore it is difficult to expose them to epistemic authority with any degree of success, for either they do not regard it as authority, or it has to be "their" authority, i.e., a higher-ranking person holding the same faulty beliefs and vices.

2.1.5. Revision of Virtue

Another peril we are facing when considering revision of vice is its exact opposite – revision of virtue. This is a concept that has not been discussed much, perhaps at all, in virtue and vice epistemology, but still does seem to be an everyday occurrence. If revising vice means becoming aware of one's vicious conduct and working on replacing that vice with a corresponding virtue, revising virtue is a more accidental process. In revising virtue, an agent wrongly recognises one of her virtues as a vice and revises it with what she considers virtue and is actually a vice.

This is quite commonly seen when people make a "return to nature" or start engaging in self-care. These latter concepts are often associated with spiritualism, astrology, and various other kinds of pseudo-science. Individuals can then commonly become convinced that their reasonable confidence in scientific knowledge is dogmatic and begin to revise it in favour of gullibility or naiveness. In context of Covid-19 pandemic we have seen numerous cases of individuals who were, before the Covid vaccine controversy, supporters of vaccines or at least had nothing against them, turning into anti-vaxxers. We can observe that in this process, they began to consider their trust in medical professionals, which is completely justified trust in proper epistemic authority, as naïve and replaced it with conspiratorial thinking that they misperceived as curiosity, open-mindedness, or even some form of epistemic courage.

The underlaying phenomenon is that people who revise virtue in favour of vice in this sense tend to do so because they accepted false epistemic authority as proper one, say a writer of a spirituality-based self-help book when a proper authority would have been an acknowledged psychotherapist, or an article written about vaccines by a concerned stay-at-

home mother when they should have trusted an immunologist. Another common occurrence of this is when people who agree with certain subset of usually moderate principles of a politician start revising their beliefs and intellectual conduct to accommodate the whole set of even the most radical principle of said politician, if they begin considering them an epistemic authority. Once such a revision is made, revising the newly acquired vice again in favour of virtue is at least as difficult as revising any vice in the first place.

3. Knowing as a Social Process

We can see above that in acquisition, as well as in most proposed processes that lead to revision, the agent doing the acquiring or revising has to rely on others in some shape or form; when acquiring, it is important what kind of virtue or vice her inner social circles tend to exercise, how she receives beliefs and knowledge (in a virtuous or vicious manner) – when triggering a revision, testimonies of others play a major role, the way others are perceived (as an authority or not) is important as well, and lastly, it is also relevant how an agent's decisions affect others around her, as that is what ultimately triggers traumatic and other transformative experiences. The latter seems a bit far-fetched, but even Ebenezer Scrooge, even when utterly selfish, is partially influenced by consequences of his actions for others, like the potential death of Tiny Tim and the grief of his family.

This leads us to a not exactly new but often disregarded idea that knowing is a social process. Grasswick (2019) goes so far as to argue that it is in fact *deeply* social, as we are, as social agents, constantly relying on testimonies of others, may they be epistemic authorities or merely peers, possess different background assumptions depending on our social circles, and abide by different standards of evidence depending on those background assumptions. As mentioned earlier, if an epistemic agent is raised in a traditional, religiously dogmatic household, her standards of evidence and background assumptions are going to be vastly different that from someone who grew up with, say, atheist scientists.

Indeed, some go so far as to say that certain groups of people can be considered epistemic agents themselves. For example, Lahroodi (2007) proposes a view of this called summativism, that states that a group A can be said to have property P, if most of its members have property P. Thus, for a family that supports and teaches open-mindedness and where most if not all members are open-minded, we can say that the family is open-

minded. In such groups, the property P gets transferred between members of the group, observably mainly from senior to junior members. This may end in the majority of the members of the group internalising the property as a sub-personal value, in our context virtue or vice, or it may lead to some interesting fringe cases akin to the Abilene Paradox.

In the Abilene Paradox a family decides to visit the town of Abilene for lunch, undertaking a 50-mile one way drive in scorching heat of summer to do so; the father who proposes the idea would rather not go, but proposes it because he believe the other family members are bored, while the others accept the suggestion because they believe that the father (and the other members of the family) wishes to go to Abilene, while all of them do not wish to go – they choose to go because they believe it is what the other members of their family want (Harvey, 1974). Similarly, in a group that is an epistemic agent with the property P, members might act accordingly with the property P because it is supposed to be their group's property, while less than a majority or even only very few members of the group actually hold P.

Such fringe cases are possible but are generally avoided by exercising virtues such as honesty and truthfulness, which Kawall (2002) dubs "other-regarding" virtues, i.e., virtues that help others obtain knowledge through one's truthful and honest testimony. It is, of course, possible that an agent acting virtuously in this "other-regarding" aspect is vicious in ways she obtained knowledge in the first place (but oblivious to that viciousness), and unintentionally spreading untrue or unjustified beliefs unto the peers in her group. However, if she is acting virtuously in acquisition of knowledge, as well as in dissemination of said knowledge, she assures that the knowledge the others then possess is justified via her own virtuousness.

It seems possible that this notion can be extended to virtues themselves, and that one can, in acting virtuously towards others, instil into those others similar virtuous conduct. If we consider Cassam's view that virtues and vices are sub-personal, this seems to be somewhat implausible, but we must note here that we are not discussing the nature of virtue and vice as ways of arriving at or obstructing access to knowledge respectively, we are discussing the mechanisms of acquisition of vice in social context. While we may agree with Cassam's claim that virtues and vices are sub-personal and largely socially and structurally independent once acquired, the processes of acquiring and revising them are not, and are, given what we summarised above about knowing as a social process and transfer of epistemic properties within groups as epistemic agents, largely socially conditioned.

4. Provision Responsibility

We have discussed, in section 2, why acquisition of virtue is often obstructed due to social and environmental factors, why revision often fails, and how triggering the need for revision tends to involve morally unfavourable outcomes of epistemically vicious conduct that should be avoided. In section 3 we have hinted that triggers of the need for revision also have a notable social component and that indeed the process of obtaining knowledge, and thus of acquiring virtue and vice, is deeply social and that methods of obtaining knowledge largely depend on our social environment, in other words, on the epistemic group we of which we are a part and from whose other members we receive example and testimonies.

If we are attempting to avoid, as we argued we should, the need for revision as well as to prevent epistemic bad luck in acquisition, we may rely on those social mechanisms to do so. Since we cannot blame the epistemic agent herself for acquiring a vice because that vice is commonplace in other agents in her environment of origin, or for acquiring a vice directly from an epistemic authority (subjective one from her point of view, such as parent, teacher, etc.) who exercised that vice, where do we assign blame and responsibility? The answer is to those other persons in her environment. A third party can be blamed for an epistemic agent's involuntary acquisition of vice, and in turn has a responsibility towards providing virtue to the agent.

We call that third type of responsibility *provision responsibility* and will now dedicate to it the core of this paper. First, it is most obvious that parents present an epistemic authority to their children, especially when the latter are still developing, as well as do teachers and other adults present in their upbringing, as people develop in relation to others (Grasswick, 2019; Baier, 1985). These we will call the primary (parents) and secondary (educational and other authorities) level of acquisition, or both together the developmental level of acquisition. There is also a tertiary level of acquisition, which we can also call constitutive level (Grasswick, 2019; Baier, 1985) of acquisition, where provision responsibility falls on other figures of epistemic authority such as scientists, politicians, leaders, bosses, etc., who may influence adult (developed) epistemic agents. This last level's function is mostly preventing revision of virtue by providing virtue and avoiding displaying vice that could replace another's virtue. We have outlined three types of provision that falls under provision responsibility, which we will explain in some detail in following subsections.

4.1. Active (Explanatory) Knowledge Provision

The first manner of provision is the Active (Explanatory) Knowledge Provision (A(E)KP), which demands that when in position of epistemic authority of some kind – that can mean a parent, a teacher, a public-lecturing scientist, or even a peer with a higher level of knowledge on a particular topic – an epistemic agent should always provide only knowledge, in the sense where knowledge is justified true belief that fulfils Cassam's condition that one should have reasonable confidence as to whether their knowledge is indeed knowledge. Additionally, the A(E)KP demands there to be an explanatory component, in line with the notion of Paul (2000) we have mentioned earlier that superficial learning of bare "facts" without being provided an explanation as to why they are true and how they are justified leads to intellectual arrogance in a variety of ways.

This requirement prohibits various types of "because we said so" scenarios. We maintain that there is nothing wrong with justifying a rule or a sanction in a "because we said so" manner to a child who is, by a reasonable estimate of the parent or teacher, too young to understand the justifications if provided. However, we argue that failing to properly justify decisions to an older child or a young teenager, for example, can be epistemically harmful. Say, should a parent prohibit their child from attending a peer activity or an event, they have to provide sufficient reason why that is so (state any presentable dangers, risks, etc.) in a non-dogmatic fashion (should not, for example, prohibit a teenage child from going to a concert due to risk of drug consumption when there is no concrete justification of such risk). This is relevant for the development of justified decision-making about self and others by the child in question.

More concretely pertaining to knowledge; any and all knowledge given by an epistemic authority should be given with proper explanation of what justifies it as knowledge, and not simply from a position of authority presenting as unquestionable. If a physics teacher is attempting to teach her students about the function of gravity, she should be expected to explain the observations that led to the first formulation of the concept by Newton, how we can know the principle to be universal, and so forth, instead of, as it is unfortunately often done, merely describing the concept briefly and demanding the students to memorise its constants and relevant formulae. Prior to such, if a parent wishes a child to be respectful and tolerant to those different from her, it does not suffice, however morally good the goal, to tautologically say "because it is the right thing to do", or something along those lines. Rather, it is expected that the parent will explain why there is nothing about those people that would make them inferior, the challenges they may be facing, and how her own child can relate to them even though they may be different.

Extending that to what we call the tertiary or constitutive level is somewhat more difficult, as it is difficult to justify why someone with a doctoral degree in immunology should have to explain details of a vaccine to a firm conspiracy theorist. The answer is that they do not have to do so, individually, but such details should be made publicly available without prejudice. The real provision responsibility lies with those who contradict such scientific materials, and who should be able to explain in careful detail with proper justification why they are contradicting established fact. Here, we must note that such provision of explanation is unlikely to occur as those that practice the vice of, say, conspiratorial thinking, have acquired it at an earlier stage, or acquired it from someone who had acquired it at an earlier stage. On the tertiary level, the A(E)KP is mostly relevant for politicians and other figures with broad public influence, who should refrain from sensationalistic statements on affairs not immediately political (science, ethics, etc.), but instead, should they make statements on these affairs at all, support them with explanations of how they are justified as true, or at the very least support them by reference to an appropriate authority (scientists and scholars) who can provide such explanations if necessary.

For the most part, all of the above applies to a junior agent's acquisition of knowledge, and not directly to acquisition of virtue, but it does also offer a mechanism by which the junior agent can establish how beliefs she wishes to convey should be justified and that she should refrain from conveying them if she is unable to justify them in such way.

4.2. Passive Virtue Provision (PVP)

Another way for the Passive Virtue Provision may be provision of virtue by example. As we noted before (by Grasswick, 2019; Baier, 1985) people develop in relation to others and never independently. This means that being actively provided knowledge does not adequately help one develop her faculties, similarly as Mary in the black-and-white room has something new to learn upon seeing a red object despite knowing everything about redness in theory. By that we mean, if Mary's friend Jane was in another room, one where things are in colour, but where she has no contact with other people and no video or audio materials portraying interpersonal relations, yet she is provided all sorts of encyclopaedias on a variety of topics, including sociology and psychology, she would still learn plenty when eventually released into the world where she would observe people actually interacting and where she would be expected to interact with them as well.

Indeed, people seem to learn more from observing senior agents' actions than they do from being told how to act, and that applies also to moral as well as epistemic agency. For example, it is known that experiencing abuse as a child will likely lead to the adoption of similar abusive behaviour. Likewise, if a parent is acting openly intolerant to minorities, the child will likely adopt the same intolerance, unless she observes tolerant behaviour by other senior agents sufficiently often and sufficiently early in her development. There is no reason to claim that epistemic conduct would be learned any differently, thus we might surmise that observing virtuous behaviour in senior agents leads to adoption of virtue in junior agent, as does observing vice lead to adoption of vice.

This leads to a responsibility to act virtuously and refrain from acting viciously when observed by those who regard one as an epistemic authority. Mainly this applies to primary and secondary levels of acquisition, wherein parents and teachers should be expected to 'practice what they preach', employing open-mindedness, intellectual humility, critical thinking, and other virtuous faculties in the presence of children.

On the tertiary level of acquisition, the PVP takes on a somewhat different form as well. Again, epistemic authorities such as politicians, employers, scientists, and a number of others are to be expected to act virtuously, but not because it may cause such passive adoption of vice. Rather it is because certain agents may value epistemic authorities' conduct over their own, thus adopting it purposely or semi-purposely, leading to virtue revision. Another possibility is that someone in a position of epistemic authority may preach virtue, but practice vice, in which case she may be regarded as hypocritical and have her professed virtue ignored due to vicious conduct; or some may misconstrue her virtuous words as vicious due to the vicious conduct and elect to act in contradiction with them, resulting in viciousness. The latter case may be seen in some politicians during the Covid-19 pandemic; certain politicians who have been, in general, acting viciously in the sense of practicing corruption, violations of rights, etc., have chosen to take the virtuous stance of being pro-vaccination. However, because of their general viciousness, many have assumed their latest stance must be vicious as well, choosing consequently to take the opposite, indeed vicious stance, while the former would have been virtuous, and while it is likely they would have taken the former in other circumstances.

4.3. Active (Theoretical) Virtue Provision (A(T)VP)

The last variety of provision pertains to direct teaching of what epistemic virtues and vices are, how they are acquired, what characterises them, how to recognise them, revise them, and so forth.

The starting point of this is the necessity expressed by some to teach critical thinking as a tool for building virtue. Kilby (2004) for example considers teaching of critical thinking as a gateway to open-mindedness, claiming that the ability to think critically is fundamental to overcoming social obstacles; someone who is adroit in critical thinking will, when faced a situation in which she is expected or even preconditioned to react in a certain way, recognise faults in such reaction and consider whether such reaction is appropriate prior to reacting. This is applicable to situations where epistemic engagement is required; somebody versed in critical thinking will be quicker to question information given to her by others, asking herself whether it is indeed knowledge and how it is justified. Furthermore, she is more likely to question intellectual behaviour of herself and others in a similar way, being able to recognise, if not in so many words, virtuous and vicious conduct.

However, this does continue towards teaching the theory of epistemic virtue and vice itself, and there is a direct requirement for teaching such theory to the general population (Battaly, 2016; Curren, 2019). Awareness of the existence of the concepts of epistemic vice, epistemic virtue, and how they function, raises awareness of an agent's own virtuous and vicious conduct, as well as of such conducts present in others. Moreover it provides *phronesis*, the practical wisdom responsible for emerging virtuous conducts without external influence, which Zagzebski at some point describes as "the excellence in deciding what to do" (Zagzebski, 2019; Curren, 2019; Wright, 2019).

For example, if an epistemic agent is aware that, as per Cassam (2019), conspiratorial thinking can be both vicious and virtuous – virtuous in conspiracy-poor and vicious in conspiracy-rich environments – she is well predisposed to critically reflect on her own thinking in light of her current environment and can 'catch herself' thinking in line with a common conspiracy, which she can then revise, or can on the other hand observe her thinking, albeit conspiratorial in nature, to be unique to herself and pursue it further with due caution. In addition to self-reflection, *phronesis* developed through theoretical knowledge of virtue and vice significantly improves one's ability to determine which epistemic authority is proper and which is not, as well as the ability to distinguish valid evidence and justifications from invalid ones.

Needless to say, this A(T)VP is entirely inapplicable to the primary level of acquisition, and most prominently applicable in the second or educational level, mainly on higher levels of education when the students have already developed suitable formal thinking. However, it can also be argued that A(T)VP can be conducted on the tertiary level through adult classes, workshops, and other means of raising awareness.

Conclusions

There are two major issues that remain unaddressed and unanswered by the notion of provision responsibility as outlined in this paper. The first of these is the problem of voluntary vices, such as epistemic corruption; so far it seems that no amount of education, good examples and theoretical knowledge can prevent an agent from voluntarily exercising an epistemic vice, knowing full well her conduct is vicious, for her personal gain or due to some other motivation. The second one is that there is no distinct starting point to provision responsibility. There do not seem to be any epistemically perfect agents, although each individual person is differently distant to the concept. A parent or a teacher unaware of her own virtuousness cannot be held acquisition or revision responsible, thus cannot be expected to be provision responsible, and so goes the cycle. One possibility is to suggest starting at A(T)VP at tertiary level, targeting primarily educators, then expanding that to parents, emphasising, through A(T)VP the importance of PVP and A(E)KP, and so forth. It is also possible that this issue is where a philosopher's job stops and a pedagogue's begins, but I shall not attempt to clear this up on the spot, as it is too demanding a task.

Outside of outstanding issues, we conclude here that accepting the notion of provision responsibility provides a reasonable answer to the question of where the responsibility for reprehensible vices lies when the agent cannot be acquisition responsible because of involuntariness of the acquisition and at the same time cannot be revision responsible because of continued obliviousness of her own viciousness.

Further, successfully implementing knowledge and virtue provision at primary and secondary levels of acquisition can guarantee proper acquisition of epistemically and intellectually virtuous conduct, which prevents the need for vice revision later on, while implementing A(E)KP and PVP on the tertiary level may prevent virtue revision, and implementing A(T)VP on all levels provides one with uninhibited reflective thinking on knowledge and her own virtuousness/viciousness, which in turn enables self-reflection to the point when the need for revision can be identified and revision can be undertaken without any morally catastrophic, personally traumatic, or other adversely-natured triggers.

References:

Baier, A. (1985). *Postures of the Mind: Essays on Mind and Morals,* Minneapolis, MN: University of Minnesota Press.

Battaly, H. (2008). »Virtue Epistemology,« *Philosophy Compass,* vol. 3, No. 4, pp. 639–663.

Battaly, H. (2016). "Responsibilist Virtues in Reliabilist Classrooms," in J. Baehr (ed.): *Intellectual Virtues and Education*, New York: Routledge, pp. 163–183.

Battaly, H. (2017). "Testimonial Injustice, Epistemic Vice, and Vice Epistemology," in Kidd, I. J., Medina, J. and Pohlhaus, G. (eds.): *The Routledge Handbook of Epistemic Injustice*, New York: Routledge, pp. 223–232.

Corlett, J. A. (2008). "Epistemic Responsibility," in *International Journal of Philosophical Studies,* 16(2), pp. 179–200.

Curren, R. (2019). "Virtue Epistemology and Education," in Battaly, H. (ed.): *Routledge Handbook of Virtue Epistemology,* New York: Routledge, pp. 470–482.

Elgin, C. Z. (2013). "Epistemic Agency," in *Theory and Research in Education,* 11(2), pp. 135–152.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing.* Oxford: Oxford University Press.

Grasswick, H. (2019). »Epistemic Autonomy in a Social World of Knowing,« in Battaly, H. (ed.): *Routledge Handbook of Virtue Epistemology,* New York: Routledge, pp. 196–208.

Greco, J. (1999). "Agent Reliabilism" in Tomberlin, J. (ed.): *Philosophical Perspectives 13: Epistemology.* Atascadero, CA: Ridgeview.

Harvey, J. B. (1974). "The Abilene Paradox: The management of agreement," in *Organizational Dynamics,* 3(1), pp. 63–80.

Kawall, J. (2002). "Other-Regarding Epistemic Virtues," in *Ratio*, 15(3), pp. 257–275.

Kilby, R. J. (2004). „Critical Thinking, Epistemic Virtue, and the Significance of Inclusion: Reflections on Harvey Siegel's Theory of Rationality," in *Educational Theory,* 54(3), pp. 299–313.

Lahroodi, R. (2007). "Collective Epistemic Virtues," in *Social Epistemology,* 21(3), pp. 281–297.

Montmarquet, J. (1992). "Epistemic Virtue and Doxastic Responsibility," in *American Philosophical Quarterly,* 29(4), pp. 331–341.

Paul, L. A. (2014). *Transformative Experience.* Oxford: Oxford University Press.

Paul, R. (2000). "Critical Thinking, Moral Integrity, and Citizenship: Teaching for the Ethical Virtues," in Axtell, G. (ed.): *Knowledge, Belief, and Character: Readings in virtue epistemology,* Lanham: Rowman and Littlefield Publishers, pp. 163–176.

Quassam, C. (2019). *Vices of the Mind: From the Intellectual to the Political.* Oxford: Oxford University Press.

Sosa, E. (1991). *Knowledge in Perspective.* Cambridge: Cambridge University Press.

Sosa, E. (2007). *Apt Belief and Reflective Knowledge, Volume 1: A Virtue Epistemology.* Oxford: Oxford University Press.

Wright, S. 2019. "Are Epistemic Virtues a Kind of Skill?" in Battaly, H. (ur.): *The Routledge Handbook of Virtue Epistemology.* New York: Routledge, pp. 58–68.

Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge: Cambridge University Press.

Zagzebski, L. (2019). "Intellectual Virtues: Admirable Traits of Character" in Battaly, H. (ed.): *The Routledge Handbook of Virtue Epistemology,* New York: Routledge, pp. 26–35.

Mašan Bogdanovski*

ULOGA MISAONIH EKSPERIMENATA U REŠAVANJU KRIZA

Sažetak: U prvom delu članka analizirane su neke od filozofskih osnova programa stručnog usavršavanja nastavnika „Primena misaonih eksperimenata u nastavi", koji je autor držao u školama. Izložena je skica prirode i motivacije misaonih eksperimenata, sa naglaskom na ulozi imaginacije. Imaginacija u misaonim eksperimentima izuzetno je važna komponenta razrešavanja ekonomskih i političkih kriza. Imaginacija ima ključnu ulogu u fenomenu koji sam nazvao uviđanjem. Uviđanje je heterogen fenomen i njegovi različiti vidovi su raščlanjeni u tekstu. Tema centralnog dela teksta su načini na koje uviđanje u raznim oblicima dovodi do novog znanja i, što je još značajnije, do razumevanja prirodnih i društvenih pojava koje omogućava rešenje društvenih kriza. Misaoni eksperimenti dovode do uviđanja i novog znanja bez novog iskustva. Uviđanje je preovlađujući epistemički cilj misaonih eksperimenata u procesu rešavanja kriza.

Ključne reči: misaoni eksperimenti, rešavanje kriza, imaginacija, uviđanje, znanje

Pandemija virusa korona prekinula je program stručnog usavršavanja nastavnika „Primena misaonih eksperimenata u nastavi", koji je akreditovao Zavod za unapređenje obrazovanja i vaspitanja. Kolega Ivan Umeljić iz Centra za promociju nauke i ja održali smo u Beogradu i još desetak gradova u Srbiji, u osnovnim i srednjim školama i u lokalnim centrima za stručno usavršavanje obuku za najširi profil nastavnika. Njen cilj je bilo osposobljavanje za korišćenje metoda misaone eksperimentacije u izlaganju i savladanju gradiva iz različitih predmeta i nastavnih jedinica.

Za realizaciju tog programa stručnog usavršavanja nastavnika od velike koristi bila je okolnost da je Centar za promociju nauke izdao prevod knjige *A šta ako...* Peg Titl (Tittle) (Titl 2018), tako da su polaznici na

* Odeljenje za filozofiju, Filozofski fakultet Univerziteta u Beogradu, *mbogdan1@f.bg.ac.rs*.

raspolaganju imali literaturu iz koje su mogli da se upoznaju sa „sabranim misaonim eksperimentima u filozofiji", kako glasi podnaslov te knjige koju je preveo autor ovog teksta. Filozofija je „rodno mesto" misaonih eksperimenata, a misaona eksperimentacija i proučavanje prirode misaonih eksperimenata spadaju u domen filozofije, iako je bez misaonih eksperimenata očigledno nemoguće razumeti, na primer, gradivo čak i elementarne, a kamoli napredne fizike. Tokom „obučavanja" nastavnika trudili smo se da pokažemo koliko misaoni eksperimenti mogu da budu korisni u savladanju nastavnih jedinica iz biologije, istorije, stranog jezika, sociologije ili građanskog vaspitanja, dakle i u prirodnim i u društvenim naukama. Filozofija i književnost se, svaka na svoj način, sastoje od misaonih eksperimenata.

Umesto da zalazim u posebna pitanja pojedinih nastavnih jedinica iz školskih predmeta, čemu ovde nije mesto, svoje izlaganje ću početi objašnjenjem nekih filozofskih pretpostavki na kojima je naš program zasnovan. To su klasična filozofska stanovišta, oblikovana uglavnom ranom modernom filozofijom jer nije bilo potrebe da ulazimo u problematizaciju ili savremenu filozofsku kritiku tih stanovišta. Ona su sasvim dovoljna da misaoni eksperimenti obave svoju nastavnu funkciju. Takav pristup podrazumeva izvesna pojednostavljenja filozofskih eksplikacija misaonih eksperimenata. Naime, iz te perspektive, misaoni eksperimenti zadržavaju sve karakteristike koje imaju „pravi" eksperimenti, koji se izvode u „stvarnosti", samo što su ovi smešteni u laboratoriju ljudskog uma, odvijaju se u našoj glavi. U skladu s tim, najčešća motivacija za izvođenje misaonog eksperimenta jeste nemogućnost da bude izveden u stvarnosti, bilo u doslovnom smislu fizičke nemogućnosti, bilo zato što bi neko morao da bude povređen ili da mu se nanese neprihvatljiva šteta. Misaoni eksperimenti mogu da opovrgavaju određene hipoteze i stvaraju potrebu za formulisanjem novih teorija, čime doprinose rastu ljudskog znanja, ako znanje shvatimo kumulativistički. Ishod neke zamišljene situacije ili događaja može ozbiljno da uzdrma neka uverenja kojih se čvrsto držimo. Slično tome, misaoni eksperiment preispituje i sadržaje naših pojmova. Zamišljeni objekat, stanje stvari ili proces govori nam šta ulazi, a šta ne ulazi u sadržaj pojma koji pokušavamo da odredimo, tako što nam govori koja obeležja ulaze, a koja ne ulaze u određeni pojam.

Imaginacija je mentalna sposobnost koja nam omogućava da misaonim eksperimentima ostvarimo takve rezultate. Pokazalo se da polaznici kurseva najlakše razumeju u kom smislu ovde govorimo o imaginaciji kada razmotrimo pitanje šta su granice imaginacije. Na njih ćemo naići u onome što filozofi nazivaju analitičkim sudovima. U savremenijoj verziji kantovske distinkcije između analitičkog i sintetičkog, analitički sudovi su oni koji su istiniti samo na osnovu svog značenja. Epistemološki po-

smatrano, možemo da znamo da su ti sudovi istiniti isključivo na osnovu znanja njihovog značenja. Znam da je iskaz „Svi bećari su neoženjeni ljudi" istinit jer znam šta termin bećar znači, a kada to ne bih znao, ne bih znao ni značenje te rečenice, pa u skladu s tim ne bih mogao da odredim ni istinosnu vrednost odgovarajućeg iskaza. Ako znam šta znači reč „bećar", onda ne mogu da zamislim da taj iskaz nije istinit. To je granica moje imaginacije. Ne mogu da zamislim da su bećari oženjeni, ne mogu da zamislim mogući svet ili protivčinjeničku situaciju u kojoj su bećari oženjeni ljudi. Štaviše, kada bih pretpostavio da taj iskaz nije istinit, pretpostavio bih nešto samoprotivrečno. Rečenica „Svi bećari su neoženjeni ljudi" istinita je u svim mogućim svetovima jer njena istinitost ne zavisi od činjenica. Kada govorimo o činjenicama u hjumovskom smislu, o onome što uvek možemo da zamislimo da je drugačije, uvek možemo da zamislimo negaciju odgovarajućeg iskaza. Istinitost iskaza „Bećari vode neuredan život" ne možemo da znamo isključivo na osnovu značenja tog iskaza, ma koliko nam se on činio prihvatljiv i ubedljiv. Treba da znamo neke činjenice o bećarima. Uvek možemo da zamislimo protivčinjeničku situaciju ili mogući svet u kojem bećari vode uredne živote. Nema ničeg samoprotivrečnog u tome da bećari ne vode neuredan život, da imaju sasvim uredne živote.

Zamislivost o kojoj ovde govorim može se pronaći u klasičnim Kripkeovim radovima o semantici mogućih svetova (Kripke 1980). O mogućim svetovima ne govorimo kao o udaljenim planetama već zamišljamo protivčinjeničke situacije, svetove koji su veoma slični našem, ali se razlikuju u nekom specifičnom relevantnom pogledu. Relevantnost te posebne razlike tiče se načina na koji takav mogući svet, u okviru misaonog eksperimenta koji se sastoji od uvođenja takve razlike, može da podrži ili opovrgne neko stanovište ili verovanje. Misaoni eksperiment, na taj način, predstavlja zamišljanje mogućeg sveta koji se od našeg, aktuelnog sveta, razlikuje samo u relevantnim aspektima. Izbegavanje da se upustim u suptilnosti Kripkeove analize pojmovnih parova *a priori* – nužno i *a posteriori* – kontingentno sasvim je namerno. Ono bi nepotrebno iskomplikovalo ovu pojednostavljenu sliku koju smo imali u vidu kada smo pristupali misaonim eksperimentima u kontekstu nastave. Međutim, morali smo da upozorimo polaznike da u nastavi filozofije postoji misaoni eksperiment koji krši to pravilo relevantnosti. Naime, Dekartov misaoni eksperiment sa Zlim demonom podrazumeva upotrebu imaginacije u svrhu konstruisanja najradikalnijeg misaonog eksperimenta, zamišljanja mogućeg sveta koji je u svim aspektima različit od našeg. Zamišljamo mogućnost da se varamo u svim verovanjima koja imamo o aktuelnom svetu usled delovanja najmoćnijeg mogućeg obmanjivača, Zlog demona, koji je sposoban da nas vara u bilo kojem našem verovanju, ali i svim uzetim zajedno. Ako je

tako nešto moguće, da li onda možemo da zaključimo da ni o čemu nikada ništa ne možemo da znamo? To je jedan od fundamentalnih misaonih eksperimenata u filozofiji i predstavlja izuzetak i odstupanje od modela relevantnih razlika.

Istorijskofilozofski posmatrano, značajna posledica Kripkeovih istraživanja bila je da misaoni eksperimenti mogu da podrže i esencijalističke intuicije o tome šta sačinjava suštinska svojstva neke stvari. Imam jaku intuiciju da radijator, koji koristim za dogrevanje u doba energetske krize i loše snabdevenosti toplana gorivom, nije mogao da bude napravljen od leda. I to ne zato što bi se topio ili prosto zato što takav uređaj ne bismo mogli da napravimo od leda. Možemo da zamislimo da je takve poteškoće moguće tehnološki prevazići. Štaviše, možda ću jednog dana moći da kupim i na isto mesto stavim nešto tako bizarno kao što je ledeni radijator. Stvar je zapravo u tome da mi moje intuicije govore da, kada bi bio napravljen od leda, to ne bi bio ovaj ovde radijator. Ne postoji mogući svet u kojem je ovaj ovde radijator napravljen od leda. U tom mogućem svetu bio bi to neki drugi radijator. Mogao bi da se zapali i eksplodira i tako potpuno promeni svoj izgled, ali bi u tim mogućim svetovima i dalje bio ovaj radijator. Međutim, u trenutku kada je napravljen, ovaj ovde radijator nije napravljen od leda.

Veća poteškoća sa takvim misaonim eksperimentima je to što je onda matematika potpuno izolovana od potencijalnih područja na kojima se sprovode misaoni eksperimenti. Ukoliko su oni ograničeni na područje činjeničkog, aposteriornog i kontingentnog, na područje onoga što uvek može da bude drugačije i što je istinito na osnovu činjenica, a ne na osnovu značenja, onda matematika ostaje s druge strane granice eksperimentatorske imaginacije. Matematika je (sem za Kanta i kantovce) analitička, apriorna i nužna, bar u ovoj simplifikovanoj filozofskoj geografiji pojmova. Ipak, odmah nam se nameću neki najopštiji i najtemeljniji misaoni eksperimenti, kao što bi bilo zamišljanje susreta sa inteligentnim bićima koja imaju matematiku radikalno različitu od naše. Svakako ne bi trebalo da se slepo držimo nekog modela i tamo gde postoje očigledni protivprimeri, a matematika, kao apriorna intelektualna disciplina, ionako je izuzetak u školskom programu.

Na kraju krajeva ni sam ne verujem da je fundamentalna distinkcija u ovoj slici filozofskih osnova misaonih eksperimenata održiva, a to je distinkcija između analitičkih i sintetičkih sudova. Na tragu Kvajna, smatram da ne postoje iskazi koji su imuni na reviziju, da su svi iskazi manje ili više sintetički i da ne postoje oni čija je empirijska komponenta ravna nuli. Međutim, upuštanje u odbranu takvog stanovišta je irelevantno za ovu raspravu. Tako pojednostavljeno shvatanje filozofske pozadine misaonih

eksperimenata ne menja ništa u njihovom pedagoškom efektu. U samoj upotrebi misaonih eksperimenata pokazuje se još jednom da su filozofske suptilnosti irelevantne za određene praktične svrhe.

O ulozi misaonih eksperimenata u obrazovanju govorim zbog toga što je uloga misaonih eksperimenata u rešavanju kriza tesno povezana sa načinom na koji oni dovode do novog znanja u obrazovnom procesu. Kada posmatramo epistemološku literaturu o misaonim eksperimentima, vidimo da je ona usredsređena na pitanje kako misaoni eksperimenti proizvode novo empirijsko znanje bez novog iskustva. Sprovodeći misaoni eksperiment, ne raspolažemo novim iskustvom već rekombinujemo staro, ali on svejedno proizvodi novo empirijsko znanje. Reprezentativan primer takvog pristupa je knjiga Tamar Gendler (Gendler 2000), koja je inspirisala mnoštvo radova na ovu temu u 21. veku. Povrh toga, baveći se pomenutim pitanjem o tome kako misaoni eksperimenti proizvode novo znanje bez novog iskustva, nekoliko poznatih epistemologa, kao što su Ketrin (Catherine) Elgin (Elgin 2006), Džonatan (Jonathan) Kvanvig (Kvanvig 2009) i Dankan (Dunkan) Pričard (Pritchard 2010) s punim pravom su istakli da je neophodno fokusirati se na fenomen koji ću nazvati uviđanjem i za koji smatram da igra ključnu ulogu u procesu rešavanja kriznih situacija. U stvari, uviđanje je preovlađujući epistemički cilj misaonih eksperimenata u procesu rešavanja kriza.

Iz njihovog doprinosa ovoj problematici izdvojiću nekoliko ključnih aspekata koji će mi omogućiti da istražim funkciju uviđanja u rešavanju kriza. Naime, najupadljivija odlika takvih eksperimenata je da ilustruju neko rešenje ili neku značajnu tezu iz tog rešenja. Mnoge intuicije koje imaju tvorci rešenja kriznih situacija na taj način postaju pristupačnije za širu javnost. Štaviše, misaoni eksperimenti pružaju hipotetička objašnjenja koja nam omogućavaju da uvidimo na koji način rivalska rešenja dovode do različitih ishoda i pomažu nam da se opredelimo između rivalskih rešenja. U slučaju krize izazvane pandemijom virusa korona, kao građani koji manje ili više utiču na političke odluke, mogli smo da razmatramo dva hipotetička scenarija. U jednom bi se rešenje sastojalo u strogom zatvaranju, pa čak i uvođenju policijskog časa. Njime bi se, u takvom mogućem svetu koji zamišljamo, radikalno smanjili kontakti između ljudi i tako presekli putevi kojima se virus širi. Na samom početku pandemije nismo raspolagali empirijskom evidencijom i preciznijim podacima o tome u kojoj meri se takvim zatvaranjem zaista smanjuje broj zaraženih virusom korona. Pretpostavka o smanjenju tog broja zasnivala se na zamišljanju situacije u kojoj prekid kontakata dovodi do prekida u lancu zaražavanja.

S druge strane, na raspolaganju nam je bila raznolikost zamišljenih scenarija od nešto popustljivijih oblika zatvaranja do potpune otvoreno-

sti i bez promena u svakodnevnom životu, uključujući možda i izostanak standardnih epidemioloških higijenskih zahteva kao što su nošenje maske i održavanje distance. Misaono eksperimentisanje sa raznim stepenima otvorenosti bilo je privlačno, između ostalog, i na osnovu našeg pozadinskog znanja o tome da prokuženost stanovništva može da proizvede kolektivni imunitet u okolnostima u kojima je vakcinacija nedostupna. Razradu tog scenarija, međutim, pratila je zabrinutost da će bolnički kapaciteti biti nedovoljni za ogroman broj zaraženih koji bi se u takvim okolnostima pojavio. Da se zadržim samo na najpoznatijim primerima, ni prvi scenario nije bio lišen problema. Zatvaranje u sopstvene domove i rad kod kuće može da izazove mnoge druge zdravstvene probleme, uključujući i psihološke, da ne pominjem ekonomske poteškoće koje takav zamišljeni model rešenja proizvodi. Grubo pominjanje poteškoća koje sagledavamo na osnovu zamišljanja rešenja otkriva nam način na koji možemo da govorimo o nečemu što bi predstavljalo kvalitet našeg saznajnog odnosa sa svetom, a misaoni eksperimenti bi bili nešto što poboljšava kvalitet tog odnosa sa svetom, i to bez povećavanja broja iskaza koje znamo.

Kada govorim o pukom ilustrovanju rešenja krize o kojoj je ovde reč, moram da naglasim da ilustrovanje nije isto što i opravdanje tog rešenja i ni na koji način ne bi trebalo da nas navodi na zaključak da je to stvarno efikasno rešenje. Ona je samo ilustracija, a može i samo da skreće pažnju na nešto. Iako ne poričem da iz misaonih eksperimenata često proističe novo znanje, ovde mi je važno da ukažem na epistemički potencijal misaonih eksperimenata koji je lako prevideti i na to da u njima postoji nešto što prethodi proizvođenju znanja i što može da bude nezavisno od njega, a što nazivam uviđanjem.

Podela uviđanja bi ovde mogla da bude od velike koristi. Kao prvo, možemo da govorimo o eksplanatornom uviđanju (Pritchard 2010: 74), koje vodi objašnjenju zbog čega je nešto slučaj. Izraz „zašto je nešto slučaj" može da navede na pogrešan trag da se eksplanatorno uviđanje odnosi samo na iskaze, da je ono isključivo propozicionalno. Međutim, kako upozorava Grim (Grimm 2006: 531), u eksplanatornom uviđanju možemo da razlikujemo podvrstu ostenzivnog uviđanja, kao kada bi mi, na primer, neko pokazivao pretrpane bolnice da bi mi objasnio zašto otvaranje u primeru epidemije virusa korona ne funkcioniše. Drugo, nasuprot eksplanatornom, mogli bismo da razlikujemo objektno uviđanje, kao razumevanje suštinskih karakteristika neke stvari, načina na koji je povezan skup nekih stvari ili uviđanje u čemu se sastoji predmet nekog istraživanja. Objektno uviđanje je sagledavanje relacija koje neki složeni skup informacija čine koherentnim i, u skladu s tim,sklonosti ka prihvatanju određenog rešenja krizne situacije. Da bismo shvatili o čemu je tu reč, moram da upozorim na to da se razumevanje neke činjenice, tehnike, zakona ili otkrića u ve-

likoj meri sastoji od uviđanja kako se oni uklapaju i kako funkcionišu u mreži verovanja koja sačinjavaju prihvaćenu nauku. Baumberger (Baumberger 2011: 77) daje odličan primer rešavanja kriza izazvanih klimatskim promenama i globalnim zagrevanjem. Ono uključuje uviđanje kakve posledice po prirodne i društvene sisteme ima globalno zagrevanje, kako je povezano sa sečom šuma i sagorevanjem fosilnih goriva i u kakvom je odnosu sa njihovim posledicama, kao što je uništavanje ozona u stratosferi. Povrh toga, u taj splet fenomena uključene su i emisije gasova koji stvaraju efekat staklene bašte i rast prosečnih vrednosti temperatura u budućnosti. Dakle, prekid nekih praksi, kao što su sagorevanje fosilnih goriva, emitovanje gasova koji proizvode staklenu baštu i seča šuma, doprinosi rešavanju problema klimatskih promena.

Vidimo da, osim o objektnom i eksplanatornom uviđanju, možemo da govorimo i o praktičnom uviđanju. Baumberger je dao ovde izloženu ilustraciju pokušavajući da svede objektno i eksplanatorno uviđanje na praktično uviđanje. Međutim, kada kažemo da smo, na primer, tokom pandemije virusa korona naučili dete kako da pravilno pere ruke, imamo početnu intuiciju da je to posebna, praktična vrsta uviđanja, koja nije svodljiva na druge dve. Ono nije objektno uviđanje jer se ne tiče objekata već tehnika i vodi „znanju kako". Takođe, nije ni eksplanatorno uviđanje da tehnika pranja ruku ne zavisi ni od kakvog teorijskog objašnjenja o prirodi virusa i bakterija i dete ne mora da bude u stanju da objasni vezu između pranja ruku i prevencije zaražavanja da bi vladalo tom tehnikom.

Ako pogledamo šta bi o eksplanatornom uviđanju mogla da nam kaže najkorišćenija literatura iz filozofije nauke, videćemo da Van Frasen (van Fraassen 1980), na primer, u svojoj analizi pojma objašnjenja govori o njemu kao nečemu čime odgovaramo na pitanja „zašto?" tako što izlažemo kontrastnu klasu. Prikazivanje kontrastne klase znači da odgovaramo na pitanje zašto se nešto događa a ne nešto drugo, zašto se nešto dešava nasuprot nečemu drugom. Misaono eksperimentisanje sa različitim pristupima pandemiji virusa korona dobar je primer takvog postupka. Objašnjenje zbog čega se u određenom zamišljenom scenariju smanjuje broj zaraženih sastoji se u kontrastiranju sa scenarijima koji ne dovode do željenog rezultata. Tako je insistiranje na strogom zatvaranju tokom pandemije bilo utemeljeno na misaonim eksperimentima koji su spekulisali o rezultatima različitih stepena otvorenosti, uobičajenog ponašanja građana i većeg broja međuljudskih kontakata. U obzir su uzimane i različite situacije: da li se ti kontakti odvijaju na otvorenom ili zatvorenom prostoru, koliko vremena ljudi provode zajedno i slične. To objašnjenje je slično poznatom primeru iz istorije nauke, kada je Galilej eksperimentisao sa ubrzanjem tela u slobodnom padu. Objašnjenje zašto sva tela padaju istom brzinom počiva na kontrastu sa zamišljenim scenariom u kojem ubrzanje i, shodno

tome, brzina pada tela zavise od njihove mase, to jest u kojem je ubrzanje direktno proporcionalno njihovoj masi.

Protivčinjenički iskazi, kada ih shvatimo kao uzročne protivčinjeničke iskaze, čine misaone eksperimente nezaobilaznim u objašnjenjima zato što su nam oni neophodni da bismo uopšte procenjivali protivčinjeničke iskaze. Vilijamson (Williamson 2004) navodi da su misaoni eksperimenti zbog toga glavni metod u istraživanju protivčinjeničkih tvrdnji o međusobnoj zavisnosti pojava i stanja, a time i za odgovaranje na pitanja „a šta ako..." iz naslova knjige Peg Titl. Takav zaključak se savršeno uklapa sa stanovištem Greka (Greco 2014) i Hilsa (Hills 2015) da je dolazak do dobrog objašnjenja posredstvom misaonih eksperimenata praktično jednak onome što sam ovde nazvao eksplanatornim uviđanjem. To se najbolje vidi na primeru sagledavanja uzroka političkih kriza. Da bismo istražili međusobnu povezanost događaja koji su uzrokovali konflikt, formulišemo protivčinjeničke iskaze o alternativnim tokovima istorije. „A šta ako..." postaje neizbežno pitanje koje od misaonog eksperimenta pravi nezaobilazan metod za dolaženje do odgovora o pitanju nastanka određene političke krize.

Objektno uviđanje testiramo u situacijama u kojima od neke osobe zahtevamo da definiše neki termin, da nešto izrazi svojim rečima ili da popuni neke praznine u rečenicama. Te situacije mogu doslovno da predstavljaju testove, ali mogu da budu shvaćene krajnje metaforično ili u najširem smislu kao provere kompetencija osoba koje pretenduju da poseduju znanje relevantno za prevazilaženje neke krizne situacije. Neka osoba prolazi takvu vrstu testa ako je formirala semantičke veze između neke ideje koja je deo sadržaja tog znanja i drugih verovanja, pojmova i sposobnosti. Štaviše, te veze čine datu ideju smislenom. Još je Koare (Koyré 1968) ukazivao na to da misaoni eksperimenti daju smisao pojmovima tako što ih povezuju sa našim iskustvom te omogućavaju naučnicima da misaonom eksperimentacijom premoste jaz između teorijskih pojmova i empirijskih činjenica, pružajući empirijski semantički sadržaj za određene delove teorija. Na taj način nam omogućavaju da stvorimo semantičke veze između novih modela rešavanja kriza, s jedne strane, i iskustava koja smo već imali ili ćemo ih tek imati, kao i postojećih znanja i sposobnosti, s druge strane.

Misaoni eksperimenti doprinose i plodnosti. Naime, misaoni eksperiment omogućava novi uvid ako posle izvršenja tog eksperimenta možemo da uradimo nešto što nismo mogli ranije; da napravimo predviđanje, dođemo do objašnjenja, koristimo neki model rešenja ili prosto izvedemo neki zaključak (Velentzas, Halkia 2013). Na taj način, povećanjem plodnosti, misaoni eksperimenti doprinose praktičnom uviđanju.

Značaj maštovitosti, kao mentalne sposobnosti, za imaginaciju u misaonim eksperimentima, na kojoj sam od početka toliko insistirao, uklapa se u čitavu jednu tradiciju razmišljanja o mašti kao izuzetno važnoj komponenti razrešavanja političkih kriza. Na taj način imaginacija u misaonim eksperimentima može da ima i političku dimenziju, koja prevazilazi uobičajeno razmišljanje o njoj u specifično emancipatorskom smislu sredstva za razbijanje predrasuda i rušenje dogmi. Pritom ne mislim na klasične misaone eksperimente u političkoj filozofiji, kao što su Rolsov sa velom neznanja, zatvorenikova dilema ili Lokov sa žirovima i jabukama, da navedem samo neke primere. Imaginacija je ovde predstavljena kao zamišljanje mogućih svetova, a to može da bude i zamišljanje svetova u kojima se prevazilaze situacije i razvoji događaja koji izazivaju potčinjenost, diskriminaciju i nepravdu. Naravno, zajedno sa tim zamišljanjem ide i zamišljanje šta bi moglo da krene naopako u odgovarajućem misaonom socijalnom eksperimentu. Neuspesi velikih socijalnih eksperimenata u pokušajima da se reše velike društvene krize nameću dodatni značaj razvijenim imaginativnim moćima u ovom drugom smislu. To je smisao u kojem bih govorio o „konzervativnoj imaginaciji". Moć konzervativne imaginacije je moć sagledavanja stranputica u koje nas mogu odvesti radikalni zahvati koji su, pak, posledica izuzetne i hvale vredne maštovitosti u procesu rešavanja kriza. Zbog toga poznati misaoni eksperimenti u kojima sebe stavljamo pred izbor da li ćemo skrenuti pomahnitali tramvaj i time usmrtiti jednu osobu, spasivši na taj način sigurne smrti pet drugih osoba na pruzi (ili da li ćemo, možda, baciti debelog čoveka sa mostića i tako zaustaviti tramvaj) ne igraju isključivo podsticajnu ili stimulativnu ulogu za bavljenje apstraktnim filozofskim pitanjima kao što je odluka između utilitarističkih i deontoloških političkih teorija. Mnogo više od toga, oni nam pokazuju kako se rešenja kriznih situacija mogu sastojati (i obično se sastoje) od kontroverznih poteza, čiju nam prirodu može razotkriti samo postupak misaonog eksperimentisanja.

Bez imalo šale, misaoni eksperimenti mogu da imaju spasonosnu ulogu u vremenima velikih finansijskih kriza. Raderford (Rutherford) je, zahvaljujući izvesnoj slavi koju je vremenom stekao i kreiranju čuvenog modela atoma, mogao da računa na velika finansijska sredstva za svoja eksperimentalna istraživanja. Međutim, taj novac je u jednom trenutku presušio i njegova laboratorija se našla u nezavidnom položaju. U tom momentu, Raderford je uskliknuo: „Gospodo, ostali smo bez para, vreme je da počnemo da mislimo!" (Jones 1962). Na sličan način ćemo možda i u srpskom školstvu, koje permanentno kuburi sa učilima i sredstvima za modernizaciju nastave, ponekad morati da pribegnemo misaonom eksperimentisanju kao jedinom održivom obliku nastave. Ovaj tekst pruža bar jedan razlog zbog kojeg to i nije toliko loše.

Bibliografija:

Baumberger, Christoph (2011). Types of understanding: Their nature and their relation to knowledge, *Conceptus* 40: 67–88.

Elgin, Catherine (2006). From knowledge to understanding. In: Stephen Hetherington (ed.), *Epistemology Futures*. Oxford: Oxford University Press.

Gendler, Tamar (2000). *Thought Experiment: On the Powers and Limits of Imaginary Cases*. New York: Garland Press.

Greco, John (2014). Episteme: Knowledge and Understanding". In: Kevin Timpe, Craig Boyd (eds.). *Virtues and Their Vices*. Oxford: Oxford University Press.

Grimm, Stephen (2006). Is understanding a species of knowledge?. *British Journal for the Philosophy of Science* 57: 515–535.

Haddock Adrian, Millar Alen, Pritchard Duncan (eds.) (2009). *The Nature and Value of Knowledge: Three investigations*. Oxford: Oxford University Press.

Haddock Adrian, Millar Alen, Pritchard Duncan (eds.) (2010). *Epistemic Value*. Oxford: Oxford University Press.

Hetherington, Stephen (ed.) (2006). *Epistemology Futures*. Oxford: Oxford University Press.

Hills, Alison (2015). Understanding why. *Nous* 49 (2): 661–688.

Jones, R. V. (1962). Brunel Lecture. *The Bulletin of the Institute of Physics and the Physical Society* 13.

Koyré, Alexandre (1968). *Metaphysics and Measurement*. Harvard: Harvard University Press.

Kripke, Saul (1980). *Naming and Necessity*. Harvard: Harvard University Press.

Kvanvig, Jonathan (2009). The value of understanding. In: Adrian Haddock, Alan Millar, Duncan Pritchard (eds.). *Epistemic Value*. Oxford: Oxford University Press.

Pritchard, Duncan (2010). Knowledge, understanding and epistemic value. In: Adrian Haddock, Alan Millar, Duncan Pritchard (eds.). *The Nature and Value of Knowledge: Three investigations*. Oxford: Oxford University Press.

Timpe Kevin, Boyd Craig (eds.) (2014). *Virtues and Their Vices*. Oxford: Oxford University Press.

Titl, Peg (2018). *A šta ako... Sabrani misaoni eksperimenti u filozofiji*. Beograd: Centar za promociju nauke.

van Frassen, Bas (1980). *The Scientific Image*. Oxford: Clarendon.

Velentzas Athanasion, Halkia Krystallia (2013). From Earth to Heaven: Using "Newton's cannon" thought experiment for teaching sattelite physics. *Science and Education* 22: 2621–2640.

Williamson, Timothy (2004). Philosophical "intuitions" and scepticism about judgement. *Dialectica* 58: 109–153.

Mašan Bogdanovski

The Role of Thought Experiments in Crises Resolutions

Summary: In the first part, the article analyzes some of the philosophical underpinnings of the teacher training program „Application of thought experiments in teaching", which the author held in schools. An outline of the nature and motivation of thought experiments is presented, with an emphasis on the role of imagination. Imagination in thought experiments is a particularly significant component of resolving economic and political crises. Imagination plays a key role in the phenomenon I have called insight. Insight is a heterogeneous phenomenon and its various aspects are analyzed in the text. The central part of the text deals with the ways in which various forms of insight yield new knowledge and, even more importantly, the understanding of natural and social phenomena that enables the resolution of social crises. Thought experiments lead to insight and new knowledge without new experience. Insight is the predominant epistemic goal of thought experiments in the process of crisis resolution.

Keywords: thought experiments, crisis resolution, imagination, insight, knowledge

Ivana Janković*

NUDGING AND DELIBERATION: INDIVIDUAL AUTONOMY, EPISTEMIC VICES AND VIRTUES**

Abstract: Recent findings on cognitive deficits and motivational-cognitive biases in human behavior and decision-making are well confirmed. There are different approaches for solving the problems these deficiencies lead to in decision-making. Since there is a growing worldwide trend of using behavioral sciences to inform public policy decisions, this paper aims to consider and critically review two strategies for improving people's behavior and decision-making in public space: nudging and public deliberation. Should policymakers develop mechanisms for guiding the choices of their citizens or support and encourage them to make better decisions by themselves? In other words, the question is whether governments can influence people to make better decisions without violating their freedom and autonomy. The debate about "libertarian paternalism" has raised many questions about the possibility of reconciling the basic assumptions of these two concepts into one. This position entails the creation of public policies by using nudging to help people make better decisions (related to health, wealth, and happiness) without limiting their freedom of choice. We will consider the arguments for and against this intervention, its consequences for personal autonomy, and the development of epistemic vices. Following the fundamental values of modern democracies, we will argue that although nudging can serve as a tool for changing people's behavior for the better, public deliberation is a better long-term strategy. When applied to public policy, the nudge strategy risks ignoring or diminishing the personal abilities and institutional requirements necessary for the fruitful exercise of democratic citizenship. Alternatively, public deliberation can improve decision-making by successfully addressing cognitive deficits while promoting civic virtues without violating liberty. It can also preserve personal autonomy and develop epistemic virtues. Finally, we will argue that, regardless of this general conclusion,

* Department of Philosophy, Faculty of Philosophy, University of Belgrade, ivana.jankovic@f.bg.ac.rs

** The paper is based on research conducted within research project *Man and Society in the Time of Crisis*, financed by the Faculty of Philosophy, University of Belgrade.

a nudging strategy can be morally permissible in crises, when quick solutions are needed.

Keywords: nudge, deliberation, decision-making, deliberative democracy, individual autonomy, epistemic vices, epistemic virtues

In the last three decades, cognitive science, behavioral economics, and social epistemology have unequivocally confirmed deviations from the classical model of rationality. The model starts from the image of man as a perfect *homo economicus* – an ideal, omniscient decision-maker, endowed with perfect rationality, unlimited cognitive capacity, perfect access to information, an invariable set of preferences, and consistent, self-interested goals. In reality, far from this ideal, human decision-making and behavior are biased by what is considered to be – from the perspective of rationality – irrelevant features of the decision-making context. There is convincing evidence in today's dominant approaches to the study of decision-making, judgment, and inference (dual process theories) of the existence of cognitive limitations and numerous errors. What is more important is that those errors are systematic, identical, and predictable for most members of the human race. The deviations of actual behavior from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the normative system (Tversky & Kahneman, 1988, p. 167). These deviations are called cognitive illusions or biases, and at the basis of their manifestation are individual heuristics, mental encapsulated algorithms that are faster and cognitively "cheaper" but also riskier in terms of outcomes (Tversky & Kahneman, 1974; Kahneman, Slovic & Tversky, 1982; Kahneman, 2011).

The phrase *bounded rationality*, which refers both to limitations in knowledge and to human capacities for data processing, describes today's unquestionable fact about numerous "mistakes" in our cognitive functioning (Simon, 1990). These insights into ways people make decisions show us how irrelevant characteristics can lead us to fail to achieve what we want (or would want if we were well-informed and perfectly rational). In other words, empirical findings show that the widely held thesis about the perfect rationality of *homo economicus* is unjustified and give us a more realistic picture of how human beings truly behave and think.

The findings of behavioral economics and psychology on the existence of cognitive limitations and biases in human reasoning had a great

impact on other areas and contexts of human decision-making.[1] Ignoring these findings and proceeding (and sticking to the previously established economic assumptions about equality in power and consumer sovereignty) can lead us astray in understanding the political and social reality in which we live. These insights show us how these constraints can systematically lead us to deviate from what would be consistent with our desired goals and intentions if they were well-informed and perfectly rational.

Paying more attention to cognitive biases, the imperfection of human judgment, and the bad decisions that result from them would help provide the key to effectively dealing with important social challenges (such as global warming, the coronavirus pandemic, obesity epidemic) and poor economic decision-making. People make systematic errors that cannot be ignored. The bad, predictable and irrational behavior then leads to the conclusion that it is necessary to introduce some paternalistic and epistemic interventions that aim at changing human behavior, which would then lead to better decisions. In other words, since human behavior is not always rational, people could benefit from some external, paternalistic intervention.

1. Nudging as a solution to the problem of the imperfection of human decision-making

People too often make poor decisions and behave in harmful ways for themselves, their families, friends, and society, even when the preconditions for rational decision-making exist. Therefore, external intervention may be desirable in certain situations. Richard Thaller and Cass Sunstein are the most famous advocates of one type of intervention – the nudging strategy. In their renowned book *Nudge: Improving decisions about Health, Wealth and Happiness,* these two authors consider many examples of successful nudging in various areas of human activity. Their illustrations of nudging promote the desired behavior change in various fields such as health care, financial structures, pension plan making, environment protection, and reducing air pollution. They also cite successful strategies for mass organ donation, increasing healthier food consumption in cafeterias,

1 In the domain of political reasoning, some authors consider the incompetence of ordinary citizens concerning the biases, information, and knowledge they (do not) possess as one of the strongest findings that social science has produced (see, for example, Carpini & Keeter, (1996), *What Americans Know about Politics and why it Matters,* Yale University Press; Caplan, (2011) *The myth of the rational voter: Why democracies choose bad policies,* Princeton University Press; Brennan, J. (2017) *Against Democracy,* Princeton University Press).

wiser investment in pension funds, and improvements in school choice decisions. According to their definition, *a nudge* is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaller & Sunstein, 2008, p. 6). These statements are not directives or dramatic incentives, but, as these authors state, simple and easy interventions (with low costs and simple implementation) that we can easily ignore. They do not ban anything. All the previous options remain available to the persons whose behavior is being directed to guide them toward a more beneficial outcome gently. What essentially separates nudging from prohibition is that the initial set of available options remains unchanged. "Putting the fruit at eye level counts as a nudge. Banning junk food does not." (Thaller & Sunstein, 2008, p. 6) In other words, they define nudges as gentle pushes in a particular direction, which preserve freedom of choice while still influencing individual behavior.

Nudge, as a type of epistemic intervention, presents one of the most famous forms of *epistemic paternalism*.[2] Epistemic paternalism refers to the practice of interfering with the inquiry of another, without prior consultation with him, for the sake of his own epistemic good (Ahlstrom-Vij, 2013). Thus, the practice of nudging is one of the ways someone interferes with an individual's ability to conduct their own research. It influences someone's behavior without giving them a reason for it or using force. Given the imperfection of human cognitive functioning, the nudges considered and defended by Thaller and Sunstein shape people's choices for their benefit. In order to be successful, well-designed nudging must start from known cognitive patterns and predictable errors.[3] The goal of those who regulate the context of our choices is, therefore, to promote behavior that is in our best interest and the best interest of the entire society.

According to defenders of the nudge strategy, it is possible to solve many social problems and improve people's lives. Insisting that such interventions, which are relatively weak, cheap, and practical, are not restrictive concerning the initial set of options (and thus do not limit freedom of choice) leads Thaller and Sunstein to label this type of intervention

2 The difference between ordinary and epistemic paternalism is that generally speaking, first represents the limitation of the freedom of an individual due to the interference of the state or other individuals, against her will and for the sake of her good (see Dworkin, G (2020). "Paternalism". Stanford Encyclopedia of Philosophy, Zalta, E. N. ed.).

3 Thaller and Sunstein discuss this in terms of dual process theories of human psychology. They cite many errors in human reasoning, such as unrealistic optimism in scheduling commitments, status quo bias, power of inertia, overconfidence, anchoring, confirmation, framing effects, and loss aversion.

as "libertarian paternalism." The first part of the phrase refers to the absence of obstacles for any individual to choose what she prefers – nudges leave the set of possible options intact. It implies a lack of restrictions on freedom of choice, the non-imposition of inconvenience, and costs concerning the time that needs to be invested in the decision-making. The attribute "paternalistic" indicates an attempt to influence the choices of individuals in a way that will make their lives better, healthier, and longer (Thaller & Sunstein, 2008).

The example Thaller and Sunstein use to illustrate what they have in mind when they talk about this specific type of paternalism is the one that shows how the placement of *food* in a *cafeteria* affects the choices people make. Caroline, a nutritionist and the person in charge of food in the school system, is responsible for the thousands of children who eat in her cafeterias daily. One day, an idea for an experiment comes to her mind. Without making the slightest change in the menu, she rearranges the food available to the children. She wanted to test whether the way the food was arranged would influence the children's choices. She set food differently in different schools: in some cafeterias, dessert was the first in line and within easy reach; in others, fruits, somewhere sweets were placed further or were served at the end, somewhere in separate rows. The results of her intervention were dramatic (with this procedure, she was able to increase or decrease the consumption of some foods by 25%). Choices made by the children were greatly influenced by how the food was served. Caroline, the designer and creator of the food arrangement, by simply moving the food (without changing the menu!), significantly influences the students' choices and encourages them to make healthier choices (Thaller & Sunstein, 2008, pp. 1–4). This type of intervention is "paternalistic" because there is a conscious intention to influence the behavior of school children in their best interest without prior consultation. At the same time, it is considered "libertarian" since the freedom of choice regarding food is still being maintained – those with junk food preferences can still choose to eat fries and sweets.

In addition to this example, Thaller and Sunstein give many others that show how the government nudges its citizens and produce desired forms of behavior – a state that uses human inertia and procrastination to automatically register adults as organ donors[4] (2008:175–183); simply changing the size and shape of the plates helps to reduce the calories that

4 Default is presumed consent, and then there is an option to change it if you disagree; instead, the default option can be that they are not organ donors, and then to have the option to round off if they want to be. The first strategy results in more organ donors than the second, even though the options are the same.

people consume when they pour food onto those plates (2008: 44,82); carving a picture of houseflies into public urinals will contribute to cleaner public toilets and less urine spilling out of the urinals[5] (2008:4); there are countless opportunities to improve people's health through the framing of the risk that the doctor presents to the patient – people are more likely to go for a preventive check-up if the doctor presents them with an increased risk of disease if they do not than if he tells them how much they can reduce their health risk if they do they go to preventive examinations on time (2008:157)[6].

2. Critique of the nudging strategy: personal autonomy and epistemic vices

We saw that the behavior guidance strategy relies heavily on psychological research of the decision-making process, using the imperfections of human decision-making abilities to influence people to make better decisions without limiting their freedom of choice. Thaller and Sunstein, therefore, claim that "libertarian paternalism" is not an oxymoron (although it seems so at first glance!) and that it is possible to guide people in the direction of better decision-making and, at the same time, preserve the freedom of choice they had before. Another important fact is that what at first appears to be an insignificant detail in creating the context of choice (e.g., the order and manner in which food is served in the canteen) in social situations has a significant effect on people's behavior and how they make their choices. "Choice architecture, both good and bad, is pervasive and unavoidable, and it greatly affects our decisions" (Thaller & Sunstein, 2008, p. 252). Additionally, Thaller and Sunstein argue that our choices are always context-sensitive and are never neutral. Having that in mind, their proposed strategy can only help us find more creative and better ways of making decisions in today's highly polarized modern democracies. These authors support nudging, which presents a beneficial strategy to people and serves as a general-purpose tool for reaching sound decisions. They argue that with institutional help, it should be present and used in today's

5 Males typically do not focus on where they are urinating, which can often lead to a mess. However, they are much more accurate if they see a specific target.

6 Four years after the publication of Thaller and Sunstein's book, the nudging strategy was widely used in American and English democracies. Sunstein was an adviser to US President Barack Obama, and Thaller was an adviser to the Prime Minister of Great Britain, David Cameron, in the Nudge Unit, which works in partnership with national, regional, and local authorities to improve the lives of people and communities. (Hansen & Jespersen, 2013, p. 4)

societies. However, even when successful, the nudging strategy has been subject to various criticisms from the start (Bovens, 2009; Hausman & Welch, 2010; Hansen & Jespersen, 2013; Barton & Grüne-Yanoff, 2015; Riley, 2017; Meehan, 2020).

2.1. Epistemic justice, autonomy and individual freedom

One of the first questions is whether the strategy of nudging, considering how it works (even when it is gently implemented and benefits people), is generally morally acceptable in our world. This issue is of great importance for interventions carried out by the government and public sector (primarily because the question of morality and the determined limit of persuasive strategies are challenging to apply to the private sector). Riley (2017) believes that guiding behavior, even when it brings accurate information to the individual on whom the intervention is performed, is not designed as an invitation to careful critical consideration and reasoning. Such interventions aim to direct people's behavior, without coercion, in a particular direction (even if it is beneficial to them and the entire society) but not to engage their critical capacities of a higher order by inviting them to open deliberation and rational persuasion. In that case, they can be considered manipulations that carry the risk of epistemic injustice (Riley, 2017, p. 600). Critical capacities and the possibility of reasoned discussion are not something that is given to us and is something we are born with. They are developed and nurtured. Riley believes that denying opportunities, means, and support for developing these capacities through the exercise of other strategies that prevent or hinder the development of these higher epistemic capacities necessary for the development of an epistemically mature person is simply *unjust.*

The question for libertarian paternalism, as well as for the practice of nudging, is whether this kind of intervention respects the person's autonomy and whether it respects the right of the individual to make his own decisions. If we understand autonomy in a sense given in the example of the school cafeteria – as the freedom to choose options or preferences (without additional inconveniences, social sanctions, or investing more time and energy than before) – then the answer is positive. On the other hand, if we understand autonomy and liberty in the broader sense as an independent capacity for critical and open practical reasoning, then the answer is not so obvious. "Their freedom, in the sense of what alternatives can be chosen, is virtually unaffected, but when this "pushing" does not take the form of rational persuasion, their autonomy—the extent to which they have control over their evaluations and deliberation—is diminished" (Hausman & Welch, 2010, p. 128). In other words, when people are not

allowed to be rationally persuaded by others but only pushed in some direction of action, their freedom to choose is reduced.

Suppose, therefore, autonomy requires the ability to actively think critically about one's various goals, attitudes, desires, and beliefs, but also the possibility of exercising these capacities and the existence of a person's authority concerning these beliefs, attitudes, desires, and one's life and character. In that case, the question is whether nudging violates personal autonomy (Riley, 2017, p. 606). Nudging practice does not aim to engage these critical capacities. It is not so by accident but with the clear intention of exploiting imperfections in our cognitive functioning and correcting "mistakes" in human reasoning by using "shortcuts" (an efficient, quick way that does not require the effort and time of those whose behavior is being directed). Having that in mind, it is clear why nudging fails to develop and nurture the capacity for proper judgment. In this light, Riley claims that, in general (not in every individual case), nudging tends to violate epistemic justice and the personal autonomy of the person on whom the nudging intervention is performed. Although freedom, understood in the narrow sense of the options they can choose, is not limited (no intervention was made regarding the set of alternatives available to individuals), personal autonomy, understood as having control over the evaluation of alternatives and reasoning about them and the direction of further action, is reduced.

If nudging does not inform people and does not use "rational persuasion" (which implies careful thinking and discussion when making decisions) but instead takes the form of "pushing" in a particular direction, perhaps a more precise term for this type of intervention is *manipulation* (Bovens, 2009; Hausman & Welch, 2010). "Paternalistic actions either coerce people or use imperfections in their deliberative abilities to shape their choices." (Hausman & Welch, 2010, p. 129) Authors such as Hausman and Welch emphasize that there is a significant difference when someone tries to convince us using only valid arguments and facts and when, for example, he tries to convince us using our inattention or one of the well-known and widespread biases. "To the extent that they are attempts to undermine that individual's control over her own deliberation, as well as her ability to assess for herself her alternatives, they are prima facie as threatening to liberty, broadly understood, as is overt coercion." (Hausman & Welch 2010:131) These authors even underline that some types of state-led paternalistic interventions (e.g., mandatory seat belt laws), even with a reduced set of available options [7], are less of a threat to freedom from what Thaller and Sunstein refer to as a weak

7 There is no longer a legitimate option for an individual not to wear a seat belt.

and unobtrusive type of paternalism[8]. This is so because this kind of paternalism implies interventions whose effects on behavior people are not aware of.

Thaller and Sunstein's point is that citizens are *always* influenced by *some decision-making context*. Keeping that in mind, they state that nudging is compatible with preserving freedom and acceptable if guided by libertarian paternalism and Rawls' publicity principle (Rawls, 1971). This principle is essential because it makes a clear distinction between a real threat to citizens' freedom through manipulation means such as subliminal messages and the nudges they favor. The first is morally objectionable precisely because the government would be unable or unwilling to defend such actions to its citizens publicly. Subliminal messages undoubtedly count as manipulations precisely because they are invisible, and we have no control over them (Thaller & Sunstein, 2008, pp. 244, 245). However, the problem with the "publicity principle" is that it does not fundamentally resolve the complaint about the violation of individual freedom, understood in a broader sense. Additionally, the nudging effect would most likely disappear if this intervention were transparent (Bovens, 2009, p. 219). Furthermore, the government could use subliminal messages and, at the same time, *inform the people* that it will use them as a means to achieve specific (beneficial) ends and thereby defend them. However, Thaller and Sunstein reject the justification of such interventions. That means that more than the publicity principle is required. In that case, it needs to be clarified what the difference is, given that such interventions are very effective, do not limit the possible alternatives available to people, nor do they require higher costs in material resources or time (Hausman & Welch, 2010, p. 132).

As already stated, Thaller and Sunstein often emphasize that external factors constantly shape our choices. There is no such thing as a "neutral design" – the cafeteria managers have to arrange the food in a particular order, the doctor has to present alternative treatments to the patient in some way, and those who design the ballots have to sort the candidates in some order, and so on (Thaller & Sunstein, 2008, p. 3). However, there is a significant difference between whether this "shaping" is intentional. The order of options must always be of some sort, and something must be the starting position. Some choice of architecture will inevitably be established. It is essential whether it is carried out on purpose.[9] If the latter is

8 It is so because we consciously accepted that choice due to the force of a coercive, binding law.

9 Thaller and Sunstein argue that the context always influences our choices and that they can be directed in more manipulative and subtle ways (i.e. marketing).

the case, such an intervention must be morally justified. "Even when unshaped choices would have been just as strongly influenced by deliberative flaws, calculated shaping of choices still imposes the will of one agent on another." (Hausman & Welch, 2010, p. 133) It seems that the democratic system and the values on which it rests stand in opposition to the idea that democratic government should influence the choices of their citizens. That is so at least to the extent that nudging strategies do not use techniques based on public, open, reasoned, and strictly rational persuasion.

2.2. Epistemic vices

We know that human cognition and human behavior are far from perfect. When an individual finds herself in everyday life situations, forced to make various decisions, she relies on a series of prejudices, biases, false beliefs, misinformation, and stereotypes rather than facts and logic. From an epistemological point of view, epistemic (intellectual) vices deserve as much attention as epistemic virtues. Those are the subject of a relatively new discipline within social epistemology called *vice epistemology*. Within this discipline, the focus is no longer on the truth value of statements but on the subject's epistemic vices. Epistemic vices include, e.g. credulity, intellectual arrogance, dogmatism, prejudice, closed-mindedness, and carelessness. Epistemic vices are character traits that systematically hinder the acquisition, transmission, and retention of knowledge (Cassam, 2019, p. 1).[10] Epistemology of vices focuses on "the nature, identity, and epistemological significance of intellectual vices" (Cassam, 2016, p. 160). These intellectual epistemic vices hinder knowledge acquisition and rational action following that knowledge.

As in nudge strategy, vice epistemology aims to guide and direct human inquiry. Alvin Goldman states, "if we wish to raise our intellectual performance, it behooves us to identify those traits which are most in need of improvement" (Goldman, 1978, p. 511). Overcoming certain weaknesses in deficiencies in human reasoning (and, consequently, decision-making) is the point of intersection between nudging and vice epistemology. If the nudge strategy "fixes" epistemic vices, then it can be considered a useful tool from the point of view of vice epistemology. If this is not the case, or if it additionally causes some epistemic vices, then this is a problem for this strategy. Suppose epistemic vices are strongly rooted in social circumstances and/or deep psychological dispositions (Cassam, 2019). In that case, the nudge strategy does not help untangle or mitigate

10 Unlike cognitive biases, which tell us how the human brain operates, epistemic vices refer to character traits that operate on a personal level.

them but merely masks them. The vase has the disposition of fragility, and "by protecting the vase with bubble wrap, we have not vitiated its fragility" (Meehan, 2020, p. 256). In other words, the disposition is only disguised. Likewise, nudging does not "correct" and does not remove epistemic vices in the long run, just as bubble wrap did not eliminate the fragility of the vase. Nudging works at the time of its intervention but does not fix or solve any long-term problem with human reasoning. When there is no nudging, people go back to their old behavior.

Another problem observed by vice epistemologists, critics of epistemic paternalism and nudge strategy, is the problem of epistemic laziness that these practices can produce in the long run (Kidd, 2017; Kidd, Battaly & Cassam, 2020). Epistemic laziness occurs when something does not allow us to practice and exercise our epistemic capacities. The practice of nudging acts as such activity and can guide these epistemic vice. Therefore, this non-use of one's epistemic capacities occurs when people too often rely on this kind of epistemic strategy in a large number of situations. As we saw in the previous chapter, an autonomous person with fully developed epistemic capacities must possess and exercise capacities for rational reasoning and critical thinking. If something prevents the development of these capacities, or when they already exist, discourages them from their usage, it leads to the emergence of epistemic vices such as epistemic laziness (Meehan, 2020, p. 257). And so, even when in certain situations nudging leads to (epistemically) better outcomes, it does so in a superficial and short-term way, thus reducing the individual's capacity to develop good epistemic traits, which in the long run would enable the disappearance or reduction of epistemic vices.

Practices that support epistemic laziness and uncritical thinking fail to treat people as rational beings capable of reasoning and actively and critically considering their options. Making people dependent on the practice of nudging over time leads to the atrophy of specific epistemic capacities (Meehan, 2020). In other words, if epistemic laziness and uncritical thinking are systematically developed rather than undermined, people are not treated as reasonable human beings. Healthy, mature human beings can play an active role in developing their own mental life. When people manage and control aspects of their mental lives, including their epistemic virtues and vices, they are actively involved with the relevant feature of their own mental life to promote development towards some desired goal[11] (Debus, 2016).

11 Someone's active involvement with their mental life is often directed towards a goal. For example, if someone wants to become better at remembering things, they will actively try to improve their memory by practicing things like memory games or mnemonic devices.

3. Deliberation as a strategy for improving decision making

Citizens today live in a complex world characterized by deep disagreement, the complexity of social problems and political decision-making. Given their cognitive limitations and capacities, limited information and time available to process a great deal of information, they use different social signals that help them decide what to do. We have seen that nudging is one strategy that a government can use to change the behavior of its citizens and the problems it can produce. Nevertheless, nudging is one of many strategies to help make better decisions. Libertarians question the government's authority to nudge citizens in any direction, even if policymakers believe it is in the individual's best interest. They believe that this behavior manipulation is a form of government overreach. However, libertarians cannot provide a full critique of nudging or an alternative vision of addressing collective action problems because they do not consider the possible effects of nudging on citizens or the democratic conditions needed for legitimate governance. If governments and other public institutions want to understand how citizens think, encourage them to change their behavior, and come to decisions that would be better for them and society, organizing public debates and deliberations can be a successful strategy that avoids previously mentioned problems.

Deliberative democrats advocate wider public engagement of citizens in the decision-making process. They claim that *if* citizens had a chance to become better informed, and had enough time and appropriate context, they would come to what is best for them and their community (Fishkin & Lushkin, 2005; Gray 2009). This can be achieved through the development and implementation of institutions designed to override cognitive and emotional biases and to bring about the outcomes that theorists of deliberative democracy and their creators predict and strive for (Warren, Pearse 2008; Siu 2009; Mercier, Landemore 2012; Gerber et al. 2018).

In recent years, public deliberation has received increasing attention as a means to improve the quality of decision-making in public and private institutions. The basic idea is that it is possible to bring together different interested parties to discuss problems and possible solutions with mutual respect, which will then lead them to better decisions (Cohen, 1989; Gutmann, Thompson, 1996; Dryzek, 2000; Bächtiger et al. ., eds., 2018). Proponents of this strategy argue that it is possible to get citizens to think about complex and complicated issues related to their lives and the

life of the community in an innovative way that allows taking into account different opinions, information, evidence, and perspectives of all involved in the decision-making process, which results in better outcomes. They think that citizens can identify, frame, and address public problems together with other diverse actors with the help of innovative institutions. Unlike nudging, which represents a non-educational strategy (by changing the way choices are presented, making the default choice something more likely to lead to the desired outcome or other similar methods), deliberation relies on a strongly educative approach. Through developing new skills and providing information, deliberation aims to make citizens more competent decision-makers. Both strategies present different ways of coping with bounded rationality.

The idea of civic capacity refers to both individual competence and virtues and institutional conditions that allow for communication and decision-making with others in a civic-minded way. These essential aspects of political life could be overlooked or bypassed if we rely too heavily on the expert-led nudging strategy. Nudge, as a form of paternalism, relies on the top-down, expert-led approach to decision-making. This kind of decision-making is at odds with the ideals of deliberative democracy. Any epistemic paternalism presupposes that there is a single correct way to view the world and that those in positions of power should use their knowledge to guide others toward this correct way of thinking. This idea contrasts the deliberative democratic ideal, which holds that collective decision-making should be based on open and reasoned debate between equals. Critics of the nudge strategy argue that nudging is a form of manipulation that infringes on people's freedom to make choices (Mitchell, 2004; Hausman & Welch, 2010). They are, as we previously said, concerned about attempts to manipulate people's behavior without engaging their reflective, conscious thought, which, as we saw, leads to a loss of liberty and autonomy. On the other side, in deliberative democracy, citizens are empowered to make decisions based on reasoned deliberation rather than being nudged into a particular course by those in a position of power. This idea applies to conceptions of personal autonomy, freedom, and epistemic virtues, as well as widely accepted democratic values, principles, and practices. It means that the ability of a society to self-govern democratically relies on the civic capacities of its citizens. If the conditions for direct public deliberation exist, the full development of each person's capacity for participation in political life and self-management is ensured (Dryzek, 2009).

The claim that freedom of choice is maintained to "mere" nudges is of great importance to the defenders of libertarian paternalism. This claim

means that even though the government may be changing people's behavior by slightly pushing them in a certain direction, those people still have the freedom to choose what they want to do. However, this is a *non-political*, private kind of freedom that cannot satisfy *political freedom* and democratic principles of consent and legitimacy (Button, 2018). What deliberative democrats claim is that for freedom to be meaningful, people should have equal opportunity to participate in the decisions that affect them and reflect freely on their political preferences (Dryzek, 2009; Mansbridge et al., 2010).

The difference between designing behavior (as in the case of the nudge strategy) and developing civic and deliberative capacities is the difference between a focus on external mechanisms and one focused on civic (toleration, mutual respect, reciprocity, practical reasoning) and ethical virtues. Deliberative democracy is a way of making decisions, promoting the ability to make thoughtful choices. The aim is to create a space where people can come together and openly, critically, and respectfully discuss the issues that matter to them. Since deliberative democracy relies on the participation of autonomous individuals who are willing to engage in open and respectful dialogue, autonomy is only possible when we have a say in the decisions that affect our lives.[12] In a collective decision, our choices will affect others. Therefore, we must be able to justify our choices to others in order to respect their agency and autonomy. Deliberative democracy can cultivate autonomy by encouraging participants to make their reasoning public, increasing the availability of information, and by ensuring all opinions are included and heard by all (Elstub, 2008).

As we said, deliberative democrats aim at powering citizens to determine – collectively and publicly – what *they* consider desirable in terms of policy responses to pressing social and political challenges. However, sometimes it can be difficult for people to participate because it costs them time and energy. Nevertheless, there are ways to reduce these costs so that more people can participate. It can be achieved by designing effective democratic deliberative practices and institutions that compensate for well-known cognitive and emotional biases, manage and reduce information costs for participants[13] and give effects that are in line with theory assumptions, as intended by their designers (Fishkin et al., 2005; Warren, Pearse 2008; Mercier, Landemore 2012; List et al., 2012; Gerber et al., 2018).

12 For deliberative democracy and autonomy, see Cohen 1989, Richardson, 2002; Christman, 2005.

13 This can be done with policy guides, expert witnesses, online modules, and other tools.

3.1. Personal autonomy and epistemic virtues

Liberal tradition and Mill's conception of personal autonomy defines a *sovereign individual's* domain as the one that includes all the activities that he or she can do without negatively impacting others (Mill, 1988). Therefore, honoring an individual's autonomy is respecting her right to make choices about their actions as long as they do not hurt others. In other words, autonomy is about individuals being in control of their own decisions, beliefs, and actions. However, it also presupposes the mental capacity to understand the consequences of their choices (Raz, 1986, p. 369). It is about self-determination and self-government. "If individual decisions require private deliberation in order to be autonomous, then collective decisions require collective deliberation, including the sharing of information and reasons through public debate" (Elstub, 2008, p. 4).

Public deliberation has many benefits, including transforming citizens' preferences and promoting their betterment (Elster, 1998). It has educative and community generative power and increases the epistemic quality of the outcome (Cooke, 2000). If we define *personal autonomy* as the ability to make choices about one's life without interference from others, then deliberation, as the process of thoughtfully considering the options and making a well-reasoned decision, can be seen as a strategy that enables the preferences of the participants to become more autonomous. These processes require a commitment to making the best possible decision. They result in decisions everyone can agree on and respect everyone's ability to make their own choices. Additionally, it requires critical thinking and people to give reasons for their preferences, which encourages them to think about the opinions and needs of others. It means that people have to explain their opinions to others, better understand their opinions and interests, help to increase the availability of relevant information, and allow participants to express themselves freely, which cultivates both hearer and speaker autonomy. All this creates an atmosphere where people see each other as autonomous individuals with equal importance. "In collective decisions, our choices will affect others and, therefore, must be justified to others in order to respect their agency and autonomy. Individual decisions require rational deliberation to form intentions and, consequently, collective decisions require collective deliberation to form collective intentions" (Elstub, 2008, p. 58). Deliberative democracy also increases the availability of information and allows people to express themselves openly. That is important for personal autonomy because it supports people in developing their own opinions and values. When people are forced to make decisions without carefully considering all the options, they are more likely to accept the values of those in authority.

Furthermore, through discussion and debate, a group can generate an idea that none of the individuals would have thought of on their own. Under deliberative conditions, individuals can expand their limited and fallible perspectives and information by relying on others' knowledge, potential, and experiences. Thus, "democratic deliberation has the capacity to lessen the problem of bounded rationality – the fact that our imaginations and calculating abilities are limited and fallible" (Fearon, 1998, p. 49). This is because reasoning works best within the group (Mercier & Landemore, 2012). It also involves epistemic standards that help judge whether a given deliberative process produces better or worse outcomes and allows us to focus on the substance of the process rather than purely on the procedure. Although in politics we may never be entirely sure if we have found the right political decision[14], and can never be sure if the political decisions we make are the right ones, that does not mean we should not try to make them as good as possible. Just as there are better and worse answers to logical questions, there are also better and worse answers to political questions.

Some authors even argue that reasoning evolved for a specific function – argumentation (Mercier & Landemore, 2012). This view implies that reasoning does not have an exclusively individual function but a social one, more precisely, a polemical function (Mercier & Sperber, 2011). The primary goal of reasoning is to find and evaluate arguments to convince others and be convinced when appropriate. These authors claim that one of its functions is to produce epistemic improvement through deliberation. Suppose someone is not *truly* listening to the arguments being made by others (reasoning from one's own opinion only without considering other opinions or reasoning with like-minded people[15]), and is instead only preparing counter-arguments. In that case, that person is not engaging in *genuine deliberation*. Furthermore, people do not blindly believe what other people say. They use their cognitive abilities to evaluate the information communicated to them. Communication is crucial because it allows us to exchange information and ideas and helps us to make better decisions as a group.

14 We can usually reveal that in the future, using, for example, different standards such as GDP growth, inflation, or unemployment rate

15 On the individual level, reasoning alone and reasoning in a group of like-minded people can increase confirmation bias and overconfidence or produce more polarization – more extreme attitudes than before deliberation (Mercier & Landemore, 2012). Suppose the group is made up of people who share the same opinion. In that case, they will be less likely to use reasoning to examine the arguments put forward by other discussants critically.

However, it can also be dangerous if the information exchanged is wrong or if it is deliberately manipulated. To protect ourselves from these dangers, humans have developed the ability to be epistemically vigilant (Sperberg et al., 2010). It means we can register and eliminate potentially threatening information in the communicative process through argumentation or polemic. In other words, all cognitive biases and illusions (to which we as a species, and as already said, are systematically prone) come as a result of the use of reason in isolation, outside of the social practice and needs. People rely on what is close to them, deepening their (false) beliefs and biases. These biases interfere with reasoning and decision-making. However, communication with people with *different perspectives* can help reveal and correct mistakes (Mercier & Landemore, 2012). When we use reasoning in its normal conditions – in a deliberation – it can be expected to lead to better outcomes, consistently allowing deliberating groups to reach epistemically superior outcomes and improve their epistemic status. Instead of making people better at reasoning, we should focus on changing the situations and environment to provide more ideal conditions for deliberation. We should make these changes at the level of institutions, not individuals.

The approach that claims that in addition to political, democracy also has epistemic values defends the concept of democracy which presupposes that besides accepted procedural values, democratic collective decision-making also has some significant epistemic values and produces epistemic good (Cohen, 1986; Estlund, 2009; Landemore 2012). It means that democracy, to be justified, cannot be reduced to its procedural values but also epistemically justified by the beliefs, decisions, and solutions it produces. Epistemic good here presupposes not only knowledge and forming *true* and avoiding false beliefs (Goldman, 1986; Vij, 2013) but also developing epistemic virtues/understanding (Pritchard, 2013)[16]. This broader understanding does not focus on knowledge or truth but on epistemic virtues such as intellectual curiosity, open-mindedness, (empathetic) understanding of others, critical thinking, truthfulness, honesty, objectivity, and impartiality. Thus, epistemic virtues are characteristics that promote intellectual flourishing or cultivate character traits (Zagzebski, 1997). Those are epistemic goods that are valuable in themselves, not only as a means to come to truth. Development of these specific epistemic virtues, which, as we already said, can be accomplished through genuine deliberation, increases personal autonomy. If autonomy means being both

16 Although epistemic democracy is usually characterized in terms of "aiming at truth", epistemic value does not say that these epistemic good must be reducible to true belief.

self-governing and able to make rational choices[17], then it requires above-listed virtues.

3.2. Nudging vs. deliberation

As we saw, nudging is a way for governments to change people's behavior by taking advantage of how people usually think and make decisions. It differs from other methods like taxes or bans, which change people's behavior by making it more expensive or difficult to do. Nudging is low cost, individuals do not need to cooperate, and it is more effective than those other methods. According to its proponents, it does not take away people's freedom to choose. It works with people's natural biases instead of against them. It accepts citizens as they are and turns *them from their path or course of action to make better decisions.* On the one hand, nudges can be seen as a means for paternalistically correcting people's choices, leading them away from errors or bad decisions.

On the other hand, we can see nudges as incompatible with respect for autonomy. People have a right to make their own choices, even if those choices are not in their own best interests. Nudges interfere with people's ability to make choices, so they disrespect people's autonomy. This means that nudges can also be seen as interventions that take advantage of our cognitive and motivational deficiencies to achieve a goal. It leads to the conclusion that policymakers are incentivized not to fix these deficiencies, as doing so would make nudges less effective. The nudge strategy is based on the belief that the *state knows* what is best for citizens and that policymakers should act as experts to steer them in the right direction. The goal is to make changes that will benefit the individual and society without the individual even being aware of it.

Deliberation, as defined by proponents of deliberative democracy theorists, has the potential to help to create a more informed and virtuous citizenry. It can increase epistemic virtues and personal autonomy by encouraging citizens to engage in thoughtful and respectful discussions with one another. The deliberation strategy assumes that people can change their beliefs, better understand the issues at hand, and think more critically about the arguments and decisions being made. In addition, deliberation can help to create a more open, more aware, and inclusive public sphere where people from diverse backgrounds can come together and share their perspectives. By working together and with the help of

17 See Feinberg, J. (1989) *The moral limits of the criminal law: volume 3: harm to self.* Oxford University Press.

the right institutional design, they can find better solutions. This strategy leads to a greater exchange of ideas and a more robust deliberation process. On the other side, it is a more demanding and more costly strategy. It asks us to invest in acquiring information and then debating with others, often in a particular context, away from our typical environments (John et al. 2009). However, "these costs (of time, effort, and opportunity) are indeed a practical constraint on democratic deliberation but they are not as prohibitively costly or devoid of proven means of remediation as to warrant the retrenchment of bounded human rationality (for the many) and the further empowerment of a technocratic elite few" (Button, 2018, p. 1040). This tells us that it is expensive and challenging to make decisions democratically, but it is still possible and worth doing.

Nudge is about affecting individual choices, like in classical economics. Deliberation, by definition, cannot happen alone, even though individuals have a significant role in the process. In the first strategy, state action is focused on getting the messages right and providing low-level incentives to get the desired behavior, while the role of the policy-maker is to be an expert who designs interventions that achieve these goals (John et al., 2009). The deliberative strategy requires the state to provide institutions to help citizens deliberate and follow up on emerging recommendations. Despite being a more expensive and demanding strategy, deliberation is, as we have seen, more fruitful in the long term for developing epistemic virtues, critical thinking, and strengthening personal autonomy. On the other hand, although costless, quickest, and effortless, the nudge strategy leads, in the long run, to epistemic vice such as epistemic laziness and loss of autonomy. In other words, even though nudging may get people to change their behavior in the short term, it could have unfavorable effects in the long term by making people less capable of making their own decisions. Policy-makers are the most competent persons whose role is to develop the best course of action and design interventions to achieve goals. This asymmetric relation between what people know and what the state knows, gives power to the state to make decisions for people. In that sense, state manipulation of citizens' choices appears to be at odds with the democratic ideals of free exercise of choice, deliberation, and public dialogue. On the other side, although the deliberative approach does not deny that expertise is relevant and that not everyone in the political arena is equal in knowledge, it emphasizes that the thesis of limited rationality and cognitive biases applies to an expert's mind as well.

We already said that according to the argumentative theory of reasoning, group decision-making could compensate for individual decision-

making, judging, and reasoning limitations. Their model indicates that during public deliberation, when discussing diverse opinions, group reasoning outweighs the individual, no matter who that individual is. Deliberative democracy is an effective way of giving ordinary citizens a say in the political process. Studies have shown that when citizens are allowed to become better informed through public deliberation[18], they can genuinely contribute to finding solutions. Thus, according to this approach, it is questionable whether the experts alone are the best avenue for attaining knowledge and solving problems. Group deliberation between diverse people can (under certain conditions) improve the reliability of individual judgments due to the combination of different perspectives, interpretations, evidence, experiences, and the like (Page, 2008; Landemore, 2013). When individuals with different perspectives and knowledge come together to deliberate, they can learn from one another and generate new insights. This cognitive diversity can lead to more effective decision-making and a deeper understanding of the issue. In addition, cognitive diversity can help to challenge assumptions and uncover new perspectives. It means it can be a good tool for fighting epistemic vices and cognitive biases and developing epistemic virtues. This strategy can be especially beneficial when addressing complex problems that require creative thinking. By bringing together individuals with different ways of thinking, the chances of finding innovative solutions increase. It can lead to more meaningful exchanges of ideas and, ultimately, better decisions.

3.2.1. Is nudging ever desirable strategy for changing civic behavior?

We showed in previous sections why the nudges strategy, used in the public space, is generally at odds with democratic values, principles, and practice, as well as why they carry the risk of epistemic injustice, violation of personal autonomy, and, in the long run, the development of epistemic vices. On the other side, people are not perfectly rational beings. We have limited time, information, and cognitive abilities. This is compounded by cognitive biases and mental shortcuts that often lead us astray and make irrational judgments. In other words, people often make suboptimal decisions, even when trying to make the best choice. Concerning all this, the role of the government in shaping our choices is complex and controver-

18 This is evident when comparing the pre-deliberative and post-deliberative survey results, as it is clear that in the deliberation process, citizens become more informed about the discussed political issues and more confident about their own political opinions (Fishkin & Lushkin, 2005)

sial. Some authors, like Thaller and Sunstein, argue that the government should have a role in shaping the choices of its citizens to secure better decision-making for individuals and society as a whole. They point out that the government can use nudges to influence its citizens' behavior in a less constraining way than passing a law requiring the behavior. Others argue that the government should not have a role in shaping the choices of its citizens. Nudges can be manipulative and take away people's autonomy to make their own choices, and it is not the role of the government to tell people what to do. The government should only provide information and education so people can make informed decisions.

However, there are situations where quick and efficient solutions are needed. If by the autonomous person, we understand a sovereign individual that includes all the activities that she can do *without negatively impacting others*, but also one who can reason correctly and has the ability to make rational decisions and responsibility (toward himself and others).[19] In that case, the nudge strategy can be justified when these abilities are severely restricted and the losses enormous. In times of *crisis*, when fear and uncertainty rise, the problems caused by cognitive biases, uncritical thinking, and misinformation become more significant and more potent than usual. When a crisis occurs, people are more exposed to pseudoscience information, fake news, and unverified content (this is especially true due to social media presence). The paradigmatic example is the recent COVID crisis (climate crises can also be a good example). The COVID-19 pandemic was full of false claims, half-backed conspiracy theories, and pseudoscientific therapies regarding the virus's diagnosis, treatment, prevention, origin, and spread. Fake news was widespread on social media. Some were benign, while others were very harmful and dangerous for the individual and the people around him. That put public health and human lives at risk. Because of this, it may be necessary to take protective measures in a crisis. Along with the protective measures suggested by the government and state institutions, nudging could be a powerful tool.

As a regular strategy for changing human behavior, nudging risks epistemic injustice, violation of personal autonomy, and, in the long run, the development of epistemic vices. In a crisis, there is no "in the long run" and thus no fear of developing epistemic laziness. Also, the nudging can still be respectful of a person's autonomy, even if it is not in line

19 For more about what autonomous person implies see Croce, M. (2020). "Epistemic Paternalism, Personal Sovereignty, and One's Own Good" in *Epistemic Paternalism: Conceptions, Justifications and Implications.*

with their current choices or beliefs. It can act as a reminder or prompt to help them better reflect on their desired choices or commitments to themselves and others, especially when cognitive biases, uncritical thinking, and misinformation flourish. As long as the *nudgers* (in this case, healthcare workers, scientists, and governments) genuinely believe that the nudge will serve human well-being, not have any reason to doubt its effectiveness, work for the public and not for private interests, the presumption is that the nudge is morally permissible. The justification for such a practice rests on a comparison of overall benefits to the loss of autonomy (in the sense of respecting the right of individuals to make their own decisions, e.g., not to wear masks, not to wash hands, not to get vaccinated) and not, as Thaller and Sunstein suggest, on the view that nudges are costless. Although shaping people's choices for their own benefit seems to be alarmingly intrusive and undermines a person's individuality, it seems too much to argue that this kind of intervention is not permissible in cases where someone needs to be denied access to (or not allowed to act on) misleading or harmful information, (Ahlstrom-Vij, 2013, p. 89). It does not exclude the practice of public deliberation and promotion of sound judgment and rational decision-making that would lead, in the long run, to increased trust in science and relevant expertise and enable a better response from the state and citizens in the next crisis and everyday decisions. The need to foster and secure the conditions for widespread critical reflection, both individually and collectively, is always necessary. It just means that sometimes we have no time and need an immediate solution, so forced to promote the desired behavior at a lower cost. Sometimes biases are powerful, and stakes are high and dangerous, so using rational persuasion to influence people's behavior may not give the desired results immediately. When there are strong reasons to get citizens to behave in a certain way, shaping people's choices may be more efficient and less constraining than limiting what they can choose by introducing mandatory law requiring specific behavior, for example. Nudge can sometimes be a good strategy that does not always violate autonomy. However, as we saw, it is very problematic as a policy that we should generally declare morally desirable.

References:

Ahlstrom-Vij, K. (2013). *Epistemic Paternalism: A Defence*. Basingstoke: Palgrave Macmillan.

Bovens, L. (2009). „The ethics of nudge", *Preference change,* 207–219, Springer, Dordrecht.

Brennan J. (2011). „The Right to a Competent Electorate", *Philosophical Quarterly* 61(245), 700–24.

Button, M. E. (2018). "Bounded rationality without bounded democracy: Nudges, democratic citizenship, and pathways for building civic capacity", *Perspectives on politics*, *16*(4), 1034–1052.

Cassam, Q. (2016). „Vice epistemology", *The Monist*, *99*(2), 159–180.

Cassam, Q. (2019). *Vices of the Mind: From the Intellectual to the Political*, Oxford: Oxford University Press

Christman, J. (2005), 'Autonomy, Self-Knowledge, and Liberal Legitimacy', in J. Christman and J. Andersen (eds), Autonomy and the Challenges to Liberalism, New York: Cambridge University Press, 330–57.

Cohen, J. (1989). "Deliberation and Democratic Legitimacy", A. Hamlin and P. Pettit (eds), *The Good Polity: Normative Analysis of the State*, Oxford: Blackwell, 17–35.

Cohen, Joshua (1986) "An Epistemic Conception of Democracy", *Ethics* 97 (1), 26–38.

Cooke, M. (2018). "Five arguments for deliberative democracy", *Democracy as public deliberation* (pp. 53–87). Routledge.

Debus, D. (2016). 'Shaping Our Mental Lives: On the Possibility of Mental Self-Regulation', Proceedings of the Aristotelian Society, 116: 341–65.

Delli Carpini M.X. and Keeter S. (1996). *What Americans Know about Politics and Why It Matters.* New Haven, CT: Yale University Press.

Dryzek, J. S. (2009). „Democratization as deliberative capacity building", *Comparative political studies*, *42*(11), 1379–1402.

Elster, J. (1998). "Introduction", in J. Elster (ed.), *Deliberative Democracy, Cambridge: Cambridge University Press*, 1–19.

Elstub, S. (2008). "Deliberative Democracy and Autonomous Decision-Making" in *Towards a Deliberative and Associational Democracy* (pp. 58–97). Edinburgh University Press.

Estlund, D. (2009). "Democratic authority", in *Democratic Authority*. Princeton University Press.

Fishkin, James, and Robert C. Luskin (2005). "Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion*", Acta Politica* 40: 284–298.

Friedman J. (1998) „Introduction: Public Ignorance and Democratic Theory", *Critical Review* 12(4), 397–411.

Gerber, Marlène; Bächtiger, André; Shikano, Susumu; Reber Simon; Rohr Samuel (2018), "Deliberative Abilities and Influence in a Transnational Deliberative Poll (EuroPolis)", *British Journal of Political Science* 48(4): 1093–1118.

Goldman, A. I. (1978) „Epistemics: The regulative theory of cognition", *The Journal of Philosophy*,*75*(10), 509–523.

Hansen, P. G., & Jespersen, A. M. (2013). „Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy", *European Journal of Risk Regulation*, *4*(1), 3–28.

Hausman, D. M., & Welch, B. (2010). „Debate: To nudge or not to nudge", *Journal of Political Philosophy*,*18*(1), 123–136.

John, P., Smith, G., & Stoker, G. (2009). "Nudge nudge, think think: Two strategies for changing civic behavior", *The Political Quarterly*, *80*(3), 361–370.

Kahneman, D. (2011). *Thinking, fast and slow*, New York: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). J*udgment under uncertainty: Heuristics and biases,* Oxford University Press.

Kidd, I. J., Battaly, H., & Cassam, Q. (Eds.). (2020). *Vice epistemology*. Routledge.

Kidd, Ian James. (2017). "Capital Epistemic Vices", *Social Epistemology Review and Reply Collective*, 6 (8), 11–16.

Landemore, H. (2012) *Democratic reason: Politics, collective intelligence, and the rule of the many*, Princeton University Press.

List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2012) "Deliberation, single– peakedness, and the possibility of meaningful democracy: evidence from deliberative polls", *The Journal of Politics*, 75(1), 80–95.

Meehan, D. (2020). „Epistemic vice and epistemic nudging: a solution?", *Epistemic paternalism: Conceptions, justifications and implications*, 247–259.

Mercier, H., & Landemore, H. (2012). "Reasoning is for arguing: Understanding the successes and failures of deliberation", *Political psychology*, *33*(2), 243–258.

Mercier, H., & Sperber, D. (2011) "Why do humans reason? Arguments for an argumentative theory", *Behavioral and brain sciences*, 34(2), 57–74.

Mill, J. S. (1998). *On liberty and other essays*. Oxford University Press, USA.

Mitchell, G. (2004). Libertarian paternalism is an oxymoron. Northwestern University Law Review 99(3): 1245–77.

Pritchard, D. (2013). "Epistemic paternalism and epistemic value", *Philosophical Inquiries 1(2),* 9–37.

Rawls, J. (1971). A Theory of Justice, Cambridge, MA: Harvard University Press.

Raz, J. (1986). *The Morality of Freedom*, Oxford: Clarendon Press.

Richardson, H. (2002). *Democratic Autonomy: Public Reasoning About the Ends of Policy*, Oxford: Oxford University Press.

Riley, E. (2017). „The beneficent nudge program and epistemic injustice", *Ethical Theory and Moral Practice*, *20*(3), 597–616.

Simon, H. A. (1990). „Bounded rationality", *Utility and probability* (str. 15–18). Palgrave Macmillan, London.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). „Epistemic vigilance", *Mind & language*, *25*(4), 359–393.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press; New Haven.

Tversky, A. & Kahneman, D. (1988). "Rational choice and the framing of decisions", *Decision.making: Descriptive, normative, and prescriptive interactions* (Bell, D. E. Raiffa H. & Tversky,. A. eds.). Cambridge University Press. 167-–192.

Tversky, A., & Kahneman, D. (1974). „Judgment under uncertainty: Heuristics and biases", *Science*, 185(4157), 1124–1131.

Warren, M. E., & Pearse, H. (eds) (2008). *Designing deliberative democracy: The British Columbia citizens' assembly,* Cambridge University Press

Zagzebski, L. (1997). "Virtue in ethics and epistemology", in *Proceedings of the American Catholic Philosophical Association*, Vol. 71, 1–17.

Ivana Janković

Usmeravanje ponašanja (nudge) i deliberacija: individualna autonomija, epistemičke mane i vrline

Apstrakt: Skorašnji nalazi o kognitivnim nedostacima i motivaciono-kognitivnim pristrasnostima u donošenju odluka i ponašanja danas su dobro potvrđeni. Postoje različiti pristupi rešavanju problema do kojih ovi nedostaci vode u kontekstu donošenja odluka. S obzirom na rastuću trend korišćenja bihejvioralnih nauka za informisanje odluka u domenu javnih politika širom sveta, ovaj rad ima za cilj da razmotri i kritički preispitaja dve strategija za poboljšanje ponašanja ljudi i donošenja odluka u javnoj sferi: usmeravanje ponašanja (navođenje) i deliberaciju. Da li kreatori javnih politika treba da razvijanju mehanizme za usmeravanje izbora svojih građana ili bi trebalo da ih podrže i podstaknu da sami zajednički dodju do boljih odluka? Da li vlade mogu da utiču na to da ljudi donose bolje odluke, a da u isto vreme ne naruše njihovu slobodu i autonomiju? Debata o „libertarijanskom paternalizmu" iznedrila je mnoga pitanja o mogućnosti pomirenja i objedinjenja ova dva pojma u jedan. Ova pozicija podrazumeva kreiranje javnih politika kroz specifičnu vrstu navođenja ljudi da donose bolje odluke vezane za zdravlje, bogatstvo i sreću, a da se time istovremeno ne ograniči njihova sloboda izbora. Razmotrićemo argumente za i protiv ove vrste intervencija i njene posledice za ličnu autonomiju i razvoj epistemičkih mana. Iako je evidentno da usmeravanje ponašanja zaista može da služi kao efikasno sredstvo za dolaženje do boljih odlu-

ka, tvrdićemo da je, u skladu sa osnovnim vrednostima savremenih demokratija, javna deliberacija dugoročno bolja strategija. Kada se primeni na javne politike, strategija usmeravanje ponašanja nosi sa sobom rizik da u potpunosti zanemari kapacitete građana i institucionalne prakse neophodne za plodno razvijanje demokratskog građanstva. S' druge strane, javna deliberacija može poboljšati donošenje odluka uspešnim rešavanjem kognitivnih nedostataka uz promovisanje građanskih vrlina bez kršenja slobode. Na kraju ćemo tvrditi da, bez obzira na ovaj opšti zaključak, strategija usmeravanja ponašanja može biti moralno dozvoljena u krizama, kada su neophodna brza i efikasna rešenja.

Ključne reči: usmeravanje ponašanja, deliberacija, donošenje odluka, deliberativa demokratija, individualna autonomija, epistemičke mane, epistemičke vrline

Miljan Vasić*

THE PROCEDURAL VALUE OF EPISTEMIC VIRTUES**

Abstract: The longstanding tension between the procedural and instrumental justification of democracy has been challenged by the theories that try to combine both approaches. These theories portray epistemic features of democracy in an instrumental framework and then try to reconcile them with procedural values. In this paper, I argue that it is possible to incorporate an epistemic dimension into a justification of democracy, without resorting to instrumentalism. On the view that I advance, Peircean epistemology, when combined with intrinsically valued epistemic virtues, constitutes a purely procedural argument for democracy.

Keywords: proceduralism, instrumentalism, democracy, epistemic democracy, epistemic virtues, pragmatism.

Epistemology and democracy are often thought of as being at odds with one another. Democracy tends to simultaneously embrace and violate certain epistemic ideals. Institutions of equality, free press, and fair elections should pave the way for well-informed, deliberate, and reasoned decision-making processes. Yet, the very same institutions can become a breeding ground for propaganda groups (Stanley, 2015), demagogues (Roberts-Miller, 2017), or strategically minded voters (Moser & Scheiner, 2009), who nullify the epistemic benefits of public opinion. However, even the authors who attempt to provide a non-epistemic justification of democracy (cf. Richardson, 2003; Shapiro, 2016) agree that epistemic ideals are an important factor in democratic legitimacy – or, at the very least,

* Institute for Philosophy, Faculty of Philosophy, University of Belgrade, miljan.vasic@f.bg.ac.rs

** The paper is based on research conducted within research project Man and Society in the Time of Crisis, financed by the Faculty of Philosophy, University of Belgrade.

they do not try to claim that uninformed and incompetent citizens will benefit democracy (Talisse, 2019).

Despite this apparent tension, a growing amount of literature suggests that the reconciliation between epistemology and democracy is both possible and desirable. My paper is another small step towards that goal. It is divided into three parts. In the *first* part, I describe the central debate in democratic theory, the one between proceduralism and instrumentalism. Each of these conceptions, taken in a strict sense, faces its own problems, which is why some authors opt for a middle way between them. I will present two such views: David Estlund's epistemic proceduralism (Estlund, 2008) and Elizabeth Anderson's experimentalist model (Anderson, 2006), as well as their critiques of these two opposing conceptions. I will claim that, although most of their criticisms are in place, neither Estlund nor Anderson manages to successfully mediate between the two opposed lines of argumentation, and that both of their accounts are leaning towards instrumentalism. In the *second* part of the paper, I will present a tripartite argument against instrumentalism which (more or less) directly affects Estlund's and Anderson's views as well. Here, I will draw heavily on the criticisms put forward by Fabienne Peter (2008). The purpose of this argument is to show that the epistemic benefits of democracy, when placed inside an instrumentalist framework, are either untenable or become a dangerous tool for anti-democratic arguments. In the *third* part, I will present a procedural way of incorporating the epistemic dimension into a justification of democracy. This view combines three theoretical approaches: Peter's pure epistemic proceduralism, Peircean epistemology, and the idea of intrinsically valuable epistemic virtues.

1. Walking a Tightrope: Between Proceduralism and Instrumentalism

Main results of the social choice theory (Arrow, 1963), and the subsequent interpretations that these results show that collective decision-making is essentially devoid of meaning (Riker, 1982) posed a challenge to democratic theorists of all persuasions. Ever since these results appeared in the literature, advocates of democracy have been looking for suitable counterarguments. The current trends of deliberative and epistemic theories of democracy emerged partially as a response to this challenge. Two distinct lines of argumentation became prominent in the literature. Authors who start from the central assumption of social choice theory – that voting is the expression of individual *preferences* – have several strate-

gies at their disposal: they can try to weaken the rationality conditions presupposed by Arrow's impossibility theorem (Black, 1998 [1958], pp. 363–367); they can use the results of descriptive social choice theory to show that these negative consequences are rarely (if ever) practically realized (Regenwetter, et al., 2009); or, they can advance the view that voting should be replaced or complemented by public deliberation (Cohen, 1989).[1] Other authors chose to abandon this starting assumption, and endorse the view that, in democratic decision-making, citizens express their *beliefs* about which option is the best one, according to some procedure-independent criterion (Cohen, 1986; Goodin & Spiekermann, 2018).

These two approaches shaped the debate between two different ways of justifying democracy – procedural and instrumental. While authors on both sides of this debate claim that democratic governments are superior to non-democratic ones, they disagree about why this is so. According to proceduralism, the political outcomes are supposed to be *fair*, while in the instrumentalist view they ought to be *right* (List & Goodin, 2001). The central claim behind the proceduralist approach is that democracy is justified intrinsically, that is, without appealing to any external criteria by which we judge political outcomes. Substantive claims are to be made about procedures themselves, rather than their outcomes. If the procedural standards are met, any possible outcome is equally desirable – at least according to proceduralism. Instrumentalism claims the opposite: an advantageous feature of democracy is that it shows a tendency to produce better outcomes when compared to alternative ways of political decision-making. However, whether we should consider the outcome to be good or bad depends on the external standards which are independent of the decision-making procedure.[2]

---

1 It is, of course, possible to combine all of these approaches (e.g., Mackie, 2001, 2003, 2011).

2 A short note on terminology. The same central distinction is sometimes made between *epistemic* and *procedural* justification of democracy (cf. List & Goodin, 2001). Although epistemic democracy is traditionally framed in instrumental terms (Schwartzberg, 2015), I believe that equating "instrumentalism" with "epistemic democracy" is not feasible, especially after Estlund's theory of "epistemic proceduralism" gained prominence. Such terminology is even more misleading when we take into account positions like Peter's, which are resolutely non-instrumental, yet epistemic at the same time. For this reason, some recent papers (cf. Fuerstein, 2019) refer to these opposed lines of argumentation as "pure epistemic" and "pure procedural" conceptions of democracy. Instrumentalism, for its part, sometimes entails a *broader* conception of democracy, where both democratic procedures and some additional institutions contribute to the correctness of outcomes (Mladenović, 2020, p. 4). The upshot is that neither every epistemic theory of democracy has to be instrumental, nor do instrumental approaches have to appeal to epistemic standards. In this paper,

However, the two views are not as mutually exclusive as they may seem. Many ways of justifying democracy include both procedural and instrumental merits, although in different proportions. In fact, theories of democracy form a spectrum in terms of epistemic demands that they put before the citizens.[3] On one end of this spectrum are the theories that attach little or no importance to the correctness of individual judgments. Here, the social choice theory is a prime example. On the opposite end of the spectrum are Condorcetian theories (e.g., Goodin & Spiekermann, 2018) and what Estlund calls correctness theories. These theories claim that democracy is a mostly or completely reliable method of arriving at the correct outcomes, and they count on the epistemic prowess of individual citizens (Kelly, 2012).

Between those opposing ends are various theories that lean towards the procedural or instrumental side of the spectrum, without being fully committed to either approach. Some theorists, however, claim that neither purely instrumental nor purely procedural justification of democracy is plausible and that it is possible to justify democracy on both grounds simultaneously. In this section of the paper, I will present two such theories, Estlund's epistemic proceduralism and Anderson's Deweyan model of democracy. My aim is to explore how each of these theories finds its way between the two opposing ends of the spectrum. Although I accept the idea that justification of democracy must include both procedural and epistemic components, I will conclude that, despite their promising and influential approach, neither of the two theories manages to position itself on the middle of this spectrum.

Estlund's Epistemic Proceduralism

Estlund presents the conflict between proceduralism and instrumentalism in the form of the Euthyphro dilemma: are good democratic decisions good because they are democratically made, or are they democratically made because they are good? Criticizing the first horn of the dilemma – a view that the value of democratic decisions lies in the mere fact that they are democratic – Estlund claims that a common feature of all forms of proceduralism is the "flight from substance" (Estlund, 2008,

---

by "instrumentalism" I mean those epistemic conceptions of democracy that presuppose a procedure-independent standard of correctness. This technically makes it a case of *epistemic* instrumentalism (Peter, 2016, p. 138); but since I will not take into account any broader notion of it, just "instrumentalism" will suffice.

3 Jamie Terrence Kelly (2012) is the first to propose the term "spectrum of epistemic demands", but the general idea was introduced by Estlund (1997, p. 182).

p. 65), i.e., the idea that democratic procedures have an intrinsic value, independent of any substantive standards that lie beyond the procedure itself. Estlund distinguishes between three variants of this view – fair proceduralism, normative social choice theory, and (procedural account of) deliberative democracy – and offers a series of criticisms aimed to show why such a flight is impossible.

Firstly, Estlund is critical of insistence on majority rule as a *fair procedure* that gives all citizens an equal right to vote. If that is enough to make a procedure fair, he claims, it would be equally fair to choose between the possible options by tossing a coin, since that procedure would also give everyone the same amount of say (Estlund, 2008, p. 82). Estlund, however, points out that this is an absurd proposal in a political context; and one that would hardly be acceptable to the theorists of proceduralism who would nevertheless insist that voting is a preferable way to make decisions. But in that case, their position no longer flees from substance, as they acknowledge that fair proceduralism must include some non-procedural values that make voting fundamentally different from random selection (Estlund, 2008, pp. 82–83). Estlund admits that this is a very "thin" version of proceduralism that is easy to refute, but he finds it important to immediately lay bare what a theory that completely rejects procedure-independent standards look like (Estlund, 2008, p. 83).

Estlund subjects the *social choice theory* to a similar line of criticism. The crucial aspect of this theory is what Estlund calls the condition of aggregativity: if a collection of individual preferences (understood in a broad sense as a set of ends, aims, or choices) leads to some procedural outcome, then individual changes in preferences should result in a change of the final outcome (Estlund, 2008, p. 73). It is intuitively clear how majority voting fulfills the condition of aggregativity – if some citizens had voted differently, the outcome could have been different. Random selection, on the other hand, violates this condition. Estlund, however, makes an interesting point here: the aggregativity condition says nothing about whether the correlation between individual preferences and the final outcome should be positive or negative. If we imagine a situation in which an option gains popularity among individual voters, but then scores poorer in the elections, the condition of aggregativity is still met. Moreover, the normative social choice theory is often aimed precisely at studying and interpreting such cases. Claiming that the correlation *should* be positive requires additional non-procedural reasons. Thus, Estlund concludes, although nominally focused on procedural conditions, the social choice theory includes additional substantive standards independent of the procedure itself (Estlund, 2008, pp. 74–75).

When criticizing the deliberative theory of democracy, Estlund combines the previous two arguments. He distinguishes between two proceduralist forms of deliberative democracy: deep deliberative democracy and fair deliberative proceduralism.[4] *Deep deliberative democracy* arises with the rejection of the basic assumptions of social choice theory. Social choice theory revolves around the idea that it is possible to aggregate individual preferences into a coherent choice of an entire group but ignores the fact that preferences themselves can arise as a product of false information and manipulation. Therefore, deep deliberative democracy focuses on idealized hypothetical procedures of public deliberation, but it (presumably) rejects any standards independent from the procedure, just as social choice theory does (Estlund, 2008, p. 88). However, as Estlund believes, this theory runs into the same problem as a social choice theory – by asserting that there is a correspondence between outcomes and ideally understood individual interests, deep deliberative democracy cannot avoid invoking any substantive standard (Estlund, 2008, p. 92). *Fair deliberative proceduralism* (which Estlund considers an unstable hybrid theory), in turn, rests on the view that the advantage of public deliberation is that it allows a large number of people to express their views, whatever they may be. This theory claims that it puts no emphasis on the epistemic value of deliberation, just on a fair representation of citizens' views. Estlund, however, considers this claim to be unsustainable: if the purpose of deliberation is to transform brute preferences into informed ones, then it plays an epistemic role nevertheless (Estlund, 2008, p. 94). Although this is *a different kind* of epistemic role when compared with theories that claim that deliberation allows better outcomes (from a procedure-independent point of view), Estlund still regards this as a substantive claim. Otherwise, fair deliberative proceduralism, just like its non-deliberative counterpart, could not explain why deliberation is a superior tool for decision-making when compared with randomly chosen outcomes (Estlund, 2008, pp. 94–95). Therefore, in Estlund's view, all standard forms of proceduralism are incoherent. Although proceduralism calls for a flight from substance, procedural accounts of democracy either explicitly make substantive claims or cannot explain why democracy does better than a coin toss without resorting to substantive claims.[5]

4 Although the deliberative theory emphasizes the social process by which individual attitudes are formed (in contrast to the social choice theory which takes preferences as simply given), Estlund posits that it is still strictly proceduralist in its original forms (Estlund, 2008, p. 87).

5 The continuing problem with Estlund's analysis of proceduralism is his failure to distinguish between two different notions of proceduralism. Ivan Mladenović argues that proceduralism can be understood in a narrower and wider sense. A narrow sense of proceduralism deals with the normative conditions that a democratic decision-

The second horn of Eutyphro dilemma presupposes non-proceduralist epistemic approaches to democracy which Estlund calls *correctness theories*. According to them, political decisions are legitimate only if they are correct by some procedure-independent standard, and democratic procedures are considered sufficiently accurate to make collective decisions correct (Estlund, 2008, p. 102). Here the *locus classicus* is Rousseau's notion of general will. For Rousseau, outcomes are legitimate because they are correct – and when they are incorrect, they are illegitimate – but this legitimacy has nothing to do with any procedural reason; it is the general will that gives legitimacy to political decisions. Estlund does not object to Rousseau's view that the outcomes should be obeyed. What he finds problematic in Rousseau's theory is the claim that those who are in minority must admit that they were wrong and that their notion of general will was a faulty one. This is what Estlund calls "the problem of deference", and it is his main reason for rejecting correctness theories (Estlund, 2008, pp. 103–104).

Estlund introduces epistemic proceduralism as an alternative to both (purely) procedural and instrumental justification of democracy. It is the theory that combines some elements of both lines of argumentation. Epistemic proceduralism is epistemic since it asserts that democracy tends to produce correct decisions (Estlund, 2008, p. 8); but it is at the same time procedural since it claims that legitimacy (its coercive power) and authoritativeness (its moral commitments) of democracy stems from the fact that democratic procedures are acceptable to all qualified points of view (Estlund, 2008, pp. 41–42). Correctness theories have a too strong epistemic claim: every legitimate decision must be correct. According to epistemic proceduralism, the outcome is legitimate even if it is incorrect, given that procedural reasons are met.

Epistemic proceduralism does not face the problem of deference, since it does not claim that the democratic outcome constitutes a rea-

making procedure should satisfy to be considered justified. The broader understanding of proceduralism takes a decision-making procedure *in a general sense* as the basis of the justification of democracy (Mladenović, 2019, pp. 166–167). Some authors, like Riker (1982), may use procedural claims in a narrow sense to draw broader conclusions about the procedural justification of democracy, but we must keep in mind that two senses of proceduralism deal with different normative problems. With that distinction in mind, we can see that substantive standards which Estlund attributes to proceduralists are procedural claims in a narrow sense which he interprets as a broader claim about the justification of democracy (Mladenović, 2019, p. 175). Furthermore, Estlund thereby undermines his own position: if epistemic proceduralism is, as Estlund claims, a form of proceduralism, then it too must deal with normative conditions that the decision-making procedure must satisfy (Mladenović, 2019, pp. 179–180).

son for a belief about the correctness of the said outcome. According to Estlund, democracy gives its citizens *moral* reasons to comply instead of epistemic reasons to believe (Estlund, 2008, p. 106). Thus, epistemic proceduralism can generate more legitimacy with less demanding epistemic claims. It is important to note that epistemic proceduralism differs from fair proceduralism only in cases where *there are* independent moral standards (more on this later) according to which some outcome is correct. In such cases, epistemic proceduralism is the view that democracy can be procedurally impartial among citizens' opinions, and tend to produce correct decisions at a better-than-random rate (Estlund, 2008, p. 107–108). Therefore, Estlund believes that epistemic proceduralism occupies a perfect place between the theories that are not epistemic enough (since they ignore moral standards even when they should not be ignored), and those that are too epistemic (and thus face the problem of deference) (Estlund, 2008, p. 102).

Anderson's Experimentalist Model

Just like Estlund, Anderson is not satisfied with the prevailing dichotomy between proceduralism and instrumentalism. She claims that such dichotomy is neither desirable nor plausible. Proceduralism, in her view, merely requires that the decision-making process is fair for democracy to be justified. However, if fairness is the only standard we should adhere to, we cannot draw a meaningful difference between a coin flip and other decision-making procedures. Justification of democracy needs more than that: we believe that citizens confer legitimacy to a certain decision with their very participation in decision-making, thus proclaiming that a given problem is of public interest. However, we are thereby complying with external criteria (Anderson, 2006, pp. 9–10). On the other hand, whether a particular problem is of public interest or not becomes clear only when citizens (or their representatives) put it into consideration through procedurally fair decision-making, which is determined internally. So, as Anderson concludes, an adequate conception of democracy must include both internal and external criteria (Anderson, 2006, p. 10).

Anderson views democracy as a process of collective problem-solving, the success of which depends on the criteria stated above. Additionally, the satisfactory model of democracy has to incorporate three constitutive features of democracy. It needs to take advantage of the epistemic *diversity* of individuals; it must model *discussion* as an epistemically productive process; and it must be *dynamic*, which means that it must provide feedback mechanisms for improving its epistemic results; I call this the DDD

conception of democracy. Anderson aims to determine the most adequate account of democracy by comparing three competing models of epistemic democracy: Condorcet's jury theorem (CJT), the diversity-trumps-ability (DTA) model by Lu Hong and Scott Page, and Dewey's experimentalist model.

CJT fails to satisfy any of the three components of DDD conception. Firstly, the main result of CJT holds regardless of the group's internal diversity. What is more, Condorcet's original formulation of the theorem (1976 [1785]) presupposes homogenous groups. While having diverse groups will not necessarily *harm* the theorem's optimistic result (Grofman, et al., 1983), diversity plays no role in CJT whatsoever. Secondly, CJT supposes that group members vote independently of one another. As with diversity, there are extensions of CJT that claim that pre-voting discussion is not necessarily harmful to the theorem (Goodin & Spiekermann, 2018, pp. 67–73). However, Anderson rightly points out that it is unclear whether CJT holds under the actual modern democratic practices, where free press and public discussions are constitutive, and not merely accidental features of democracy. Lastly, CJT cannot capture the dynamic features of democracy. Since this model suggests that the majority of voters are (nearly) infallible from the start, there is no need to revise any of the previous decisions (Anderson, 2006, pp. 11–12).

Things look brighter with the DTA model. In Hong and Page's (2004) computational experiment, diversity plays a crucial role in collective problem-solving. Additionally, Anderson claims that this model presents discussion, not as a hindrance, but as an epistemically productive factor. She finds this approach much more promising in comparison to the limited assumptions of CJT. The DTA model explicitly states that the problems agents are trying to solve are complex, which is important for the democratic interpretation of the model. As Anderson notes, one of the shortcomings of autocratic regimes is that they can solve only the simplest of problems – like catching a murderer – but perform far worse than democratic governments when the problem is politically complex. Nevertheless, the DTA model cannot comply with the last element of DDD conception – dynamics. Just like CJT, the model does not support any feedback mechanisms that could alter the decisions regardless of their consequences (Anderson, 2006, pp. 12–13).

Anderson claims that the only model of democracy which succeeds in grasping all three components of the DDD conception is Deweyan experimentalist account of democracy. According to this view, deliberation should be conceived as a type of thought experiment which aims to predict the consequences of implementing proposed solutions (Anderson,

2006, p. 13; Dewey, 1922). When citizens reach a decision, they act upon the agreed solution to see its actual consequences. If the results are unfavorable (e.g., the problem was not solved, or its solution produced some additional problems that could not be foreseen), the implemented solution is refuted, as in science, and the problem-solving process returns to the deliberative phase. Anderson asserts that Deweyan model is the only one that manages to provide a satisfactory feedback mechanism: its dynamism encourages the institutions of regular elections, free media, petitions, protests, and public reaction to proposed legislation. Moreover, this model envisages that diversity and discussion are fostered through institutions of civil society where members of certain social groups can work collectively to address common concerns (cf. Dewey, 1946, pp. 206–210) and those institutions are parties and civic associations (Anderson, 2006, p. 14). The Deweyan account is, therefore, a model of democracy that, according to Anderson, manages to fulfill both procedural and instrumental criteria.

The Persisting Dichotomy

Previously, I have introduced a view that different theories of democracy can be presented as particular positions on the spectrum of epistemic demands. On this view, the social choice theory is the least epistemically demanding account of democracy, while the correctness theories sit on the other end of the spectrum. The question is not whether one can endorse a theory that is positioned close to the middle of this spectrum. Such theories are well known (two of them are outlined above) and, I believe, are more plausible than any of the "extremes". The real issue is whether any theory of democracy can claim that it has found its place *precisely* on the middle of this spectrum.

However, two now-classical theories that I have presented claim exactly that. Despite the differences in their approaches, Estlund's and Anderson's views share some general assumptions. As both authors remark, the accounts of democracy on either end of the spectrum are unsatisfactory, which is why they try to reconcile the two approaches, and I consider that a remarkable endeavor. But they consequently endorse the view that the dichotomy between proceduralism and instrumentalism cannot be sustained, and this is where I disagree.

A justification of democracy that combines procedural and epistemic elements is superior to those which are, as Peter puts it, *monistic*. The instrumentalist approaches that reduce democratic legitimacy to a single dimension of correctness are monistic; but equally monistic are the proce-

duralist views that reduce democratic legitimacy to the dimension of political fairness (Peter, 2008, p. 35). Neither Estlund's nor Anderson's theory is monistic in this sense, which makes them much more appealing. Hence, I accept that both procedural and epistemic components are necessary for a robust justification of democracy. Nevertheless, I believe that any particular line of justification *must* fall into one of the two general categories.

To put it bluntly, I do not believe that it is possible to walk this tightrope without falling to either side of the chasm. The threshold is this: a justification of democracy either makes an appeal to procedure-independent standards or it does not. If it does, it is instrumental – even though its instrumentalism may vary in degrees. Estlund believes that the "gray area" between proceduralism and instrumentalism is big enough to fit an entire theory there; but any theory that accepts that democratic outcomes should be (fully or partially) judged according to some external standard of correctness is, at its crux, instrumental. Thus, despite being labeled "proceduralism", I consider Estlund's theory to be a moderate form of instrumentalism.[6] For Estlund, democracy is still (at least in certain cases) a truth-tracking process.

In Anderson's view, democracy is not a truth-tracking, but a problem-solving process. On the Deweyan account that she endorses, we cannot know things as they exist independently of our inquiry. This view is thus not veritistic, unlike Estlund's. But it is, nevertheless, equally consequentialist, since it assumes some shared goals that direct the problem-solving process and gives judgment about the consequences of different proposals (Peter, 2008, pp. 42–45). Thus, Anderson's theory is also a form of weak instrumentalism since democratic outcomes are subject to external criteria of evaluation.

Even if, as Anderson believes, every internalist justification of democracy must include some external criteria and *vice versa*, it does not follow that dichotomy itself is non-existing or misleading. Justifications of democracy can include internal and external criteria in varying degrees, in such a way that every particular theory is either (slightly) more procedural or instrumental. The latter is the case with Estlund's and Anderson's theories. I consider this their biggest weakness, which I will address in the second part.

6 Of course, there are commenters (cf. Prijić Samaržija, 2020, p. 58) who claim that, despite being called "epistemic", epistemic proceduralism is still a form of proceduralism. While I disagree with that particular verdict, such comments nevertheless show just how close Estlund's theory is to the demarcation line between procedural and instrumental accounts. But it also speaks in favor of my view that the middle ground is practically unattainable.

2. A Gorgian Argument against Instrumentalism

In this section of the paper, I present what I call a *Gorgian argument*[7] against instrumentalism. It is not a conclusive argument by any means; but it is a set of objections that, when taken together, I consider sufficient for rejecting epistemic instrumentalism. This sort of argument is primarily aimed against fully-fledged instrumentalists like Robert Goodin and Kai Spiekermann (2018) who defend epistemic democracy in the CJT framework, or Hélène Landemore (2013) who makes use of a wider variety of such models. However, I believe that it consequently compromises Estlund's and Anderson's positions, since they are committed to the very same basic assumptions. The argument goes as follows:

1. There is no procedure-independent standard of correctness.
2. Even if there is such a standard, we cannot distinguish between correct and incorrect decisions.
3. Even if we can make that distinction, what decisions are correct is known only by a select few.

1) In denying the existence of any independent standard of correctness, one ostensibly invokes the old objection that has its roots in Hume's sharp division between facts and values (2011 [1748]). According to this notion, value judgments (that form the core of religious, aesthetical, ethical, and, presumably, political views) cannot be true in the same sense as scientific facts are. In the context of democratic decision-making, the very same objection was raised by social choice theorists (Black, 1998 [1958], p. 196; Little, 1952, p. 427). Delving too deep into this problem exceeds the aims and scope of this paper.[8] I believe, however, that one can reasonably deny that there is an independent standard without full commitment to this dichotomy.

While commenting on the fact/value dichotomy (2018, pp. 39–42), Goodin and Spiekermann introduce the idea of *moral majoritarianism* (which they tend to reject). According to this view, political statements can neither be true nor false according to any external standard. Instead,

7 Ancient Greek sophist Gorgias famously defended skepticism with a tripartite argument. He claimed that 1) nothing exists; 2) even if something exists, it cannot be comprehended by anybody; 3) even if it is comprehensible, it is still incapable of being communicated to others (DK 82B3). By "Gorgian" here I mean the argument which shares the general form with this type of reasoning. Even if we accept some questionable statement to be true (for the sake of argument), it just opens another difficulty for the position we are arguing against. Thus, to defend against criticisms, one must address all of these objections concurrently.

8 See Putnam (2002) for an exhaustive discussion.

different people hold different values within the community, and thus different answers are "correct" from these different points of view. If a solution is to be adopted for the community, this view holds that from a democratic standpoint it should be the one that is "correct" from the perspective of the larger segment of the people (Goodin & Spiekermann, 2018, pp. 41–42). While I personally find this outlook promising[9], Goodin and Spiekermann are right to claim that it is not very suitable within the CJT framework. Instead, they simply opt for views like moral realism or moral conventionalism. This is tantamount to what Peter calls "naïve instrumentalism", which is the assumption that there is a way of identifying the ideal outcome that does not require any democratic participation. Deweyan account of democracy, which emphasizes the constructive functions of democracy, does better in this regard, but Estlund's theory is on the brink of being naïvely instrumental too (Peter, 2008, p. 37).

Similar to Goodin and Spiekermann, Estlund asserts that *there are* true procedure-independent standards by which we judge political outcomes and claims that this position is very difficult to deny (2008, pp. 30–31).[10] Those who nevertheless claim that there are no (even minimal) standards, Estlund accuses of political nihilism. A nihilist stance is dangerous, he maintains, because it calls into question any kind of political activism (Estlund, p. 25–26). However, I believe that the morally majoritarian view does not entail political nihilism (nor the claim that value judgments cannot be true), while it still rejects procedure-independent standards. I think that procedurally inclined citizens may nonetheless be very politically active. For example, protestors who claim that an election was rigged are not necessarily appealing to any procedure-independent standard; their attitude about political life is far from nihilistic, and yet they may still reject any such standard.[11]

Let us, however, assume that naïve instrumentalism is right and that *there is* a procedure-independent standard of correctness. It is one thing

9 There is the obvious concern that this view might entail tyranny of the majority. However, most of the potential issues (like putting to vote who should have the right to vote, or depriving certain minority groups of their rights) can be countered on purely procedural grounds, for they undermine the very idea of political participation.

10 Here, Estlund introduces the idea of a "minimal" kind of moral truth, where "*x* is *F*" is true in at least the minimal sense if *x* is indeed *F* (2008, pp. 5; 25). Gerald Gaus objects that by this implicit committing to redundancy theory of truth, Estlund does not solve any problems, since this definition of truth is either too broad or too narrow for the purposes it is supposed to fulfill (2011, pp. 275–276).

11 This is another instance of Estlund failure to distinguish between two senses of proceduralism (see footnote 7).

to assert that such a standard exists *in principle*, but it is a completely different feat to put your finger on what it is. Whether this standard is conceived as general will (Cohen, 1986), public reason (Estlund, 2008), or truth (Prijić Samaržija, 2020), there is bound to be a disagreement about what outcomes are supposed to be considered "correct". Liberal argues that the priority is to avoid coercion; Marxist argues that the priority is to eliminate structural injustices in the economy; pacifist argues that correctness means avoiding war no matter what, etc. There is simply no way to apply a correctness standard without giving priority to a certain set of moral commitments, which themselves are legitimately contested (Fuerstein, 2019, p. 381).

For Peter, this is the biggest drawback of instrumentalism.[12] She claims that instrumentalism fails to respect Rawlsian "fact of reasonable pluralism" (Rawls, 1996, pp. 66–67). Instrumentalists, however, do not deny this fact – they choose to ignore it, hoping against hope that it will not come back to haunt them. Nonetheless, it matters not how deep the pluralism of values is; the real issue with instrumentalism is its inability to recognize that the respect of reasonable pluralism implies that people's possibility to participate in deciding between alternative social states is a constitutive part of democratic legitimacy (Peter, 2008, p. 36).[13]

Estlund's response to this type of objection is a strong one. By focusing on the list of "primary bads"[14] (an inversion of the Rawlsian theory of primary goods) Estlund reconciles the value pluralism with the general instrumentalist approach. He limits the situations where the correctness standard may be applied to a short list of (presumably) universally accepted beliefs. By defining the standard negatively, he avoids any positive claims about standards of justice or public goods, thereby respecting the pluralist argument (Estlund, 2008, p. 163–165).

That being said, I am not entirely convinced that Estlund's endeavor is successful. First, it is surprising that, after claiming that democratic legitimacy lies in the fact that democracy achieves correct answers at better

12 "Here is why I think that we ought to reject instrumentalism. First, I take it as a premise that the interests and perspectives of the members of the democratic constituency inevitably diverge and that they have different views – with good reasons – about what social state is best" (Peter, 2008, p. 36).

13 Peter also claims (and I agree) that Estlund's theory is not guilty of this misconception. He explicitly denies that a decision has to be correct in order to be legitimate (this is the feature of correctness theories). Thus, Estlund's theory can explain why procedures are constitutive for legitimacy (Peter, 2008, p. 40). It is Anderson's Deweyan approach that fails in this regard (Peter, 2008, p. 45).

14 The list goes as follows: "war, famine, economic collapse, political collapse, epidemic, and genocide." (Estlund, 2008, p. 163).

than random rate (Estlund, 2008, p. 98), Estlund admits that the core of the shared goals is simply avoiding bad outcomes (Gaus, 2011, p. 293). This is not a very encouraging outlook, since it tells us nothing about everyday political decisions that do not entail wars or famines; here we are supposed to fall back to purely procedural grounds (Estlund, 2008, p. 107). Second, even if such a list is possible it is bound to cause disagreement about a) which primary bads are to be included and b) avoidance of what bads has a lexical priority over avoidance of others. Estlund is aware that sometimes even primary bad must be allowed for the sake of avoiding even greater evil (2008, p. 163). This hierarchical relationship between primary bads opens up a second objection that I address in this argument.

2) Let us assume that there is a procedure-independent standard of correctness *and* that we can unambiguously determine what that standard is. How do we apply such a standard to judge the actual decisions and label them to be "correct" or "incorrect"? Since the correctness is determined externally, it cannot depend on the outcome. Yet, in one important aspect, I will argue, it must be either determined internally, or not determined at all, and both readings are undesirable from an instrumentalist point of view.

Whether we think of democracy as a (moral) truth-tracking (Estlund, 2008; Goodin & Spiekermann, 2018) or a collective problem-solving (Anderson, 2006; Landemore, 2013), we are obliged to claim that some options/decisions/solutions are better than others in that particular regard. The question is: *what others?* Is any conceivable option/decision/solution a subject of a comparison, or only those existing on the agenda?[15] To illustrate my point, I will use the following imaginary case of decision-making. Suppose the voters[16] are considering two options, *A* and *B*, where *B* is the correct one according to the independent standard. Now imagine that a third option is added to the agenda, but everything else remains the same. The new option, call it *C*, represents an even more superior candidate/policy/social state according to the very same standard. Would, in the revised scenario, option *B* still be the correct choice? There are two possibilities, and in my view both are untenable.

Possibility 1: B is the correct option in both scenarios. According to this view, the correctness of an option is completely exogenous, and cannot depend on the correctness of any other option. In this case, we must admit that there is some kind of "threshold of correctness" that options either

15 Similar objections are raised in: Gaus, 2011; Fuerstein, 2019.

16 For simplicity's sake, I will presume the case of majority voting; however, any standard decision-making procedure faces the same issue.

pass or not. There are two issues with such an interpretation. First, who determines where the threshold is? Suppose that, in terms of correctness, option *A* is as far from option *B* as option *B* from option *C*. What are the reasons for the threshold being placed between one pair of options rather than the other? It follows that if an independent standard determines the relative "correctness" of two options, it requires *yet another standard* for distinguishing correct options from those that are incorrect.

Second, what if *all* available options are above, or they are all below the threshold of correctness?[17] While the first case might sound like the best scenario ever, it faces us with the unpleasant Orwellian conclusion: all options are correct, but some are more correct than others. And if the threshold of correctness is comparably low, then instrumentalism is, well, *instrumental* only in a very limited number of cases. But wherever the threshold is, there are no guarantees that the correct answer will ever be present on the agenda. Thus, it is much likelier that we might face a pessimistic case, where all the options are incorrect. In that case, we must either accept that no good decision is available, or we are forced to treat the "least incorrect" option as "the correct one"[18]. But, if we follow that line, we are abandoning the first possibility.

Possibility 2: B was the correct option in the first scenario, but not in the second. In this case, the correctness of an option depends on the entire set of options it is compared with. In other words, correctness is agenda-sensitive and there is no exogenous threshold of correctness. Yet, this makes correctness the subject of *internal* evaluation. In that case, we can always visualize a scenario where another (slightly) more correct option is added to the agenda, rendering the previously correct option incorrect. If we follow this interpretation, we are forced to admit that, strictly speaking, there are no correct options.[19] Agenda-sensitive correctness thus gives rise to a sorites paradox that can only be resolved by a fixed threshold; but in

17 There is also a third undesirable possibility, and that is the case when there are multiple correct options, and only one (or significantly smaller number of) incorrect. This is especially troublesome for Condorcetian framework of Goodin and Spiekermann. In such case, the "competent" voters may spread their votes among the correct options, thereby allowing incompetent voters to win the day. The counterargument offered by the authors is that this scenario is unlikely since "there are usually a great many more ways to be wrong than to be right" (2018, p. 45), but they admit that this is not much of a solution.

18 Goodin and Spiekermann argue for this strategy: "[T]he object of the CJT exercise is not really one of finding the needle in the haystack of the 'one truly correct option out there in the world'. Rather, the object of the exercise is then to select the best alternative among the alternatives offered for choice." (2018, p. 44)

19 Perhaps the very usage of the terms "correct" and "incorrect" entails that *the* correctness, in the strict sense, is unattainable (cf. Unger, 1971).

that case, we are going back to the first possibility.[20] Thus, whatever the supposed procedure-independent standard of correctness is, it cannot be used to determine the (in)correctness of any particular decision. It matters not if we are talking about different policies, social states, or simply electoral candidates; it is also irrelevant what particular decision-making procedure we have in mind, be it voting or public deliberation – the result is the same: correctness is not a binary case. Instead, it comes in a gradable form; and with it come several tangled issues, none of which is satisfactorily resolved.

3) Even if we grant that there is a reliable method of resolving these issues *and* that there are indeed undeniably correct decisions, one glaring problem remains. What if the correctness of a decision is better discerned by a small number of citizens who possess the necessary knowledge? In that case, good outcomes will be achieved more reliably if we concede political decision-making to those citizens. This makes democracy inferior to "epistocracy", i.e., the rule of the experts. Epistocrats can (and do) use the general results of models like the CJT or the DTA and turn their results upside-down (Brennan, 2016). The mere fact that these models come with a built-in peculiarity that they, whenever interpreted as a mechanism of democracy, simultaneously become a mechanism of epistocracy, speaks for itself.[21] This is the unavoidable consequence of the uncomfortable reality: epistocracy and epistemic instrumentalism share the same commitment to correctness standards.

Estlund, who coined the term "epistocracy", agrees that both epistocracy and his account of democracy begin with this same basic assumption (2018, p. 30). Yet, at the same time, he holds that the alleged authorita-

20 It may seem that Estlund's view is compatible with the first possibility. For Estlund, the threshold is avoiding primary bads. Thus, if we are facing multiple decisions where none entails any of the primary bads, then the epistemic approach can only get us so far. From there on, we must appeal to procedural values. But at the same time, Estlund admits that there must be some kind of hierarchy between primary bads: "For example, famine, epidemic, and genocide are evidently always great disasters. On the other hand, I assume that war, economic collapse, and political collapse might be necessary evils in some extreme cases" (2008, p. 163). Let us return to my hypothetical example and suppose that the options relate to the primary bads in the following way: *A* – entails greater primary bad; *B* – entails lesser primary bad; *C* – entails no primary bad. In that case, Estlund would be compelled to admit that *B* is correct in the original, but incorrect in the revised scenario.

21 The belief that the CJT can easily lead to pessimistic results if we allow mass participation was held by Condorcet himself. Some experiments show that if we try to apply the DTA model in a less abstract sense (precisely in order to capture the notion of expertise which is supposedly beaten by diversity) the results might prove the opposite; that, in fact, the ability that trumps diversity (Grim et al., 2019).

tiveness of epistocracy can be refuted on epistemic grounds. While discussing Mill's epistocratic proposal of plural voting (2001 [1861], pp. 174–183), Estlund offers the *demographic objection* towards such policy.[22] This objection states that the educated citizens may nevertheless possess epistemically damaging features which disproportionately affect the epistemic benefits of education. Education was (and still is) the privilege of certain demographic groups. Giving extra votes to those groups gives them more leeway to act on their biases, thereby damaging the epistemic quality of political decisions (Estlund, 2008, p. 215).[23] Estlund could have easily made a similar counterclaim on moral grounds, or by pleading to procedural fairness. Yet, he insists that the demographic objection is an epistemic argument against epistocracy. Since he realizes that his partially instrumental approach favors epistocracy (at least) as much as democracy, a relevant *epistemic* advantage of democracy must be put forward.

However, I believe that with epistemically framed demographic objection Estlund confuses correlation with causation. Cyril Hédoin objects that the "epistemically damaging features" that Estlund writes about have nothing to do with epistocracy as such.[24] Epistocratic institutions (such as plural voting) only reflect the preexistent social injustices and non-legitimate domination relationships within the society (Hédoin, 2021, p. 510). Since those institutions are not a *cause* of those injustices, they may still be preferable from a purely epistemic perspective. In other words, they cannot make an already bleak situation any worse, from an epistemic point of view. Thus, demographic objection, understood as an epistemic objection, completely misses its intended mark.

---

22 Before he introduces a demographic objection, Estlund considers the *deference objection*, that is, the claim that people might reasonably refuse to submit to the rule of the educated. Yet, Estlund realizes the threat that such a claim might simultaneously undermine the view that a good education promotes wise rule in a democracy. He concludes that: "If Mill's plural voting loses on these grounds, perhaps the whole epistemic dimension of political argumentation loses, too" (2008, p. 213). I wholeheartedly agree that this and similar epistemic attacks on epistocracy backfire dreadfully. However, I disagree with the verdict that the *whole* epistemic dimension is lost. It is only its instrumental component that comes under fire.

23 Note that Estlund does not claim that diversity *per se* is desirable for instrumental reasons (cf. Landemore, 2018). He only suggests that the lack of it may be epistemically harmful: "Exactly what is meant by bias here, and how it leads to increased collective error, would need more careful explanation, but I accept this as a powerful objection." (2008, p. 215).

24 Hédoin claims that mechanisms of epistemic avoidance and epistemic domination are the actual threat that Estlund alludes to. Epistemic avoidance refers to the fact that persons who belong to socially advantaged groups (willingly or unwillingly) avoid engaging with the problems of socially less-advantaged groups (2021, p. 509). This circumstance may lead to epistemic dominance, where the policies that favor disadvantaged groups are mostly ignored.

This is just one example of a general viewpoint that I subscribe to. The viewpoint is this: all instrumental arguments for democracy are implicit arguments for epistocracy (Gunn, 2019), and all epistemic arguments against epistocracy are inevitably arguments against (instrumentally devised) epistemic democracy. To avoid falling into this trap, we should always make a case against epistocracy on moral/political grounds (Hédoin, 2021). Estlund's theory, for the most part, does exactly that. But its unfortunate commitment to independent standard causes this particular argument against epistocracy to fail.

* * *

The purpose of a Gorgian argument was to offer what I believe are compelling reasons for rejecting (both naïve and not-so-naïve) instrumentalism. Proceduralism, however, comes through all these objections unscathed. Since it presupposes no procedure-independent standard, it is compatible with the fact of reasonable pluralism and does not turn the political arena into a breeding ground for epistocracy. As for the particular theories of Estlund and Anderson, I consider Estlund's epistemic proceduralism to be a step in the right direction, as I believe that an adequate theory of democracy must include both epistemic and procedural values. It is, however, a step too long, since it also included some problematic instrumental claims which made this theory, to use Estlund's own words, "an unstable hybrid". A Gorgian argument also affects the last "D" of Anderson's DDD conception, since the dynamism of a Deweyan account rests solely on its consequentialist outlook which does not acknowledge the fact of reasonable pluralism. In the next section, I will offer a different justification of democracy which does better in this regard.

3. Drawing the Target around the Arrow

So far, my paper has been mostly critical of the classical account of epistemic democracy which seemingly always includes at least some instrumental claims. In this section, I put forward an alternative way of reconciling procedural and epistemic features of democracy. This view rests on the idea that *epistemic virtues* have procedural epistemic value, in contrast to instrumental value that is usually ascribed to them (e.g., Landemore, 2018).

If we reject instrumental justification of democracy, but at the same time wish to retain the view that democracy must be (at least partially)

justified on epistemic grounds, we are left with some form of "pure epistemic proceduralism", which is a position that Peter advances (2008, 2013, 2016). Although Estlund sometimes (quite misleadingly) calls his conception "purely procedural" (Estlund, 2008, p. 116), it should be noted that in Rawlsian terminology, a pure proceduralist conception is one that makes no reference to procedure-independent standards (Rawls, 1999 [1971], p. 75). In the case of Estlund's epistemic proceduralism, these standards play their part in the selection of the legitimacy-generating procedure. As such, his conception has the structure of imperfect proceduralism[25] (Peter, 2008, p. 39). Since epistemic proceduralism claims that citizens may rationally believe that majority is mistaken but must nevertheless obey the mistaken law, Estlund makes an analogy to the judicial system: citizens obey the verdict of a fair trial, not because it necessarily produces correct decision, but because it shows a tendency to do so (2008, p. 108). However, this is the very Rawlsian example of an imperfect procedure (Rawls, 1999 [1971], pp. 76–77), which indicates that Estlund's position is far from purely procedural. Peter develops an alternative to Estlund's view, which she calls (for reasons just stated) *pure epistemic proceduralism*. For Peter, public deliberation has a procedural epistemic value, not because it leads to more or less accurate beliefs, but because it fosters mutual accountability among participants, provided that deliberation is properly conducted (Peter, 2013). In contrast to veritist or consequentialist views, proceduralist political epistemology drops the idea that procedure-independent standards are necessary to judge the quality of political decisions (Peter, 2008, p. 45).[26] Pure epistemic proceduralism differs from purely procedural non-epistemic accounts for it includes criteria of epistemic fairness. It also differs from Estlund's view, and that of correctness theories, because it excludes the veritistic quality of outcomes (Peter, 2008, pp. 49–50). Finally, despite sharing multiple focal points with Anderson's theory (like insistence on diversity and discussion), pure epistemic proceduralism diverges from Deweyan account by putting the process of deliberative inquiry in the center, rather than its outcomes (Peter, 2008, p. 51). Thus, on Peter's view, deliberative democratic decision-making has epistemic value even in those cases where its effect diminishes the accuracy of the participants' beliefs (Peter, 2016, p. 142).

25 Unlike correctness theories which, on this account, are an instance of perfect proceduralism.

26 Peter's view relies on Helen Longino's social epistemology which emphasizes the socially-realized criteria for scientific objectivity (Longino, 1990, pp. 76–79). On this view, knowledge-producing is a social practice and has no relation to procedure-independent ideas of truth. Peter expands on this theory and presents public deliberation as analogous to Longino's account of scientific inquiry.

I believe that Peter's account of democracy successfully reconciles epistemic and procedural claims in justification of democracy, and at the same time avoids the biggest pitfalls of instrumentalism. However, since it focuses on knowledge-producing practices of public deliberation, pure epistemic proceduralism may still be vulnerable to some epistocratic counterarguments. That is, even if we abandon the veritistic view in favor of epistemic fairness as Peter suggests, the *quality* of deliberative inquiry may still vary, despite not being assessed by its truth-tracking potential. Thus, one may argue that a small community of experts is in a better place to arrive at high-quality decisions through their own internal deliberations (Fuerstein, 2019, p. 383). To address this last concern, I will propose, not an alternative, but what I consider to be an extension of Peter's view. This extension is based on two different sources: 1) Peircean justification of democracy advocated by Cheryl Misak (1999, 2008, 2009) and Robert Talisse (2005, 2011a, 2011b) and 2) James Montmarquet's intrinsic evaluation of epistemic virtues (1987).

Peircean pragmatist epistemic argument for democracy can be reconstructed in the following way: i) few fundamental epistemic principles cannot be denied and they ii) entail several epistemic commitments which iii) justify democracy in a deliberative sense (Erman & Möller, 2016). In line with Peter's view, Talisse claims that the biggest drawback of Deweyan democracy is its incompatibility with Rawlsian idea of reasonable pluralism (Talisse, 2011b, pp. 558–562). Due to its consequentialist and perfectionist nature, Deweyan democracy not only allows but *entails* state coercion in order to foster those values and attitudes that are deemed necessary for human flourishing. Talisse also remarks that Anderson chooses to gloss over the less pleasant parts of Dewey's theory, and instead adopts a restrained view of it; so restrained that it is questionable whether it should even be called Deweyan (Talisse, 2011a, pp. 518–519).

Talisse believes that Peircean justification of democracy permits the fact of reasonable pluralism while remaining distinctively pragmatist (2011a, p. 519).[27] Drawing on Peirce's methods of fixing beliefs (CP 5.377–5.387), he presents a set of norms that are internal to our beliefs.[28] In Peircean account, it is believing itself which motivates one to engage

27 Talisse is not bothered by the fact that Peirce never wrote *anything* on political theory. He believes that certain political claims are implicitly present in his writings; Misak adopts a similar view (Misak, 2008, p. 94).

28 Talisse's list of basic epistemic principles goes as follows:

(1) To believe p is to hold that p is true.

(2) To hold that p is true is to hold that p would be able to withstand the challenge of ongoing scrutiny as new reasons, arguments and evidence are brought to bear.

in inquiry. Thus, these cognitive norms are not imposed externally but are instead an articulation of cognitive commitments that "*we already endorse*, regardless of the content of our beliefs" (Talisse, 2011a, p. 520; original emphasis).

The crucial point is that, in Peircean view, any knowledge-seeking process requires the notion of *community* (CP 5.311). Yet, this community lives in an ever-evolving world; and even though a process of inquiry must necessarily converge on a specific point, that point is always just provisional. For Peirce, aiming for truth is not shooting at a fixed target, but a moving one (Burch, 2022, §4). Thus, in Peircean outlook, epistemic commitments must always be interpersonal. Talisse and Misak take this to mean that believers must be committed to different epistemic virtues, such as honesty, modesty, charity, integrity (Talisse, 2005, pp. 112–113), as well as open-mindedness, courage, willingness to listen to the others' views, etc. (Misak, 2008, p. 103). Those who want their beliefs to be governed by reasons are required to expose their beliefs to different perspectives and arguments (Misak, 1999, p. 106). From here, an adherence to democracy follows naturally: if we are to live up to our epistemic commitments, we must endorse the only political order which allows us to do so. Thus, the Peircean process of inquiry requires the institutions of equality, free speech, freedom of information, open debate, and access to decision-making (Talisse, 2011a, p. 520).

Once deliberation is understood in terms of epistemic virtues, it can complement Peter's view and overcome the possible epistocratic counter-argument. Since everybody is a potential contributor to political deliberation, there is no identifiable pool of epistemic experts. There may be people who are *better* at exchanging reasons, but it is not obvious that any special education could make somebody trained to do so (Misak, 2009, p. 35). In other words, there is no epistocracy of the virtuous. Yet, at the same time, Misak's justification of democratic legitimacy is distinctly instrumentalist: "Democratically produced decisions are legitimate because they are produced by a procedure with a tendency to get things right" (Misak, 2008, p. 95).[29] She endorses a reliabilist virtue epistemology, where virtue is justified if it is a constituent part of a reliable method that is likely to lead us to a true belief (Misak, 2009, p. 36). This strikes me as a step in

(3) To hold that a belief would meet such challenges is to commit to the project of justifying one's belief, what Peirce called 'inquiry'.

(4) The project of squaring one's beliefs with reasons and evidence is an ongoing social endeavor that requires participation in a community of inquiry (Talisse, 2011a, p. 519–520).

29 However, she does not necessarily ascribe truthiness to the outcomes, only legitimacy (Misak, 1999, p. 7), which makes her view very close to Estlund's.

the wrong direction, but one which has an easy remedy. Instead of sneaking instrumentalism in through the back door, I will adopt Montmarquet's account of intrinsically valued epistemic virtues.

Montmarquet accepts the view that the commitment to truth is the supreme epistemic virtue, yet at the same time explicitly denies that epistemic virtues require reliability. To defend this claim, he imagines a Cartesian evil demon who, without our knowledge, made our world in such a way that the truth is best achieved by demonstrating a wide variety of epistemic vices, such as dogmatism or epistemic laziness. Montmarquet's view is that traits like open-mindedness would still be considered virtues, even in a demonic world. Conversely, if the words of a mad prophet suddenly turned out to be completely true, that would not make those who blindly followed him epistemically virtuous (Montmarquet, 1987, p. 482–485). Thus, according to Montmarquet, some epistemic trait should be regarded as a virtue, not for its reliability, but because it is desirable for those who want the truth. Thus, the virtues are not valuable as instruments for attaining truths, but because the very motivation for the truth has an intrinsic value (Battaly, 2008, p. 649).

I believe that Montmarquet's account of epistemic virtues is a much better supplement to Peircean argument for democracy than a reliabilist view – it is the last piece of the proceduralist puzzle. To summarize, the most adequate way of defending democratic legitimacy is the position of pure epistemic proceduralism, where deliberative procedures are entailed by the set of intrinsically valued epistemic virtues. According to this view, democracy is neither a truth-tracking nor a problem-solving process. It is, in fact, a process of *truth-seeking*. On this interpretation, citizens are like archers who shoot, not even at a moving target, but at an empty wall. And whenever the arrow successfully lands, we can call it bullseye and draw the target around it. Thus, truth is both the source and the aim of our democratic concerns, and not just an elusive superficial entity.

The last major point of Peircean view is the fact that it invokes no moral claims. People can have all sorts of beliefs about the good life, or the meaning of human existence, or the value of community – but they will all have a reason to endorse a democracy simply because they hold some beliefs.[30] This is why Peircean version of pragmatist democracy, unlike Deweyan, acknowledges the fact of reasonable pluralism (Talisse, 2011a,

30 I believe that, on Peircean account, moral disagreements can be settled on the grounds of a morally majoritarian view without abandoning the notion of moral truthiness. It is Peirce's rejection of the correspondence theory of truth (CP 5.416) which, I believe, recommends his epistemology to a non-procedural justification of democracy.

p. 521).[31] Because of this, Peircean model of democracy is the superior one according to Anderson's DDD conception. It embraces both epistemic *diversity* and the value of *discussion*. A Peircean believer has an epistemic motivation to actively seek out partners in inquiry who advocate different views from her own. And, on this account, to say "I believe that *p*, but have discussed the matter with no one" reveals an epistemic deficiency. But, most importantly, it can model the *dynamism* of democracy, since Peircean epistemology sees inquiry as an ever-ongoing process. Thus, democracy based on this model requires that channels of dissent and feedback are open after any collective decision is reached (Talise, 2011a, pp. 522–523).[32]

* * *

The ongoing trends in democratic theory reveal the prevailing view that epistemic justifications of democracy are superior to non-epistemic ones. However, most epistemic justifications tend to (fully or partly) justify democracy on instrumental grounds. This approach entails several theoretical problems and also makes an implicit claim for epistocratic governments. I, too, have advanced the view that the epistemic component is not only important but mandatory for an adequate justification of democracy, as the lack of such a component could only deepen the tense relationship between epistemology and democracy. Instead of an instrumentalist outlook, however, I have argued for a procedural justification of democracy. My view was that a virtue-oriented account of deliberative

31 One may argue that Peircean political epistemology is still monistic since it emphasizes the truth as the one and only epistemic good, while, in fact, the truth may be just one of the many epistemic goods that we want to be promoted by our political system. For example, we may prefer procedures that have the "ease of deliberative use" over those which foster the search for truth (Lever & Chin, 2017, p. 2). In that case, the problem of reasonable pluralism can be applied to epistemology quite as much as to morality (Lever & Chin, 2017, p. 3). However, this objection misses its mark since the epistemic virtues that make the core of Peircean democracy do not constitute any *distinct* comprehensive epistemology. They are instead, as Talisse points out, the commitments of any well-developed epistemology (Talisse, 2011a, p. 522).

32 Critics may object that, even if we accept that the epistemically virtuous process of truth-seeking requires some kind of democratic practice, it still does not entitle the citizens to participate in the process of political decision-making (Erman & Möller, 2019). I do not consider this objection to be particularly strong. First, as I aimed to defend a procedural and not purely epistemic view of democracy, I think that Peircean account of democracy can be accompanied by usual procedural appeals to fairness. Second, it seems to me that this objection rests on a superfluous distinction; a Peircean democrat can simply adopt a neutrally monistic view – for all intents and purposes, truth-seeking *is* decision-making.

democracy can overcome most of the problems that more traditional versions of epistemic democracy face. On this account, epistemic virtues that are valued for their truth-seeking instead of truth-producing potential are the required epistemic component that makes democracy intrinsically justified.

References:

Arrow, K. (1963). *Social Choice and Individual Values, 2nd Edition*. New York: John Wiley & Sons.

Anderson, E. (2006). The Epistemology of Democracy. *Episteme* 3(1–2). 8–22.

Battaly, H. (2008). Virtue Epistemology. *Philosophy Compass* 3/4. 639–663.

Black, D. (1998 [1958]). *The Theory of Committees and Elections*. New York: Springer Science + Business Media, LLC.

Brennan, J. (2016). *Against Democracy*. Princeton: Princeton University Press.

Burch, R. (2022). Charles Sanders Peirce. *The Stanford Encyclopedia of Philosophy* (Zalta, E. N. ed.) URL = https://plato.stanford.edu/archives/sum2022/entries/peirce/

Cohen. J (1986). An Epistemic Conception of Democracy. *Ethics* 97(1). 26–38.

Cohen. J (1989). Deliberation and Democratic Legitimacy. *The Good Polity* (Hamlin, A. & Pettit, P. eds). Oxford: Basil Blackwell. 17–34.

Condorcet, M. J. A. N. d. C., Marquis de. (1976 [1785]). Essay on the Application of Mathematics to the Theory of Decision-Making. *Condorcet: Selected Writings* (Baker, K. M. ed.) Indianapolis: Bobbs-Merrill Company, Inc. 34–72.

Dewey, J. (1922). Valuation and Experimental Knowledge. *The Philosophical Review* 31(4). 325–351.

Dewey, J. (1946). *The Public and its Problems*. Chicago: Gateway Books.

Erman, E. & Möller, N. (2016). Why Democracy Cannot Be Grounded in Epistemic Principles. *Social Theory and Practice* 42(3). 449–473.

Erman, E. & Möller, N. (2019). Pragmatism and Epistemic Democracy. *The Routledge Handbook of Social Epistemology* (Fricker, M.; Graham, P. J.; Henderson D.; Pedersen, N. J. L. L.). New York: Routledge. 367–376.

Estlund, D. (1997). Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority. *Deliberative Democracy: Essays on Reason and Politics* (Bohman, J. & Rehg, W. eds.). Cambridge, MA and London: The MIT Press. 173–204.

Estlund, D. (2008). *Democratic Authority: A Philosophical Framework*. Princeton and Oxford: Princeton University Press.

Fuerstein, M. (2019). Epistemic Proceduralism. *The Routledge Handbook of Social Epistemolgy* (Fricker, M.; Graham, P. J.; Henderson D.; Pedersen, N. J. L. L.). New York: Routledge. 377–385.

Gaus, G. (2011). On Seeking the Truth (Whatever That Is) through Democracy: Estlund's Case for the Qualified Epistemic Claim. *Ethics* 121(2). 270–300.

Goodin, R. & Spiekermann, K. (2018). *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.

Grim, P., Singer, D. J., Bramson, A., Holman, B., McGeehan, S., Berger, W. J. (2019). Diversity, Ability, and Expertise in Epistemic Communities. *Philosophy of Science* 86(1). 98–123.

Grofman, B., Owen, G., Feld, S. L. (1983). Thirteen Theorems on Search of the Truth. *Theory and Decision* 15. 261–278.

Gunn, P. (2019). Against Epistocracy. *Critical Review* 36(1). 26–82.

Hédoin, C. (2021). The 'Epistemic Critique' of Epistocracy and Its Inadequacy. *Social Epistemology* 35(5). 502–514.

Hong, L. & Page, S. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences* 101(46). 16385–16389.

Hume, D. (2011 [1748]). *An Enquiry Concerning Human Understanding*. In: *Hume: The Essential Philosophical Works*. Hertfordshire: Wordsworth Classics of World Literature. 571–706.

Kelly, J. T. (2012). *Framing Democracy: A Behavioral Approach to Democratic Theory*. Princeton: Princeton University Press.

Landemore, H. (2013). *Democratic Reason: Politics, Collective Intelligence and the Rule of the Many*. Princeton: Princeton University Press.

Landemore, H. (2018). What Does It Mean to Take Diversity Seriously? On Open-Mindedness as a Civic Virtue. *Georgetown Journal of Law & Public Policy* 16. 795–805.

Lever, A. & Chin, C. (2017). Democratic Epistemology and Democratic Morality: The Appeal and Challenges of Peircean Pragmatism. *Critical Review of International Social and Political Philosophy*. DOI: 10.1080/13698230.2017.1357411

List, C. & Goodin, R. (2001). Epistemic Democracy: Generalizing the Condorcet Jury Theorem. *Journal of Political Philosophy* 9(3). 277–306.

Little, I. M. D. (1952). Social Choice and Individual Values. *Journal of Political Economy* 60(5). 422–432

Longino, H. E. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Mackie, G. (2003). *Democracy Defended*. Cambridge: Cambridge University Press.

Mackie, G. (2011). Deliberation, but Voting Too. *Approaching Deliberative Democracy: Theory and Practice* (Cavalier, R. ed). Pittsburgh, PA: Carnegie Mellon University Press. 75–99.

Mackie. G. (2001). Is Democracy Impossible?: Riker's Mistaken Accounts of Antebellum Politics. URL = https://core.ac.uk/download/pdf/156617035.pdf

Mill, J. S. (2001 [1861]). *Considerations on Representative Government*. London: Elecbook Classics.

Misak, C. (1999). *Truth, Politics, Morality: Pragmatism and Deliberation*. London and New York: Routledge.

Misak, C. (2008). A Culture of Justification: The Pragmatist's Epistemic Argument for Democracy. *Episteme* 5(1). 94–105.

Misak, C. (2009). Truth and Democracy: Pragmatism and the Deliberative Virtues. *Does Truth Matter? Democracy and Public Space* (Geenens, R. & Tinnevelt, R. eds.). Dordrecht: Springer. 29–39.

Mladenović, I. (2019). *Javni um i deliberativna demokratija* [*Public Reason and Deliberative Democracy*]. Belgrade: Institut za filozofiju i društvenu teoriju.

Mladenović, I. (2020). Deliberative Epistemic Instrumentalism, or Something Near Enough. *Philosophy and Society* 31(1). 3–11.

Montmarquet, J. A. (1987). Epistemic Virtue. *Mind* 96(384). 482–497.

Moser, R.G. & Scheiner, E. (2009). Strategic Voting in Established and New Democracies: Ticket Splitting in Mixed-Member Electoral Systems. *Electoral Studies* 28. 51–61.

Peirce, C. S. (1994) *The Collected Papers of Charles Sanders Peirce I-VIII* (Deely, J. ed.). Cambridge, MA: Harvard University Press.

Peter, F. (2008). Pure Epistemic Proceduralism. *Episteme* 5. 33–55.

Peter, F. (2016). The Epistemic Circumstances of Democracy. *The Epistemic Life of Groups: Essays in Epistemology of Collectives* (Brady, M. S. & Fricker, M. eds.). New York: Oxford University Press. 133–149.

Peter. F. (2013). The Procedural Epistemic Value of Deliberation. *Synthese* 190. 1253–1266.

Prijić Samaržija (2020). The Epistemology of Democracy: The Epistemic Virtues of Democracy. *Philosophy and Society* 31(1). 56–70.

Putnam, H. (2002). *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA and London: Harvard University Press.

Rawls, J. (1996). *Political Liberalism*. New York: Columbia University Press.

Rawls, J. (1999 [1971]). *A Theory of Justice*. Cambridge, MA. Harvard University Press.

Regenwetter, M., Grofman, B., Popova, A., Messner, W., Davis-Strober, C. P., Cavagnaro, D. R. (2009). Behavioural Social Choice: A Status Report. *Philosophical Transactions of The Royal Society B* 364(1518). 833–843.

Richardson, H. (2003). *Democratic Autonomy: Public Reasoning about the Ends of Policy*. Oxford: Oxford University Press.

Riker, W. (1982). *Liberalism Against Populism: A confrontation Between the Theory of Democracy and the Theory of Social Choice*. Long Grove, Illinois: Waveland Press, Inc.

Roberts-Miller, P. (2017). *Demagoguery and Democracy*. New York: The Experiment.

Schwartzberg, M. (2015). Epistemic Democracy and Its Challenges, *Annual Review of Political Science* 18. 187–203.

Shapiro, I. (2016). *Politics Against Domination*. Cambridge, MA: Harvard University Press.

Sprague, R. K. (ed.) (2001). *The Older Sophists*. Indianapolis and Cambridge: Hackett Publishing Company, Inc.

Stanley, J. (2015). *How Propaganda Works*. Princeton: Princeton University Press.

Talisse, R. (2005). *Democracy After Liberalism*. New York: Routledge.

Talisse, R. (2019). The Epistemology of Democracy. *The Routledge Handbook of Social Epistemology* (Fricker, M.; Graham, P. J.; Henderson D.; Pedersen, N. J. L. L.). New York: Routledge. 357–366.

Talisse, R. B. (2011a). A Farewell to Deweyan Democracy. *Political Studies* 59. 509–526.

Talisse, R. B. (2011b). Toward a New Pragmatist Politics. *Metaphilosophy* 42(5). 552–571.

Unger, P. (1971). A Defense of Skepticism. *The Philosophical Review* 80(2). 198–219.

Miljan Vasić

Proceduralna vrednost epistemičkih vrlina

Apstrakt: Dugo prisutna tenzija između proceduralnog i instrumentalnog opravdanja demokratije dovedena je u pitanje pojavom teorija koje pokušavaju da objedine oba pristupa. Ove teorije predstavljaju epistemičke odlike demokratije u instrumentalnom okviru, a potom pokušavaju da ih pomire sa proceduralnim vrednostima. U ovom tekstu tvrdim da je moguće uključiti epistemičku dimenziju u opravdanje demokratije bez obavezivanja na instrumentalizam. Prema gledištu koje zastupam, persovska epistemologija spojena sa epistemičkim vrlinama kojima se pripisuje intrinsična vrednost daje čisto proceduralni argument u prilog demokratije.

Ključne reči: proceduralizam, instrumentalizam, demokratija, epistemička demokratija, epistemičke vrline, pragmatizam.

Mirjana Sokić*

IS HAPPINESS IN THE HEAD?

Abstract: This paper examines the philosophical implications of Nozick's thought experiment, specifically focusing on the assumption that most people would not want to be plugged into the experience machine. I present an "inverted" experience machine scenario in order to argue that this assumption is incorrect and that the scenario raises important philosophical questions about our purported unwillingness to be plugged in. The paper concludes that the "inverted" experience machine scenario is compatible with the central thesis of hedonism and other internalist theories of well-being, and provides strong support for the idea that happiness is truly in the head.

Keywords: Nozick's thought experiment, anti-hedonistic argument, experience machine, internalist theories of well-being, subjective aspect of experience

1. Introduction

Robert Nozick's experience machine thought experiment, which appears in his book *Anarchy, State and Utopia* (1974), was originally intended to make a point about the morally unacceptable treatment of animals (Weijers 2011a).[1] However, shortly after the book was published, many philosophers took Nozick's thought experiment as one of the strongest objections to hedonism, and possibly to all positions that view our well-being or welfare as exclusively dependent on the *subjective aspect* of experience. According to this popular opinion, Nozick's thought experiment

* The Institute for Philosophy, Faculty of Philosophy University of Belgrade, mirjana-sokic19@gmail.com

1 Nozick argues that until we can explain why most people would not want to be plugged into the experience machine and show that the same reasons do not apply to animals, we cannot claim that only the conscious experiences of animals determine the limits of acceptable behaviour towards them (Nozick 1974: 43). His point is intended to refute the common argument against ethical veganism, which suggests that an animal's pleasant life justifies its mutilation, restriction of freedom, and eventual killing. For a defence of this anti-vegan argument, see Zangwill (2021).

raises important questions about the nature of happiness, the value of our experiences, and the limits of hedonism.[2] It suggests that many people, despite the allure of the experience machine, would not choose to be plugged in, because they value something more than their own subjective experiences.[3] This challenges the hedonist view that all that matters is how our lives feel from the inside, and implies that there are other factors that contribute to our well-being. To better understand the philosophical implications of this thought experiment, let us consider it in its entirety:

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences, selecting your life's experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to select the experiences of your next two years. Of course, while in the tank you won't know that you're there; you'll think it's all actually happening. Others can also plug in to have the experiences they want, so there's no need to stay unplugged to serve them. (Ignore problems such as who will service the machines if everyone plugs in.) Would you plug in? *What else can matter to us, other than how our lives feel from the inside?* (Nozick 1974: 42–43)[4]

As we can see, Nozick describes a fantastic scenario in which the reader is asked to imagine that they are given the choice to be plugged into an "experience machine" that would provide them with any experiences they desire, without any negative consequences or limitations. The machine would allow them to live out their wildest dreams, have the most thrilling experiences, and achieve all their goals, without any effort or risk. The only

2 The term "hedonism" is somewhat ambiguous (see, Weijers 2011b), but the experience machine thought experiment is typically used to challenge the generic version of this view, which holds that pleasure (hēdonē [ἡδονή]) and only pleasure intrinsically contributes positively to well-being (Weijers 2014: 514). For more information on other objections that have been raised against hedonism, see Shafer-Landau (2018).

3 In the relevant literature, positions that hold that subjective experience is the only thing that has intrinsic value are often referred to as "the internalist mental state theories of well-being" (Weijers 2014).

4 For an updated version of this scenario, see Nozick (1989: 104).

catch is that all their experiences would be artificially created and simulated. Nozick asks the reader whether they would choose to be plugged in or prefer to live a "real" life, with all its challenges and hardships.

Nozick's answer to the questions at the end of the paragraph is that people's intuitive reaction to the presented dilemma would be to stay unplugged (see, De Brigard 2010: 43), and that, more importantly, the fact that the vast majority of people would want to stay in reality – even though this reality might turn out to be less pleasurable for them – shows that, contrary to the central thesis of hedonism, there are other things that matter to us in addition to our experiences or, to use Nozick's phrase, how our lives feel from the inside. Yet, despite the influence that the quoted paragraph has on the philosophical debate about the plausibility of hedonism, numerous authors now reject Nozick's conclusion (see, Linn 2016: 315–316). Thus, for example, Harriet Baber (2008: 133–8), Matthew Silverstein (2000: 279–300), Jason Kawall (1999: 381–87), Sharon Hewitt (2010: 331–49), Alex Barber (2011: 271), Torbjörn Tännsjö (2007), and many others, maintain that Nozick's thought experiment does not make a compelling case against hedonism. The primary goal of this paper is to provide insight into how to resolve this debate.

I will introduce and analyse recent formulations of Nozick's anti-hedonistic argument in the relevant philosophical literature. This analysis will reveal that the main weakness of the thought experiment is the assumption that most, if not all, people would not want to plug into the experience machine. By presenting the "inverted" experience machine scenario, I will demonstrate that this assumption is incorrect. In the conclusion, I will argue that the "inverted" experience machine scenario raises important philosophical questions about our purported unwillingness to plug into the experience machine, and that it is compatible with the central thesis of hedonism and other internalist theories of well-being that view the subjective aspect of experience as the only thing with intrinsic value. I will thus conclude that the "inverted" experience machine scenario provides strong support for the idea that happiness is truly in the head.

2. Formulations of Nozick's argument against hedonism

As previously mentioned, many philosophers believe that Nozick's thought experiment provides a powerful and persuasive argument against hedonism and other internalist theories of well-being. However, it should be noted that pleasure and well-being were not explicitly mentioned in

the original formulation of the scenario. This means that in order for the thought experiment to refute hedonism, it must be restated appropriately. In this section, I will examine some versions of Nozick's anti-hedonistic argument that have been put forward in recent literature. One simple version of the argument is presented by Ben Bramble (2016: 137):

[A] Plugging in would not be best for one.

[B] Hedonism entails that plugging in would be best for one.

[C] Therefore, hedonism is false.

In order to see where this argument goes wrong, we should recall that, strictly speaking, Nozick nowhere suggests that plugging into the experience machine 'would not be best for one' – as is stated in premise [A] – but rather makes an *assumption* that the intuitive reaction of most people would be to stay in reality, even if plugging in would turn out to be significantly more pleasurable. As Bramble explains, a possible rationale for premise [A] emerges from the following reasoning: the psychological fact that, given the option, most if not all people would decide not to plug into the experience machine represents the most reliable indicator that plugging in would not be best for them (see, Bramble 2016: 137). Notwithstanding the initial plausibility of this reasoning, the point to keep in mind is that to identify the content of people's wants or choices with what is good for them is problematic to say the least (see e.g., Kawall 1999; Silverstein 2000; Hewitt 2010). At any rate, without some additional and conceptually independent elaboration on how we can identify *what is desired with what is desirable*, the question-begging character of this argument is evident (see, Baber 2010). Given this difficulty, I think we should sidestep the entire debate about whether Nozick, in fact, incorporates a claim similar to premise [A] in the original version of his scenario, by considering what I take to be a significantly superior formulation of Nozick's anti-hedonistic argument. The formulation in question was proposed by Dan Weijers (2011a: 229–231), and it runs as follows:

1. Plugging into an Experience Machine would make the rest of your life dramatically more pleasurable and less painful than it would otherwise have been (stipulated in thought experiment).
2. Given the choice to plug into an Experience Machine for the rest of your life, ignoring any responsibilities you might have to others, you would decline (appeal to readers' judgment).
3. If, ignoring any responsibilities you might have to others, you would decline the chance to plug into an Experience Machine for the rest of your life, then pleasure and pain are not the only things of intrinsic value (or disvalue) in a life.

4. Pleasure and pain are not the only things of intrinsic value (or disvalue) in a life (modus ponens, [2], [3]).
5. If hedonism is true, then pleasure and pain are the only things of intrinsic value (or disvalue) in a life.
6. Hedonism is false (modus tollens, [4], [5]).

A great deal could be said about this argument. First, there is no doubt that it is valid, for the premises adequately support the conclusion. Moreover, observe that premise [3] successfully avoids all of the problems that we have encountered with the previous argument; namely, it does not say that plugging in is *not best* for people; rather, it is a conditional according to which if it truly is the case that, when presented with Nozick's scenario, people would decide not to plug in – i.e., if they would choose not to abandon their current life in reality in favour of a much more pleasant virtual life – then pleasure could not be the *only* thing that is valuable in itself, as hedonists argue. The question that remains here, on the other hand, is whether the antecedent is true. In the following section, I intend to examine if Nozick was correct in assuming that most people would refuse to plug into the experience machine.

3. "Real" pleasure and "illusory" displeasure

As we have seen, the biggest challenge with Nozick's anti-hedonistic argument is the assumption that people's intuitive response to the opportunity to plug into the experience machine would be to stay in reality. Some philosophers believe that this response is due to confusions about the concept of reality, as well as misunderstandings about the implications of the perfect illusion that the experience machine is supposed to create. However, most authors argue that the negative response to Nozick's scenario is due to the so-called *status quo bias*, which is "an inappropriate preference for things to remain the same" (Weijers 2014: 530; see also Bostrom & Ord 2006; De Brigard 2010: 44). This bias typically manifests as a preference for the source of one's experiences to remain the same, regardless of whether those experiences are virtual or real. To overcome the problem posed by the status quo bias, Weijers created a scenario in which neither reality nor the experience machine is presented as the status quo (Weijers 2014: 252). In addition, unlike in the original Nozick scenario, the purpose of Weijers's version of the experience machine thought experiment is to determine people's intuition about whether it would be best for *someone else*, named Boris, to plug into the experience machine for the rest of their life. Interestingly, Weijers found that 55% of 77 partici-

pants said that the best option for Boris would be to plug into the experience machine for the rest of his life. These results indicate that Nozick's assumption in premise [2] – that the vast majority of people would prefer to remain in reality – is factually incorrect.

Weijers's scenario with Boris, while involving a decision about someone else rather than oneself, shares an important similarity with Nozick's scenario. Namely, Eden Lin (2015) notes that one of the most problematic aspects of Nozick's scenarios is that they typically involve a life within the experience machine that is hedonically superior to real life. In contrast, Lin proposes testing hedonism by considering two lives that are "experientially and thus hedonically identical" (2015: 320). Lin asks us to compare the life of Adam, who lives in the real world, with the life of Bill, who was plugged into an experience machine immediately after birth. Lin stipulates that the lives of Adam and Bill are identical "with respect to the qualitative features, durations, and temporal distribution of the pleasures and pains they contain" (2015: 321), which should mean that they are equal in welfare as well. However, when comparing these two lives, we may feel that there is something pitiful about Bill's life, but not Adam's. This suggests that Bill's life is somehow lower in welfare than Adam's, despite being hedonically identical. In short, Lin's example shows that if we feel bad for Bill, it can only be because Adam's life is higher in total welfare than Bill's, and since their lives contain the same amount of pleasure, we can conclude that the central thesis of hedonism must be incorrect. Yet, it is possible to object that many people believe that there is something pitiful about Bill's life because they have a conceptual confusion and bias towards the concept of "reality". They tend to believe that if something is real, it is *more valuable* than something that is virtual, even if the virtual thing provides the same experiences. This line of thinking becomes clear when we ask *why* people would feel bad for someone who has a life filled with pleasant experiences, but those experiences are part of a perfect simulation within an experience machine. The answer, that the experiences are not real and therefore less valuable, only begs the question against hedonism. So, where do people get the idea that a perfect illusion or simulation of reality is in any way worse than a real experience that is indistinguishable from the illusion?

The answer to this question can be found in Bart Engelen's interesting paper, in which he discusses the philosophical implications of Nozick's thought experiment and uses the film *Open Your Eyes* (*Abre los Ojos* 1997), directed by Alejandro Amenábar, as a reference. Engelen provides a thorough and insightful analysis of the thought experiment and its relevance to debates about the nature of reality and our experiences. The movie tells the story of César, a wealthy and handsome young man who

is disfigured in a car crash. He has a series of horrifying experiences and eventually learns that he committed suicide after the crash, but signed a contract with *Life Extension* to be cryogenically preserved until technology could revive him and attach him to a machine that would replace some of his memories. This machine, just like Nozick's experience machine, would allow him to live a virtual life of his choosing. However, César's machine is not functioning properly, leading to a nightmare-like existence.[5] The only way to disconnect from the machine is to commit suicide, which César eventually does.

In addition to its impressive cinematic qualities, *Open Your Eyes* raises serious questions relevant to Nozick's thought experiment. Engelen points out that the perfect illusion created by the experience machine is indistinguishable from reality. This leads us to ask: are someone's horrifying experiences less dreadful and disturbing because they are not happening independently or outside of their experiential perspective? A modified version of Lin's example with Adam and Bill can help answer this question. Suppose that Adam and Bill both have lives filled with horrors, tragedies, and unpleasantness, and the only difference is that Adam lives in reality while Bill's conscious experiences are the result of being plugged into the experience machine. In this case, we would feel bad for Adam, but the key question here is whether Bill's life would warrant *less* pity since his experiences are not real. I admit that it is notoriously difficult to provide a definitive answer to the question of whether the person connected to the experience machine, inside of which she suffers from *virtual* depression, is better off than the person suffering from depression that has *natural* or *real* causes. However, based on the fact that the perfect illusion is indistinguishable from reality, I am inclined to think that the answer to this question is negative.

To further support this point, let us consider a scenario in which two people, person *A* and person *B*, experience the same tragic event. After the tragedy, person *A* passes away without any additional complications, while person *B* suffers from severe depression. Despite all objective factors and circumstances being the same for both people, it is hard to deny that person *B*'s life is worse than person *A*'s, as depression is a factor that we take into account when evaluating the quality of someone's life, even though its effects are limited to the individual's subjective experience. Now, consider a hypothetical individual, *C*, who experiences the same situation as individuals *A* and *B*, but who is immediately connected to an experience machine that provides her with the same experiences as individual *B*, with the only difference being that person *B*'s depression

5 Engelen notes that Amenábar wrote the script for the film *Open Your Eyes* after experiencing a series of unpleasant hallucinations due to a high fever (Engelen 2010: 44, note 1).

was natural, while person *C*'s was the result of the experience machine. Is it accurate to assert that person *C*'s circumstances are "better" – in the sense of deserving less pity – than those of person *B*? I personally tend to answer this question with a resounding no, but it would be interesting to see if most people's common-sense intuitions agree with this answer. This thought experiment illustrates how our evaluation of the quality of someone's life is not solely based on objective factors, but also takes into account the individual's subjective experiences.

4. The reversed scenario: Neo's dilemma

Adam Kolber's thought experiment modifies one aspect of Nozick's scenario while keeping the other aspects consistent and focusing on the same issue (Kolber 1994: 15). In his scenario, the reader is already hooked up to an experience machine and is asked if they would like to remain connected or go to reality. Kolber argues that more people would choose to stay connected to the experience machine in his reversed scenario than would agree to be hooked up in Nozick's original scenario (1994: 15). This thought experiment explores the implications of being in an experience machine and the choices we might make if given the option to remain in it or leave. One way to present this reversed scenario is as follows:

Imagine that you are currently hooked up to an experience machine. All the beings you have interacted with so far, including your family, friends, acquaintances, and pets, are part of the perfect illusion created by the machine. Your entire life, which you thought was real, is actually just a carefully designed program. If you disconnect from the machine, you will meet *real* people, form *real* friendships, find *real* partners and pets, and so on. However, you have been warned that if you disconnect from the machine, you will permanently lose contact with all the people and things you believed to be real while you were hooked up. Given this information, *would you choose to disconnect from the experience machine?*

This thought experiment raises questions about the nature of reality, our relationships and experiences, and the value we place on them. It also challenges our assumptions about what it means to be "real" and whether the reality we perceive is the only one that matters. The dilemma faced by Neo in *The Matrix* (1999) is similar to the dilemma presented in the above hypothetical scenario. Neo finds out that his previous life was an illusion created by a computer program, and he must choose between taking the red pill and leaving the illusory world for a real life, or taking the blue pill and continuing to live in ignorance in the illusory world. In the movie, Neo's life is shown to be very unfulfilling, and there is no information

about his relationships with others. This presentation of his life justifies his decision to take the red pill, which is consistent with Nozick's conclusion.

However, it is worth considering whether we would act like Neo in this situation. In other words, with Engelen's assertion that the perfect illusion is *phenomenologically indistinguishable* from reality in mind, it is questionable whether we would choose to abandon our previous lives (including friends, family, partners, and pets) if we were told that these entities do not actually exist independently of the experiences created in us by a computer program. This thought experiment challenges us to consider the extent to which our relationships and experiences are valuable to us, and whether we would be willing to give them up for the sake of "reality". This raises further questions about the value and meaning of our relationships, experiences, and emotions, and whether they are ultimately based on a real or an illusory reality. It also prompts us to think about what it means to be "real" and whether the *reality that we perceive* is the only one that matters. In Nozick's thought experiment, the reader is presented with the dilemma of choosing between a life in reality or a life in a perfect illusion created by an experience machine. While Nozick argues that most people would choose to remain in reality, we have seen that many philosophers have criticized this assumption and pointed out that people's decisions in such scenarios may be influenced by irrational factors, such as the desire to maintain the status quo and the continuity of their experiential identity. Despite the potential problems with Nozick's thought experiment, it still raises important philosophical questions about the value of our experiential perspective and the role of reality in determining the quality of our lives.

5. Concluding remarks

The results of this paper indicate that our understanding of the concept of reality and the role of our experiential perspective is fraught with confusing intuitions.[6] As technology continues to advance and the devel-

6 Philosophical discussions often involve confused intuitions about the concept of reality and the role of our experiential perspective. For instance, Derek Parfit's (1984) thought experiment with teleportation challenges our everyday understanding of personal identity. Imagine a situation in which scientists have developed a teleportation machine that can scan a person's entire body (destroying it in the process), transmit the information to a distant location (such as Mars), and recreate a new, qualitatively identical body from the same particles. The question posed is whether this procedure allows a person to travel at the speed of light, or whether it simply kills one person and creates another, qualitatively identical one at the destination.

opment of virtual reality becomes increasingly possible, scenarios like Nozick's thought experiment become not only a theoretical possibility but a potential reality. In order to properly evaluate the value of our experiential perspective in relation to what is considered "real", it is necessary to examine how variations of the experience machine thought experiment affect human intuition and determine whether these intuitions are influenced by irrational or extraneous factors.[7] In conclusion, my final answer to the central question of this paper – "Is happiness in the head?" – is that happiness appears to be in the head, at least in the sense that people's experiential perspectives and conscious experiences play a crucial role in determining the quality of their lives and the overall amount of welfare. This answer aligns with the fact that no one would agree that a person leads a happy life based solely on objective circumstances, while ignoring the person's own subjective experiences and overall situation. This is true regardless of whether those experiences are the product of an experience machine or are real.

Also, we have seen that Nozick's assumption that people would have a negative intuitive response to being plugged into an experience machine is largely accurate, but only in the sense that, due to various irrational factors such as conceptual confusions, irrational fears, and the status quo bias, people do not always choose happiness. While Nozick's scenario may seem to be against hedonism, the question remains whether it poses a conclusive challenge to the philosophy, considering that the most common reason people give for refusing to plug into the experience machine is based on irrational and irrelevant considerations, as well as confusion regarding key concepts such as *happiness* and *reality*. It is, thus, reasonable to adopt Feldman's conclusion that, even if most people would refuse to plug into the experience machine, Nozick's position against hedonism and other internalist theories of well-being does not hold (see, Feldman 2011: 67–70). While I recognize that such a "hybrid" solution – which attempts to reconcile various viewpoints and theses (despite their popularity and increasing prevalence in contemporary philosophical literature) – is not

---

Many people tend to view the described procedure as a kind of advanced killing, distinct from regular killing in that it creates a person who is qualitatively the same as the one who was killed. Even if we are confident that the teleportation machine will work perfectly, would we be willing to let our loved ones "travel" in this way, knowing that their original body will be destroyed and replaced with a numerically different one? Parfit believes that any opposition to such a procedure is irrational and based on our prejudices regarding numerical identity.

7 Something similar is the case with the famous philosophical problem known as "the trolley problem". This problem has many different variants and formulations that are used to examine which factors determine our reactions. For more on this issue, see Edmonds (2014) and Kamm (2015).

particularly satisfactory, it is currently the only solution that appears to me to consistently and philosophically accurately consider all of the arguments and objections made in recent decades to Nozick's views in his well-known passage.

References:

Baber, H. (2008). "The Experience Machine Deconstructed." *Philosophy in the Contemporary World* 15, 133–8.

Bostrom, N., & Ord, T. (2006). "The Reversal Test: Eliminating Status Quo Bias in Applied Ethics." *Ethics* 116, 656–679.

Bramble, B. (2016). "The Experience Machine." *Philosophy Compass* 11, 136–145.

De Brigard, F. (2010). "If You Like It, Does It Matter If It's Real?" *Philosophical Psychology* 23, 43–57

Edmonds, D. (2014). *Would You Kill the Fat Man?* Princeton University Press.

Engelen, B. (2010). "Open Your Eyes? Why Nozick's Experience Machine Does Not Refute Hedonism." *Film and Philosophy* 14, 33–46.

Feldman, F. (2011). "What We Learn from the Experience Machine." In Ralf M. Bader and John Meadowcroft (ed.), *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*. Cambridge University Press, 59–86.

Hewitt, S. (2010). "What Do Our Intuitions about the Experience Machine Really Tell Us about Hedonism?" *Philosophical Studies* 151, 331–49.

Kamm, F. M. (2015). *The Trolley Problem Mysteries*. Oxford University Press.

Kawall, J. (1999). "The Experience Machine and Mental State Theories of Well-Being." *Journal of Value Inquiry* 33, 381–87.

Kolber, A. J. (1994). "Mental Statism and the Experience Machine." *Bard Journal of Social Sciences* 3, 10–17.

Lin, E. (2016). "How to Use the Experience Machine." *Utilitas* 28, 314–333.

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Nozick, R. (1989). *The Examined Life*. New York: Simon and Schuster.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Shafer-Landau, R. (2018). *The Fundamentals of Ethics*. Oxford University Press.

Silverstein, M. (2000). "In Defense of Happiness: A Defense of the Experience Machine." *Social Theory and Practice* 26, 279–300.

Tännsjö, T. (2007). "Narrow Hedonism." *The Journal of Happiness Studies* 8, 79–98.

Weijers, D. (2011a). "The Experience Machine Objection to Hedonism," In Michael Bruce & Steven Barbone (eds.), *Just the Arguments: 100 of the Most Important Arguments in Western Philosophy*. Wiley-Blackwell. 229–231.

Weijers, D. (2011b). "Hedonism." *Internet Encyclopedia of Philosophy*. URL= https://iep.utm.edu/hedonism/

Weijers, D. (2014). "Nozick's Experience Machine Is Dead, Long Live the Experience Machine!" *Philosophical Psychology* 27, 513–35.

Zangwill, N. (2021). "Our Moral Duty to Eat Meat." *Journal of the American Philosophical Association* 7, 295–311.

2. THE GOOD, THE BAD, AND THE SENTIMENTAL – EXPLORING THE MANY FACES OF VIRTUE ETHICS

Nenad Cekić*

KANT NA RASKRŠĆU DUŽNOSTI I VRLINE? NE.

Apstrakt: Autor u ovom radu pokušava da reinterpretira Kantovu filozofiju morala u duhu savremene etike vrline. Analiza tih pokušaja počinje iznošenjem glavnih stavova pobornika etike vrline, koja se karakteriše i kao „etika zasnovana na delatniku" („agent-based"), „etika zasnovana na motivima („motive-based") i „etika zasnovana na karakternim crtama" („trait-based"). Autor potom prelazi na ekspoziciju „minimuma Kantove ortodoksije" kako bi pokazao granice koje ne sme da pređe bilo koja reinterpretacija. Glavno pitanje glasi: „Šta to etičari vrline u Kantovoj filozofiji pokušavaju da pronađu?" Autor pokazuje da pobornici etike vrline koja je „zasnovana na delatniku i motivima" pokušavaju da u pojmovima moralnog delatnika i određenja same vrline pronađu sličnosti koje ih povezuju sa Kantom. Takvi napori, prema autorovom sudu, nisu plodotvorni jer se Kantova etika, zasnovana na potpuno formalnom kriterijumu „kategoričkog imperativa", ne može dovoljno približiti etici koja u svoj centar stavlja empirijski („sadržinski") određenog moralnog delatnika.

Ključne reči: moralni delatnik, motivi, vrlina, etika vrline, kategorički imperativ

Literatura o Kantu je šarolika i, slobodno se može reći, nepregledna. Zbog toga su se interpretacije Kantove etike tokom istorije menjale u skladu sa određenim obrascima koji su „u trenutnoj modi". Vrlo dugo je školska i široko prihvaćena interpretacija Kantove etike bila prevashodno zasnovana na osnovnim idejama iz njegovog temeljnog etičkog spisa *Zasnivanje metafizike morala*. *Kritika praktičkog uma* je više tretirana kao dopuna celovite Kantove zamisli „kritike" svekolikog saznanja i ideja iznetih u *Zasnivanju* nego kao spis koji bi mogao da služi kao polazna interpretacija Kantovih osnovnih etičkih zamisli. I tu se manje-više „ortodoksna" interpretacija Kantove etike završavala.

* Odeljenje za filozofiju, Filozofski fakultet Univerziteta u Beogradu, *ncekic@f.bg.ac.rs.*

Do pojave savremene etike vrline, „etike koja je zasnovana ili fokusirana na delatniku", malo ko je bio spreman da Kantovo razmatranje vrline koje se u razvijenijem (mada i prilično opskurnom) obliku pojavljuje tek u njegovom poznom spisu, *Metafizici morala*,[1] razmatra kao bilo šta drugo sem kao egzotični dodatak „ortodoksiji" iz *Zasnivanja metafizike* i njegove druge *Kritike.* Međutim, savremeni etičari vrline nastoje da pokažu da supstancijalno određeni delatnik na osnovu vrline baš na osnovu „Učenja o vrlini", drugog dela *Metafizike morala*, može dobiti bitniju ulogu u moralnim procenama i tako ublažiti opšti Kantov formalizam koji moralnu vrednost pripisuje isključivo „bezličnoj" vrednosti postupka.

1. Savremena etika vrline: fokusiranje na delatnika i (ne) određenost postupaka

Da bismo razumeli za čime to etičari vrline tragaju u Kantovim delima, treba bar donekle razjasniti šta sama „etika vrline" u savremenoj filozofiji morala podrazumeva. Pobornici etike vrline svoj pristup često deklarišu kao „etiku motiva". Takvo određenje je donekle nejasno jer se ne vidi šta tačno termin „motiv" (sve) podrazumeva, a može biti i teorijski pristrasno. Naime, u savremenoj terminologiji etike vrline termin „motiv" je, bez posebnih obrazloženja, sužen na emotivnu ili afektivnu stranu delatnikove prirode. Mogući razlog takvog pojednostavljenja je rasprostranjena prihvaćenost tzv. hjumovske teorije motivacije. U toj teoriji, kojoj su posebno skloni uticajni naturalistički orijentisani etičari, zaista se tvrdi da je razum u delanju potpuno inertan. Stoga bi rezervisanje termina „motiv" za emocije i afekte (Hjum kaže „strasti") za pobornike *te* teorije bilo sasvim opravdano.[2] Međutim, teorijska popularnost Hjumovog stava ne bi smela da ima bilo kakve veze sa *značenjem* opšteg i netehničkog termina „motiv". Etimološki, reč „motiv" upućuje na nekakvog „pokretača", bez posebnih specifikacija. Dakle, značenjski gledano, termin „motiv" ničim nije „privezan" za emotivnost ili čulnost. On je, naprosto, oznaka *svega* što

1 *Metafiziku morala*, kao Kantov pozniji spis (napisan 1797. godine), treba razlikovati od mnogo ranije napisanog *Zasnivanja metafizike morala* iz 1785. godine. Kantovi radovi će se citirati kako je to uobičajeno u literaturi o Kantu – na osnovu rednog broja toma i strane u Kantovim *Sabranim delima* Pruske akademije nauka (*Kant's gesammelte Schriften,* vid. bibliografiju). Reference su date u zagradama, a slova ispred broja stranice upućuju na naslov citiranog dela u nemačkom i srpskom izdanju. Korišćeni su i srpski prevodi Kantovih dela navedeni u bibliografiji.

2 Vid. D. Hume, *A Treatise of Human Nature*, [1740], Merchant Books, La Verge, TN USA, 2011, Bk. II, Part III, Sect. III, 305–306. Up. N. Cekić, *Metaetika: problemi i tradicije*, Akademska knjiga, Novi Sad 2013, 116–118.

čoveka može da pokrene na delanje. U motive se, bez ikakve opasnosti od jezičke konfuzije, mogu svrstati i *racionalni* razlozi za delanje.

I kod Kanta, koga bi savremeni etičari vrline rado videli u svom taboru, sasvim se komotno može govoriti o „moralnoj motivaciji". Međutim, nasuprot etičarima vrline, on vrednom smatra samo racionalnu ili, njegovim rečnikom rečeno, „umsku" motivaciju, a ne emotivne pokretače, navikom stvorene nastrojenosti („moralne dispozicije") ili druge „materijalne" podsticaje. Centralni Kantov pojam – pojam dužnosti – može se sagledati isključivo kao *racionalni razlog* za delanje, a ne kao pojam izveden iz osobina čoveka kao prirodnog bića, ma koliko nam se te osobine činile dobrim. Zato nije pogrešno reći da je Kantova „etika dužnosti" zapravo „etika *racionalne* motivacije". S druge strane, današnja etika vrline ne samo da se ne može svesti na racionalnost već ona moralnu vrednost nastoji da zasnuje na „valjanim emocijama" i suštinski empirijskim karakteristikama delatnika. Naime, sami njeni pobornici etiku vrline ne smatraju samo etikom motiva već i „etikom zasnovanom na karakternim crtama" (delatnika). To bi značilo da „dobro" izgrađene *deskriptivne* osobine čoveka kao afektivnog bića (što su „vrline" u uobičajenom smislu) u moralu imaju presudnu ulogu.

Opet nasuprot Kantu, etičari vrline, paradoksalno želeći da ga prisvoje, u osnovi tvrde i da vrednost postupka ne može biti samosvojna već je derivativna: „Pristup etike vrline koji je zasnovan na delatniku (*agent-based*) tretira moralni ili etički status postupaka kao potpuno derivativan iz nezavisnih i fundamentalno aretičkih[3] (kao suprotnost deontičkim)[4] etičkih karakterizacija motiva, karakternih crta ili pojedinaca."[5] Kad se to pomalo nejasno određenje rastumači, onda dobijamo stav da bi ispravni postupci bili postupci tačno određene vrste ljudi sa tačno određenim motivima, u smislu odgovarajućih emotivnih reakcija. To bi opet značilo da u pogledu morala zavisimo od empirijskih svojstava čoveka i situacija u kojima se on zatiče, koje, baš zato što su iskustvene, moraju imati određenu vrstu nasumičnosti. Ključ za razumevanje morala leži i u specifičnim „moralnim emocijama" i dispozicijama. Međutim, one se, kako to i kaže etički „antiteoretičar"[6] Bernard Vilijams (Bernard Williams), naprosto ne mogu *propisati* na osnovu bilo kakve etičke teorije koja obuhvata *objektivnu* proceduru odlučivanja.[7] Ako je to dobra slika morala i moralnog delat-

3 „Vrlinskih" – poteklih iz vrline. Namerno smo iskoristili oblik „aretički", a ne „aretaički".

4 „Dužnosnih" - poteklih iz dužnosti.

5 M. Slote, Agent based virtue ethics, in: R. Crisp, M. Slote, *Virtue Ethics*, 1997, Oxford University Press, Oxford 239–262.

6 Savremeni „antiteoretičari" bliski su etičarima vrline jer smatraju da su emocije u moralu presudne. Međutim, baš zbog toga sistematska etička teorija nije ni moguća.

7 Vid. B. Williams, *Moral Luck*, Cambridge University Press, Cambridge (UK) 1981, Preface, x.

nika, onda za valjano moralno delanje treba „imati moralne sreće (*luck*)". Taj stav ne odgovara Kantovom aprioristićkom viđenju morala, koje podrazumeva da čovek uvek bar načelno može da zna šta mu je činiti i da na osnovu toga, nasuprot „sklonostima" (derivatima želja i emocija), može slobodno *odlučiti* da postupi moralno.

Kao što je rečeno, za Kanta su postupci nosioci moralne vrednosti, a oni svoju vrednost imaju isključivo na osnovu racionalne motivacije koja se ogleda u izboru valjanih principa (maksima) delanja. Nasuprot tome, etičari vrline smatraju da postupci zapravo nisu samostalni nosioci moralne vrednosti već da njihova vrednost proističe iz vrednosti samog delatnika. U tom duhu, na primer, poznati savremeni etičar vrline Majkl Slot (Michael Slote) kaže da standardi i zahtevi koji se stavljaju pred postupke „delaju i obavezuju iznutra".[8] To bi značilo da ispravnost delanja izvire iz vrline kao svojstva samog delatnika. Slot nudi primer benevolentne osobe, pa kaže: „... benevolentna osoba može, kao sredstvo, da odredi da li je neki postupak dozvoljen ili ga treba izvesti, razmatrati da li je njegov postupak motivisan benevolencijom... jer se u okviru moralnosti koja je zasnovana na delatniku, kako bi utvrdio da li neki postupak treba izvesti, vrli delatnik može pozvati na ono što postupak čini ispravnim".[9]

Do sada je sve relativno jasno. Komplikacije nastaju kada Slot dalje kaže da teorija zasnovana na delatniku „dozvoljava da razlozi za postupke obuhvataju same činjenice koje postupak čine plemenitim, zadivljujućim ili ispravnim".[10] Ta dva različita određenja izvora vrednosti postupka – „iznutra" (motivi i vrednost samog delatnika) i „spolja" (činjenice koje postupke čine vrednim) – ne deluju usklađeno. Čini se zato da je Slotovo obrazloženje vrednovanja postupaka putem aretičkih pojmova nejasno, ako ne i konfuzno. Uostalom, čak i sam Slot dalje kaže da nije dovoljno samo reći da je postupak ispravan jer ga je učinio „vrli" čovek. U njima postoji „nešto što ih čini ispravnim". Međutim, ako delatnik mora da se zapita šta to postupak čini ispravnim, onda vrlina nije izvor vrednosti postupka.

Te nedoumice možemo ostaviti samim etičarima vrline, pa nastaviti dalje. Pojednostavljeno rečeno, osnovni stav etičara vrline sastoji se u sugestiji da moralna ispravnost postupaka nekako „izvire" iz „vrlog delatnika", to jest njegovog karaktera kao skupa vrlina. Takav stav je pomalo nejasan jer se suprotstavlja zdravorazumskoj moralnosti koja podrazumeva da čak i zlikovci ponekada mogu učiniti nešto što je ispravno ili dobro. Na to jasno ukazuje i sam Kant (V. G 4.454, ZMM 114). Međutim, pred ta-

8 M. Slote, 244.

9 M. Baron, Ph. Petit, M. Slote, *Three Methods of Ethics*, Blackwell, 1997, 272–273. Citat je iz dela knjige koji je napisao Slot.

10 Posebno je pitanje koliko „ispravno" bez razjašnjenja ide uz „zadivljujuće".

kvom koncepcijom „derivativnosti" vrednosti postupka iz vrline ili karaktera stoji jedan mnogo veći problem, na koji, pozivajući se na Mila (Mill), ukazuje poznati interpretator Kantove etike Alen Vud (Allen Wood). On „problem derivativnosti" (postupaka iz vrline) formuliše na sledeći način: „Pretpostavimo da nekog slavnog čoveka smatramo osobom velike hrabrosti, moralne mudrosti i dobronamernosti – jednim od heroja doba u kojem živimo. Međutim, potom doznajemo da je on plagirao neke od svojih akademskih radova i da je više puta bio neveran svojoj ženi. Još uvek ga možemo smatrati vrlim koliko to ljudsko biće ikada može biti, ali naše istrajavanje na tom sudu ne treba da nas vodi zaključku da su njegovi postupci plagiranja i preljube moralno ispravni."[11]

Slot smatra da bi se, na izneti „milovski" prigovor moglo odgovoriti da delatnik uopšteno može biti čovek od vrline, iako ponekada postupa rđavo. Navodno, samo je potrebno dovoljno pažnje obratiti na motivaciju: „Postupci će se smatrati pogrešnim ili suprotnim obavezama ako pokazuju lošu ili manjkavu motivaciju."[12] Pod „motivacijom", Slot ovde podrazumeva valjan *emotivni* stav. Naravno, sledilo bi pitanje na osnovu čega se, bez pozivanja na postupke koji su po pretpostavci „izvedeni", motivi uopšte mogu *zasebno* procenjivati. Načelni odgovor bi glasio da „etika zasnovana na delatniku" nije doslovno zasnovana na delatniku, već je ona pre „etika zasnovana na *karakternim crtama*" (*trait-based*). Međutim, to značajno komplikuje procene intuitivno očigledne vrednosti postupaka kao što je to, recimo, Kantov omiljeni primer zabrane (moralne osude) davanja lažnog obećanja.

2. Minimum Kantove ortodoksije

Vrlo je teško Kantovo učenje prilagođavati bilo kojoj drugoj etičkoj tradiciji. Razlog pre svega leži u činjenici da je „ortodoksija" Kantovog učenja, obuhvata mnoštvo međusobno suptilno ukrštenih pojmova i specifičnu terminologiju koja je neprevodiva na jezik vrline ili, recimo, jezik utilitarističke korisnosti.[13] Zato ćemo ovde skicirati nekoliko osnovnih Kantovih teza sa kojima svako ko želi da „prisvoji" delove Kantove etike, pa bio to i etičar vrline, mora da se teoretski izbori.

11 A. Wood, Kant and agent-oriented ethics. In: L. Jost, J. Wuerth (eds.), *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*, Cambridge University Press, Cambridge 2011, 62.

12 M. Slote, Agent-based virtue ethics. In: R. Crisp, M. Slote (eds.), *Virtue Ethics*, Oxford University Press, Oxford 1997, 62.

13 Podrazumevamo da je čitalac saglasan sa uobičajenim stavom da normativna etika ima tri glavne tradicije, a to su etika vrline, kantovstvo i konsekvecijalizam (utilitarizam u širokom smislu).

2.1. Moral je forma

Moglo bi se reći da je za Kanta moralno znanje analogno logici, a ne empirijskoj nauci. Kant u objašnjenju morala kreće od jedne jednostavne intuicije koju prepoznaje u „običnom moralnom saznanju". Najkraće rečeno: moralni principe zamišljamo kao *univerzalne* jer moral, kakvim ga mi doživljavamo, jednostavno *ne trpi izuzetke*. Ova intuicija ima jednu na prvi pogled nevidljivu pretpostavku. Naime, striktno logički gledano, „univerzalno" može biti samo ono što nema sadržaj. Šaroliki svakodnevni iskustveni sadržaji na kojima se temelje kantovski čulni „podsticaji" i „sklonosti" ne garantuju nikakvu opštost i nužnost svog pojavljivanja. To dalje znači da celokupno saznanje o moralu ne može imati „sadržaj" već je kantovski „čisto", to jest neiskustveno ili *apriorno* (G 4.391, ZMM 9).[14]

Kako ovo shvatiti? Recimo, logički zakon nekontradikcije zabranjuje sve iskaze tipa koji imaju opšti oblik (formu) „p i ne-p". Međutim, on u sebi ne sadrži ništa „materijalno" (iskustveno, empirijsko). Mi ne moramo znati šta je „p" iako sigurno znamo da nije moguće da važi „p i ne-p". Kako onda znamo da „nikada nije slučaj da p i ne-p"? Tako što se Kantov *Um* može zamisliti kao logička („formalna") mašinerija čiji je osnov zakon nekontradikcije (vid. npr. A 151/B 191). Um „zabranjuje" kontradikcije pa zato *znamo* da „nikada nije slučaj da p i ne-p". Moralna pravila su analogna logičkim jer takođe podležu zahtevima *Uma*. Štaviše, sam „praktički um" nije posebna moć već samo „čisti um u praktičkoj upotrebi". To znači da i sam moralni zakon mora da ima istu formu kao i prirodni zakoni – a to je *forma opštosti i nužnosti*, odnosno *univerzalnosti*. *Um* po sopstvenoj prirodi ima univerzalno pravo „veta" na sve principe koji generišu kontradikcije, pa i na one koje se tiču praktičkog delanja.[15]

2.2. Sloboda je uslov moralnosti i svakog delanja

Iako Kant to nigde ne kaže tako eksplicitno, on podrazumeva da pojam morala, analitički sagledano, *obuhvata* pojam slobode. Štaviše, bez slobode ne bi bilo *nikakvog* (pa ni moralnog) delanja već bi na „svetskoj sceni" bila jedino bezlična fizički nužna „zbivanja". Zato Kant i kaže da je sloboda *ratio essendi* (suštinski preduslov, suštinska pretpostavka) moralnog zakona. S druge strane, čovek biva svestan slobode uvek kada (moralno) *odlučuje*. Kant ovde na umu ima svoju zamisao da se sloboda kao

14 O Kantovom shvatanju moralnog saznanja, više u: N. Cekić, Šta pokazuje Kantov „kompas"?, *Theoria* 4, 2020, 17–35.

15 J. B. Schneewind, Autonomy, obligations and virtue: An overview of Kant's moral philosophy. In: Gayer, P. (ed.), *The Cambridge Companion To Kant*, Cambridge University Press, Cambridge (UK) 1992, 323.

efektivna moć može suprotstaviti čulnim podsticajima kao manifestacijama prirodne nužnosti. Ako *možemo* da obuzdamo sklonosti, onda smo svesni i toga da jesmo slobodni. Stoga je moralni zakon *ratio cognoscendi* (uslov svesti o..., uslov saznanja) slobode (KPV 5.5, KPU 26). Naravno, ta svest o slobodi ipak nije „pravo" već praktičko saznanje jer nam je nepoznat mehanizam funkcionisanja slobode kao uzročnosti koja se „nadmeće" sa fizičkom.

U *Zasnivanju metafizike morala*, Kant slobodu karakteriše pre svega kao „autonomiju". Ideja autonomije ima dva aspekta. Prva je da delatnik može biti slobodan i od podsticaja čulnosti. To je negativna sloboda. Međutim, sama „sloboda od" nije garant mogućnosti moralnog delanja. Čovek lišen uticaja sklonosti može biti inertan. Zato Kant smatra da čovek može i da utvrđuje sopstvene *razloge* („predstave zakona") na osnovu kojih dela. Čovek ne samo da može da se oslobodi uticaja sklonosti već i da pozitivno odredi principe svog delanja.

Jedini način da zaista pokažemo sopstvenu slobodu (*autonomiju* kao *samo*određenje[16]) jeste slobodno podvrgavanje moralnom zakonu. U suprotnom, zarobljeni smo u prirodnoj kauzalnosti. Naša volja tada nije autonomna već *heteronomna*[17] jer njen „zakon" ne određujem ja već *nešto drugo* – podsticaji i sklonosti koji su deo prirodnog lanca uzroka i posledica. Distinkcijom *autonomija–heteronomija* Kant zasebno ukazuje na nužnost etičkog formalizma: ukoliko bi moralni zakon morao da ima neki sadržaj, taj sadržaj bi dolazio iz čula. Ono što dolazi iz čula je „raznovrsnost" kojom vladaju empirijski zakoni, a ne „zakon slobode". To znači da svi „sadržajni" etički principi (npr. hedonizam) nužno vode u heteronomiju.

Kant jasno stavlja do znanja da sloboda nije isto što i proizvoljnost. Istina, jedan deo slobode volje („moć izbora", *Willkür*, *arbitrum*)[18] sastoji se baš u tome što možemo da činimo šta god da nam se prohte. Međutim, *nije* podudarna sa slobodom iz *Zasnivanja* koja podrazumeva poštovanje moralnog zakona. Ipak, ovde stvari ne treba mistifikovati jer je Kantova ideja u osnovi jednostavna: da bismo mogli da delamo moralno, mi moramo biti u stanju i da delamo nemoralno, što je stvar „moći izbora". U suprotnom bismo bili nepogrešivi „moralni automati" (u suštini neka vrsta *stvari*, a ne ličnosti) ili bismo imali nepogrešivu „svetu volju". Svetu volju mogu imati samo bića koja nemaju nikakve sklonosti ili bilo šta analogno „čulnim podsticajima". Bića sa svetom voljom ne mogu moralno da zastrane. S druge strane, čovek koji je izložen podsticajima čulnosti pogrešiv je.

16 Od grč. *autos* – ja, lično i *nomos* – zakon.

17 Od grč. *heteros* – drugi, različiti i *nomos* – zakon.

18 „Moć izbora" iz *Metafizike morala* je termin koji Kant koristi u tom delu, ali ne i u *Zasnivanju*. Razlog je verovatno to što je neophodno razlikovati moralnu slobodu od puke moći da se uradi šta god se poželi.

2.3. „Postupak" podrazumeva postojanje *principa* delanja („maksima")

Potreba za delanjem do volje stiže pomoću tzv. maksima. Šta označava taj Kantov tehnički termin? Maksima je, kratko rečeno, *specifičan* (možda je bolje reći *specifikovan*) subjektivni racionalni princip postupanja.[19] Naime, Kant naprosto podrazumeva da je racionalno delanje uvek delanje prema *nekom* principu. Naši postupci, iako nisu uvek preduzeti iz dobrih razloga, uvek su preduzeti na osnovu *nekih* razloga. Istina, postoje i empirijska „ponašanja", poput emotivnih reakcija, koja imaju fizički uzrok, ali ne i „razlog" ili „princip".[20] Takvo shvatanje ljudskog delanja očigledno se ne može neposredno pomiriti sa delanjem na osnovu vrline kao stečene veštine. U moralu nema „navika" već samo i uvek „odluka".

Maksime kao subjektivni (sadržinski) principi jesu lični planovi postupanja koji obuhvataju delatnikove razloge za postupke, uklopljenost u okolnosti i dovoljno indikacija na koje nas to postupke *Um* poziva. One su ponekada „eliptične" i u određenom smislu elastične. Ta elastičnost se ne odnosi na formu već na okolnosti postupanja. Međutim, maksime, čak i u eliptičnom obliku, uvek specifikuju i *materijalnu* svrhu delanja: „Puna maksima naprosto ovo čini eksplicitnim."[21] Kratko rečeno, kantovski shvaćena maksima treba da pruži jasan *načelan* i *opšti* odgovor na pitanje: „A šta ti to (nameravaš da) (u)radiš i zašto?" Maksima pritom sadrži i informaciju o subjektivnom „stanju delatnika" – često i o njegovim neznanju i sklonostima (G 4.421n, ZMM 60n).[22]

Da bi neki postupak bio ispravan, nije dovoljno da bude u skladu sa moralnim zakonom već i da bar deo delatnikove motivacije bude poštovanje zakona. Da bi imao moralnu vrednost, postupak mora biti učinjen „iz dužnosti", a ne samo „shodno dužnosti". U *Metafizici morala* Kant dodaje da „slaganje jedne radnje sa zakonom dužnosti jeste *legalitet* (legalitas), a slaganje maksime radnje sa zakonom njena *moralnost* (moralitas)" (MS 6.225, MM 27).[23] Da bismo doznali da li je nešto učinjeno „iz dužnosti", neophodan je uvid u motivaciju delatnika. Ali, ona nije dostupna spoljnjem posmatraču (ponekada ni samom delatniku), pa je „legalitet" jedini kriterijum koji se može primeniti „spolja". „Iz sklonosti", ako ne postupamo

19 „Maksima" je za Kanta „subjektivni princip htenja" (G 4.401; ZMM 27), ali i „subjektivni princip postupka" (G 4.421, 425, 429; ZMM 60, 68, 73).

20 Barbara Herman to Kantovo gledište razjašnjava stavom da su ljudski postupci „od vrha do dna" određeni *razlozima* koji su po definiciji opšti. Vid. B. Herman, *The Practice of Moral Judgment*, Harvard University Press, Cambridge (MA), 229

21 J. B. Schneewind, 319.

22 Up. N. Cekić, Šta pokazuje Kantov kompas?, 28–29.

23 Ta razlika je sasvim paralelna razlici legalno–moralno, ali se ti termini pojavljuju tek u *Metafizici morala*, a ne u *Zasnivanju*.

nemoralno, možemo postupati jedino legalno, to jest ne kršeći moralni zakon, što je dozvoljeno ali nema moralnu vrednost. Ipak, za pravu moralnost je nužno poštovanje zakona bez obzira na sklonosti ili *uprkos* njima. To je razlog zbog kojeg Kant kaže da je za vrlinu kao snagu volje koja stremi ispunjenju dužnosti nužna „apatija" (MS 408–409, MM 209–210).

2.4. Moralni *princip* je zakon koji zapoveda kategorički

Hedonizam i etika vrline koji su istorijski prethodnici Kantovoj etičkoj teoriji, svako na svoj način, fokusirani su na pojam „intrinsične vrednosti" (nekog „dobra") koja leži van domena slobode. Intrinsična vrednost zahteva *hipotetičku* moralnost jer se moralna procena svodi na relaciju sredstvo–cilj. Princip je jednostavan: „Ako hoćeš da postigneš dobro, *treba* da učiniš to i to." S druge strane, Kant je ubeđen da se mi, u susretu sa dužnostima, neposredno uveravamo da moralni zahtev *nije* hipotetičan. Dužnost (= ispravnost postupka), kakvom je doživljavamo, sama po sebi je neposredan i presudan razlog za delanje. To bi značilo da nikakvo „dobro" ne prethodi ispravnom kao njegov osnov.

Kad zađemo u Kantovu tehničku terminologiju, onda možemo reći da on razaznaje dva modusa čovekove racionalnosti („umnosti") u delanju. *Volja*, koja obuhvata i „moć izbora", može da nešto „hoće" hipotetički i kategorički. Oba htenja se izražavaju rečju „treba". Svako „treba", smatra Kant, ukazuje na neku vrstu „imperativa". Hipotetički imperativ se odnosi na postupke koji su *uslovljeni* nekim poželjnim (materijalnim, empirijskim) „svrhama", među koje se mogu svrstati i zahtevi *sadržinski* određene vrline. Princip hipotetičkog imperativa osnov je instrumentalne racionalnosti i glasi: „Ko hoće cilj, on hoće i sredstvo koje se... nalazi u njegovoj vlasti" (G 4.417, ZMM 54). Dakle, za hipotetičke imperative nije dovoljno samo maštati o nekom cilju. Naprotiv, hipotetički imperativ kaže da se mora hteti i *sve što je potrebno* da se do njega dođe. Ako delatnik neće sredstva ili hoće neostvarljiv cilj (za čije ostvarenje ne postoje sredstva), onda je on praktički iracionalan.

Kada bi moral bio „hipotetički", isti postupci bi, zavisno od cilja, u jednoj situaciji mogli biti ispravni, a neispravni u drugoj. Kant to smatra i očiglednim i neprihvatljivim. Zato on smatra da osnov morala leži u nekom „treba" koje *ne zavisi* od cilja (svrhe) delanja. Sledi da imperativ moralnosti ne sme biti uslovljen „svrhom", odnosno da on mora važiti „bezuslovno". Takav imperativ se naziva „kategoričkim". Taj imperativ odgovara predstavi o moralu koji „ne dopušta izuzetke", koji važi univerzalno. Iako se to možda ne vidi na prvi pogled, taj stav podrazumeva da imperativ moralnosti mora biti *čisto formalan* jer „kategorično" (tj. bezuslovno i univerzalno) može važiti samo princip koji *nije* oslonjen na empirijski promenljiv „sadržaj".

„Kategoričnost" morala Kant neposredno ilustruje primerom negativnih dužnosti koje se javljaju u formi *zabrane*. Na nemoralne „maksime" (subjektivne principe delanja) reakcija, da pozajmimo taj Kantov izraz, „običnog ljudskog razuma" bila bi jednostavna: „Ne, ti to ne možeš hteti!" Zašto? Zato što nas takav postupak uvek upliće u kontradikcije. Kratko rečeno, sam razum, odnosno *razumevanje* pojma morala, nam kaže: „... ja nikada *ne treba* da postupam drukčije nego tako *da mogu takođe hteti da moja maksima treba da postane jedan opšti zakon*" (G 4.401, ZMM 29). To je u literaturi prilično zapostavljena „negativna" formula kategoričkog imperativa *kao moralnog zakona*. Ona je zapravo naličje najpoznatije „formule univerzalnog zakona": „Postupaj samo prema onoj maksimi za koju istovremeno možeš *hteti*[24] da postane jedan opšti [univerzalni][25] zakon" (G 4.421, ZMM 60).[26]

„Formulu opšteg zakona" Kant proglašava „kanonskim" oblikom jedinstvenog kategoričkog imperativa.[27] Izraženo „može hteti" u formuli opšteg zakona pokazuje da prilikom izbora maksime treba voditi računa o konzistenciji naših maksima i u logičkom i u praktičkom smislu. Sasvim je očigledno da taj „kanon" ne podrazumeva bilo kakvu procenu delatnika ili njegovih ciljeva. Moralni kriterijum se odnosi isključivo i neposredno na formu principa (*maksime*) njegovog postupanja.

2.5. Centralni pojam Kantove etike je dužnost – „ispravnost postupka"

Već smo sugerisali da Kant vrednost pripisuje samim postupcima, i to u vrlo specifičnom smislu racionalne motivacije. Vrednost imaju samo postupci učinjeni „iz dužnosti". Samu dužnost Kant vrlo opskurno određuje kao *nužnost jedne radnje* [*postupka*] *iz poštovanja prema zakonu* (G 4.400, ZMM 27). Pogledajmo šta Kant zapravo hoće da kaže.

U praktičkom postupanju, o kome se u Kantovoj definiciji dužnosti i govori, „nužnost" se ne odnosi na neizbežan kauzalitet prirode već na „*praktičku*, neuslovljenu nužnost radnje [postupka]" (G 4.425, ZMM 67). Sam pojam „dužnosti" ukazuje na umsku prinudnost morala jer čulne sklonosti uvek mogu da zavedu na pogrešan trag. Autor brojnih radova iz oblasti istorije etike Džerom Šnivind taj stav razjašnjava na sledeći način:

24 U srpskom prevodu pogrešno stoji „želeti". „Hteti" i „htenje" su sinonimi za „volju", a ne za „želeti" i „želju".

25 *Univerzalni* je ovde bolji prevod nego samo „opšti" jer spaja „opštost" i „nužnost", što jeste Kantova ideja. „Univerzalno" je ono što nipošto ne dopušta izuzetke.

26 Up. N. Cekić, Šta pokazuje Kantov „kompas"?, 19–20.

27 Kant kaže: „Mora da se *može hteti* da svaka maksima naših radnji postane jedan opšti zakon; to je *kanon moralnog ocenjivanja* naših maksima uopšte" (G 4.424; ZMM 65).

„Kada govorimo o svojoj dužnosti da učinimo nešto, mi mislimo na nužnost tog postupka, bez specifikacije koji je to postupak nužan, nazvati neki postupak dužnošću isto je što i reći da je to postupak koji je obavezan."[28] Dužnosti nisu plod generalizacije iz iskustva već su apriorne, to jest neempirijske ili „čiste" (V. G. 4.419, G 4.391, ZMM 8–9, ZMM 57–58). Moralne obaveze, zato što u svom pojmu obuhvataju logičku zabranu izuzetaka, jesu, kako je i naglašeno, *univerzalne*.

Moral nam se kao „nužnost" i jezički javlja na poseban način – u izrazu „(ne) *treba*". Prema Kantovom uverenju, izraz „treba" u našem moralnom rečniku zauzima centralno mesto jer je baš napetost između slobodnog *Uma* i empirijskih sklonosti i podsticaja (želja, emocija i sl.) centralna za naše moralno iskustvo.[29] Taj izraz, sam po sebi, značenjem nadilazi neko puko (fizički nužno) „jeste". To znači da dužnosti kao izraze „praktičke nužnosti" *ne moramo* da poštujemo već da *treba* da ih poštujemo. *Slobodno* birajući „praktičku nužnost" (umsku *prinudu*!) morala, poništavamo fizičku nužnost koja nas navodi da delamo shodno sklonostima. *Um* kao vrhovni princip reda nalaže nam da delamo prema principima („predstavama") koje *mogu* da važe *kao* zakoni, a ne nasumično. Takvo „uvođenje logičkog reda" aktivnost je samog *Uma*, sam *Um* „na delu". Moral je, dakle, „prinudan", ali ne fizički. On je *samoprinuda* koju nameće *slobodni* Um.

3. Šta savremeni etičari vrline traže u Kantovom učenju?

Savremeni kantovci često naglašavaju da različita Kantova dela treba tumačiti oprezno. Obično se naglašava potreba da se ne pobrkaju „fundacionalna" („zasnivačka") pitanja sa praktičkim normativnim problemima. Jedno je etiku utemeljiti, drugo je „zasnovane principe" primeniti. Zašto bi za etičare vrline to moglo biti važno? Zato što *Zasnivanje* jeste baš to – teorijski temelj koji treba da drži celu konstrukciju Kantove filozofije morala. Kada se temelj učvrsti, onda se navodno može preći na konkretnija moralna pitanja koja su prilagođenija realnim situacijama. Činjenica je tako da u „Učenju o vrlini" iz *Metafizike morala* kategorički imperativ, pogotovo u *Zasnivanju* naglašavana formula opšteg zakona, gotovo nestaje sa vidika. To je, uvereni su pobornici projekta „zbližavanja" Kantove etike sa etikom vrline, znak da se Kantova „primenljiva" etika nalazi baš u „Učenju o vrlini", a ne u *Zasnivanju*. Zaboravljenim marksističkim rečnikom

28 J. B. Schneewind, 318.

29 *Ibid*., 317.

rečeno, u celokupnoj Kantovoj teoriji morala *Zasnivanje* i druga *Kritika* čine „bazu", a *Metafizika morala* „nadgradnju" te teorije. To je vrlo smela teza koja odmah povlači pitanje: „Može li se 'nadgradnja' ikako odvojiti od 'baze' koju čini već izložena Kantova 'ortodoksija'?"

Kako bi uopšte izgledalo to „uranjanje" krutog okvira *Zasnivanja* u „moralnu stvarnost"? Interpretatorka Kantove etike Barbara Herman (Barbara Herman) smatra da pojmovi „svrhe koje su dužnosti" i „dužnosti vrline" koji se pojavljuje u „Učenju o vrlini" mogu da nas navedu na pomisao da je Kantova etika ipak „etika ciljeva".[30] Alen Vud, koji je takođe kantovac, u sličnom maniru dodaje: „Pretpostavimo da u obzir uzmemo mogućnost... činjenicu da su *Zasnivanje*, pa i *Kritika praktičkog uma* samo radovi na zasnivanju etike, a ne izlaganje same Kantove etičke teorije. Onda bismo mogli da iza ovih radova bacimo pogled na *Metafiziku morala* koja bi nam rekla kakva bi etička teorija mogla biti. U tom slučaju, otkrili bismo da su osnovni pojmovi: dužnosti (uske i široke, savršene i nesvršene, dužnosti poštovanja i dužnosti ljubavi, ciljevi (obavezni ciljevi, a to su sopstveno savršenstvo i tuđa sreća) i možda iznad svega vrlina, zamišljena naravno na kantovski način, kao moralna snaga karaktera, jačina dobrih maksima da se obezbedi poštovanje dužnosti... Kant bi čak dozvolio pluralitet vrlina kojem odgovaraju i pluralitet ciljeva i pluralitet maksima dužnosti."[31] Iako ovo nabrajanje nekih Kantovih pojmova naizgled ide u prilog etičarima vrline, ono zapravo ne afirmiše njihove reinterpretacije. Naime, sam Vud jasno naglašava da, i pored svih navedenih „zavodljivih pojmova", nastojanje da se Kant „pretvori" u etičara vrline nema utemeljenja. Bazični formalizam Kantove etike to ne dopušta.

Još uvek nije sasvim jasno za čim sve to u Kantovim delima savremeni etičari vrline konkretno tragaju. Postoje stvari koje su manje-više jasne. Pre svega, iako Kant naglašava nužnu formalnost moralnog zakona, etičari vrline tragaju za *nekakvim* „sadržajem" koji bi navodno ispunio prazninu etičkog formalizma. Iz mnoštva predloga, izdvajamo zamisli *sadržinski* određenog *delatnika* (tip karaktera) i samu *vrlinu* kao deskriptivno svojstvo moralnog delatnika.

3.1. „Vrli" delatnik i Kantova „dobra volja"

Kantovski orijentisana Marša Baron (Marcia Baron) smatra da „vrlog delatnika" treba potražiti u Kantovom pojmu „dobre volje". Podsetimo se, u *Zasnivanju* je „dobra volja" okarakterisana kao „dobra bez ograničenja". Taj zavodljivi izraz često služi kao osnov za tezu da Kant suštinski

30 B. Herman, The difference that ends make. In: *Perfecting Virtue*, str. 95.

31 A. Wood, Kant and agent-oriented ethics, 59

razmatra i *vrstu* moralnog delatnika, a ne, kako on sam često naglašava, isključivo vrednost postupaka.[32] Kada bi ta teza bila tačna, fokus etike bi se donekle mogao pomeriti sa vrednosti postupaka na vrednost delatnika. Ipak, malo je osnova za tvrdnju da pojam dobre volje upućuje na bilo kakvu deskriptivno određenu vrstu delatnika ili karaktera. Naime, Kantova „dobra volja" se mora posmatrati pre svega kao tehnički termin.

Pogledajmo prvo šta Kant tačno kaže o volji uopšte: „Volja se zamišlja kao sposobnost da se sami od sebe odlučujemo na delanje *shodno predstavi izvesnih zakona*" (G 4.427, ZMM 71). „Predstave zakona nisu ništa drugo nego 'principi'. Stoga je volja zapravo isto što i *upotreba* (praktičkog) *uma* i izvođenje principa iz njega" (V. G 4.412, ZMM 46–47). Kant sugeriše da slobodan čovek zapravo nikada ne bira pojedinačne postupke već uvek principe delanja, istina „subjektivne", odnosno „maksime". Postupak, da bi bio postupak a ne „fizičko zbivanje", mora da ima neki *razlog* koji je uvek opšti ili može da bude opšti.

Šta bi onda bila „dobra volja"? Pre svega, treba naglasiti da, i pored Kantovih poetičnih pohvala, „dobra volja" *nije* čak ni bazični pojam njegove teorije. Naime, pojam dobre volje već je ugrađen u nedvosmisleno temeljni pojam *dužnosti* koji „obuhvata pojam dobre volje, mada pod izvesnim subjektivnim ograničenijima i preprekama" (G 4.397, ZMM 21–22). Pojednostavljeno rečeno, dobra volja se razotkriva onda kada neko postupa prema onome što smatra ispravnim ili obaveznim, a to čini *samo zato* što to smatra ispravnim ili obaveznim, bez obzira na empirijska svojstva delatnika ili izazvane posledice.[33] Zato „dobra volja" nije isto što i empirijski motivisana ljubaznost ili velikodušnost. Ona uopšte nije „dobra" u smislu nekakvih „dobrih" karakteristika delatnika već predstavlja „*ispravnu* nastrojenost volje", koja se praktički ogleda u „ispravnom postupanju", to jest *ispunjenju dužnosti* (V. ZMM 36, G 4.406).[34] „Dobra volja" je naprosto akt usvajanja „dobrih" (= ispravnih) maksima. To, na kraju krajeva, znači da taj specifičan Kantov izraz odnosi se isključivo na moralno *delanje*, a *ne* na vrstu moralnog *delatnika* obdarenog vrlinom.

Da dodamo još i ovo. U tumačenju Kantovog shvatanja (dobre) volje treba na umu imati tri različita izraza: „volja", „dobra volja" i „sveta volja". Status samog moralnog delatnika kao *nužno razapetog* između čulnih „partikularija" i umskog naloga univerzalnosti principa može se ilustrovati poređenjem „dobre volje" sa „svetom voljom". „Sveta volja" je volja koja ne bi *mogla* moralno da zgreši. (Zamišljena „sveta bića" po svom određe-

32 M. Baron, Ph. Petit, M. Slote, 39–40.

33 J. B. Schneewind, 325.

34 Ona je „dobra" po sopstvenoj usmerenosti („htenju") (G 4.394, ZMM 17). U delanju „iz dužnosti" *Um* mora da bude „određujući razlog volje".

nju nemaju pogrešive *čulne* opažaje već nepogrešive *intelektualne* opažaje.) Pomoću pojma svete volje može se dalje rasvetliti Kantova ideja dužnosti kao „nužnosti". Na postojanje dužnosti ovde signalizira izraz „treba". Dužnost se u razjašnjenju (kategoričkog) „treba" može odrediti i tako što ćemo reći da šta god da bi sveta volja ili savršeno racionalna volja uradila, mi kao pogrešivi moralni delatnici *treba* da uradimo (G 4.413, ZMM 49; KPV 5.125, KPU 144; MS 6.394–395, MM 196–197). To znači da „čovek dobre volje" nije empirijsko biće *naviknuto* da radi „dobre stvari". „Dobra volja" je produkt *Uma* koji uvek ima *novi* zadatak da savlada sklonosti, a ne uvežbana navika.

3.2. Dva dela morala?

Ovde moramo zastati jer stižemo do pojmova i objašnjenja koji se nalaze samo u *Metafizici morala*, ali ne i u ostalim Kantovim etičkim spisima. Kada u „Učenju o vrlini" iz *Metafizike morala* Kant govori o „dužnostima vrline", on misli na *jednu* vrstu dužnosti, tzv. nesavršene ili široke dužnosti iz ranijih spisa. U tom spisu se pojavljuje podela koje nema u ranijim Kantovim etičkim spisima. Celokupni „moral" obuhvata i „pravo" [Recht] [35] i „vrlinu". Za dva odvojena dela opšteg učenja o moralu karakteristične su dve vrste dužnosti – juridičke, koje su „uske"[36] i usmerene prema drugima, i „dužnosti vrline", koje su „široke".[37]

Domen zakona i prava obuhvata se pre svega maksime koje kategorički imperativ *logički* zabranjuje. Striktna prava drugih ljudi su „naličje" juridičkih dužnosti. Ta prava se mogu iznuditi opravdanom prisilom. Tako mi ne smemo ometati ostvarenje dozvoljenih ličnih ciljeva drugih ljudi, ne možemo im oduzimati imovinu itd. Novost u „Učenju o pravu" je Kantov eksplicitan stav da se u oblasti prava direktno zabranjuju i sami *postupci*, a ne tek maksime. Ipak, ovde ne treba tragati za skrivenim namerama. Naime, maksime koje su protivne pravu primeri su kršenja „uskih" (striktnih, u *Zasnivanju* „savršenih") dužnosti. Njih je najbolje zamisliti kao striktne

35 Pojam „prava" Kant uobičajeno koristi na dvosmislen način – i kao oblast „spoljnjeg zakonodavstva", u kojem postoji i „ovlašćenje prinude", i u smislu individualnih prava, to jest prava na nešto.

36 U *Zasnivanju* tu vrstu dužnosti naziva i „savršenom", ali nije jasno da li je Kant u „Učenju o vrlini" zapravo donekle rekonfigurisao podelu dužnosti iz *Zasnivanja*,ili je čak potpuno izmenio.

37 Nesavršene iz podele koja je navedena u *Zasnivanju*. Nije sasvim jasno da li su „savršene" ili „uske" dužnosti prema drugima uvek i nužno juridičke. Takođe nije sasvim jasno da li su široke dužnosti obavezno i dužnosti vrline. Kršenje savršenih dužnosti vodi u logičke kontradikcije. „Široke" ili „savršene" dužnosti nisu predmet logičke zabrane već se u maksimama koje podrazumevaju njihovo zanemarivanje javlja protivrečnost u *volji*.

zabrane koje ne dozvoljavaju nikakvu „širinu" delanja.[38] *Zbog toga* se juridičke zabrane mogu odnositi i na jasno određene postupke. Ipak, da ne bude zabune, i ispunjenje juridičkih dužnosti deo je morala. Naime, sasvim je sigurno da je kršenje pravnih obaveza – nemoralno.

„Dužnosti vrline" Kant naziva i „svrhama koje su ujedno dužnost". To su „vlastito savršenstvo" i „tuđa sreća". Iako u tom smislu ima mnoštvo pokušaja, „dužnosti vrline" koje su „svrhe koje su ujedno dužnost" ne daju mnogo prostora za razrađivanje argumenata kojima se pobija teza da su te svrhe „materijalne", odnosno da se Kantova etika može pretvoriti i u etiku ciljeva (V. MS 391–394, MM 193–196).

Međutim, čini se da je bitno da uvođenjem „svrha koje su dužnosti" Kant prilično neodređeno redefiniše „nesavršene dužnosti" iz *Zasnivanja* u nešto što bi *moglo* da se shvati kao „cilj". Možda iznenađujuće, on to ne čini u skladu sa idejom „čoveštva kao svrhe po sebi", koja igra značajnu ulogu u *Zasnivanju*. Humanitet je u *Zasnivanju* samo limitirajući faktor delanja i ne odnosi se ni na šta empirijsko. U „Učenju o vrlini" pojavljuje se (tuđa) „sreća", koja je, da podsetimo, za čoveka uopšte *prirodna* (ali ne i moralna) svrha. Međutim, vrednost naizgled empirijskih ciljeva (vlastito savršenstvo, tuđa sreća) nije „intrinsična" već je zasnovana na vrednosti (nesavršenih) dužnosti koja je na njih usmerena. Uostalom, Kant nigde ne poriče vrednost ciljeva, čak ni empirijskih. On samo istrajava na stavu da ciljevi nisu izvor moralne vrednosti.

Česta je zabluda da Kant dopušta izuzetke od nesavršenih dužnosti „u korist sklonosti", odnosno da dozvoljava narušavanje formalizma. Ta zabluda je nastala na osnovu jedne nejasno formulisane fusnote iz *Zasnivanja* i, kasnije, jedne primedbe iz *Metafizike morala*. U njoj izgleda da samo savršene dužnosti zabranjuju izuzetke „u korist sklonosti". Iz toga navodno sledi da nesavršene to dopuštaju. Međutim, kasnija razjašnjenja iz *Metafizike morala* obesmišljavaju takve interpretacije. Ako sklonost i ima neku ulogu, onda se ona odnosi samo na to *na koji ću način* (nesavršenu) dužnost obaviti, a ne da od nje mogu „u korist sklonosti" nekako uteći. „Ako zakon može da naredi samo maksimu radnje, a ne same radnje, onda je to znak da on u sleđenju (pokoravanju) ostavlja prostor (*latitudo*) za slobodan izbor, tj. ne može određeno da navede kako i koliko treba delati za svrhu koja je ujedno i dužnost. Međutim, pod širokom dužnošću ne podrazumeva se dopuštenje za izuzetke od maksime radnji, nego samo ograničenje jedne maksime drugom (npr. opšta ljubav

38 Treba imati na umu i to da u pravu, koje jeste („spoljašnji") deo moralnosti uopšte, nemamo isključivo posla sa zabranama. Na primer, baš umsko „pravo veta" na nepoštovanje zakona koje ima kategorički imperativ nalaže mi da ipak uradim nešto *pozitivno* – da platim porez.

prema bližnjem roditeljskom ljubavlju." On samo ukazuje na to da se, budući da delatniku na raspolaganju stoji spektar mogućnosti ispunjenja tih dužnosti, neke savršene dužnosti mogu „specifikovati" na osnovu drugih nesavršenih dužnosti, na primer, opšta [praktička] ljubav prema bližnjem roditeljskom ljubavlju (MS 6.390, MM 192; Up. G 4.421, ZMM 61). To znači da delatnik ima određenu slobodu *kada i kako* će ispuniti nesavršene dužnosti, ali nikako ne znači da je moralni nalog u slučaju nesavršenih dužnosti „slabiji". U Kantovoj etici zapravo nema mesta „stepenovanju" ili kvantifikaciji moralnosti, iako etičari vrline baš to nastoje u svojim reinterpretacijama Kanta.

I ispunjenju svih dužnosti moralno je vredno samo to što je *Um*, a ne sklonosti ili empirijske osobine delatnika, presudni „motivator". U ispunjenju dužnosti, pa i u ispunjenju širokih „dužnosti vrline", *Um* se uvek potencijalno bori sa podsticajima i trajnijim „sklonostima" koji čoveka mogu da odvrate od moralnosti. Takav stav se etičarima vrline ne dopada. Zato oni kažu: „Postupati iz vrline nije, kako Kant misli, postupanje suprotno sklonostima. Ona je delanje iz sklonosti koje su stvorene kultivacijom vrlina."[39]

3.3. Vrlina kao dispozicija?

Etičari vrline u traganju za Kantovom „etikom vrline" samu vrlinu sagledavaju kao Aristotelovu karakternu vrlinu, to jest kao *moralnu dispoziciju*, uvežbanu veštinu adekvatnog afektivnog reagovanja i delanja.[40] Savremenom tehničkom terminologijom rečeno, vrlina bi morala biti „dispozicionalno svojstvo". To znači da je ona izdržljiva delatnikova tendencija da *oseća* i *dela* na određen način. Da li se ta ideja može prepoznati kod Kanta?

Kant zaista pominje „dispozicije", i to na dva različita načina.

Prvi pojam su opšte „moralne dispozicije", kao empirijske subjektivne sklonosti. One ne pogoduju smislu za kojim etičari vrline tragaju. Naime, kantovske „moralne dispozicije" nisu nužno povezane sa vrlinom. Dispozicija („naklonjenost"), kao *empirijska* stvar, može biti pogrešiva pa voditi vrlini ili nasuprot njoj (KPV 5.84; KPU 104–105). Nasuprot tome, sama vrlina je „čvrsto utemeljen stav da se dužnost ispuni striktno" (REL 6.23, R 23).

Na drugom mestu, Kant samu vrlinu ipak opisuje i kao „moralnu dispoziciju", ali „u borbi" (*im Kampfe*, KPV 5.84; KPU 105). Ta nejasna zamisao donekle se može rasvetliti Kantovim opisom vrline kao prirodno stečene dispozicije volje koja nije sveta (KPV 5.33, KPU 56–57). To

39 A. MacIntyre, *After Virtue* [1981], 2nd ed., University of Notre Dam Press, Notre Dam (IN) 1984, 149.

40 *Loc. cit*.

bi značilo da je vrlina zapravo *vanmoralna* delatna sposobnost koja se ispoljava dok borba sa sklonostima traje. Ona sama po sebi *nije* motiv, ali jeste „ispomoć" u ispunjenju dužnosti. „Vrlina" pritom nije automatizovana veština. Zato Kant u svojoj *Antropologiji* kaže: „Mi ne možemo vrlinu definisati kao *stečenu sklonost* ka slobodnim zakonitim postupcima jer bi onda ona bila puki mehanizam upotrebe naših moći. Naprotiv, vrlina je *moralna snaga* u ispunjenju naših dužnosti, koja nikad neće postati [prirodna] navika već uvek treba da proističe, *uvek nova i originalna*, iz našeg načina mišljenja" (A 7.147).[41] Mi namećemo moralni zakon sami sebi, a zakon čini nužnim obavezu, nužnost da se dela na određene načine. Dakle, moralnost *ne* proističe iz „vrlih dispozicija" ili navika koje nas (na primer) teraju da poželimo da pomažemo drugima.[42] I sasvim ravnodušan i „neutreniran" čovek može biti moralan.

3.4. Vrlina, moral i dužnost

Nije moguće shvatiti šta Kant pod vrlinom podrazumeva bez njenog dovođenja u vezu sa pojmom dužnosti. Pre svega, treba imati u vidu da Kant na raznim mestima varira ideju dužnosti kao prisile. Taj stav se često naglašava tvrdnjom da ljudska volja nije „sveta". Čovekova „ne-sveta" volja je suočena sa izazovima čulnosti (sklonostima i podsticajima) kojih „nadljudsko" biće nema. To je razlog zbog kojeg se moral čoveku pojavljuje kao prisila. U *Metafizici morala*, čini se, (samo)prisila se posebno naglašava i gotovo izjednačava sa vrlinom. Naime, tu i nalazimo određenje da je vrlina „moralna prisila putem čovekovog sopstvenog zakonodavnog uma, ukoliko sam um konstituiše silu koja *sprovodi* zakon" (MS 6.405, MM 206). Iz toga bismo mogli zaključiti da *Um* ne samo da zabranjuje nemoralne maksime već i delatno *podstiče* na moralno delanje. Međutim, i dalje nije jasno kako vrlina kao „sprovođenje zakona" zapravo funkcioniše. Šta je vrlina u odnosu na dužnost koja je osnov morala?

Nije naodmet vratiti se malo unazad. Može delovati iznenađujuće to što se još u *Zasnivanju* može pronaći jedno malo pominjano određenje vrline koje vrlinu gotovo da izjednačava sa samim moralom: „Ugledati vrlinu u njenom pravom obliku ne znači ništa drugo nego predstaviti moral u njegovoj potpunoj čistoti, to jest kao lišenog svake primese čulnosti i svakog lažnog ukrašavanja nagrađivanjem ili lišenog svake primese samoljublja" (G 4.426, ZMM 69n). Budući da se moral za Kanta nedvosmisleno svodi na ispunjenje dužnosti, ne vidi se po čemu bi se vrlina kao samoprisilno (ali slobodno) ispunjenje dužnosti uopšte razlikovala od samog

41 Up. M. Baron, *Kantian Ethics almost without Apology*, Cornell University Press, Ithaca (NY) 1995, 79.

42 Vid. J. B. Schneewind, 310.

moralnog *delanja*. Uostalom, moral je, to već iz *Zasnivanja* znamo, i autonomija, to jest „samozakonodavstvo", koje prinudno sputava sklonosti i čulne podsticaje. Šta je u „prisilnosti" vrline iz *Metafizike morala* u odnosu na „prisilnost" same dužnosti iz *Zasnivanja* zapravo novo ili različito, ne vidi se tačno.

Nešto jasniju sliku vrline možemo pronaći u njenom određenju kao „moralne snage volje", tj. „moralne snage *čoveka* u ispunjenju njegove *dužnosti*."[43] Ovde bismo dužnost mogli da vidimo kao potencijal za ispravno postupanje, a vrlinu kao snagu koja taj potencijal realizuje. Međutim, ta snaga nije nikakav kontinuitet navike (dispozicija) jer, da bi postupak bio moralan, delatnik uvek mora biti slobodan da odlučuje, pa i da pogreši. U tom smislu, vrlina bi se i ovde možda mogla shvatiti kao nekakva empirijska pomoć u ispunjenju dužnosti, ali nije jasno kako bi ta pomoć uopšte bila *moralno* relevantna. Jedini izvor moralne vrednosti je *slobodno umsko određenje volje*, a ono se uopšte ne odvija u sferi empirijskog. Samo bi u empirijskoj realnosti navike ili svojstva delatnika mogla biti relevantna, ali moralno odlučivanje („određivanje volje") nije empirijsko zbivanje.

3.5. Vrlina kao „snaga maksime"

Još uvek ima vrlo uticajnih interpretatora koji jasno sugerišu da je Kant u *Metafizici morala* odstupio od minimalizma koje je sugerisano u *Zasnivanju*. Argument je sledeći. Domen prava i domen vrline razlikuju se po tome što delanje iz vrline podrazumeva da ili delam u korist drugih ili radi sopstvenog usavršavanja. U domenu prava nije bitno šta ja radim, sve dok ne narušim prava drugih. U domenu prava nema moralne zasluge, a delatnik može biti prinuđen da obavi dužnost. Čini se da Kant ovde podrazumeva da su *sve* savršene (striktne, uske) dužnosti *prema drugima* zapravo „juridičke", to jest podložne spoljnom propisivanju. Na njihovo ispunjenje mogu biti prinuđen.[44] U oblasti prava, ispunjavajući dužnost „samo ne prljam ruke". S druge strane, u domenu vrline dužnost nas „navodi na svrhe", ali ni na šta ne mogu biti „spolja" prinuđen (MS 6.381, MM 183). Mi autonomno biramo „svrhe koje su dužnosti" kojima odgovaraju „dužnosti vrline". Zbog toga je ispunjenje „dužnosti vrline" zasluga.

---

43 Pregled svih Kantovih različitih određenja vrlina može se naći u A. Wood, Kant and agent-oriented ethics, 69.

44 U *Metafizici morala* postoji zbunjujuć detalj. Dužnost da se ne laže i na njoj druge zasnovane dužnosti u *Zasnivanju* i *Kritici praktičkog uma* korišćene su kao *paradigmatični* primeri „savršenih dužnosti *prema drugima*" (vid. npr. G 4.420–422, ZMM 61–63). Na ovom mestu Kant najavljuje jasnije razvrstavanje dužnosti koje ostavlja za buduću *Metafiziku morala*. Međutim, u *Metafizici morala* zabrana laganja se pojavljuje kao „potpuna (= savršena) dužnost *prema sebi*" (MS 6.429–431, MM 229–231) .

Zato se te dužnosti nazivaju (uz ostala imena) i „zaslužnim". Međutim, to ne znači da je ispunjenje juridičkih dužnosti na osnovu umske motivacije bezvredno. Budući da je striktno određeno, da je određeno na jedinstven način, ono se samo ne može stepenovati prema težini savladanih prepreka, odnosno „zasluga".

Sada treba imati na umu da su dužnosti vrline („svrhe koje su dužnosti" – vlastito savršenstvo i tuđa sreća) „nesavršene". To znači samo to da se one mogu ispunjavati na *više* različitih načina. Prema kome i kako treba biti dobrotvor, ne može se unapred utvrditi. Takođe, nije moguće utvrditi dokle se i kako može napredovati u sopstvenom samousavršavanju. Na primer, prilikom usavršavanja sopstvenih talenata treba uzeti u obzir i to da neko može imati više talenata, više načina za usavršavanje, više prilika i sl. (up. MS 6.392, MM 194). Ovde bi se moglo naći mesta za zamisao da je vrlina zapravo svojstvo izabrane maksime. Naime, budući da se „dužnosti vrline" mogu ispunjavati na nebrojeno mnogo različitih načina, među njima se mogu naći lakši i teži putevi. U večitoj borbi sa sklonostima, izabrati teži put (zahtevniju maksimu) znači pokazati vrlinu.

Ovde stižemo do određenja da vrlina zapravo nije dispozicionalno svojstvo delatnika. Ona postaje „svojstvo principa (*maksime*) na osnovu koje delatnik postupa (ili delatnikovog podsticaja u delanju shodno njemu), čak i onda kada delatnik postupa iznenadno i izuzetno nekarakteristično".[45] Takvo shvatanje vrline kao konkretne „jačine" *slobodno* odabranog principa, podrazumeva da ona ne može sama po sebi da bude deo dužnosti pa, samim tim, ni morala. Uostalom, i sam Kant kaže: „Ona [vrlina] sama, ili imati je, nije dužnost, zato što bi onda moralo postojati obavezivanje na dužnost. Nego, ona zapoveda i prati svoju zapovest moralnom (prema zakonima unutrašnje slobode mogućom) prinudom. Ali, budući da ona treba da bude neodoljiva, za to je potrebna snaga koja se može proceniti samo na osnovu stepena veličine prepreka koje čovek svojim sklonostima sam sebi stvara. [...] Poroci jesu... čudovišta protiv kojih se treba boriti" (MS 406, MM 206). Baš u borbi sa porocima koji su plod strasti vrlina se pokazuje kao snaga i u tom smislu je *jedinstvena* – borba protiv poroka ne zavisi od vrste poroka već od snage delatnika. „Moralna snaga", kaže dalje Kant, u saglasnosti je sa antičkom tradicijom, *ujedno* je i hrabrost i mudrost.[46] Ako se ona zamisli kao dovršena, što je empirijski nemoguće, onda „čovek ne bi posedovao vrlinu, već vrlina čoveka" (MS 407, MM 207). Ovo bi se moglo rastumačiti kao stav da je vrlina *ideal* snage volje u borbi sa nemoralnim sklonostima.[47]

45 A. Wood, 61.

46 Kant ovde verovatno misli na Aristotelovu „praktičku mudrost", grč. *phronesis*.

47 Nisu sve sklonosti za Kanta rđave. One jednostavno nisu osnov morala. Štaviše, sreća, koja je izvor podsticaja i sklonosti, jeste i „prirodna svrha" koja sama po sebi nije ni

Posedovati vrlinu, dakle, uopšte nije dužnost jer je samo na osnovu nekog stepena vrline moguće staviti sebe pod samoprinudu dužnosti (MS 6.405, MM 206). Vud te Kantove reči tumači tako što kaže da, baš kao što nemate striktnu („usku", „strogu", „savršenu") dužnost da imate vrlinu u što većem stepenu, nemate ni *striktnu* („savršenu") dužnost da činite zaslužne postupke koji su predmet „dužnosti vrline".[48] Ispunjenje tih dužnosti je stvar slobode.

Kant dozvoljava da se govori o stepenovanju *vrline*, iako sam *moral* zapravo stepenovanje ne dozvoljava. Srž Kantove etike sastoji se u stavu da neki postupak jeste ili nije moralan (istina, može biti i moralno irelevantan), ali suštinski nema smisla reći da je jedan postupak „moralniji" od drugog. Kad *Um* odredi volju – postupak ima *moralnu* vrednost. S druge strane, „veća" vrlina je usavršavanje volje (koja je podložna i sklonostima), pa zato postoji široka ili zaslužna dužnost da se u tom pogledu usavršavamo (MS 6.446, MM 247). Međutim, nema nikakve striktne („uske") dužnosti da se postigne bilo koji specifičan stepen vrline. Sve to je zapravo naznaka shvatanja vrline kao empirijskog „pomoćnika" morala u vidu ispunjenja dužnosti. Da bi se postupilo moralno, nije nužno prethodno se uvežbavati u vrlini, iako to može biti od koristi za samog delatnika. Štaviše, ako bi se delatnik oslonio samo na vrlinu shvaćenu kao empirijska navika, onda njegov postupak ne bi imao moralnu vrednost (mada ne bi bio nemoralan) jer bi bio suštinski *heteronoman*.

Vrlina je, videli smo, i „samoprinuda" na osnovu slobodne izabrane valjane maksime. Šta to tačno znači? Čovek je, za razliku od Boga, pogrešivo biće. Zato se moral pojavljuje u obliku obaveze, to jest slobodne prisile na delanje. „Sveta volja" po svojoj prirodi ne može ni da pogreši, pa zato ne može imati „vrlinu" jer za prisilu mesta nema. Pogrešivo ljudsko biće može pokazivati samo vrlinu kao snagu u borbi sa strastima. Za to je potrebna „apatija", odnosno „bezafektivnost". Vrlina je uvek „napredovanje", ali, budući da *Um* u moralu mora biti zakonodavac, ona ne počinje od zatečenog već uvek „počinje ispočetka" (MS 405–407, MM 206–208). Zato nas Šnivind i upozorava: „Samo bića koja smatraju da je biti moralan teško i koja razviju otpornost u borbi sa izazovima mogu imati vrlinu. Mi kao konačna bića nikada nećemo stići do tačke u kojoj nam neće biti potrebna snaga kako bismo se oduprli željama. Mi nismo ni anđeli ni životinje. Vrlina je naše pravo mesto unutar univerzuma."[49]

---

moralna ni nemoralna (MS 6. 389, MM 190). Kant se tog stava dosledno drži u svim svojim glavnim etičkim spisima (V. G 4.415, 4.430; ZMM 51, 76; KPV 5.25; KPU 47–48).

48 Vid. A. Wood, How a Kantian decides what to do. In: M. C. Altman (ed.), *The Palgrave Kant Handbook*, Palgrave Macmillan, Ellensburg (WA) 2018, 273.

49 J. B. Schneewind, 318. (vid. MS 6.405–9)

4. Umesto zaključka: Kant i Aristotel

Mnogo je različitih načina na koje su razni filozofi pokušali da „dovuku" Kanta u svoj tabor. „Konsekvencijalizaciju" Kantove etike pokušao je još Mil,[50] potom i neki savremeni konsekvencijalisti. Etičari vrline imaju sličnu nameru. Pa ipak, Kantova izvorna etika naprosto ne popušta pred takvim pokušajima. Da bi se Kant pretvorio u nešto drugo, neophodno bi bilo prebrisati sve što je on napisao o „sadržinskim" etikama, u koje se mogu ubrojati i klasični utilitarizam i savremena etika vrline. Selektivno biranje pojmova nije dobar metod interpretacije. Reinterpretatori Kantove etike, ma kako nastrojeni, uvek mogu naići na iznenađenja. Primera radi, etičarima vrline o kojima smo govorili teško pada Kantov stav da, iako postoje dužnosti vrline, niko nema obavezu da poseduje samu vrlinu. Tako je i sa drugima pojmovima koji navodno vode ka centralnosti vrline. Alen Vud nas zato i upozorava: „Bez sumnje... htenje, dobra volja, maksime i moralna dispozicija jesu moralno relevantna svojstva delatnika; ali, nijedna od njih sama po sebi nije vrsta svojstva za koju je etika vrline obično zainteresovana."[51] Razlog za to je činjenica da Kant nigde ne povezuje težnju ka moralnom savršenstvu sa empirijskim delatnicima. Savršenstvo se ne nalazi u empirijskom delatniku već u principu koji glasi: „Najveće moralno savršenstvo čoveka jeste: čini svoju dužnost i to iz dužnosti" (MS 392, MM 194). To je linija argumentacije koja prati argumente iznesene u *Zasnivanju*, u kojem se jasno razlikuju postupci izvedeni „iz dužnosti" i postupci koji su samo „u skladu sa dužnošću".

Možda najjači argument za odbacivanje mogućnosti pretvaranja Kanta u etičara vrline leži u činjenici da je on sam odbacio mogućnost približavanja Aristotelu, koga njegovi savremeni tumači ujedno i prisvajaju i reinterpretiraju. Istina, postoje delovi *Metafizike morala* u kojima Kant zvuči aristotelovski. Može se, primera radi, učiniti da Kant u svom kratkom osvrtu na Aristotela dozvoljava da vrlina ne mora biti samo „stvar principa" već i „stvar delatnika". On, na primer, kaže: „Moralna vrlina kao veština [i navika] (habitus) jeste lakoća delanja i subjektivno savršenstvo moći izbora" (MS 6.407, MM 208). Ta rečenica sasvim podseća na Aristotelove pasaže u kojima se (karakterna) vrlina shvata kao vrsta navikom steknute veštine (težnje, dispozicije, grč. *hexis*).[52] Veća vrlina podrazumeva veću unutrašnju snagu delatnika u odupiranju izazovima sklonosti koje ga zavode na kršenje dužnosti. Zato „stanju vrline ono što je inače teško

50 Vid. J. S. Mill [1862], *Utilitarianism*, Hackett, Idianapolis (IN) 2001, 4 (poglavlje i pasus 1.4.)

51 A. Wood, 61.

52 Aristotel, *Nikomahova etika*, 1103a–1105.

postaje lako". Međutim, Kant dodaje i da vrlina nije „puka navika" u smislu „jednoobraznosti delanja koje je postalo *nužnost*[53] pomoću ponavljanja" (MS 6.407, MM 208). Vrlina je uvek i snaga, a snaga se meri sposobnošću savladavanja otpora. „Postupati moralno po navici bio bi gubitak jer bi bio gubitak slobode" (MS 409, MM 210). Ako je, kako Aristotel to smatra, vežbom stečena „navika", onda je ona, insistira Kant, *slobodna navika*.

Ni sam Aristotel karakternu vrlinu ne vidi kao puku naviku. Zapravo, i Kant i Aristotel smatraju da se vrlina razotkriva u *racionalnim* postupcima koji su preduzeti jer se procenjuju kao sami po sebi vredni.[54] Kod Aristotela, za karakternu vrlinu potrebno je i dejstvo posebne *saznajne* vrline, praktičke mudrosti (grč. *phronesis*) kao racionalne moći procene. S druge strane, Kant sasvim odbacuje čulnost kao izvor morala. (Ona mu samo stavlja prepreke.) Za Aristotela vrlina mora podrazumevati postojanje neke želje za ostvarenjem valjanih svrha, koju prati zadovoljstvo ili bol.[55] Kod Kanta takve motivacije nema. Kantova vrlina je snaga *u borbi* sa sklonostima. Međutim, reč je o borbi sa sklonostima koje se *opiru* dužnosti. Štaviše, Kant ne kaže da vrlina nikada ne obuhvata *i* delanje iz simpatetičkih sklonosti jer neke od njih uvećavaju naš kapacitet da ispunjavamo dužnosti, pa pripadaju ili bar pripomažu vrlini u njenoj „borbi" (vid. MS 6.456, MM 258). Pohvaljive su i vrlina i „dobre" dispozicije i karakterne crte, ali ne u *moralnom* smislu jer mi nismo odgovorni za svoj prirodni karakter. Međutim, Kant je potpuno siguran u to da mi jesmo *slobodni*. Zato je do nas, kako to kaže Marša Baron, da *delamo*, a ne da *menjamo sebe*.[56]

Bibliografija

Aristotel, *Nikomahova etika*, prev. Radmila Šalabalić, BIGZ, Beograd 1980.

Baron, M., *Kantian Ethics almost without Apology*, Cornell University Press, Ithaca (NY) 1995.

Baron, M., Petit, Ph., Slote, M., *Three Methods of Ethics*, Blackwell, Oxford (UK) 1997.

Cekić, N., *Metaetika: problemi i tradicije*, Akademska knjiga, Novi Sad 2013.

Cekić, N., Šta pokazuje Kantov „kompas"?, *Theoria* 4, 2020, 17–35.

Herman, B., The difference that ends make". In: L. Jost, J. Wuerth (eds.), *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*, Cambridge University Press, Cambridge (UK) 2011.

53 Naš kurziv.

54 A. Wood, 70.

55 Aristotel, 1104–1105.

56 M. Baron, *Kantian ethics*, 79.

Herman, B., *The Practice of Moral Judgment*, Harvard University Press Cambridge (MA) 1993.

Hume, D., *A Treatise of Human Nature* [1740], Merchant Books, La Verge (TN) 2011.

Kant, I., *Kant's gesammelte Schriften, herausgegeben von der Deutschen /Könliglichen Preuissischen Akademie der Wissenschaften*, 29 vol, Berlin: Walter de Gruyter et al., 1902ff. G – *Grundlegung zur Metaphysik der Sitten* (1785), KPV – *Kritik der praktischen Vernunft* (1788), REL – *Die Religion innerhalb der Grenzen der blossen Vernunft* (1793), MS – *Die Metaphysik der Sitten* (1795), A – *Anthropologie in pragmatischer Hinsicht* (1798). Navodimo i dostupne prevode na koje se odnose reference u tekstu; ovi prevodi su konsultovani, ali obično nisu doslovno citirani:

Kritika praktičkog uma (KPU), prev. Danilo Basta, BIGZ, Beograd 1990.

Zasnivanje metafizike morala (ZMM), prev. Nikola Popović, 2004.

Metafizika morala (MM), prev. Danilo Basta, Izdavačka knjižarnica Zorana Stojanovića, Novi Sad 1993.

MacIntyre, A., *After virtue* [1981], 2nd ed., University of Notre Dam Press, Notre Dam (IN) 1984.

Mill, J. S., *Utilitarianism* [1862], Hackett, Indianapolis (IN) 2001.

Religija unutar granica čistog uma (R), prev. Aleksa Buha, BIGZ, Beograd 1990.

Schneewind, J. B., Autonomy, obligations and virtue: An overview of Kant's moral philosophy. In: P. Gayer (ed.), *The Cambridge Companion To Kant*, Cambridge University Press, Cambridge (UK) 1992.

Slote, M., Agent-based virtue ethics. In: R. Crisp, M. Slote (eds.), *Virtue Ethics*, Oxford University Press, Oxford (UK) 1997.

Williams, B., *Moral Luck*, Cambridge University Press, Cambridge (UK) 1981.

Wood, A., Kant and agent-oriented ethics. In: *Perfecting virtue...*

Nenad Cekić

Kant at the crossroads of duty and virtue? No.

Abstract: This paper deals with some attempts to reinterpret Kant's moral philosophy in the spirit of modern virtue ethics. The analysis begins with the presentation of the central claims of the supporters of virtue ethics, which is also characterized as „agent-based," „motive-based," and „trait-based" ethics. The author then exposes the „minimum of Kant's orthodoxy" to show boundaries that any reinterpretation must not exceed. The main question is: „What are virtue ethicists in Kant's philosophy trying to find?" The author shows that protagonists of

the „agent-based" (or „motive-based") based virtue ethics in the notions of the moral agent and the definition of virtue seek similarities between Kantian ethics and their own approach. Such efforts, according to the author's judgment, are not fruitful because Kant's ethics, based on the utterly formal criterion of the „categorical imperative," cannot be made sufficiently close to the ethics that put at their center an empirically determined („material") moral agent.

Keywords: moral agent, motives, virtue, virtue ethics, the categorical imperative

Marijana Kolednjak*

MARTHA NUSSBAUM AND VIRTUE ETHICS

Abstract: Martha Nussbaum argues that the current tendency to teach that there is any single approach such as "virtue ethics" is a big mistake. This is, first of all, a category error of an elementary kind, given that many people write and think about virtue within the Kantian and utilitarian traditions. Virtue ethics cannot, therefore, be an alternative to these traditions.

Keywords: Martha Nussbaum, Aristotle, Kant, utilitarianism, virtue ethics

Introduction

Martha (Craven) Nussbaum (b. 1947) is the Ernst Freund Distinguished Service Professor of Law and Ethics, appointed in the Philosophy Department and the Law School of the University of Chicago. In 2015, she was awarded the Inamori Ethics Award for exemplary ethical leadership. It is an award that pays tribute to outstanding international ethical leaders whose actions and influence have greatly improved the condition of humanity. She has received both the 2016 Kyoto Prize in Arts and Philosophy, regarded as the most prestigious award available in fields not eligible for a Nobel, and the 2017 Don M. Randel Award for Achievement in the Humanities from American Academy of Arts and Sciences. In 2018, she received the Berggruen Prize for Philosophy and Culture, an award to thinkers whose ideas have deeply shaped human self-understanding and progress in a rapidly changing world. Motivated by the desire to understand the conditions for wellbeing in light of the complexity of human existence, she has used the power of literature to reveal and explore the central place of the emotions: vulnerability, anger and fear in moral and political life ("Berggruen Prize", internet). In 2021, she was awarded the

* University of Zagreb, Faculty of Philosophy and Religious Studies, marijana.kolednjak@ffrz.hr

Holberg Prize for her groundbreaking contribution to research in philosophy, law and related fields. The Holberg Prize is awarded annually to a scientist who has made an outstanding contribution to research in the humanities, social sciences, law or theology (either in one of these fields or through interdisciplinary work).

Nussbaum is an American philosopher and legal scholar known for her wide-ranging work in ancient Greek and Roman philosophy, the philosophy of law, moral psychology, ethics, philosophical feminism, political philosophy, the philosophy of education, and aesthetics and for her philosophically informed contributions to contemporary debates on human rights, social and transnational justice, economic development, political feminism and women's rights, LGBTQ rights, economic inequality, multiculturalism, the value of education in the liberal arts or humanities, and animal rights (Duignan, internet).

Nussbaum has written dozens of books, including *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy* (1986, updated edition 2000), *Love's Knowledge* (1990*), Sex and Social Justice* (1999), *Upheavals of Thought: The Intelligence of Emotions* (2001), *Hiding From Humanity: Disgust, Shame, and the Law* (2004), *Not For Profit: Why Democracy Needs the Humanities* (2010), *Anger and Forgiveness: Resentment, Generosity, Justice* (2016), *and Aging Thoughtfully: Conversations about Retirement, Romance, Wrinkles, and Regret* (with Saul Levmore, 2017), *The Monarchy of Fear: A Philosopher Looks at Our Political Crisis* (2018).

She believes that the purpose of philosophy is to give a concrete contribution and applicable answers to the questions and challenges that citizens, ordinary people, face in their daily lives. She tries to achieve this in her philosophy as well, and the fact that she has been recognized by society is evidenced by the many awards she has been given outside of academia.

The Kantian approach vs the Utilitarian approach

Nussbaum is one of many contemporary philosophers who endeavor to reanimate the Aristotelian idea of virtues.[1] She is very critical of the

1 Whoever deals with Aristotle's theory of virtues cannot neglect Martha Nussbaum as one of the most authoritative interpreters of Aristotle. For Aristotle, each of the virtues is an organized way of cherishing a particular end that has intrinsic value. Taken together, the virtues, and their orderly arrangement, represent a set of commitments to cherish all the valuable things, and to organize them all together, insofar as one can. Courage, justice, moderation – all these virtues deal with our need for externals; that is why, as Aristotle said, we cannot imagine needless gods having the virtues

modern project of Enlightenment especially promoted by Immanuel Kant (Adiprasetya, 2016: 2). In her article *Virtue ethics: a misleading category*? – what will be the basis of this content – she writes that virtue ethics is standardly taught and discussed as a distinctive approach to the major questions of ethics, a third major position alongside Utilitarian and Kantian ethics. She argues that this taxonomy is a confusion because both Utilitarianism and Kantianism contain treatments of virtue. Hence, virtue ethics cannot possibly be a separate approach contrasted with those approaches (Nussbaum, 1999: 163).

One group of modern virtue-theorists, Nussbaum argues, are primarily anti-Utilitarians, concerned with the plurality of value and the susceptibility of passions to social cultivation. These theorists want to enlarge the place of reason in ethics. They hold that reason can deliberate about ends as well as means, and that reason can modify the passions themselves. Another group of virtue theorists are primarily anti-Kantians. They believe that reason plays too dominant a role in most philosophical accounts of ethics, and that a larger place should be given to sentiments and passions – which they typically construe in a less reason-based way than does the first group (Nussbaum, 1999: 163).

Philosophy, or better yet ethics is turning from an ethics based on Enlightenment ideals of universality to an ethics based on tradition and particularity; from an ethics based on principle to an ethics based on virtue; from an ethics dedicated to the elaboration of systematic theories to an ethics suspicious of theory and respectful of the wisdom embodied in local practices; from an ethics based on the individual to an ethics based on affiliation and care; from an ahistorical detached ethics to an ethics rooted in the particularity of historical communities (Nussbaum, 1999: 164).

Nussbaum claims that "virtue ethics" is regularly presented as a major genus of ethical approach. In the typical class in medical ethics in the US, for example, young doctors learn that there are three approaches to deciding an ethical question: the Kantian[2] approach, the Utilitarian approach, and the "virtue ethics" approach. A similar trichotomy increasingly makes

---

(Nussbaum 2001:373). Aristotle objected that in the ideal city, where there was no private ownership and therefore no inequality of property, there would be no room for the moral virtue of generosity. Against this idea, Kant's point is a forceful one: what need do we have for these moral virtues themselves, if their role in human life is simply to correct a bad state of affairs and if we can and should correct the bad state of affairs antecedently, by means of laws? (Nussbaum 2001: 404). Aristotle exercises surveillance over every aspect of life, whereas Kant lets the passions go, so long as they don't interfere with the will.

2 Kant thought that virtue must always be a matter of strength, as the will learns to keep a lid on inappropriate inclinations, rather like a good cook holding down the

its appearance in high-level works of academic moral philosophy (Nussbaum, 1999: 164). She expressly claims that this increasingly popular way of talking is an obvious category mistake. Immanuel Kant has a theory of virtue, and devotes a great deal of attention to its exposition. Although The Doctrine of Virtue was at one time a relatively neglected part of Kant's moral philosophy, read only by specialists, it is now widely discussed, and widely recognized as central. Nobody can any longer think of Kant's view as obsessed with duty and principle to the exclusion of character-formation and the training of the passions. We are well aware that he offers a general account of virtue, in terms of the strength of the will in overcoming wayward and selfish inclinations; that he offers detailed analyses of standard virtues such as courage and self-control, and of vices, such as avarice, mendacity, servility, and pride; that, although in general he portrays inclination as inimical to virtue, he also recognizes that sympathetic inclinations offer crucial support to virtue, and urges their deliberate cultivation. In short, his account of virtue covers most of the same topics as do classical Greek accounts. Although the substance of his theory of virtue differs from the Greek theories, particularly in its non-cognitive account of passion, it is a theory about the same things, and it bears sustained comparison with those theories. Moreover, the rediscovery of Kant's theory of virtue has also led to serious reevaluation of the substantive positions of his other ethical writings, as scholars depict a Kant who is less rigorist and more flexible, less concerned with abstract principle and more concerned with the exercise of moral judgment, than the Kant of previous generations (Nussbaum, 1999: 165).

As for the British Utilitarians (Henry Sidgwick's *The Methods of Ethics* (1907), Jeremy Bentham's *The Principles of Morals and Legislation* (1780), John Stuart Mill's *The Subjection of Women* (1869)), they do not neglect the virtues either: virtue is among its primary topics. Thinking about Utilitarianism is ultimately the best way of understanding the common notions of virtue. Within the Utilitarian tradition there were extensive discussion of the extent to which Utilitarians should teach the Utilitarian principle to the young, as opposed to the (closely related) principles of common-sense virtue. It would be quite implausible to oppose virtue ethics to Utilitarianism, given the fact that two of the three major Utilitarian thinkers set themselves squarely in the ancient Greek virtue-theoretic tradition, and owe a considerable positive debt to earlier analyses of virtue, in both cases using these analyses to propose improvements in the social thinking of their time.

---

lid on a boiling pot rather like a good cook holding down the lid on a boiling pot (Nussbaum 2001:172).

How, then, could "virtue ethics" be a thing on its own, opposed to Kantianism and Utilitarianism, when it is so obviously an important element of both of those ethical theories, as it is also a department inside other ethical theories, such as those of Thomas Aquinas and Aristotle? (Nussbaum 1999: 166).

"Virtue ethics"

What, if anything at all, "virtue ethics"[3] is? Nussbaum[4] argues that there is some genuine unity to the set of concerns that led all these thinkers, and many others, to take an interest in the category of virtue, and to turn to the Greeks, as many have, for illumination on this topic. But this area of agreement, though philosophically significant, is thin. It does not demarcate a distinctive approach that can usefully be contrasted with Kantian and Utilitarian ethics (Nussbaum, 1999: 168). Many of the major defenders of a return to virtue have quarrels with either Kantian or Utilitarian ethics – in a few cases with both. They see a turn to Greek conceptions of virtue as helping them to solve the problems that they find in these Enlightenment moral theories. Nussbaum points out that some "virtue theorists" are best understood as motivated by a dissatisfaction with Utilitarianism. In particular, they question its neglect of the plurality of goods; its narrowly technical conception of reason, which holds that reason can deliberate only about means and not about ends; and the non-cognitive conception of emotion and desire that has frequently been taken for granted in Utilitarian thought both in philosophy and, even more obviously, in economics. These "virtue theorists" are friends of reason. On the whole they want to give reason and deliberation a larger role in our moral and political life than Utilitarians usually concede it. They are keen on the criticism of entrenched satisfactions and habits. They like the idea that not only our beliefs, but also our passions and desires, can be enlightened by the critical work of practical reason. These "virtue theorists" are likely to turn to Aristotle[5], or a certain

3 Concepts such as belief and consciousness, virtue and justice, look far more difficult to specify in any interestingly unified way. And yet this has not stopped philosophers from investigating commonalities and saying things that are genuinely illuminating as a result (Nussbaum 2001:9).

4 She draws her examples from the Anglo-American debate.

5 In the *Nicomachean Ethics*, Aristotle discusses how virtue (courage, generosity) is a matter of feeling the right thing. A brave individual, therefore, is neither fearless nor overwhelmed by fear in a dangerous situation. He further argues that we can shape our emotions through education and habit. The entirety of the *Nicomachean Ethics* is an extended example of rational deliberation about ultimate ends. In *Rhetoric*, he

reading of Aristotle, to elaborate their picture of a deliberative political life. They are not hostile to Kant, and they may even desire a synthesis of Aristotle and Kant. They value Aristotle's theoretical ambitions, and they see these as inseparable from the critical work of philosophy. They are likely to be universalist and anti-relativist.

Other "virtue theorists", by contrast, begin from a dissatisfaction with Kantian ethics. They question the dominant role Kant gives to reason in human affairs, and the type of Kantian rationalism that they judge to be dominant in contemporary ethical theory. They also question Kantian universalism, together with Kant's idea that practical judgment should be based on principles that abstract from particular local features of the agent's situation. These theorists want more recognition of "non-rational" elements in our makeup, and they take emotions and desires to be such elements. On the whole, they believe that our social life would go better if it were less deliberative and less critical, more the outgrowth of entrenched habits of desire and entrenched features of social position. They are hostile to universal theorizing in ethics, and they are likely to have some sympathy with cultural relativism, although they do not all endorse it (Nussbaum, 1999: 168–169).

The anti-Utilitarian group needs further demarcation in Nussbaum's opinion. It contains a group of thinkers who focus on moral awareness and are relatively indifferent to politics; and it contains a group of critical political thinkers. It also contains different views about the moral work involved in perfecting our emotions and desires. The anti-Kantian group contains different positions with regard to the possibility of ethical theorizing, and its relation to political theorizing.

The Common Ground

Nussbaum claims that if there is a common denominator among defenders of "virtue ethics"[6], it can be reduced to the following three claims:

---

points out that what characterizes many emotions is a strong moral conviction about how others should behave. The circumstances of situations, namely, do not look the same to people who love and to those who hate, the perspectives of an angry person or a person of a mild nature are completely different. Since he advocated the idea that every virtue lies between two extremes, Aristotle was an advocate of moderation in emotions as well as in action. The Aristotelian agent's entire personality can be enlightened by reason. Virtue is a mean concerning both passion and action, because Aristotle expects that the passions, as well as choice, can be crafted by reason until they themselves embody virtue.

6 Reasoning is, according to Nussbaum, an ability in virtue of which we commit ourselves to a view of the way things really are (Nussbaum 2001:35).

1. Moral philosophy should be concerned with the agent, as well as with choice and action.
2. Moral philosophy should therefore concern itself with motive and intention, emotion and desire: in general, with the character of the inner moral life, and with settled patterns of motive, emotion, and reasoning that lead us to call someone a person of a certain sort (courageous, generous, moderate, just, etc.).
3. Moral philosophy should focus not only on isolated acts of choice, but also, and more importantly, on the whole course of the agent's moral life, its patterns of commitment, conduct, and also passion (Nussbaum, 1999: 170).

In the sixties of the 20th century the competing normative theories competed to give the best account of how one ought to choose in a complex situation, and the competing metaethical theories vied to give the best account of what ethical discourse and reasoning aimed at choice really were. Little or nothing was said about reliable patterns of motivation and choice that might or might not be present in the agent. Little was said about the agent's emotions and desires, and virtually nothing about alternative analyses of what emotion and desire are. And, given the focus on the context of choice, little was said about the overall ethical life of the agent, the way in which choice both expresses and builds traits of character that have a complex connection with overall ethical and personal goals (Nussbaum, 1999: 171).

There was much, so Nussbaum, to be criticized. Even though a concern for motive, intention, character, and the whole course of life was not in principle alien to Kantian and Utilitarian philosophy, it was certainly alien to most British and American Kantians and Utilitarians of the period. Not surprisingly, scholars in Greek philosophy, often moral philosophers of distinction themselves, were in a position to make a valuable contribution. What these virtue thinkers did was to insist that we cannot adequately assess the ethical performance of the agent without knowing quite a lot about the agent's moral life, both in and outside of the immediate context of choice. In the immediate context, we need to know with what motives and intentions the agent chooses and acts; with what quality of deliberation and reflection; and with what reactive emotions. Does he/she do the just action for its own sake, or for gain? Does he/she think about it, making it her own, or just do what parents and teachers have taught him/her? And does he/she do it with strain, as if it goes against the grain, or easily, as if her whole personality approves of this action? Outside the immediate context, we need to ask how the choice fits into patterns of choice and response that this person has (or has not) cultivated. Does her life in gen-

eral how a commitment to justice, or is this act an isolated performance? (Nussbaum, 1999: 173).

One further element in the rise of virtue ethics should now be mentioned. It is the rise of feminism, together with the entry of significant numbers of women into the profession. It is in retrospect hardly surprising that among the major defenders of virtue ethics a substantial number have been women: Murdoch, Foot, Elizabeth Anscombe Diamond, Baier, Annas, Sherman, Homiak, Nussbaum, and others. Nussbaum thinks that women's[7] experiences have sometimes suggested questions and emphases that have been lacking from the dominant male tradition of moral philosophy (Nussbaum, 1999: 175–176).

Women's typical lives, in short, led them to want to investigate the role of reason in charting the whole course of life, and the problems reason encounters when values are plural and the world makes it difficult to organize them. This is the common ground. It does not imply the rejection of moral theory. Indeed, partisans of virtue ethics frequently notice that all its major proponents in the ancient Greco-Roman[8] world were strongly pro-theory. Nor does the common ground imply the rejection of universality in ethics, asking us to cling to local norms and traditions. Nor does the common ground imply a rejection of the guidance of rules. Rules are different from theories: theories give overall explanations, showing the point and purpose of a prescription, whereas rules are frequently obtuse. But that does not mean that rules are not frequently valuable in the agent's deliberations. Nor, finally, does the common ground imply that we should rely less on reason and more on non-rational sources of guidance, such as emotion and desire (if we should construe them as non-rational), and habit, and tradition. The thing one should notice about these ancient thinkers

7 It is not very surprising that women have been in the forefront of the move to make moral psychology and the study of emotion and desire central in philosophy. One reason for this emphasis is reactive: women have frequently been denigrated on account of their allegedly greater emotional nature, so one way of responding to that would be to understand these elements of the personality better – and, for example, to argue that they are not brutish but highly discerning, not devoid of thought but infused with thought. Another reason for the emphasis is that on balance women have more often been encouraged by society to attend to, cultivate, and label their emotions. This means that they are often better placed to undertake such an inquiry. Finally, women have often spent more time than men caring for young children, an occupation that both confronts one every day with a tremendous range of emotions, both in the child and in oneself, and requires one to deal with these responsibly and perceptively (Nussbaum 1999:176).

8 The way those ancient thinkers typically defended the value of philosophy as against other pursuits that claimed to produce virtue was to emphasize the central importance of reflection and theory in planning a virtuous life.

is that they live in a culture suffused with talk of the virtues. What they offer as philosophers is a specific conception of what it is to pursue the life of virtue, and instruction in that conception (Nussbaum 1999: 177–178).

Among the dissident virtue theorists, then, one can identify one large group that is motivated, above all, by a dissatisfaction with Utilitarianism, especially as formulated in the social sciences and public policy. These thinkers in general wish to give reason a larger role in human affairs than the instrumental and merely technical role given it in versions of neoclassical economics that see preferences as exogenous and impervious to reasoning. They tend to share the following four views:

1. The goods that human beings pursue are plural and qualitatively heterogeneous; it is a distortion to represent them as simply different quantities of the same thing.
2. Because the goods are plural and because they need to be both harmonized with one another and further specified, reason plays a central role not only in choosing means to ends, but also in deliberating about the ends themselves of a human life, which ones to include with which other ones, and what specification of a given end is the best.
3. Emotion and desire are not simply mindless pushes, but complex forms of intentionality infused with object-directed thought; they can be significantly shaped by reasoning about the good.
4. Existing social ideas about the good form defective passions and judgments; we should criticize these deficiencies, and this rational critique can be expected to inform the passions themselves (Nussbaum 1999: 180).

Conclusion

Martha C. Nussbaum, philosopher, classicist, political theorist, and public intellectual. For her, philosophy is a tool that brings clarity to people, not only in the academic community, but also to "ordinary" citizens related to numerous questions, difficulties, challenges they face both in their professional and personal lives. She believes that philosophy enables respect and appreciation of another person, his/her dignity, his/her rationality, even when she represents completely different views, pointing out that his view of philosophy is Socratic and democratic (Nussbaum 2012:3). By means of philosophical arguments, but also through the right to contribute to the discussion of every topic and area to which she devotes himself. Nussbaum believes that all those who deal with philosophy

should give their concrete contribution to socially current issues and encourage citizens to reflect and form their own critical attitude.

Nussbaum agrees with Aristotle that the philosopher must be someone who's attentive to and almost humble before the variety of human life and its great richness. But at the same time one who is committed to giving explanations, one who is committed to mapping that richness in a perspicuous way. In every area Aristotle strikes a kind of balance: between oversimplifying theorizing that takes philosophy too far from the richness and complexity and even messiness of ordinary discourse and ordinary life, and, on the other hand, a kind of negative or deflationary philosophizing that says theorizing is all houses of cards and there's no point in asking for and giving explanations. Aristotle[9] has found the right balance, and has probably the best conception of the philosophical task that one can give to a student (Magee 1987:54).

The current tendency to teach that there is any such unitary approach as "virtue ethics" is a big mistake, so Nussbaum. It is, first of all, a category mistake of an elementary kind, given that lots of people are, and have long been, writing and thinking about virtue within the Kantian and Utilitarian traditions. Virtue ethics cannot, then, be an alternative to those traditions.

What Nussbaum has called the "common ground" is significant: but it can be pursued within Kantianism, within Utilitarianism, and within neo-Aristotelian and neo-Humean projects of many different sorts. She proposes that we do away with the category of "virtue ethics" in teaching and writing. If we need to have some categories we should speak of Neo-Humeans and Neo-Aristotelians, of anti-Utilitarians and anti-Kantians (Nussbaum 1999: 200–201).

Nussbaum's cognitive view, by including a developmental dimension, makes room for the mysterious and ungoverned aspects of the emotional life in a way that many such views do not. This has consequences, as well, for the picture of character the view will support. All cognitive views of emotion entail that emotions can be modified by a change in the way one evaluates objects. This means that for such views virtue need not be construed (as Kant construes it) as a matter of strength, the will simply holding down the brutish impulsive elements of the personality. Instead, we can imagine reason extending all the way down into the personality, enlightening it through and through. If a person harbors misogynistic an-

9 Aristotelianism focuses on the worldly conditions for a good human life. Virtue can be realized in connection with material conditions as well as with education. Often these conditions are beyond the individual's control. Aristotle asks politics to provide essential conditions for people so that everyone, just everyone, can live a good, fruitful life.

ger and hatred, the hope is held out that a change in thought will lead to changes not just in behavior but also in emotion itself, since emotion is a value-laden way of seeing. Clearly this view has important implications for moral education, in the area, for example, of emotions toward members of other races and religions: we can hope to foster good ways of seeing that will simply prevent hatred from arising, and we don't have to rely on the idea that we must at all times suppress an innate aggressive tendency (Nussbaum 2001:232–233).

Nussbaum maintains that the main and necessary condition of a minimally just society is to protect a set of central human possibilities to some appropriate level. She believes that political justice, at the same time, offers a thorough transformation of moral emotions (both in the personal and in the public sphere). In the political realm, the primary virtue, according to Nussbaum, is impartial justice – it is a benevolent virtue that looks to the common good. That would, first of all, be a virtue of the institution, but also a virtue of the people who support the institutions.

It is very hard it is, even with the best intentions, to live a virtuous life.

Bibliography:

"Berggruen Prize", (internet) available at https://www.berggruen.org/prize/ (viewed December 2, 2022).

Adiprasetya, Joas (2016), "Alasdair Macintyre and Martha Nussbaum on Virtue Ethics", *Diskurs* 15: 1–22.

Duignan, Brian, "Martha Nussbaum", (internet) available at: https://www.britannica.com/biography/Martha-Nussbaum (viewed December 3, 2022).

Magee, Bryan (2000), *The great philosophers: an introduction to Western philosophy*. Oxford: Oxford University Press.

Nussbaum, Martha (1999), "Virtue Ethics: A Misleading Category?", *The Journal of Ethics* 3:163–201.

Nussbaum, Martha C. (2001), *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.

Nussbaum, Martha C. (2012), *Philosophical Interventions: Reviews 1986–2011*. Oxford: Oxford University Press.

Nussbaum, Martha C. (2016), *Anger and Forgiveness: Resentment, Generosity, Justice*. Oxford: Oxford University Press.

Monika Jovanović[*]
Andrija Šoć[**]

VRLINA I INTEGRITET U KANTOVOJ ETICI[***]

Apstrakt: U ovom radu cilj nam je da osvetlimo i međusobno povežemo Kantove teze o vrlini koje on iznosi u svojoj *Metafizici morala*. Vodeći se Kantovim tekstom, usredsredićemo se na odnos vrline i sklonosti, i odnos vrline i dužnosti. Tvrdićemo da pojmu vrline, kako je Kant shvata, odgovara savremeni pojam moralnog integriteta i nastojaćemo da pokažemo da Kantovo implicitno shvatanje integriteta može parirati nekim današnjim shvatanjima.

Ključne reči: Kant, vrlina, sklonost, dužnost, integritet

Uvod

Da li u Kantovoj etici pojam vrline ima supstantivan značaj? Iz ugla *Zasnivanja metafizike morala* i *Kritike praktičnog uma*, dva glavna spisa Kantove etike, jezgro Kantovog stanovišta čine pojmovi dužnosti, kategoričkog imperativa i poštovanja moralnog zakona.[1] S obzirom na to, ne čudi što se Kantova etika tradicionalno smatrala deontološkom.[2] Pa ipak,

* Odeljenje za filozofiju, Filozofski fakultet Univerziteta u Beogradu, *mojovano@f.bg.ac.rs*.

** Institut za filozofiju, Filozofski fakultet Univerziteta u Beogradu, *andrija.soc@ f.bg.ac.rs.*

*** Rad je nastao u okviru projekta „Čovek i društvo u vreme krize" Filozofskog fakulteta Univerziteta u Beogradu. Realizaciju tog istraživanja finansijski je podržalo Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije u sklopu finansiranja naučnoistraživačkog rada na Univerzitetu u Beogradu – Filozofskom fakultetu (broj ugovora: 451-03-68/2022-14/200163).

1 U ovom radu ćemo se služiti kembričkim izdanjem *Zasnivanja metafizike morala*, *Kritike praktičkog uma* i *Metafizike morala*, objednjenih u knjizi *Practical Philosophy* (Kant 1996a, 1996b i 1996c).

2 O podeli etičkih teorija na deontološke i teleološke, vid., na primer, Jovanović 2019, 53–54.

u literaturi nailazimo i na relativizacije ovog, dominantnog shvatanja. Barbara Herman, tako, ističe: „Ovakva slika Kantove etike ima osnova u onome što Kant kaže, ali delovi teksta koji se uzimaju kao potvrda deontološkog čitanja često nemaju značaj koji im se pridaje, niti igraju ulogu koja im je pripisana" (Herman 1993, vii) i potom dodaje (1993, x) da pojmovi vrline i karaktera nisu periferni u Kantovoj etici. Gajer ide i dalje od toga, tvrdeći da bi „Kanta, možda pre nego bilo koga drugog, trebalo smatrati najboljim modelom za etiku vrline" (Guyer 2011, 214).

Koliko su ovakve opaske, koje afirmišu ulogu vrline u Kantovoj etici, zaista utemeljene? Značaj vrline u Kantovoj filozofiji morala ispitaćemo tako što ćemo osvetliti na koji način se vrlina, kako je Kant shvata, odnosi prema drugim elemenatima njegove etike i kako je Kantova doktrina vrine relevantna za jednu, još uvek aktuelnu etičku debatu. U prvom delu rada razmotrićemo kako Kant u svoju etiku uvodi pojam vrline i zašto ključnu ulogu u tom kontekstu igraju sklonosti. U drugom delu rada govorićemo o odnosu pojma vrline i pojma dužnosti. Videćemo da Kant vrlinu definiše dvojako, formalno i materijalno. U trećem delu rada pokušaćemo da pokažemo na koji način se Kantova doktrina vrline (pre svega Kantovo formalno shvatanje vrline) može dovesti u vezu sa savremenim shvatanjima moralnog integriteta.

Vrlina i sklonosti

Kant o vrlini najdetaljnije govori u *Metafizici morala*, u drugom delu knjige, koji nosi naziv „Učenje o vrlini". Na samom početku tog odeljka Kant govori o vrlini kao o nečemu što čoveku treba da omogući da „savlada sklonosti iz kojih se rađaju poroci" (Kant 1996c, 6: 376). „Snaga se", kako Kant pronicljivo primećuje, „zahteva u stepenu koji se može meriti isključivo veličinom prepreka koje ljudsko biće samo sebi postavlja preko svojih sklonosti" (Kant 1996c, 6: 405). Da bismo razumeli ulogu vrline u Kantovoj etici, potrebno je, po svemu sudeći, razumeti kako Kant shvata sklonosti. On smatra da ljudska bića imaju prirodne sklonosti, koje ih „mogu navesti da prekrše moralni zakon, čak i onda kada prepoznaju njegov autoritet", i u nastavku dodaje da „čak i kada se povinuju zakonu, to čine *nevoljno* (uprkos protivljenju sopstvenih sklonosti)" (Kant 1996c, 6: 380).

Postojanje sklonosti koje idu nasuprot zapovesti moralnog zakona moguće je zahvaljujući činjenici da ljudska bića nemaju svetu već čistu volju, što je distinkcija koja igra važnu ulogu u drugoj *Kritici*. Za Kanta, sveta volja „nije sposobna da formuliše bilo koju maksimu koja bi bila u sukobu sa moralnim zakonom" (Kant 1996b, 5: 32). Za razliku od bića koje ima svetu volju, ljudska bića su aficirana „potrebama i čulnim motivima":

> Moralni zakon je, stoga, za njih *imperativ* koji zapoveda kategorički pošto je zakon bezuslovan; odnos takve volje prema zakonu je *zavisnost* pod imenom obaveze koja, iako samo putem uma i njegovog objektivnog zakona, označava nužnost postupka koji se zove *dužnost* pošto je izbor patološki aficiran (iako ne time determinisan, te je dakle i dalje slobodan) i sa sobom nosi želju koja proističe iz subjektivnih uzroka (Kant 1996b, 5: 32).

Na tom mestu Kant, čini se, zastupa stav da ljudska motivacija nije isključivo racionalna, ako se pod time podrazumeva moralno postupanje u kojem sklonosti ne igraju baš nikakvu ulogu. Čak i kada čovek postupa i u skladu s dužnošću i iz dužnosti, u njegovoj motivaciji izgleda moraju da postoje „patološki" elementi. Pred kraj citiranog pasusa, Kant govori o vezi između tog shvatanja i uloge koju vrlina ima u moralnom postupanju:

> Svetost volje je ipak jedna praktička *ideja*, koja nužno mora poslužiti kao *model* kojem se sva konačna umna bića mogu samo približavati u beskraj, i koju čist moralni zakon, koji se sam smatra svetim zbog toga, neprestano i ispravno drži pred svojim očima; najviše što konačan praktični um može postići je da osigura da ovaj progres naših maksima u pravcu tog modela bude konstantan, da, drugim rečima osigura vrlinu; a vrlina sama, zauzvrat, bar kao prirodno stečena sposobnost, ne može se nikada ispuniti, jer sigurnost u ovakvom slučaju nikada ne postaje apodiktička izvesnost, a kao ubeđenje je veoma opasna (Kant 1996b, 5: 33).

U ovom pasusu, koji dolazi na kraju sedmog paragrafa *Kritike praktičnog uma*, Kant upotrebljava pojam vrline u kontekstu moralnog progresa koji ljudska bića treba da pokušaju da naprave kako bi se, u svom moralnom postupanju, približila svetosti volje. Vrlina se, kao što vidimo, izjednačava sa konstantnim nastojanjem da se čovek oslobodi patoloških, odnosno afektivnih elemenata u moralnom domenu.

U *Metafizici morala* Kant tu tezu razvija tvrdeći da se prirodni impulsi ponekad direktno suprotstavljaju dužnostima[3]:

> [Čovek] mora prosuditi da je sposoban da se odupre [ovim silama] i da ih pokori ne nekad u budućnosti već odmah (u trenutku kada razmišlja o dužnosti): on mora prosuditi da *može* da učini ono što mu zakon bezuslovno govori da *treba* da učini. Sposobnost i promišljena odlučnost da se izdrži jak ali nepravedan protivnik je *snaga* (*fortitudo*) i, s obzirom na ono što se *u nama* sukobljava sa moralnom dispozicijom, jeste vrlina (*virtus*, *fortitudo moralis*). (Kant 1996c, 6:380)

---

3 Kant i u *Predavanjima iz etike* vrlinu dovodi u vezu sa postojanjem prepreka moralnom postupanju. Vid. Kant 1997, 27: 27.

Na ovom mestu se jasno vidi kako Kant shvata ulogu vrline. Ona, naime, predstavlja snagu u odupiranju sklonostima i ukoliko je ta snaga dovoljna, čovek će, uprkos konstantnom postojanju afektivnih elemenata u moralnom odlučivanju, ipak postupati i u skladu sa dužnošću i iz dužnosti. Premda bi moglo delovati da je takvo shvatanje vrline nalik onom iz druge *Kritike*, nije sasvim tako. Naime, u prethodno navedenom pasusu Kant je govorio o vrlini kao o sposobnosti čije je posedovanje neophodno za napredak ka postepenom oslobađanju od sklonosti. U *Metafizici morala* vrlina je, kao što smo videli, mera snage karaktera da se u svakom pojedinačnom slučaju moralnog postupanja aktivno odupre sklonostima koje otežavaju postupanje u skladu s dužnošću i da se osigura da se učini ono što kategorički imperativ nalaže. Sličan stav se nalazi i u *Predavanjima iz etike*, gde Kant tvrdi sledeće:

> Kada je reč o prirodnoj sklonosti, trebalo bi, ipak, primetiti da što se više borimo sa njom, utoliko više zaslužujemo pohvalu; stoga se vrlina više može pripisati nama nego anđelima, jer oni nemaju toliko prepreka kao mi (Kant 1997, 27: 292).

Kao što vidimo, mera uspešnog suprotstavljanja sklonostima za Kanta je mera vrline koju poseduju ljudska bića.[4] Pa ipak, za potpuno razumevanje Kantovog shvatanja vrline, pored ovog, uslovno rečeno, negativnog određenja pojma vrline, potrebno je, ništa manje, uzeti u obzir i Kantovo pozitivno određenje prema kojem je vrlina karakterna čvrstina u „ispunjavanju sopstvenih dužnosti" (Kant 1996c, 6: 394).

Vrlina i dužnosti

U *Zasnivanju metafizike morala* dužnost se definiše kao „nužnost jedne radnje iz poštovanja prema zakonu" (Kant 1996a, 4: 400). Dužnosti se dalje dele na savršene i nesavršene, dužnosti prema sebi i dužnosti prema drugima (Kant 1996a, 4: 421).[5] U *Metafizici morala* Kant daje nešto drugačiju podelu dužnosti:

> Svakoj dužnosti odgovara *neko* pravo u smislu *dopuštenja* da se nešto učini; ali nije slučaj da svakoj dužnosti odgovara nečije *pravo* da nekog primora. Takve dužnosti se, specifično, zovu *dužnosti prava*. Slično tome, svakoj etičkoj obavezi odgovara pojam vrline, ali nisu sve etičke dužnosti samim tim dužnosti vrline. One dužnosti koje nisu toliko povezane sa određenom svrhom (materijom, predmetom izbora) koliko sa onim *što je formalno* u

4 Uporedi: Kant 1996c, 6: 405.

5 Detaljnije o ovoj podeli, kao i mestu koje centralni etički pojmovi imaju u Kantovom sistemu, vid., na primer, u Šoć 2021, gl. 5.

> moralnom određenju volje (to jest, da radnja učinjena u skladu sa dužnošću mora takođe biti učinjena *iz dužnosti*) nisu dužnosti vrline. Samo *svrha koja je takođe dužnost* se može zvati dužnošću vrline (Kant 1996c, 6: 383).

Ovako shvaćene dužnosti vrline predstavljaju meru unutrašnjeg samoograničavanja koje je neophodno, kako Kant ubrzo posle navedenog pasusa ističe, zato što ljudi nisu sveta bića (Kant 1996c, *ibid*.). Vrlina igra integralnu ulogu u moralnom postupanju jer bi ljudska bića bez nje bila preslaba da se odupru uticaju sklonosti. Vrlina ima etičku dimenziju koja se ne javlja ni u odgovoru na pitanje šta je moralno, ni u odgovoru na pitanje šta nas pokreće na moralno delanje. Možemo biti veoma motivisani da postupimo moralno, pa da ipak to ne učinimo: možda su nas naše sklonosti odvukle na suprotnu stranu, možda smo pomislili da je u datim okolnostima moralan postupak nešto što bi bilo suviše teško učiniti. Kada se to desi, treba odgovoriti na pitanje zašto se to dogodilo, odnosno zbog čega, i pored ispravnog sagledavanja toga koji postupak je moralno ispravan, i pored postojanja motivacije da datu stvar učinimo, to ipak nismo učinili. Kantov odgovor je da je to slučaj zato što nismo posedovali vrlinu. (Kant 1996c, 6:380)

Kant izdvaja dve dužnosti vrline: sopstveno savršenstvo i sreću drugih (Kant 1996c, 6: 386). Te dužnosti kao da odgovaraju nesavršenim dužnostima o kojima Kant govori u *Zasnivanju metafizike morala*: „kao racionalno biće, nužno ćemo hteti da sve naše sposobnosti budu razvijene, jer nam one služe i date su nam za razne moguće svrhe" (Kant 1996a, 4: 423). Kako Kant u nastavku dodaje, nemoguće je hteti da princip po kojem nam nije stalo do tuđeg blagostanja ni do toga da ljudima pomognemo u nevolji postane opšti zakon.[6] To što su te dužnosti nesavršene za Kanta naprosto znači da, iako ne postoji „unutrašnja nemogućnost" formulisanja maksima u skladu sa opštim prirodnim zakonom (kao u slučaju samoubistva ili laganja), ipak je nemoguće *hteti* univerzalnost maksime po kojoj svoje talente ne bismo razvijali, odnosno po kojoj ne bismo hteli da pomažemo drugima (Kant 1996a, 4: 424): za nesavršene dužnosti Kant kaže da su široke, premda na tom mestu ne elaborira šta pod time podrazumeva.

Podela dužnosti na uske i široke ima ključnu važnost za razumevanje dve dužnosti vrline o kojima Kant govori u *Metafizici morala*. Kako Kant ističe u naslovu sedmog poglavlja „Učenja o vrlini" (Kant 1996c, 6: 390), etičke dužnosti su dužnosti „široke obaveze" (za razliku od dužnosti prava koje imaju usku obavezu). Osnov te distinkcije leži u tome što se, smatra Kant, u pravu propisuju postupci, dok se u moralu propisuju samo maksime postupaka (Kant 1996c, 6: 389). To znači da postoji širok prostor za ispunjenje etičkih dužnosti (i specifično, dužnosti vrline) te da je izbor

6 Vid. Kant 1996a, 4: 423–424.

u načinu na koji će naši postupci biti u skladu sa moralnim zakonom na nama. Povezujući distinkciju „usko/široko" sa distinkcijom „savršeno/nesavršeno", Kant ističe:

> Što je šira dužnost, to je nesavršenija naša obaveza da postupimo na određen način; ipak, kako se maksimama kojima se saglašavamo sa širokom dužnošću (u dispoziciji) približavamo *uskoj* dužnosti (dužnosti prava), u toj meri je savršeniji naš vrli postupak (Kant 1996c, 6: 390).[7]

Kant o vrlini govori na dva načina: formalno i sadržinski. Određenje vrline kao snage karaktera je formalno, dok je određenje vrline pozivanjem na dužnosti vrline materijalno odnosno sadržinsko (Kant 1996c, 6: 395). U kakvom su tačno odnosu ta dva određenja? Kako bismo svako od njih mogli dovesti u vezu sa različitim vrlinama kao što su hrabrost, dostojanstvo, zahvalnost (o kojima Kant govori na više mesta u *Učenju o vrlini*, *Antropologiji* i drugim spisima)?

Kada Kant vrlinu definiše kao „snagu karaktera",[8] reč je o formalnom shvatanju vrline. Posedovanje tako shvaćene vrline čoveka čini sposobnim da dela moralno, uspešno prevazilazeći izazove prirodnih sklonosti. Sledeća ravan razmatranja vrline je materijalna; u njoj se nudi odgovor na pitanje koje dužnosti ima čovek koji poseduje formalno shvaćenu vrlinu. Kao što smo videli, postoje dve takve dužnosti – unapređenje sopstvenih talenata (dužnost prema sebi) i negovanje tuđe sreće (dužnost prema drugima). U Kantovom tekstu postoji i treća ravan razmatranja: ona koja se tiče konkretnih karakternih crta koje figurišu u pojedinačnim postupcima. Naime, kako se za Kanta vrlina tiče svrha koje sebi postavljamo, pošto postoji više različitih svrha, mora postojati i više vrlina (Kant 1996c, 6: 395).

Budući da za Kanta vrlina, po svemu sudeći, nije karakterna crta, kao što je u istoriji etike praktično uvek bilo sučaj, postavlja se pitanje šta je ona suštinski. Premda Kant ne daje konkretan odgovor na to pitanje, čini se da na ovom mestu možemo, držeći se njegovog shvatanja, primeniti pojam koji je postojao u Kantovo vreme, ali je postao prominentan deo etike u proteklih nekoliko decenija. Reč je, naime, o pojmu integriteta. Polazeći od Kantovog shvatanja vrline, izgleda da bi se moglo tvrditi da je vrlina za njega zapravo integritet: posedovanje vrline u Kantovom smislu bi se tako

7 Vid. takođe Sherman 1997, 331–332; Timmons 2021, 103–109.

8 Kant daje više sličnih određenja: „moralna snaga ljudskog bića" (Kant 1996c, 6: 405 i 6: 393), „duševna snaga" (6: 384), „snaga odlučnosti" (6: 390), „snaga u sprovođenju maksima" (6: 394), „samoograničavanje u skladu sa principom unutrašnje slobode" (6: 394). Premda postoje i druge interpretacije (Merritt 2018) prema kojima je vrlina za Kanta zapravo veština, iz navedenih pasusa je, čini se, jasno da je vrlina za njega pre svega odlika karaktera.

moglo izjednačiti sa posedovanjem integriteta u savremenom smislu.[9] U narednom odeljku probaćemo da pokažemo kako, ako kantovsku vrlinu shvatimo kao integritet, njegovu etiku možemo učiniti relevantnom i za savremena razmatranja integriteta.

Vrlina kao integritet

Da bismo razumeli zašto je Kantovo implicitno shvatanje integriteta još uvek relevantno, pogledajmo neke od vodećih teorija integriteta, onako kako ih klasifikuju Koks, La Kaz i Levin (Cox, La Caze and Levine 2021). Prema jednoj grupi shvatanja, integritet se posmatra kao „samointegracija", odnosno kao „integrisanje različitih strana ličnosti u harmoničnu, netaknutu celinu". Tom stanovištu je bliska teorija „samokonstitucije". Za tu teoriju je specifično to što se „integritet ne shvata toliko kao uslov izvrsnosti kojoj težimo, koliko kao preduslov da uopšte budemo delatnici" (Cox, La Caze and Levine 2021, gl. 3). Stavovi Kristin Korsgard se u toj klasifikaciji svrstavaju u ovu drugu grupu stanovišta. Međutim, u sledećem citatu pomenuti autori sugerišu da je njeno shvatanje bolje tumačiti kao kombinaciju ta dva viđenja: „Svoje [neprikladne želje] moramo potisnuti kako bismo bili jedno, bili objedinjeni, bili celoviti... i osoba koja uspe u tome je dobra – ne zato što teži da bude dobra, nego zato što teži da bude objedinjena, da bude celovita" (Korsgaard 2009, 26). Integritet se ponekad razume i kao stvar očuvanja identiteta, odnosno „nepokolebljivo delanje u skladu sa svojim stavovima" (Cox, La Caze and Levine, gl. 2).

Zajedničko za ta gledišta je, primećuju Koks, La Kaz i Levin, to što se njihovi zastupnici usredsređuju na individualni aspekt integriteta, to jest na ono što se tiče samog moralnog delatnika, zanemarujući njegov društveni aspekt, koji se tiče delatnikovog odnosa prema drugim ljudima (Cox, La Caze and Levine, gl. 4). Još više upada u oči činjenica da su navedena shvatanja primarno psihološka, pa čak i metafizička, a tek sekundarno etička, dok je etička dimenzija integriteta kod Kanta u prvom planu.

Prema Kantovom implicitnom shvatanju, integritet podrazumeva istrajnost u postupanju i vernost osnovnim principima na kojima počiva njegova etika, kao što su delanje iz dužnosti, poštovanje moralnog zakona, autonomija i tretiranje drugih ljudi kao svrhe po sebi. Takvo, kantovsko shvatanje integriteta je obuhvatno jer kao relevantne uzima i unutrašnje

9 Ovom tezom ne tvrdimo da je integritet jedna od karakternih odnosno moralnih vrlina. Takva teza bi se mogla razviti polazeći od nekog etičkog stanovišta koje pripada etici vrline. O jednom takvom shvatanju, teoriji vrline Rozalind Hersthaus, vid. više u Jovanović 2011.

principe delatnikovog postupanja i način na koji se on odnosi prema drugima; etička dimenzija je naglašena. Intuitivno gledano, svako adekvatno shvatanje integriteta bi, bar u određenoj meri, trebalo da sadrži obe pomenute komponente integriteta. Ali, to je važno i iz teorijskih razloga.

Jedan od gavnih izazova za shvatanje integriteta tiče se pitanja da li postoje slučajevi u kojima se može reći da je neka osoba imala integritet, ali da njeno postupanje nije bilo moralno. Iako se u određenim shvatanjima integriteta eksplicitno naglašava moralna komponenta integriteta,[10] ako je objašnjenje koje se pritom daje suviše apstraktno ili formalno, otvara se mogućnost da to shvatanje bude u skladu sa različitim problematičnim stavovima o tome šta je moralno, što je intuitivno neprihvatljivo budući da pojam integriteta ima pozitivnu vrednosnu konotaciju.

Za razliku od vrednosno neutralnih stanovišta koja su izložena ovoj kritici, Kantovo obuhvatno shvatanje integriteta ima nedvosmisleno pozitivnu vrednosnu polarnost i vrlo određenu sadržinsku komponentu. Ta komponenta se ogleda u njegovom učenju o dužnostima vrline: dužnosti samousavršavanja i dužnosti unapređenja tuđe sreće. Takvo shvatanje odnosa vrline i dužnosti utemeljeno je u Kantovoj formulaciji kategoričkog imperativa, prema kojoj treba „delati tako da čoveštvo kako u sebi, tako i u drugima posmatramo uvek i kao cilj, a nikada samo kao sredstvo" (Kant 1996a, 4: 429). Pa ipak, tek kada formalne elemente Kantovog shvatanja dovedemo u vezu sa sadržinskim, Kantovo implicitno shvatanje integriteta dobija pravu eksplanatornu snagu.

To je zato što tri elementa svakog adekvatnog shvatanja integriteta – obuhvatnost, vrednosni karakter i sadržinska komponenta – kod Kanta padaju ujedno. Da bi nam se mogao pripisati integritet, nije dovoljno biti dosledan sebi ili svojim principima i nepokolebljiv u odnosu prema drugima već postupci, motivi i moralna načela moraju biti jasno vrednosno određeni. Negovanje sopstvenih talenata, kao i unapređenje tuđe sreće, u čemu ni sebe ni druge nećemo tretirati kao puka sredstva, onemogućava bilo kakvu fanatičnost ili doslednost nemoralnim principima koje bi neko mogao smatrati moralno ispravnim. Nezavisno od toga kako bi se takvo shvatanje integriteta moglo kritikovati, ono, kao što vidimo, može doprineti savremenim debatama.

Pa ipak, izgleda da se neki, a možda i mnogi autori ne bi složili sa glavnim tezama iznetim u ovom radu. Štaviše, čini se da je među interpretatorima prisutan, pa čak i dominantan stav da se ne može plauzibilno govoriti o Kantovom shvatanju integriteta budući da on relativno malo pažnje posvećuje vrlini i karakteru:

10 Vid., na primer, McFall 1987.

> Jedna značajna i uticajna linija argumentacije, koju je prvo razvio Bernard Vilijams, nastoji da pokaže da određene morane teorije ne uzimaju dovoljno u obzir integritet moralnih delatnika. (Vid. Williams 1973 & 1981.) Ovo je postao značajan tip kritike modernih moranih teorija. (Vid., na primer, Scheffler 1993 i Lomasky 1987.) Moderne moralne teorije, čiji su najprominentniji predstavnici utilitarizam i kantovska moralna teorija, ne bave se direktno vrlinom i karakterom. Umesto toga, one se primarno bave opisom moralno ispravnog postupka. (Cox, La Caze i Levine 2021, gl. 8)

Na jednom mestu Vilijams o tome tvrdi sledeće:

> Jednom kada razmislimo o tome šta podrazumeva posedovanje karaktera, možemo videti da to što Kantovci izostavljaju karakter predstavlja uslov njihovog krajnjeg insistiranja na zahtevima nepristrasne moralnosti, a to je i razlog iz kog je njihovo objašnjenje individue neadekvatno (Williams 1981, 14).

Stav da se Kantu ne može pripisati nikakvo, pa čak ni implicitno shvatanje integritata, čini se, počiva na dve pretpostavke. Prema prvoj, pojam moranog integriteta podrazumeva određeno shvatanje karaktera.[11] Prema drugoj, Kant u svojoj etici ne pridaje odgovarajući značaj karakteru. Dok je prva pretpostavka nesumnjivo ispravna, druga je očigledno pogrešna. Karakter je za Kanta relevantan, i kada je reč o moralnoj motivaciji, i kada je reč o pojmu čoveštva (koji ima središnji značaj u jednoj od formulacija kategoričkog imperativa). Povrh toga, vrlina predstavlja značajan faktor u opisu moralnih delatnika jer, osim samog postupka (koji mora biti u skladu s dužnošću) i motivacije moralnog delatnika (koji treba da postupa motivisan samom dužnošću odnosno poštovanjem prema zakonu), mora postojati nešto što će delatniku dati snagu da se suprotstavi sklonostima koje ga mogu odvući na suprotnu stranu. U tom smislu, druga pretpostavka Kantovih kritičara je pogrešna, a utoliko i njihova kritika Kanta.

Mada neki autori, kao što su Barbara Herman i Hening Jensen (Herman 1983/1993, Jensen 1989), Kanta brane od kritičara kao što je Vilijams, oni nedovoljno ističu da je način na koji Kant shvata vrlinu i mesto koje joj daje ključni za odbranu i od kritike da u njegovoj etici nema mesta za razmatranje karaktera i integriteta. Rasprava između onih koji (poput Vilijamsa) kritikuju Kanta i onih koji ga brane (kao što to čine Herman i Jensen) tiče se pitanja da li su za Kanta postupci iz dužnosti motivisani emocijama ponekad preferabilniji od postupaka koji su motivisani isključivo poštovanjem zakona (kao što tvrde Herman i Jensen) ili to nikada

11 Hening Jensen, na primer, govoreći o Vilijamsovoj kritici, sugeriše da: „Bernard Vilijams [...] tvrdi da je to što Kant preferira [postupke učinjene iz dužnosti u odnosu na postupke učinjene iz motiva kao što su ljubav ili saosećanje] dokaz da odbacuje lične aspekte moralne reakcije, kao i ulogu integriteta". (Jensen 1993, 193)

nije slučaj (tu tezu zastupa Vilijams i na osnovu nje kritikuje Kanta). Oni pokušavaju da pokažu da kod Kanta ima mesta za etička razmatranja koja se tiču karaktera, ali ne uzimaju u obzir ključnu ulogu koju u opisu moralnog delatnika igra Kantovo shvatanje vrline i njegovo implicitno shvatanje integriteta.

Zaključak

U ovom radu cilj nam je bio da osvetlimo i međusobno povežemo Kantove teze o vrlini koje on iznosi u svojoj *Metafizici morala*. Vodeći se Kantovim tekstom, akcenat smo stavili, s jedne strane, na odnos vrline i sklonosti, a sa druge, na odnos vrline i dužnosti. Tvrdili smo da pojmu vrline, kako je Kant shvata, odgovara savremeni pojam moralnog integriteta i nastojali da pokažemo da njegovo implicitno shvatanje integriteta može parirati nekim današnjim shvatanjima.

Shvatanje integriteta koje se može iščitati iz Kantovog teksta, čak i u nerazvijenoj i implicitnoj formi, ima inherentne prednosti u odnosu na danas prominentna shvatanja, i to zahvaljujući svojoj (1) obuhvatnosti, (2) vrednosnom karakteru i (3) sadržajnosti. Zahvaljujući tim odlikama, ono može da odgovori na pitanje kako objasniti slučajeve u kojima se odgovarajućim subjektima pripisuje integritet iako im se ne može pripisati moralnost.

Kantovo shvatanje je obuhvatno: ono se podjednako tiče unutrašnje (individualne) i spoljašnje (društvene) strane ljudske moralnosti. Ono je, štaviše, obuhvatno i na drugi način: vrlina je, za Kanta, neka vrsta moralne čvrstine, to jest vernosti moralu (pozitivna teza), ali i sposobnost odupiranja patološkim sklonostima (negativna teza). Ako vrlinu shvatimo kao integritet, Kantovo implicitno shvatanje integriteta teško da može biti vrednosno neutralno (bar ako vrlinu ne shvatimo kao veštinu, što retko ko čini).[12] Sadržajnost Kantovog stanovišta ogleda se u njegovom učenju o dužnostima vrline. Pomalo paradoksalno, s obzirom na njegovu kritiku sadržinskih etika,[13] Kant ovde insistira na konkretnim svrhama i vrednostima kao što su samousavršavanje i sreća drugih ljudi.

Za razliku od nekih shvatanja koja smo pominjali, Kantovo implicitno shvatanje integriteta je sasvim na liniji leksičkih određenja tog pojma. Tako, na primer, ako pogledamo *American Heritage*, videćemo da se integritet na prvom mestu definiše kao „nepokolebljivo sleđenje striktnog moralnog ili etičkog kodeksa", dok se u *Meriam-Vebster*-u, slično, integri-

12 Vid., na primer, Merritt 2018.

13 Osnov za tu kritiku možemo videti, na primer, u: Kant 1996a, 4: 390.

tet određuje kao „čvrsto pridržavanje kodeksa prevashodno moralnih ili umetničkih vrednosti".

Postoji nekoliko razloga da Kantovo shvatanje vrline razumemo kao deo jezgra njegove etike. Osim toga što za to nalazimo tekstualno potkrepljenje u „Učenju o vrlini", takav pristup nam omogućava da Kanta odbranimo od kritičara koji tvrde da je njegova etika suviše apstraktna i formalistička.[14] Drugo, on nam omogućava da Kantovu etiku posmatramo kao danas relevantnu onako kako se obično ne čini. Treće, kada osvetlimo mesto vrline (odnosno integriteta) u Kantovoj etici, moći ćemo bolje da razumemo različite implikacije Kantovih stavova u etici i da, u krajnjoj liniji, bolje sagledamo u kakvoj je vezi Kantovo shvatanje morala i moralnih delatnika sa drugim aspektima njegovog sistema, kao što su, na primer, religija, politika[15], umetnost i istorija.

Literatura

Cox, Damian, La Caze, Marguerite, Levine, Michael (2021). Integrity. *Stanford Encyclopedia of Philosophy.*

Guyer, Paul (2011). Kantian Perfectionism. In: Lawrence Jost, Julian Wuerth. *Perfecting Virtue: New Essays on Kantian Ethics*. Cambridge: Cambridge University Press, 194–214.

Hegel, G. W. F. (2008). *Outlines of the Philosophy of Right*, prev. T. M. Knox. Oxford: Oxford University Press.

Herman, Barbara (1993). *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.

Herman, Barbara (1983). Integrity and Impartiality. *Monist* 66: 233–250. (Ovde navedeno prema: Herman 1993, gl. 2).

Jensen, Henning (1989). Kant and Moral Integrity. *Philosophical Studies* 57: 193–205.

Jovanović, Monika (2011). Vrlina i *eudaimonia* u filozofiji morala Rozalind Hersthaus. *Theoria*, Beograd 54: 37-50.

Jovanović, Monika (2019). *Prima facie* dužnosti i struktura etičkog objašnjenja. *Theoria*, Beograd 62: 53-64.

Kant, Immanuel (1996a). Groundwork of the Metaphysics of Morals. In: Immanuel Kant, *Practical Philosophy*, prev. i ur. Mary J. Gregor. Cambridge: Cambridge University Press, 37–108.

Kant, Immanuel (1996b). Critique of Practical Reason. In: *Practical Philosophy*, 133–272.

14 Pomenuli smo Bernarda Vilijamsa kao jednog od najpoznatijih savremenih kritičara, ali ta kritika tipski seže bar do Hegela. Vid., na primer, Hegel 2008, §135.

15 Veza između etike i politike posebno je relevantna za Kantovo razmatranje osnova na kojima treba da počiva građansko društvo, i, još konkretnije, za njegovo razmatranje pitanja odnosa suverena i građana. O ovoj temi vid., na primer, Šoć 2013.

Kant, Immanuel (1996c). The Metaphysics of Morals. In: *Practical Philosophy*, 353–604.

Kant, Immanuel (1997). *Lectures on Ethics*, ed. Peter Heath, J. B. Schneewind. Cambridge: Cambridge University Press.

Korsgaard, Christine M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

McFall, Lynne (1987). Integrity. *Ethics* 98: 5–20.

Merritt, Melissa (2018). *Kant on Reflection and Virtue*. Oxford: Oxford University Press.

Sherman, Nancy (1997). *Making a Necessity of Virtue: Aristotle and Kant on Virtue*. Cambridge: Cambridge University Press.

Šoć, Andrija (2013). Kant i legitimnost pobune protiv suverena. *Theoria*, Beograd 56: 63-77.

Šoć, Andrija (2021). *Transcendentalni idealizam*. Beograd: Clio.

Timmons, Mark (2021). *Kant's Doctrine of Virtue*. Oxford: Oxford University Press.

Williams, Bernard (1981). *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.

Monika Jovanović
Andrija Šoć

Virtue and Integrity in Kant's Ethics

Summary: Our goal in this paper is to shed light on and elaborate upon the connections between Kant's theses on virtue, which he discusses in his *Metaphysics of Morals*. Leaning on Kant's text, we will focus on the relation between virtue and inclination, and the relation between virtue and duty. We will claim that the concept of virtue, as viewed by Kant, corresponds to the contemporary concept of moral integrity, and we will attempt to show that Kant's implicit view on integrity can rival some of today's views.

Keywords: Kant, virtue, inclination, duty, integrity

Stefan Mićić*

SENTIMENTALISTIČKO SHVATANJE VRLINE: HJUM, HAČESON, SLOT

Apstrakt: Cilj ovoga rada je da se ispita kako shvataju vrlinu oni koje smatramo paradigmatskim predstavnicima sentimentalizma u filozofiji morala. Sentimentalizam je pravac koji u centar istraživanja morala stavlja posebnu vrstu osećanja koja proističu iz onoga što su njegovi predstavnici nazivali moralnim čulima. U ovom radu bavićemo se Dejvidom Hjumom, Fransisom Hačesonom i Majklom Slotom da bismo ukazali na najvažnija obeležja moralnog sentimentalizma i da bismo ispitali da li i na koji način emocije mogu da budu konstitutivne u izgradnji normativne teorije.

Ključne reči: moralni sentimentalizam, vrlina, empatija, moralno čulo, Hjum, Hačeson, Slot

U filozofiji morala sentimentalizam se povezuje prvenstveno sa britanskim moralistima 18. veka, a kao teorijski pravac u savremenim raspravama zauzima malo ili gotovo nikakvo mesto. Ipak, ideje sentimentalizma se – više ili manje implicitno – javljaju i kod savremenih autora. U raspravama o moralnom sentimentalizmu ili teoriji moralnih čula, kako se taj pravac još i naziva, danas se raspravlja o etici vrline, u nekim verzijama o „etici brige", ili se one prosto beleže kao deo pregleda istorije etike.

Upravo zbog svega toga, počećemo od ispitivanja šta se u etici podrazumeva pod sentimentalizmom.

Etimologija reči „sentimentalizam" (lat. *sentimentum*) upućuje da je reč o pravcu koji u središte moralnog delanja i moralne motivacije stavlja osećanja (emocije). U potrazi za normativnom teorijom – principima koji konstituišu moralna načela – sentimentalisti kreću empirijski, to jest empirijski istražuju šta ljude navodi da prihvataju moralne norme koje prihvataju. Sentimentalisti time na neki način počinju tamo gde su raci-

* Filozofski fakultet Univerziteta u Beogradu, *stefan.micic@f.bg.ac.rs*.

onalisti stali. Naime, racionalistički filozofi prvo konstruišu normativne teorije koje imaju svoje razumsko poreklo, pa tek onda ispituju kako se u običnom ljudskom životu delatnici podvrgavaju autoritetu moralnih načela koja prihvataju. Za sentimentaliste, pak, norme su izrazi moralnih osećanja, što ne znači da sentimentalisti „operišu" na čisto deskriptivnom nivou:

> Način razvoja naše moralne psihologije nikada ne bi mogao da, sam po sebi, opravda naše moralne obaveze. Verovati u suprotno znači brkati empirijsko objašnjenje porekla obaveznosti koju nam vrednosti nameću sa demonstracijom istinskog normativnog autoriteta... Ipak, kada prihvatimo sentimentalističko objašnjenje naše psihologije morala – ono koje naše moralne obaveze vidi kao refleksivne izdanke bazičnih ljudskih emocija koje svi mi iz sveg srca odobravamo – predstavlja nam se distinktivno sentimentalistički metod normativnog opravdanja te obaveznosti (Frazer 2010, 8).

Kao što smo već naveli, taj pravac se, istorijski, najviše povezuje sa britanskim filozofima prosvetiteljstva – Erlom od Šaftsberija (Anthony Ashley Cooper, Third Earl of Shaftesbury), Fransisom Hačesonom (Frances Hutcheson), Dejvidom Hjumom (David Hume) i Adamom Smitom (Adam Smith). Naše istraživanje će se fokusirati na tri predstavnika sentimentalizma – Fransisa Hačesona, Dejvida Hjuma i Majkla Slota (Michael Slote).

Koji je kriterijum presudio da izaberemo ta tri autora? Majkl Slot je svakako najpoznatiji i najuticajniji savremeni predstavnik moralnog sentimentalizma; Dejvid Hjum je najpoznatiji i najuticajniji predstavnik moralnog sentimentalizma koji nije ograničen odrednicama vremenskog razdoblja u kome je živeo i stvarao; Fransis Hačeson je, pak, autor koji nije mnogo poznat široj čitalačkoj publici, ali su njegove ideje imale veliki uticaj i na Hjuma i na Slota, a time i na ukupan pravac razvoja moralnog sentimentalizma.

Iako se sentimentalizam najviše povezuje sa Dejvidom Hjumom, on nije začetnik toga pravca, ali jeste najpoznatiji. Njegovo filozofsko zaveštanje u pogledu moralnog sentimentalizma ogleda se, između ostalog, u tome što je napravio otklon, za razliku od svojih prethodnika, od teoloških osnova sentimentalizma – iako je teološka tradicija 17. i 18. veka na njega snažno uticala (up. Burton 1848, 114). To se najbolje vidi u tezi da sve što je prirodno jeste normativno jer je prirodno ono što Bog hoće da radimo (up. Lind 1993, 136). Uticajni anglikanski biskup i autor 17. veka Džozef Batler (Joseph Butler) je, primera radi, smatrao:

> [A]ko prava Priroda bilo kog stvorenja (to jest, ono što je stvarno prirodno tom stvorenju) vodi to stvorenje i prilagođena je samo određenim svrhama više nego bilo čemu drugome, onda je to razlog da se veruje da je krea-

> tor te Prirode nju namenio samo za te svrhe (Butler 1749, sec. 211). Hjum je pokušavao da de-teologizuje prirodno, to jest, želeo je da sačuva normativnu snagu prirodnog, ali time što će to da uradi bez Boga (Lind 1993, 136).

Kada je reč o Hjumu i njegovom shvatanju prirode normativnosti, treba imati u vidu sledeće:

1. Verovanja ne mogu da nas navedu na delanje.
2. Vrednosni sudovi ne mogu se ispravno izvesti iz čisto činjeničkih propozicija.
3. Moralni sudovi su samo izrazi osećanja i nemaju istinosnu vrednost (Cohon 2008, 2).

Čuvena je Hjumova teza da „razum jeste i treba da bude samo rob strastima, i nikada ne može pretendovati na bilo koju drugu ulogu osim da im se pokorava i služi" (Hume 2011, 306). Filipa Fut smatra da je Hjum pokušavao da „identifikuje postupak ili kvalitet kao vrle u kontekstu posebnog osećanja" (Foot 1978, 76). U svojoj filozofiji morala, Hjum, kao i mnogi drugi autori, polazi od opštijeg pogleda na ljudsku prirodu i ljudsko znanje. On smatra da je ljudski um prepun utisaka i ideja, a da je svaka ideja kopija nekog utiska; afekcije ili strasti (tj. emocije, osećanja) jesu utisci i neposredno su dostupni svesti (up. Cekić 2013, 115; Cohon 2008, 4).

Kada govori o moralnosti, Hjum kreće od „opšteg pogleda" (Hume 2009, 581). „Opšti pogled" na moralne sudove gleda kao na proizvod osećanja koja nas navode da odobravamo ili ne odobravamo neki postupak. Nakon toga, a na osnovu toga, dolazimo do procene karaktera neke osobe. Kada procenjujemo karakter neke osobe, to činimo iz perspektive „prvog lica", a na osnovu osećanja koja su u nama izazvana. Kod Hjuma je karakteristično to što mi moralnu procenu dajemo na osnovu očekivanja koje će emocije u nama ta osoba izazvati. Na koji način? Hjum polazi od dobro poznate činjenice da se svi ljudi nalaze u nekom odnosu, koji on naziva „užim krugom" (Hume 2009, 602), u koji spadaju porodica, prijatelji, komšije. Sud o osobi mi dajemo na osnovu emocija koje ta osoba u nama izaziva, kao što smo već naveli. Dalje, mi sud o osobi dajemo na osnovu njenog karaktera i osećanja koja se obično povezuju s tim karakternim crtama, svrstavajući ih u odnosu na očekivani efekat u određenoj situaciji, pre nego na osnovu stvarnih posledica. Hjum to naziva „opštim pravilima" (Hume 2009, 585).

Tom tezom Hjum pokušava da obezbedi opštost moralnih sudova. Ideja koja stoji u osnovi jeste da su ljudi koji imaju hvale vredne kvalitete – ponosni što te kvalitete imaju, dok ljudi kojima ti kvaliteti nedostaju osećaju stid zbog toga. Hjum dakle smatra da su karakterne crte sposobne da stvore moralne sentimente koje karakterišemo kao vrle ili loše, zavisno od toga da li proizvode zadovoljstvo ili ne.

Kada je reč o karakteru, Hjum smatra da je o njemu moguće govoriti na dva načina: u užem smislu, s obzirom na konkretne karakterne crte (kao što su hrabrost, poštenje itd.), dok u širem smislu Hjum o karakteru govori kao o skupu karakteristika koje pojedinac poseduje (Hume 2009, 575). Jasno je, pritom: kada govori o vrlinama, on govori o određenim karakternim crtama za koje kaže da su „trajni principi uma" (Hume 2009, 575). Prema Hjumovom shvatanju, za filozofiju morala je najvažnije da su naša dela indikativna trajnim kvalitetima uma da bi osoba mogla da za njih bude odgovorna (Russell 2013, 96).

Dejvid Hjum dalje naglašava da niko ne bi mogao ozbiljno da postavi pitanje da li više želimo da se nama ili ljudima koji nas okružuju dive ili da nas mrze (Hume 2006, 280). Prirodno je da želimo da povećamo osećanje zadovoljstva a smanjimo stid, te ćemo težiti da negujemo vrline a izbegavamo poroke. Vidimo, dakle, da je za Hjuma moralna osoba ona osoba koja dela u skladu sa prirodnom inklinacijom da se ponaša moralno.

Moglo bi da se postavi pitanje šta se dešava kada osobi nedostaje motivacija za moralnost i kada je ta osoba toga nedostatka svesna. I sam Hjum je bio svestan mogućnosti takvog prigovora. U tom slučaju, kaže on, osoba može da bude vođena dužnošću da neko delo izvrši, bez motiva (Hume 2009, 479). Da bi mogao da objasni taj nedostatak motivacije kod delatnika, Hjumu je bilo potrebno da uvede pojam „veštačkih vrlina". Primera radi, pravda je veštačka vrlina i samo osećanje dužnosti jeste motivišuće za određeno delanje u skladu sa tom vrlinom. Moralna obaveza prema pravdi stvara se u odnosu na osećanje odobravanja ili neodobravanja u odnosu na pravdu, to jest u odnosu na nepravdu. Ovde bi trebalo biti na oprezu. Naime, Hjum smatra da osećanje odobravanja samo po sebi nije direktan pokretač za delanje jer, kako dalje navodi, ono je samo motiv, a ne i „spoljašnja izvedba" (Hume 2009, 478).

Hjum, dakle, vrline deli na prirodne i veštačke. Veštačke vrline su stvar konvencije i odnose se prvenstveno na imovinu i obećanja, čime se uspostavljaju osnovna pravila i ističe obaveznost pravde (Russell 2013, 98).

> Jedina razlika između prirodne vrline i pravednosti leži u tome što se dobro koje proizlazi iz prvog, javlja iz svakog pojedinačnog dela, i objekat je neke prirodne strasti, dok bi neko pravedno delo, posmatrano samo po sebi, često moglo biti u suprotnosti sa javnim dobrom (Russell 2013, 100).

Prema Hjumovom mišljenju, i kada je reč o nedostatku neposredne motivacije da se postupa na određeni način, naposletku se postupanje i u takvim slučajevima oslanja na osećanje slaganja ili korisnosti. On smatra da je lični interes osnovni interes koji nas navodi da učestvujemo u „sistemu pravde". Naime, pravednost, kao veštačka vrlina, postoji u „sistemu pravde". Da bi taj sistem mogao da funkcioniše, potrebna je spona između

pravednih dela pojedinca i sistema, a ta spona je simpatija. Simpatija je osećanje zahvaljujući kome želimo da očuvamo taj sistem i zbog čega osuđujemo sva dela koja urušavaju njegovo funkcionisanje (Hume 2009, 500). Za Hjuma,

> [I]mati osećaj za vrlinu nije ništa drugo do osećanje zadovoljstva posebne vrste, koje proističe iz kontemplacije karaktera. Samo osećanje proizvodi našu pohvalu ili divljenje. Ne idemo dalje, niti dalje istražujemo uzrok toga zadovoljstva. Mi ne zaključujemo da je karakter vrl jer stvara zadovoljstvo, već u osećanju zadovoljstva posle određenog postupanja mi posledično osećamo da je vrl (Shaw 1993, 45).

Hjum ovde želi da naglasi da ne moramo da ulazimo u mehanizme ljudske prirode koji nas navode da se osećamo kako se osećamo povodom određenog postupka. Dovoljno je da (odnosno to što) to osećanje imamo.

Savremeni zastupnici sentimentalizma empatiju stavljaju u centar svojih istraživanja, smatrajući da ona koja konstituiše moralno odobravanje ili neodobravanje. Prema rečima Majkla Slota, empatija je posebno moralno čulo, o kome je govorio već Hačeson (up. Slote 2007; Slote 2010). Ako se Slot, jedan od najpoznatijih zastupnika sentimentalizma današnjice, poziva na definiciju empatije iz 18. veka, naša je obaveza da ispitamo šta je Fransis Hačeson pod tim pojmom podrazumevao.

Hačeson u suštini želi da pokaže na koji način se dolazi do moralnog odobravanja koje se oslanja na nesebiče afekcije. Njegov pogled na čoveka ne polazi iz perspektive korisnosti, nema voluntaristički karakter u bilo kom obliku niti umski određene principe. Njegovo uverenje se oslanja na ono što on shvata kao činjenicu o ljudskoj prirodi: imamo moralno čulo koje direktno objašnjava osećanje odobravanja. Za Hačesona moralno čulo je – za refleksivne akte odobravanja – najverovatnija

> Hipoteza za objašnjenje moralnog iskustva, i zadaje sebi mnogo muke da ga opravda, pomoću svakakvih analogija, na osnovu celokupne ljudske nastrojenosti. Ali kolebanje i protivrečnost njegovog načina izražavanja, koje izgleda da je protiv jedva izabrane oznake 'čulo' odmah ponovo primoran da protestuje, primetno ukazuje na teškoću koja se nalazi u samoj stvari: da se tačnije odredi suština i način delovanja ove sposobnosti koja se nalazi na sredini između osećanja i uma (Jodl 1963: 203).

Fransis Hačeson prepoznaje značaj ličnog interesa u izborima koji nas usmeravaju u vođenju života u skladu sa vrlinom. Vrli život dovodi do dobra za delatnika ali, prema Hačesonu, tu se ne iscrpljuje zašto bi trebalo da budemo vrli. Naime, osim dobra za samog delatnika, postoji i intrinzična moralna obaveza da postupamo u skladu sa vrlinom. Moralno čulo nas intrinzično obavezuje na moralnost. Na koji način?

Hačeson kaže:

> Ako pod obavezom podrazumevamo motiv iz interesa, dovoljan da odredi sve one koji ga razmatraju, i koji mudro idu ka svojoj koristi putem određenog delanja, mi možemo da imamo čulo takve obaveze time što vršimo refleksiju o toj determinaciji naše prirode da odobri vrlinu, da budemo zadovoljni i srećni kada vršimo refleksiju o tome da smo učinili vrlo delo, a da, kada se ne osećamo lagodno, jesmo svesni da smo učinili suprotno i, takođe, koliko više cenimo sreću vrline u odnosu na bilo koje drugo uživanje (Hutcheson 2008: 178).

Vidimo da Hačeson razlikuje dva smisla u kojima smo obavezani da budemo benevolentni. S jedne strane, benevolenciju odobravamo nezavisno od interesa, dok, s druge strane, benevolenciju odobravamo na osnovu interesa. Savremeni interpretator Hačesona Majkl Valšots (Michaels Walschots) tu drugu vrstu (iz interesa) dalje raščlanjuje na dve podvrste:

> 1. benevolencija omogućuje da doživimo posebnu vrstu 'pratećeg' zadovoljstva kada delamo – zadovoljstva moralnog čula, i 2. benevolencija je najpouzdaniji način da se donese prirodno dobro ili zadovoljstvo, kako sebi tako i drugima (Walschots 2022).

Vidimo da Hačeson odbacuje primat uma zarad moralnog čula. Možemo zato reći da, za Hačesona, mesto uma u moralu od konstitutivnog postaje regulativno. Uloga uma je da koordinira i raspoređuje naše čulne utiske i osećajne afekcije (up. Hutcheson 2014, poglavlje 2, paragraf V).

Sada treba pogledati šta je vrlina za savremenog sentimentalistu. Rečima Majkla Slota, „vrlina je brižnost". Slot dalje ukazuje na direktnu vezu između vrline i empatije, pojma koji zauzima centralno mesto u njegovom izlaganju ideje moralnog sentimentalizma.

1. Moralna vrlina... je svojstvo pojedinaca ili njihovih stavova ili delanja, koje mi shvatamo putem empatije kao osećaj topline.
2. Jedino svojstvo koje putem empatije može da se primeti kao toplina jeste toplina brižnosti.
3. Dakle, moralna vrlina sastoji se od pružanja tople brižnosti (Kauppinen 2017: 871).

Za Majkla Slota se, dakle, vrlina sastoji u spoljnom pokazivanju određenog svojstva karaktera, koju pojedinac poseduje. To svojstvo se oslanja na empatiju, ali se manifestuje toplinom u međusobnom odnosu. Iz toga možemo da zaključimo da je za Slota nužan (ostaje nejasno da li je i dovoljan) uslov moralnosti osećanje empatije koje mora da bude vidljivo i drugim delatnicima.

Osvrnimo se ukratko na to šta normativne teorije uzimaju za predmet moralne ocene:

> [U] uobičajenom (ne-tehničkom) smislu podrazumeva se da predmet moralne (pr)ocene, gledano kroz istoriju etičkih teorija, nesumnjivo čine: 1) karakter (sačinjen od specifičnih osobina – 'vrlina') i način života moralnog delatnika, 2) postupci (radnje), i 3) posledice postupaka (Cekić 2023).

Slot, kao etičar vrline u širem smislu, a sentimentalista u užem, kao osnovni kriterijum moralnosti vidi karakternu crtu empatiju, uz dodatak da ta karakterna crta treba da bude jasno vidljiva i drugima.

Pobornici moralnog sentimentalizma analiziraju na koji način funkcioniše empatija u stvaranju moralnog odobravanja ili neodobravanja. Kada analiziramo tezu da je empatija moralno čulo, šta ona zapravo sadrži? Zastupnici te teze žele da kažu da putem empatije saznajemo toplinu (ili nedostatak topline) delatnikovih motiva. Dalje, moralna dobrota ili vrlina znači da delatnik bude toplog srca (*warm-hearted person*). Nasuprot tome, zloba se sastoji od toga da smo indiferentni prema drugima ili, rečima sentimentalista, da smo hladnog srca (*cold-hearted*) u odnosu sa drugim ljudima. Naposletku, možemo da zaključimo da putem empatije saznamo vrlinu ili (karakterne) mane drugih.

Majkl Slot smatra da je jedini „način na koji možemo da pokažemo kako referisanje moralnih pojmova kao što je 'moralno dobro' ili 'ispravno' jeste fiksirano, taj da možemo reći da je apriori jasno da moralno dobro (ili ispravno) jeste bilo koje osećanje ili toplina usmerena ka delatnicima i ispostavljena mehanizmima empatije koja ih je uzrokovala" (Slote 2010, 61). Slot tvrdi da empatična osećanja fiksiraju referenciju moralnih pojmova, što, kada je reč o „moralnom dobru", to znači da se tvrdi „stvar koja izaziva toplinu i empatična osećanja prema delatnicima" (Kauppinen 2017, 869). Moralni sentimentalisti u centar moralnog procenjivanja stavljaju „direktnu brigu za druge ljude, za ono što bismo mogli da nazovemo prirodnom vrlinom brižnosti. Ako empatija utiče na ljude da se suzdrže od nanošenja bola ili ubijanja (ili da to dozvole), onda mnogima sa razvijenom empatijom u većini slučajeva nisu potrebni moralni principi da im kažu da je pogrešno ubiti" (Slote 2010, 94). Kao što možemo videti, Slot (u knjizi *Moralni sentimentalizam*) pokušava da uspostavi kriterijum ispravnosti postupanja koje moralnost izvodi iz toga da se jasno vidi da postupak u sebi sadrži empatiju ili empatičnu brigu za druge. U suprotnom, nedostatak vidljive empatije znači da je postupak nemoralan.

Anti Kaupinen (Antti Kauppinen), autor koji se bavio moralnim sentimentalizmom, naveo je i nekoliko primedbi koje bi mogle da se upute „empatičnom odobravanju":

a) Moguće je da odobravamo / ne odobravamo nešto bez verovanja da je to moralno dobro / loše ili pogrešno.

b) Moralno odobravanje je stav sa intencionalnim sadržajem – uključuje razumevanje nekog H kao dobro ili loše (ili, moguće, na neki povezani ne-konceptualni način).
c) Postoji razlika u mišljenju da je nešto dobro i da je nešto ispravno (da može da bude dozvoljeno).
d) Ako mislite da je nešto moralno ispravno ili pogrešno, nije vas briga kako drugi ljudi na to reaguju.
e) Normalno, moralno neodobravanje nečega motiviše da se to izbegava.
f) Barem neke forme moralnog neodobravanja zahtevaju odgovor (Kauppinen 2017, 875).

Vrlina za moralne sentimentaliste, kao što smo mogli da vidimo, znači delanje u skladu sa osećanjima koja imamo. Priroda normativnih principa je takva da moraju da pretenduju na univerzalnost. Osim toga, moglo bi se reći da moralni principi imaju prinudni karakter: sasvim je prirodno da ako nešto moralno odobravamo, želimo da se i drugi slažu sa našim (moralnim) odobravanjem toga (up. Blackburn 1998). Preferencije u pogledu, na primer, ukusa sladoleda ne utiču na našu ocenu karaktera osobe koja ima ukus drugačiji od našeg. Ali, odnos delatnika prema laganju svakako će biti bitan deo naše procene njegovog karaktera.

Jasno je da se ovde ne uspostavlja jasan kriterijum koji bi mogao da se upotrebi u proceni i oceni moralnosti određenog postupka, što znači da se suočavamo sa nekoliko problema. Šta se dešava kada delatnik prosto ne oseća empatiju prema određenom postupku ili, čak, klasi postupaka? Postoji mogućnost da neko zaista ne oseća emaptiju, pa se postavlja pitanje da li treba da je indukujemo veštački. Šta bi to, onda, tačno značilo?

Setimo se da smo nešto slično mogli da vidimo već na primeru veštačkih vrlina kod Hjuma. Indukovanje empatije bi unelo kognitivni element u odlučivanje o delanju, što bi značilo da delatnik zna da je nešto dobro i da bi trebalo da postupa u skladu s tim. Nije pritom jasno zašto nam je uopšte potrebna brižnost prema nekom drugom da bismo postupali moralno. Ako već u nekoj instanci znamo da je nešto ispravno, onda se čini redundantnim uvoditi empatiju kao nešto što, očigledno, nije nije definišući kriterijum moralnosti.

Sve u svemu, čini se da empatija može da bude deo morala, ali ne i osnov neke posebne normativne teorije. Ponekad se čak čini da je uloga osećanja da samo olakšaju delatnikove izbore, ništa više od toga. Ostaje, dakle, da moralni sentimentalisti pokažu na koji način bi moralni sentimentalizam mogao da postane normativna teorija. Za sada se čini da argumenti u prilog tome nisu ubedljivi.

Bibliografija

Blackburn, S. (1998). *Rulling Passions*. Oxford: Claredon Press.

Burton, J. H. (1846). *Life and Correspondence of David Hume*. Edinburgh: William Tait.

Butler, J. (1749). Fifteen Sermons Preached at Rolls Chapel. In: Darwall, S. (1983). *Five Sermons*. Indianapolis: Hackett.

Cekić, N. (2013). *Metaetika: problemi i tradicije*. Novi Sad: Akademska knjiga.

Cekić, N. (2023). *Etika: teorijski minimum*. Novi Sad: Akademska knjiga.

Cohon, R. (2008). *Hume's Morality: Fellings and Fabrication*. Oxford: Oxford University Press.

Foot, P. (1978). *Virtues and Vices and Other Essays in Moral Philosophy*. Berkeley: University of California Press.

Frazer, M. (2010). *The Enlightenment of Sympathy: Justice and the Moral Sentiment in the Eighteen Century and Today*. Oxford: Oxford University Press.

Hume, D. (2006). *An Enquiry Concerning the Principles of Morals*. Oxford: Oxford University Press.

Hume, D. (2009). *A Treatise of Human Nature*. Auckland: The Floating Press.

Hutcheson, F. (2008). *An Inquiry Into the Original of Our Ideas of Beauty and Virtue: In Two Treatise*. Indianapolis: Liberty Fund.

Hutcheson, F. (2014). *A System of Moral Philosophy*. Cambridge: Cambridge University Press.

Jodl, F. (1963). *Istorija etike*. Sarajevo: „Veselin Masleša".

Kauppinen, A. (2017). Empaty as the Moral Sense?. *Philosophia*, 45: 867–879.

Lind, M. (1993). Hume and Moral Emotions. 133–149. In: O. Flanagan, A. O. Rorty (eds.). *Identity*, *Character and Morality*. Cambridge: MIT Press.

Russell, P. (2013). Hume's Anatomy of Virtue. 92–123. In: Russell, D. (ed.). *Cambridge Companion to Virtue Ethics*. Cambridge: Cambridge University Press.

Shaw, D. (1993). Hume's Moral Sentimentalism. *Hume Studies* XIX, 1: 31–54.

Slote, M. (2007). *The Ethics of Care and Empathy*. Oxford: Oxford University Press.

Slote, M. (2010). *Moral Sentimentalism*. Oxford: Oxford University Press.

Walschots, M. (2022). Hutcheson's Theory of Obligation. *Journal of Scottish Philosophy* 2 (2): 121–142.

Milica Smajević Roljić*

CONVERSATIONS WITH KANT: ON THE RIGHT TO REVOLUTION

Abstract: It is often argued that Kant's understanding of the right to revolution is contradictory. On the one hand, he expresses enthusiasm for the French Revolution and the ideas on which it rests, while on the other, he openly denies the existence of a legal right to revolution. This paper aims to make Kant's position plausible by showing that he does not deny the right to revolution in all states, but only in those that fulfill the purpose for which they were created, which is to protect the rights and freedoms of all citizens.

Keywords: state, legal right, rebellion, reform, public use of reason, freedom, state of nature.

Immanuel Kant is considered one of the greatest sympathizers of the French Revolution (see: Beiser 1992:36), the father of liberalism and the Enlightenment movement, a fighter for the autonomy of each individual, and a philosopher who placed the problem of human freedom at the core of his teaching. He openly writes about the enthusiasm that the French Revolution generated among its observers, attributing it to the "moral predisposition in the human race" (SF 7:85).[1] At the same time, Kant unequivocally and vehemently rejects the right to revolution in many of his

* Department of Philosophy, Faculty of Philosophy, University of Belgrade, milica.smajevic.roljic@f.bg.ac.rs

1 The following abbreviations have been used: MS: *Die Metaphysik der Sitten;* RGV: *Die Religion innerhalb der Grenzen der bloßen Vernunft*; SF: *Der Streit der Fakultäten*; TP: *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis*; WA: *Beantwortung der Frage: Was ist Aufklärung?*; ZeF: *Zum ewigen Frieden*; Refl: *Reflexionen*. The numbers refer to volume and page in the Prussian Academy edition. Translations are from *Immanuel Kant: Practical Philosophy*. Ed. Mary J. Gregor, Cambridge: Cambridge University Press, 1996., *Immanuel Kant: Religion and Rational Theology*. Ed. Allen Wood, Cambridge: Cambridge University Press, 1996.,

published works (see: MS 6:320, ZeF 8:381, TP 8:302). He condemns all forms of rebellion[2], even those directed against an unjust ruler who violates the rights of citizens (see: TP 8:300).

Kant's views on revolution seem paradoxical and confusing to his readers and scholars, who have been trying for decades to find a solution that will make Kant's position consistent. His contradictory claims have raised and continue to raise a number of questions, such as: "If revolution is always wrong, how can the spectators of the French Revolution, including Kant himself, justify this feeling of enthusiasm?" (Surprenant 2005:151); "How are we to reconcile Kant's denial of the right to resist the sovereign with what appears to be an endorsement of the French Revolution in his essay 'The Conflict of the Faculties'?" (Zreik 2018:197); How can the human right to freedom, which Kant believed belonged to every human being (see: MS 6:238), be reconciled with the denial of the right to resist an unjust sovereign?

Various attempts have been made by Kant scholars to resolve these contradictions, but no consensus has yet been reached. Some authors suggest that Kant betrayed the basic principles of his practical philosophy and that the condemnation of the revolution in his published works was out of fear of Prussian censorship (see: Maliks 2014:113, Beiser 1992:52–53), while others have tried to reconstruct "what they take to be a more consistent Kantian view, where his basic principles would support a right of resistance" (Maliks 2014:113). Understanding his position is made even more challenging by the fact that Kant did not write a systematic and comprehensive work on revolution, and that his views were presented in several books and articles published at different times.

This text seeks to make Kant's position plausible by arguing that he believed there was no right to revolution as long as the state fulfilled its

---

and *Immanuel Kant: Notes on Metaphysics*. Ed. Paul Guyer, Cambridge: Cambridge University Press, 2005.

2 In *Metaphysics of Morals*, Kant mentions various types of civil disobedience, such as resistance, rebellion, and revolution (MS 6:320), but he never explains the difference between these terms. Peter Nicholson believes that the terms used by Kant can be roughly divided into three categories. The term "resistance" has the widest scope and refers to civil disobedience in the broadest sense; "rebellion" has a narrower meaning and presumably refers to armed resistance by which the sovereign is forced to act in a certain way or to abdicate power; and "revolution" refers to a special kind of rebellion and has the narrowest meaning of these three terms (Nicholson 1976:215). Although Nicholson's terminological explanations may be correct and useful in some cases, reading Kant's text reveals that he did not attempt to make precise distinctions between these terms, especially not between "rebellion" and "revolution" which he uses interchangeably (Nicholson 1976:216). Therefore, all of the above-mentioned terms will be treated as synonyms in this paper.

purpose: the preservation of the rights and freedoms of its citizens. However, if the state does not perform its primary function, it resembles the state of nature, and citizens have no duty to respect it, but rather to fight for a new civil society by all available means, even violent. In this way, we will show that Kant's understanding of revolution is inconsistent only at first glance, until we become acquainted with his political philosophy. In the first part of the paper, we will see on what grounds Kant rejected the right to revolt (even against imperfect rulers), as well as why he believed that returning to the state of nature is always worse than respecting the current government. The second part emphasizes the importance of the public use of reason and reform in Kant's political philosophy. According to Kant, as long as these elements are present in the state, the government should be obeyed because there is a possibility of changing and improving existing laws that citizens consider unjust. In the third part of the text, we will quote passages from Kant's works to show that he believed that in cases when the government does not respect basic human rights and freedoms, it loses legitimacy and the people have the right to revolt. In this way, we will show how, relying primarily on Kant's own words, his understanding of the revolution can be made plausible.

Is there a (legal) right to revolution?

Suppose we live in a state where the government is corrupt and the sovereign is unjust.[3] More and more citizens are dissatisfied with the ruling regime and plan to organize a rebellion against it. Let us also imagine that Kant is one of our fellow citizens, and several of our compatriots ask him to join in organizing the revolution. Kant's answer would probably be the following:

> "Any resistance to the supreme legislative power, any incitement to have the subjects' dissatisfaction become active, any insurrection that breaks out in rebellion, is the highest and most punishable crime within a commonwealth, because it destroys its foundation. And this prohibition is *unconditional*, so that even if that power or its agent, the head of state, has gone so far as to violate the original contract and has thereby, according to the subjects' concept, forfeited the right to be legislator inasmuch as he has empowered the government to proceed quite violently (tyrannically), a subject is still not permitted any resistance by way of counteracting force" (TP 8:300).

3 This assumption raises old dilemmas. Even the Roman philosopher Seneca, in his treatise *On Leisure* (*De Otio)*, argued that a wise man should not participate in any government that is corrupt (see: Plećaš & Nišavić 2022).

Kant would, therefore, unequivocally refuse to join the revolution and would reject any possibility of a legal right to resist the ruler. His argument is based on the claim that positive legislation cannot contain a law that would allow its destruction (see: MS 6:321). The constitution cannot contain any article that allows resistance to the sovereign, because if any opposition to absolute and supreme power were allowed, that power would be neither absolute nor supreme, which would create contradictions. Therefore, no legal institution based on the principles that lead to its dissolution is possible (see: MS 6:372). Revolution denies established laws and implies a return to the state of nature, which is why positive legislation unequivocally condemns it. Beck argues that we should not be surprised by Kant's argument, as is it clear, obvious, and simple. "Revolution abrogates positive law; therefore, positive law and its system condemn revolution" (Beck 1971:414). Hence, there is no legal right to rebel against a legitimate government. The ban on raising a revolution is unconditional and no exceptions are allowed.

Although we can agree with Beck that Kant's legal argument is obvious[4], it is very likely that Kant's fellow citizens would be dissatisfied with the offered answer and insist on additional explanations. Even if they agreed that the constitution could not contain a basis for its own abolition, they would probably ask: "Isn't even a return to the state of nature better than living in an unjust society?"

To understand why Kant believed that a return to the state of nature is inadmissible and that any government is better than a state of powerlessness (see: TP 8:300), we must briefly recall the basic elements of his political philosophy and explain the relationship between Kant's understanding of justice and the state. Basic human rights cannot be guaranteed in a hypothetical state of nature, which represents a state of powerlessness, which is why it is necessary to abandon it and form an orderly civil society.[5] By remaining in the state of nature, an individual cannot protect their property and their rights, which in those circumstances are only provisional (see: MS 6:257), because there is no contract "in which we reciprocally commit ourselves to guaranteeing each other's rights" (Korsgaard

4 In addition to the provided legal argument against the right to revolution, Kant offers at least two other arguments against the right to revolt in his works: the argument based on publicity (see: ZeF 8:381) and the argument based on the principle of happiness (see: ZEF 8:379). An analysis of these arguments exceeds the scope of this paper.

5 According to Kant, the state of nature is just a hypothetical, transcendental idea, which allows us to see the importance of the existence of social institutions, not a historical state that once existed and in which people lived without the rule of law and protection of their rights (see: Korsgaard 1997:303; Smajević 2020:208).

1997:302). The state of nature is always a state of injustice, or at least "a state *devoid of justice* (*status iustitia vacuus*), in which when rights are *in dispute* (*ius controversum*), there is no judge competent to render a verdict having rightful force" (MS 6:312). That is why Kant contended that each individual has the right to "impel the other by force to leave this state and enter into a rightful condition" (MS 6:312) in which institutions for a fair trial and the realization of each individual's personal freedom will be established. A legal condition can only exist within political society.

The state and justice are inextricably linked because justice can only exist in the state; the state is the source of justice. Citizens must obey the state to which they belong. The duty to form a state, as well as the duty not to resist the sovereign, is based on the need for a clear and solid legal framework that ensures the freedom and autonomy of all citizens. From all the above, we understand why Kant believed that maintaining the existing civil society (no matter how deficient it may be) was always better than returning to the state of nature. By entering civil society, the people unite under a general legislative will, embodied in government and sovereign, which has the task of protecting the rights and freedoms of all citizens. Even if the current government is corrupt and does not complete its task in the best possible way, it is still better than a state of complete anarchy and powerlessness with no legitimate judge to resolve ongoing disputes. Korsgaard stresses that "the imperfections of the actual state of affairs are no excuse for revolution – if they were, revolution would always be in order" (Korsgaard 1997:319).

Reform instead of revolution?

After hearing Kant's explanation, the fellow citizens who invited him to join them in their rebellion against the current government would most likely feel hopeless: even if they adopted Kant's argument, they would still believe that they live in an unjust society that restricts their freedom, violates their rights, and makes them unhappy. They would probably conclude that Kant believes that citizens never have the right to fight for a more just and egalitarian society, and that the established laws, however flawed, can never be changed by legal means.

However, this is by no means Kant's view. His compatriots would be surprised if Kant told them that every citizen "has complete freedom and is even called upon to communicate to the public all his carefully examined and well-intentioned thoughts about what is erroneous" (WA 8:38) and thereby incite changes in society. As Surprenant puts it: "Kant's posi-

tion is not that laws in a state are unable to be changed, but rather the legitimate mechanism for change is internal, coming from the legislators themselves, not the citizens – at least not through the use of coercive force. The method available for citizens to incite change in the policies of the government is through non-coercive means, through speech and writing for example" (Surprenant 2005:156). Although he does not justify revolution, Kant believes that every individual that sees the unfairness of the political system is called upon to speak about it publicly and thereby contribute to the necessary changes.

To explain when and where citizens can publicly express their opinion, Kant introduces a distinction between private and public use of reason, where the former must be "narrowly restricted," while the latter "must always be free" (WA 8:37). "What I call the private use of reason is that which one may make of it in a certain civil post or office with which he is entrusted" (WA 8:37). Whether a teacher, professor, clergyman, or soldier, every citizen is obliged to show obedience to the state and perform their service as prescribed by law. While performing their duty, no citizen may question the correctness of the orders received from the state. Kant says "it would be ruinous if an officer, receiving an order from his superiors, wanted while on duty to engage openly in subtle reasoning about its appropriateness or utility; he must obey" (WA 8:37). Here again we see Kant's view that the state's established legal system must be respected without exception.

However, although no citizen has the right to refuse or question their performance of official duties, every citizen, as a scholar, has not only the right but also an obligation to "publicly expresses his thoughts about the inappropriateness or even injustice" (WA 8:38) of state decrees. In Kant's words:

> "A citizen must have, with the approval of the ruler himself, the authorization to make known publicly his opinions about what it is in the ruler's arrangements that seems to him to be a wrong against the commonwealth. For, to assume that the head of state could never err or be ignorant of something would be to represent him as favored with divine inspiration and raised above humanity. Thus, *freedom of the pen* is the sole palladium of the people's rights" (TP 8:304).

If the state encourages freedom of thought, speech, and writing, the reform of the existing system and the progress of society are highly probable (see: MS 6:355). The sovereign, as a human being, is fallible and the principles on which they act may be unjust and sometimes even cruel. That is why every individual must have the right to draw attention to laws and principles that they consider incorrect, which should compel the sov-

ereign to implement reforms and amend existing laws. The reform cannot be carried out by anyone other than the holder of the legislative power, because it is the only legitimate way to achieve a just socio-political system (see: MS 6:321–322). Reform leads to progress and restoration of the state, while revolution returns us to a state of lawlessness. Kant concludes that he, unlike Hobbes, believes that "the people too has its inalienable rights against the head of state, although these cannot be coercive rights" (TP 8:304).

Therefore, although Kant would not join his fellow citizens in resisting legitimate authority and would instead draw their attention to the illegitimacy of such an act, he would not advise them to be passive and suffer injustice but rather give them clear instructions on how to try to solve the problem. He would invite them to speak and write publicly about the injustices present in society, while also drawing their attention to the fact that they must not do so in their workplace where they have the duty to respect state orders. Listening and reading the citizens' observations should make the sovereign understand the importance and necessity of changes and their implementation. "If these reforms are necessary, it is a duty for the government to undertake them, as it is the only legitimate way of realizing the highest political good" (Reiss 1956:186). Kant firmly believed that social progress can be achieved in this way, while revolution would only lead to chaos and lawlessness. However, it should be noted that "reforms can be brought about only within a considerable interval of time" (Reiss 1956:186), not in a day, week, or month. That is why Kant would probably advise his fellow citizens to be patient and persistent.

What if the government does not undertake reforms?

While publicly criticizing existing laws and/or their application is allowed in some states (according to Kant, most often in those with or aspiring to a republican system [see: ZeF 8:350]), in others it is prohibited or does not lead to the desired result: the implementation of reforms by the ruler. Public use of reason can be a good way to incite change in the first type of state, while in the second type public speaking is either prohibited or ineffective. If we imagine that Kant lived in a state where "freedom of the pen" is encouraged, then we can say that he gave good advice to his fellow citizens when he recommended public criticism of the government rather than revolution. However, if we assume that Kant and his compatriots lived in a society where public speech and writing were subject to censorship, then the public use of reason cannot bring about the desired changes and the formation of a new, more just order.

If Kant's fellow citizens said that they tried to publicly expose all the injustices of the existing system but were prevented from doing so due to the harsh censorship present in public life, how would Kant respond? Would he allow the right to revolution and on what grounds? Or would he offer another solution? We have reason to believe that Kant would allow the right to revolution at this point.[6] Several (often overlooked) passages in Kant's writings indicate that he believed that citizens were not obliged to obey the government under all circumstances. For example, in *Religion Within the Boundaries of Mere Reason*, he says that "when human beings command something that is evil in itself (directly opposed to the ethical law), we may not, and ought not, obey them" (RGV 6:100). Then, in *Reflections*, he claims that "the people cannot rebel except in the cases which cannot at all come forward in a civil union, e.g., the enforcement of a religion, compulsion to unnatural sins, assassination, etc." (Refl 19:594–595, see: Beck 1971:412).

While the first passage justifies passive disobedience, the second indicates the conditions under which resistance is justified. Kant seems to think that certain acts of the sovereign do not befit the so-called state or civil order. In other words, when the state prohibits its citizens from publicly expressing their religious, political, moral and other views, and when it imposes immoral and unjust demands upon them, the possibility of justified resistance to the ruler arises. In recent decades, Kant's scholars have begun arguing that Kant does not reject the right to revolution in all states but only within constitutional ones (see: Maliks 2013:33). The argument goes roughly as follows:

"Although revolution is always *prima facie* wrong, it is not wrong to revolt against a civil state when it has failed to create or maintain a condition of civil society" (Surprenant 2005:161). In other words, revolution is always wrong when directed against the unconditional duty to preserve

6 Some authors believe that Kant would allow the right to revolt in situations other than when basic human freedoms are not respected. For example, Byrd and Hruschka (2010) believe that there is the right to rebel against any government that is not republican because Kant argued that "the civil constitution in every state shall be republican" (ZeF 8:350). As Maliks rightly observes, this view is very difficult to defend, primarily because it is inconsistent with Kant's claim that we ought to obey even an imperfect ruler (TP 8:300, Maliks 2013:33). Other interpreters, such as Ripstein, believe that the right to revolution exists when the state does not respect fundamental human rights. "Nazi Germany is the clearest example. These are cases of human rights violation so fundamental that they undermine the organization that commits them" (Ripstein 2009:337). Then, some authors argue that the right to rebel exists only when the state has already been dissolved (see: Maliks 2013:34). All of these authors recognize that Kant allows the right to revolution in certain cases.

civil order (see: Surprenant 2005:163) but it is not wrong if we do not live in a civil society. Citizens enter civil society to form institutions that protect their rights and freedoms and they may even use force to achieve this goal. However, if the ruling regime does not fulfill "the end for which the state exists" (Maliks 2013:29), that is, if there is no constitutional regime protecting the rights and freedoms of citizens, then the current situation resembles the state of nature and citizens have the right and obligation to fight for the formation of the state even by force.

If we follow this line of interpretation[7], we can conclude that Kant would only allow the possibility of a rebellion in the absence of basic freedoms in society. The state was created to protect our rights and freedoms, and it cannot be called a state if it does not fulfill this. In such a situation, we have the right to assume that Kant would have advised his fellow citizens to revolt.

> "If there is no civil society, then there is no civil law and we may use violence to establish it" (Axinn 1971:426). "Individuals have an obligation to resist the institutions of a civil state when the de facto holders of power in that civil state have either returned them to the state of nature or kept them in a state of nature condition" (Surprenant 2005:164).

The above quotations help us understand Kant's enthusiasm for the French Revolution. Dieter Henrich underlines that for Kant this was not a revolution in the conventional sense because there was no resistance to a legitimate ruler (see: Henrich 1996). He argues that Kant believed Louis XVI "abdicated [his sovereignty] and simultaneously returned the Estates to the state of nature" (Henrich 1996:111, Surprenant 2005:152). In other words, at the start of the revolution, Louis XVI was not the legitimate holder of state power but rather a former ruler who abdicated his sovereignty.[8] Kant's approval of the French Revolution can therefore be interpreted as support to the people to leave the state of nature and form a civil society, which ceased to exist with the ruler's abdication.

We can conclude that in his published works, Kant clearly and unequivocally rejects the right to revolution in all cases, except when the rights and freedoms of citizens are threatened to the point that the society they live in can no longer be called a state. When the existing state turns

---

7 Reidar Maliks emphasizes that "Jeremy Waldron (2006), Arthur Ripstein (2009) and B. Sharon Byrd and Joachim Hruschka (2010) have maintained the view" (Maliks 2013:30).

8 Chris Surprenant claims that "Kant's position on the French Revolution clearly suffers from historical inaccuracies" (Surprenant 2005:152), which does not change the fact that Kant believed that Louis XVI had illegitimately abdicated (see: MS 6:341) and does not affect the above argument.

into a state of nature[9], the citizens have the right, but also the obligation, to use all available means to fight for the establishment of a new state order. A state that does not respect the basic rights of its citizens is not a state at all.

Concluding remarks

This paper set out to investigate whether there is a contradiction in the fact that Kant decisively rejected the right to revolution in his juridical-political writings on the one hand, while openly showing enthusiasm for the French Revolution on the other. First, we showed that Kant's legal argument against revolution is based on the claim that no constitution can contain articles that permit its own destruction. Every state was formed as a guarantor of human rights and freedoms, and therefore an attack on it would represent an attack on the freedom of each of its citizens. As a result, from a legal perspective, citizens living in a civil society never have the right to revolution. Later we showed that Kant was aware that governments often make imperfect and unjust decisions. A perfect government in which the ideals of enlightenment, education, and eternal peace are realized is a goal that has not been attained in reality.[10] For that reason, Kant encourages citizens to, through the public use of reason, point out existing injustices in society to the ruler, thereby initiating the implementation of reforms. Finally, we provided arguments in support of the thesis that Kant allows the right to revolution only in cases where the ruler does not implement reforms and the state no longer fulfills the purpose for which it was created – the protection of the rights and freedoms of its citizens. Kant believed that the French Revolution was an example of such a revolution, and therefore his enthusiasm for this event did not contradict his rejection of the legal right to revolt.

In an attempt to show the plausibility of Kant's understanding of revolution, we have only dealt with the legal aspects of the argumentation. However, it is important to note that this is only one possible defense of Kant's position. Among the scholars who have sought to make Kant's posi-

9 Although it is difficult to determine "where exactly should the line be drawn between a highly imperfect regime that is still entitled to obedience, and a regime that has crossed the line and is no longer to count as a juridical condition" (Maliks 2013:36), in this paper we have suggested that the "distinction will revolve around" two questions: "whether reform of the present regime is possible" (Maliks 2013:36) and whether the basic rights and freedoms of citizens are respected.

10 For more on the process of achieving these goals, especially those related to education, see Smajević Roljić (2021).

tion consistent are those who believed that the key to the solution lies in the separation of the *legal* right to rebellion from the *natural* (see: Haensel 1926; Maliks 2014), *moral* (see: Korsgaard 1997), or *philosophical-historical* (see: Beck 1971) rights. We leave the consideration of these possibilities for future research.

References:

Axinn, Sidney (1971), "Kant, Authority, and the French Revolution", *Journal of the History of Ideas* 32 (3): 423–432.

Beck, Lewis (1971), "Kant and the Right of Revolution", *Journal of the History of Ideas* 32 (3): 411–422.

Beiser, Frederick (1992), *Enlightenment, Revolution, and Romanticism: The Genesis of Modern German Political Thought,* Cambridge, MA: Harvard University Press.

Byrd, Sharon and Hruschka, Joachim (2010), *Kant's Doctrine of Right: A Commentary*, Cambridge: Cambridge University Press.

Dieter, Henrich (1996), "On the Meaning of Rational Action in the State", in *Kant and Political Philosophy: The Contemporary Legacy,* ed. by Ronald Beiner and William James Booth, New Haven: Yale University Press.

Kant, Immanuel (1996), *Practical Philosophy*, trans. and ed. by Mary J. Gregor, Cambridge: Cambridge University Press.

Kant, Immanuel (1996), *Religion within the Boundaries of Mere Reason,* in *Religion and Rational Theology,* trans. and ed. by Allen Wood, Cambridge: Cambridge University Press.

Kant, Immanuel (2005), *Notes on Metaphysics,* trans. and ed. by Paul Guyer, Cambridge: Cambridge University Press.

Korsgaard, Christine (1997), "Taking the Law into Our Own Hands: Kant on the Right to Revolution", in *Reclaiming the History of Ethics. Essays for John Rawls,* Reath Andrews, Herman Barbara, Korsgaard Christine (eds.), Cambridge: Cambridge University Press, 297–328.

Maliks, Reidar (2013), "Kant, the State, and Revolution", *Kantian Review* 18 (1): 29–47.

Maliks, Reidar (2014), *Kant's Politics in Context*, Oxford: Oxford University Press.

Nicholson, Peter (1976), "Kant on the Duty Never to Resist the Sovereign", *Ethics* 86 (3): 214–230.

Plećaš, Tamara & Nišavić, Ivan (2022), „Rimski stoici Seneka i Epiktet o epikurejskom hedonizmu i društvenim ulogama filozofa", *Theoria* 65 (3): 5–19.

Reiss, H.S. (1956), "Kant and the Right of Rebellion", *Journal of the History of Ideas* 17 (2): 179–192.

Ripstein, Arthur (2009), *Force and Freedom: Kant's Legal and Political Philosophy,* Cambridge, MA: Harvard University Press.

Smajević Roljić, Milica (2021), "An Interpretation of the Educational Process from the Perspective of Kant's Philosophy of History and Legal-Political Theory", in *Liberating Education: What from, What for?,* Cvejić Igor, Krstić Predrag, Lacković Nataša, Nikolić Olga (eds.), Beograd: Institut za filozofiju i društvenu teoriju: 83–100.

Smajević, Milica (2020), „Prosvećenost u Kantovoj misli – metodološke pretpostavke i aktuelnost u savremenoj filozofiji", in *Nauka i savremeni univerzitet 9,* Šaranac Stamenković Jasmina, Skrobić Ljiljana, Ilić Mirjana, Kaličanin Milena (eds.), Niš: Univerzitet u Nišu, Filozofski fakultet: 205–215.

Surprenant, Chris (2005), "A Reconciliation of Kant's Views on Revolution", *Interpretation* 32 (2): 151–169.

Zreik, Raef (2018), "Kant on Time and Revolution", *Graduate Faculty Philosophy Journal* 39 (1): 197–225.

Milica Smajević Roljić

Razgovori sa Kantom: o pravu na revoluciju

Apstrakt: Često se tvrdi da je Kantovo shvatanje prava na revoluciju kontradiktorno. Sa jedne strane, on izražava entuzijazam prema Francuskoj revoluciji i idejama na kojima ona počiva, dok sa druge strane otvoreno negira postojanje legalnog prava na revoluciju. Cilj ovog teksta je da se Kantova pozicija učini plauzibilnom tako što će se pokazati da Kant ne negira pravo na revoluciju u svim državama, već samo u onim koje ispunjavaju svrhu zbog koje su nastale, a to je zaštita prava i sloboda svih građana.

Ključne reči: država, legalno pravo, pobuna, reforma, javna upotreba uma, sloboda, prirodno stanje.

3. LESSONS FROM THE PAST – VIRTUES AND VICES FROM ANTIQUITY TO MODERN CORPORATE SCANDALS

Drago Đurić*

IGNORANCE AND THE GOOD LIFE: CARNEADES, SEXTUS EMPIRICUS, AND BLAISE PASCAL

Abstract: It is usually considered that our good and happy life depends on our knowledge. This paper explores theories according to which our beliefs and knowledge are unreliable and, therefore, can have a bad effect on some aspects of our practical life. The paper discusses the relationship between ignorance and the good life among ancient skeptics (Carneades, Sextus Empiricus) and Pascal. The presentation begins by citing the earliest thinkers who doubted the possibility of unquestionable and irrefutable knowledge. Then, on the subject of the limitations of our knowledge, the viewpoints of other classical Greek thinkers are briefly presented. The greatest attention was paid to the academic and Pyrrhonian skeptics, and above all to Carneades and Sextus Empiricus. In the end, the view of Sextus Empiricus is connected with Pascal's consideration of the so-called bets on god's existence.

Keywords: Carneades, Sextus Empiricus, Blais Pascal, ignorance, good life.

Introduction

I begin the paper by briefly discussing views on knowledge and science in ancient Greece. Some of the earliest thinkers held that we can have conclusive and irrefutable empirical knowledge. Others, for various reasons, concluded our knowledge of the world has a limited character. Still others defended the thesis that we cannot know anything for sure, while some argued we cannot even know whether we do not know that we do not know. Based on their individual views on knowledge, they all had an answer to the question of a good life. I will pay more attention to two skeptics, Carneades and Sextus Empiricus, and their understanding of practical action, as it rests on the impossibility of theoretical knowl-

* Department of Philosophy, Faculty of Philosophy, University of Belgrade, drago.djuric@f.bg.ac.rs

edge and thus on the impossibility of prediction. I will also refer to the question of the relationship between ignorance and religiosity in Sextus Empiricus and Blaise Pascal, under the conditions of the impossibility of knowledge about God's existence or non-existence. I will say something about how they practically solve the problem of the balance of knowledge or ignorance about God's existence and how they defend the thesis that piety leads to a better life, regardless of whether God exists. Based on this latter consideration, I will defend the thesis that agnosticism does not necessarily have a position that would exclude theism and atheism, and its outcome can be theistic and atheistic fideism.

Pre-Socratics

Many ancient authors were aware that their theories were not conclusively justified. For example, according to the testimony of Demetrius Phaleron, the Milesian historian Hecataeus characterized his work in the following way:

> I write these things as they seem to me to be true. For the tales told by the Greeks are, as it appears to me, many and absurd /ridiculous/ (τάδε γράφω, ὥς μοι δοκεῖ ἀληθέα εἶναι: οἱ γὰρ Ἑλλήνων λόγοι πολλοί τε καὶ γελοῖοι, ὡς ἐμοὶ φαίνονται, εἰσίν).[1]

Although the majority of ancient Greek thinkers aspired to irrefutable knowledge and believed it could be achieved, some thought it was impossible. The context of Hecataeus' claim is not preserved, and thus we have no contextual explanation of his comment, but it seems clear that he considered his claim to be more convincing than the claims of other Greeks. It is also natural to assume that he came to his conclusion because he doubted the "many and ridiculous" stories of other Greeks.

Let me give one more example. According to Plutarch's testimony, Xenophanes said the following about one of his claims or of his claims more generally: "Let's look at this as if it was originally like that /as if it were true/ (ταῦτα δεδοξάσθω μὲν ἐοικότα τοῖς ἐτύμοισι)".[2] The basis for this claim can be found in other fragments of the preserved teachings of Xenophanes. Sextus Empiricus says Xenophanes also belongs among

1 Roberts, R. W (ed.). 1902. *Demetrius On style /De elocutione/*. Cambridge: Cambridge University Press, 1.12. Collected fragments of Hecateus' teaching see in: Giussepe, N. (ed.). 1954. *Hecataei Milesii Fragmenta*. Firenze: La nouva Italia.

2 Plutarch. 1892. *Moralia* 4. Gregorius N. Bernardakis (ed.). Leipzig: Teubner, 746b. DK 21B35 (DK = Diels, H./Kranz, W. 1985. *Die Fragmente der Vorsokratiker*, Zurich/ Hildesheim: Weidmann).

those who think that there is no criterion for truth, because, according to some, he argued "All things ... are non-apprehensible /ἀκαταληπτα/", and this could be concluded, he believed, based on his thoughts:

> Neither ever was nor ever will be a man who would have undoubted/ true (σαφές) knowledge (εἰδὼς) of the gods as well as of all that I speak of. For even if he were to happen to say [something that is true], he himself would still not know it (αὐτὸς ὅμως οὐκ οἶδε): for all is swayed by opining (δόκος δ' ἐπὶ πᾶσι τέτυκται).[3]

If we have the classic Platonic definition of knowledge in mind, we could assume Xenophanes believed that merely stated statements can be true, but we cannot know this unless we have undoubted, i.e., incorrigible and irrefutable, justification. But he believed we cannot attain such knowledge and therefore concluded that we can only arrive at opinions.[4] Yet it seems Xenophanes also believed in the advancement of investigation, saying the following in one fragment:

> Certainly, the gods did not reveal everything to mortals from the beginning, but [people themselves], searching over time, saw better (οὔτοι ἀπ' ἀρχῆς πάντα θεοὶ θνητοῖς ὑπέδειξαν, ἀλλὰ χρόνῳ ζητοῦντες ἐφευρίσκουσιν ἄμεινον).[5]

Xenophanes seems to have believed that although people can only reach opinions, they can still advance cognitively through those opinions.[6]

Although it follows implicitly from Xenophanes' assertions, Parmenides makes the boundary between truth or knowledge and opinion explicitly impassable. None of the cosmological theories of his predecessors can satisfy his strict criterion of truth. All cosmological and cosmogonic theories belong to what he calls the opinions of mortals. If, according to his criterion, it is not possible to speak truthfully about coming-to-be and passing-away, past or future, changes, or multitude, and this is necessary for every cosmological theory, then such theories can only represent opinions. Although they offer some argumentation for their theories, he be-

---

3 Sexstus Empiricus. 1976. *Adversus Mathematicos*. ed. G. R. Bury. Cambridge, Mass.: Harvard University Press / London: William Heinemann, 7.1.49. DK. 21B34.

4 And Sextus believes that Xenophanes under σαφές thinks "what is true and known" (Sexstus Empiricus. *Advesus Mathematicos* 7.1.50).

5 Hereen, H. L. A. (ed.). 1742. *Ioannis Stobaeus Eclogarum Phisicarum et Ethicarum*. Gottingae: Vandenhoek et Ruprecht. 1.9.2. DK. 21B18. In DK, it says 1.8.2 incorrectly, and this error is almost regularly inherited in the literature.

6 See more about it in: Tulin, A. 1993. "Xenophanes Fr. 18 D.-K. and the Origins of the Idea of Progress". *Hermes* vol. 121. 129–138. and Lesher, J. H. 1991. "Xenophanes on Inquiry and Discovery: an Alternative to the 'Hymn to Progress' Reading of F. 18". *Ancient Philosophy* vol. 11. 229–248.

lieves these theorists do not have built-in criteria for truth; instead, they indulge their senses and their senses lead them astray. Accordingly, Zeno of Elea, the defender of Parmenides, strictly opposes logical-mathematical argumentation of opinions based on phenomena accessible to our senses, trying to convince us that the implications of those opinions lead to paradoxical conclusions.

The so-called causal pluralists similarly insist on a difference between the real and the apparent nature of the world. However, they try to reach conclusions about the real nature of the phenomena available to our senses, starting from those phenomena and trying to explain them in terms of their real nature. In all these cases, the essential nature of the world is not immediately accessible to our senses.

Plato and Aristotle

Plato argues that the real world, that is, the intelligible world, which for him is the unchanging world of ideas, about which we can have certain and irrefutable knowledge, is strictly separated from the sense-appearing world, about which we can only have opinions. Investigations of the world of appearance in Plato's dialogues often begin with Socrates' interlocutor first expressing an opinion about the phenomenon being examined then, Socrates and the interlocutor progress together through the dialogue. The consideration of these questions is related to our sensory world and can only be used to arrive at something that has the character of likely argument/claim (εἰκός λόγος) or likely story (εἰκός μῦθος). The possibility of acquiring certain knowledge is reserved for the timeless and unchanging world of ideas. Plato believes that by considering sensory phenomena, we cannot reach irrefutable knowledge, knowledge that is "completely consistent and fully precise";[7] we can only reach opinions. The obstacle in our way is that "we have human nature (φύσιν ἀνθρωπίνην ἔχομεν)", and the subject of research, the sensory world, is changeable or unstable.[8] We can only reach "accounts no less likely (εἰκότας) than any other".[9]

7 Plato. 1960. *Timaeus, Critias, Cleitophon, Menexenus, Epistles*. Cambridge, MA: Harvard University Press, 1960, 29c 6.

8 Proclus claims something similar in his commentary on *Timaeus*. He says it this way: "Timaeus has recounted that the lack of reliability and precision in the case of the account on nature has two sources, (1) the essential nature of the realities themselves ... (2) the lack of power on the part of those carrying out the investigation" (Proclus. 2008. ed. and transl. T. D. Runia and M. Share. *Commentary on the Plato's Timaeus*. Cambridge: Cambridge Univesity Press. 351.20–27).

9 Plato. *Timaeus* 29c 7.

Aristotle, unlike Plato, thinks that through our research we can, even when starting from sensory experience, arrive at irrefutable knowledge about the nature of things. Moreover, he believes our investigation of nature must start from what the majority of people or the majority of scientists think about it,[10] i.e., from their opinions based on sensory phenomena, and that during our investigation, sooner or later, we can arrive at what is first in nature, or as Aristotle puts it in his *Physics*, we can move "from what is more obscure by nature, but clearer to us, towards what is more clear and more knowable by nature (ἐκ τῶν ἀσαφεστέρων μὲν τῇ φύσει ἡμῖν δὲ σαφεστέρων ἐπὶ τὰ σαφέστερα τῇ φύσει καὶ γνωριμώτερα)".[11]

Hellenism

The central question of Epicureans, Stoics, and skeptics is the question of a good or happy life. This is basically a practical question, but the answer to it cannot be arrived at without considering physics, psychology, and above all, epistemology. In short, the question of a good life depends on our knowledge of nature. We will not, for example, be afraid of death if we come to know that after it, we cease to exist, so, in fact, there is no one to be afraid of, not even the gods – that is, if we agree with the Epicureans' conclusion that their causal connection with the world is not in accordance with God's attribute of bliss or perfect happiness. And if we agree with the Stoics, we will not be disturbed by the fact that we cannot do something if we know, based on the study of nature, that it is not possible to do it by nature. In the same way, if we accept what the skeptics say, we will not be disturbed by the fact that we do not know something if, through research, we come to the conclusion that nothing can be known for certain.

As the Epicureans saw it, we owe all our knowledge to our senses, just as all our theories must be in accordance with them. They were atomistic reductionists; as such, they believed the impressions made on our senses are caused by matter coming from the things themselves. The fact that we sometimes have the wrong impression was explained by the Epicureans as a consequence of the interference of the body that appears between us and the things themselves, such as air or sunlight. However, these distractions

10 Aristotle. *Topics* 100b 22–101a 4. (= Aristotle. 1960. *Posterior Analytics. Topica.* Cambridge, MA: Harvard University Press).

11 Aristotle *Physics.* 184a17–184a21 (=Aristotle. 1968. *Physics.* Cambridge, MA: Harvard University Press), Aristotle. 1999. introd. and transl. T. Irwin *Nicomachean Ethics*. Indianapolis: Hackett Publishing Company. 1095b 1–4.

can be removed by further repeated observations, so that we can always reach unquestionable knowledge. The Epicureans investigated nature, but for them, physical science had no goal of its own. In this understanding, theoretical research aims to free us for practical reasons from the prejudices that fuel human fears, since they disturb us and make us unhappy. They are, therefore, a main obstacle to a happy life. Rather, if we get to know nature, we will also remove those fears: first and foremost, the fear of death and the fear of the gods.

The Stoics believed we can reach unquestionable, reliable knowledge through apprehensible impressions or representations (φαντασία καταλεπτική). We are talking about such impressions when we establish through research that an impression can only come from a very specific object. Unlike the Epicureans, who rejected the detailed investigation of logic, considering that, simply put, we should not exhaust ourselves by dealing with judgments, but with the things themselves,[12] the Stoics made a significant contribution to logical research. For them, dealing with "things" was an indispensable prerequisite for exploring nature and progressing towards wisdom. Getting to know nature also had a practical purpose for the Stoics. In order to achieve an undisturbed or happy life, we must, through research, find out what is "up to us" and what is "not up to us", that is, what we can and cannot do, by nature. A sage, who, for example, knows stronger sunlight is accompanied by stronger heat, will not be disturbed by the fact that the two cannot be separated. The one who, like a child, does not know this, may be unhappy because we are not able to satisfy his request that the sunlight be maximum and the temperature be moderate.

If we can characterize the Epicureans and Stoics as epistemological optimists, perhaps we can consider the skeptics to be epistemological pessimists. Skeptics could say that the epistemological optimism of the Epicureans and Stoics is unfounded; hence, they would conclude that their considerations, as well as those of most of their predecessors, are dogmatic. Teachings are based on the basic views of the teachers, and followers are obliged to adhere to them and defend them.[13] What gives these teach-

12 According to Sextus Empiricus cyrenaics "are thought by some to embrace the ethical division only, and to dismiss physics and logic as contributing nothing to the happiness of life" (Sextus Empiricus. *Advesus Mathematicos* 7.1. 11). This extreme point of view was rejected by the stoics; they believed that both physics and logic are necessary for a happy life. The epicureans thought physics is more important, and they replaced logic with their canonic.

13 Cicero. *Academica* 2.8. (=Cicero. 1967. *De natura deorum/Academica*. Cambridge, MA: Harvard University Press).

ings confidence is their criterion of truth. Skeptics disputed this.[14] Mutual criticism of dogmatic teachings is a consequence, among other things, of the change, diversity, and unreliability of criterion. Based on the fate of dogmatic teachings, it should not be difficult to conclude that in research, we cannot arrive at final and irrefutable knowledge.

If this is so, we are left with a worry: studying nature does not contribute to the conditions for a happy life. This is why skeptics claimed doubts about the truth of theories do not have to make us unhappy: we just need to free ourselves from pretensions to wisdom and refrain from judgment on all questions of nature.

The Stoics objected to this by arguing we cannot then make decisions about our actions, and such a view will lead to inactivity (ἀπραξία). According to the Stoics, we base our decisions about what we will do on knowledge, so if knowledge is not possible, then action is not possible either. If our actions are not the result of knowing what we can and cannot do, then we will be unhappy because we will try to do what is impossible to do. This will also result in redundant or unnecessary actions. If we suspend judgment, for example, about whether someone who is claimed to have died really died, we could prepare a lunch for her, which she will not come to, or buy her clothes, which she will not wear.

Carneades and Sextus Empiricus

Pyrrhonist and academic skeptics responded differently to the remark that skepticism leads to inaction. The views of academic skeptics changed significantly from Arcesilaus to Antiochus. Here, to distinguish them from the Pyrrhonists and to point out the specific viewpoint of the New Academy, I will briefly present Carneades' teaching.[15] Carneades' skeptical attitude that nothing can be known reliably and without doubt does not

14 For a detailed critique of the criterion of truth in Greek philosophy, see: Sextus Empiricus. *Advesus Mathematicos* 7.29–446. The first book of Empiricus' work *Against the Logicians* (the seventh and eighth part of the extensive work *Advesus Matematicos*) is almost entirely devoted to that question.

15 Carneades was, by all accounts, an emblematic example of academic skepticism, and for that reason, I will briefly present his teaching. However, some believe the basic ideas he defended were already developed by Arcesilaus. What makes it difficult to contribute to the consideration of this topic is the fact that neither left written traces of his learning. For an instructive discussion of this thematic complex, see: Striker, G. 1996. *Assays on Hellenistic Epistemology and Ethics*, Cambridge: Cambridge Universuty Press, p. 92–115, and: Thorsrud, H. 2010. Arcesilaus and Carneades, in: Bett, R. ed. *Cambridge Companion to Ancient Scepticism*. Cambridge: Cambridge University Press. 58–80.

lead, testifies Sextus, to a complete abandonment of the fact that the study of nature has influence on practical life. Although Carneades believed we cannot prove that we have apprehensible impressions, we can divide non-apprehensible (ἀκατάληπτα) impressions into "persuasive (πιθανός) and unpersuasive (ἀπιθάνους)",[16] and based on those that are persuasive, we will make decisions about our actions in practical life.[17]

According to Sextus, Carneades' persuasive impressions further divide into simply persuasive (πιθανός), persuasive and tested (διεζωδευμενας), and persuasive, tested, and undeniable (without distraction, ἀπερίσπαστους). He says Carneades ranked these degrees of persuasiveness and considered they should be the criteria for making decisions about our actions. If an impression about something is persuasive, tested, and undeniable, then we must base our actions on it, and not on the impression that is simply persuasive or on the one that is persuasive and tested. Sextus tries, not quite successfully, to explain this division with examples. His example of determining the identity of a coil lying in a dark room explains the difference between simply persuasive and persuasive and tested impressions quite well, but the example with which he explains the difference between these and persuasive, tested, and undeniable impressions is based on a mythological story and is not as clear as it might be.[18]

Referring to Clitomachus, Cicero claims Carneades held there are no apprehensible impressions, but there are some of which we can approve.[19] However, although we can approve of many persuasive impressions, we must always keep in mind the dictum that there is no such thing as an apprehensible impression, and the possibility of the appearance of some other impressions of equal strength about the same thing, that cannot be distinguished from it, is never excluded. Therefore, in practical action, we should rely on a persuasive impression until some other impression appears that would call its persuasiveness into question.[20] Arguably because of this, Cicero does not even mention *undeniable* impressions. Therefore, it could be said that, unlike Sextus, Cicero thinks Carneades believes that

16 For a more detailed account of the history that led to this point of view, see: Quintero, G. C. 2022. *Academic Scepticism in Hume and Kant: A Ciceronian Critique of Metaphysics*. Cham:Springer. p. 13–49.

17 Sextus Empiricus. *Pyr.* 1.227. (=Sextus Empiricus. 1976. *Outlines of Pyrrhonism /Pyrrhoniae Hypotyposes/*. Cambridge, MA: Harvard University Press).

18 Sextus Empiricus. *Pyr.* 1.228–229). We need a better example to more clearly distinguish between the Stoics' apprehensible impressions and Carneades' persuasive, tested, and undeniable impressions. A somewhat better example, but still not entirely clear, is given by Sextus in: *Advesus Mathematicos* 7.176–177.

19 Cicero. *Academica* 2.99.

20 Cicero. *Academica* 2.99.

a wise person in practical life will always rely on *deniable* persuasive[21] impressions, and undeniable impressions are impossible for him.[22]

Sextus especially insists on the difference between the skepticism of the New Academy, or, as he characterizes it, negative dogmatism, and Pyrrhonian skepticism, presenting himself as a faithful Pyrrhonist. According to him, academics are not real skeptics, but negative dogmatists; they claim nothing can be known with certainty, but real skeptics, such as Sextus himself, would think it cannot even be known. In this context, we could use, as an interpretative model, the modern KK*p* principle, which claims we cannot know that *p* if we do not know that we know that *p*. Then, we could argue that Carneades accepts the principle in a negative form: if we do not know that *p*, then we know that we don't know that *p*. In this interpretive framework, Sextus should defend the principle: we don't know if we don't know that *p*.

Sextus believes that when they are distinguishing between good and bad in terms of practical action, academics, like Carneades, start from impressions; based on these, they consider impressions according to their degree of persuasiveness, and then make decisions about action. But this leads, he thinks, to the abandonment of the original skepticism. A Pyrrhonist, a true skeptic, behaves differently:

> [W]hen we describe a thing as good or bad we do not add it as our opinion that what we assert is persuasive (πιθανόν), but simply conform to life undogmatically that we may not be precluded from activity.[23]

Sextus is a bit more specific on this issue:

> Furthermore, as regards the end [or aim of life] we differ from the New Academy; for whereas the men who profess to conform to its doctrine use persuasiveness as the guide of life, we live in an undogmatic way by following the laws, customs, and natural affections (φυσικοῖς πάθεσιν).[24]

Many questions remain open here, and readers are deprived of a detailed explanation. In fact, it seems as if Sextus is looking for an excuse. The Pyrrhonists, he suggests, by suspending judgment, remain consistent skeptics

21 It seems Cicero translated the word "πιθανόν", used by Carneades, as "probable" and "veri simile" (see, for example, Cicero. *Academica* 2.32). For a detailed discussion of the meaning of these and similar words in Greek and Latin antiquity, see: Glucker, J. 1995. Probabile, Veri simile, and Related Terms. in: J. G. P. Powell (ed. and introd.) *Cicero the Philosopher: Twelve Papers*. Oxford: Clarendon Press. p. 115–144.

22 If this is so, his point of view in this context would correspond to Karl Popper's concept of refutability of scientific theory. This would imply that, according to Carneades, impressions can be more or less persuasive, but not irrefutable.

23 Sextus Empiricus. *Pyr*. 1.226–227.

24 Sextus Empiricus. *Pyr.* 1.231.

about sense impressions, while Carneades, by introducing an imbalance in the weighing of impressions, abandoned the original skepticism, and thus made a concession to dogmatism. However, he implicitly, without suspension of judgment, accepts practical action in accordance with inherited laws and customs. Although in both statements, he claims such an approach to actions in the practical sphere is ἀδοζάστως (undogmatic, without any special opinion or doctrine), it might seem that conformist behavior in accordance with the prevailing laws and customs is dogmatic: such an attitude could be considered to be his doctrine. If we bear in mind that laws and customs differ among different peoples, this would imply that the Pyrrhonists, if they moved to live elsewhere, would accept the laws and customs of those people without hesitation. This could certainly be interpreted as a kind of suspension of judgment.

But what does it mean to "live ... by following ... φυσικοῖς πάθεσιν"? The Greek word πάθος means not only affection, but also, among other things, "that which happens to a person or thing" or "what one has experienced, good or bad, experience"[25], and in some cases, "condition". At the beginning of *Outlines of Pyrrhonism*, Sextus says the skeptic indulges in phenomena in practical life which are unintentional and present independently of his or her will, and without knowing whether these phenomena are a true apprehension of things:

> Following phenomena, we live spontaneously according to life experience, adoxastically (ἀδοξάστως), since we cannot remain completely inactive. And it would seem that this regulation of life is fourfold, and that one part of it lies in the guidance of nature, another in the constraint of the passions, another in the tradition of laws and customs, another in the instruction of the arts.[26]

By "the guidance of nature", he means "that by which we are naturally capable of sensation and thought"; by "constraint of the passions", he has in mind "that whereby hunger drives us to food and thirst to drink"; by "the tradition of laws and customs", he thinks "that whereby we regard piety in the conduct of life as good, but impiety as evil"; finally, by "the instruction of the arts", he means "that whereby we are not inactive in such arts as we adopt". After all this clarification, Sextus concludes by saying: "But we say all these adoxastically (ταῦτα δε πάντα φαμέν ἀδοξάστως)."[27]

---

25 See the entry πάθος: Liddell, George H., Robert Scott, and H.S. Jones. *A Greek and English Lexicon*, Oxford: Clarendon Press, 1940.

26 Sextus Empiricus. *Pyr*. 1.23.

27 Sextus Empiricus. *Pyr*. 1.24.

The exposition in 1.23–24 begins and ends with the assertion that the skeptic obeys everything said ἀδοξάστως. In the rest of his presentation, Sextus tries to strengthen this by asserting that obedience to phenomena in practical life is αζήτητος, that is, without investigation, αβούλητος, that is, unconscious, spontaneous or without intention, and οὔν προσεχοντες, or without paying attention. If we keep in mind the fact that these are all tips for a relaxed, undisturbed, and happy life, their listed negative qualifications seem unconvincing, almost becoming rhetorical disclaimers of what the tips actually claim.

A question now arises: what is the essential difference between Carneades and Sextus? Sextus actually reproaches Carneades for not adhering to strict skepticism by ranking phenomena in terms of their persuasiveness, and then making decisions in the practical sphere relying on these differences. However, if we keep in mind the above, we can conclude that Sextus, regardless of all the disclaimers, must also make a distinction between phenomena. It is true, for example, that "hunger drives us to food and thirst to drink", but we must, if not know, at least have some opinion about what food is, and what liquid can satisfy our thirst. Therefore, we cannot do it and stay ἀδοξάστως, αζήτητος, αβούλητος, or do it οὔν προσεχοντες.

Sextus Empiricus and Pascal's Wager

In his *Thoughts*, Pascal presents a problem about the question of God's existence. He is skeptical about the possibility of knowing whether God exists or does not exist. We are equally unable to prove one or the other. In such situations, skeptics would advise suspension of judgment (εποχή), and this, according to them, would lead us to indifference or tranquility (ἀταραξία). We can remain agnostic as far as theoretical knowledge is concerned. It would follow that there is no reason to be either a theist or an atheist. However, Pascal thinks this cannot be the case among those familiar with Christian teaching. An integral part of this teaching is the belief that there is this world and the afterlife, and the eternal afterlife depends on whether we are believers in this life or not. The decision about whether or not we will be believers is a practical one. It does not concern the knowledge of whether God exists or not; rather, it concerns the practical consequences of one or the other possibility.

Pascal's calculation of the probability of the expected outcome, which we will not present here, indicates it is better to be a Christian believer than a skeptic or atheist. It is the result of a rational calculation. But

Pascal thinks this is not enough and assumes we will practically behave as if the rationality of that calculation does not concern us. To reap the fruits of that account and avoid the risks associated with the afterlife, we must embrace Christianity and really act as Christians: we must truly become believers. Pascal suggests a practical technique by which this may be achieved:

> Endeavour then to convince yourself, not by increase of proofs of God, but by the abatement of your passions ... Learn of those who have been bound like you ... Follow the way by which they began; by acting as if they believe, taking the holy water, having masses said, &c. Even this will naturally make you believe.[28]

Pascal believes this it will have an effect, assuming that we can't pretend to follow Christian customs and rituals for a long time and be pure agnostics or atheists. Sextus, as we have seen, suggests embracing "the tradition of laws and customs" as one of the conditions for a happy life, by which he also means that practicing piety in life is good and practicing impiety is evil.

Sextus, like Pascal, believes we cannot investigate God's existence. In the first book of *Against the Physicist*, he extensively presents the reasons given by his predecessors for religiosity, both their theistic and their atheistic arguments. He then explains his suspension of judgment by referring to different understandings of God and opposing arguments of theists and atheists:

> Well then, such are the opposing arguments alleged by the Dogmatic philosophers in favour of the existence and of the non-existence of gods. As a result of these the Skeptics' suspension of judgement is introduced, especially since they are supplemented by the divergency of the views of ordinary folk about the gods. For different people have different and discordant notions about them, so that neither are all of these notions to be trusted because of their inconsistency, nor some of them because of their equipollence (ὥστε μήτε πάσας εἶναι πιστάς διά την μάχην μήτε τινάς διά την ἰσοσθενειαν).[29]

Does suspension of judgment on God's existence implies impiety? On the contrary. Sextus is, as Thorsrud says, "intentionally provocative"[30] when he claims that the firm belief in God's existence leads to impiety. His argu-

28 Pascal, B. 1910. *Thoughts*. transl. Trotter, F. W. New York: P. F. Collier & Son Corporation, & 233.

29 Sextus Empiricus. *Advesus Mathematicos* 9.191–192. This presentation is one of the most important sources of the reasons given by his predecessors for religiosity; it provides evidence of their theistic and atheistic arguments.

30 Thorsrud, H. 2011. Sextus Empiricus on skeptical piety. Machuca, E. D. (ed.), *New Essays on Ancient Pyrrhonism*. Leiden: Brill. p. 95.

mentation rests on the consideration of the problem of evil. Namely, the one who believes there are gods and who trusts in their divine providence would have to conclude that "either that the gods are malign or that they are weak – and anyone who says this is clearly impious"[31]. It is difficult to explain the existence of evil in the world in any other way, except by attributing the responsibility for its existence to the gods.

Is it a requirement for a skeptic to be pious without believing in God's existence? In fact, Sextus says that the skeptic, in a practical sense, believes both in God's existence and in divine providence. "We Pyrrhonists," writes Sextus, "following the ordinary life ... say without opinions (ἀδοξάστως) that God exists and we are pious towards the gods and say that they are provident".[32] Researching the question of God's existence leads us in the theoretical sphere to the suspension of judgment or to agnosticism. However, the goal of the Pyrrhonist is a happy life. In order to maintain a state of tranquility, we passively, without intention and with spontaneity, rely on phenomena and on the inherited customs and rules to which most people adhere.

The same applies to religious customs and rules. We might think that this is enough for a happy, conflict-free or relaxed life. We would not be in a state of tranquility if we were "self-centered or anti-social",[33] because if we were, we would come into conflict with the customs practiced by the vast majority of people and thus into conflict with those people. But it seems Sextus thinks this conflict cannot be avoided without accepting customs and rules and, therefore, without accepting belief in God's existence and providence. Pascal, as we have seen, thinks that only the holding of, to use the words of Sextus, the "customs and rules" of Christian believers should lead to belief in God's existence. Regardless of this difference, it could be said that both Sextus and Pascal are agnostics in terms of the possibility of certain knowledge about God. But in terms of beliefs, they are, in practical life, some kind of fideist. Therefore, we could say that both are agnostic (theistic) fideists.[34]

---

31 Sextus Empiricus. *Pyr*. 3.12.

32 Sextus Empiricus. *Pyr*. 3.2; *Adversus matematicos* 9.49.

33 Bett, R. 2021. Introduction. In: Bett, R. (selec. transl. and introd.) *How to keep open Mind: An Ancient Guide to Thinking like a Skeptik* (Sextus Empiricus). Princeton and Oxford: Princeton University Press. p. 12.

34 If we look at the question of agnosticism in this way, it would not exclude either theism or atheism; we could talk about theistic, but also atheistic fideists. A man who in the time of Sextus wanted to be a Pyrrhonist would be obliged to be a theistic fideist, since it was the custom to believe in the existence of God. But it is not difficult to imagine the existence of an atheistic fideist, who, for his atheism, would not need any knowledge about the non-existence of God.

Pyrrhonist skeptics in the last instance would think the suspension of judgment will spontaneously lead to a happy life or to tranquility. The vice or the flaw of dogmatists is that they too quickly come to their dogmas or doctrines, that is, to a judgment about how things really are. Because they themselves may come to the conclusion that their theories are not entirely convincing, they may become upset or even despair and therefore be unhappy. The cure is suspension of judgment, and the way to get there is to oppose any impression or theory with another impression or theory. However, in skeptical consideration, it is necessary to ensure that the arguments in support of opposing theories are of equal strength.

There are two problems with this. First, how do we know that the arguments in support of opposing theories are of equal strength? Although some criteria may immediately come to mind, the Pyrrhonists would not allow any talk of criteria. Second, the claim that dogmatists arrive at their conclusions "rashly"[35] is problematic. Such a qualification implies that valid theory could be achieved over time, during a longer investigation. However, if this were the case, we might expect that some extensive investigation, the investigation that would not be rash, would lead to the breaking of the equilibrium between opposing viewpoints. But this would not support the idea of suspension of judgment, which is a condition for a happy life, and, as we know, a happy life is the goal of Pyrrhonist theory. Longer research could perhaps increase or refine the arguments in support of opposing views, but for a consistent Pyrrhonist, everything would have to end with arguments of equal strength, and, accordingly, with suspension of judgment. It could only be a tool in a polemic in which someone disturbs the argumentative equilibrium by offering new arguments in support of one of the opposing theses; this person would have to be countered with arguments in support of the other thesis and thus show that the arguments in support of that thesis are equally strong.

It seems that in this case, the Pyrrhonist would not be able to suspend judgment and remain carefree. According to the above argumentation, he would accept the obligation to be constantly on guard in order to re-establish the equilibrium, but he would still be safer (ασφαλέστερος) than dogmatists who reach their conclusions rashly, because he would be less exposed to objections.[36]

35 See, for example, "προπέτεια" (rashess) in *Pyr*. 3.2 and "προπετευόμενος" in *Advesus Mathematicos* 9.49, as characteristics of the dogmatic way of investigation.

36 Bett, for example, writes that the word "safe (ἀσφαλής)" for Sextus "is referring simply to intellectual safety; if you make no definite commitments, you are just less likely to be mistaken (with or without the element of worry) than if you do – in fact, you are guaranteed not to be mistaken" (Bett, R. 2015. God: M 9.13–194. in K. Algra and K. Ierodiakonou (ed.) *Sextus Empiricus and Ancient Physics*. Cambridge: Cambridge

Pascal's reasoning, by all accounts, implicitly assumes such an obligation. His initial assertions about the equal impossibility of defending theism and atheism assumes the knowledge of all previous arguments in favor of one or the other. However, if someone appeared with a new argument and claimed that it weakened his equilibrium, and if he wanted to defend the conditions for his bet, he would have to refute it. The one who would make the above remarks to Sextus and Pascal, and who would thereby worry them with the obligation to defend the equilibrium, could say that they came to it rashly.[37]

Contrary to the initial claim, according to which the acceptance of Christian customs and rules, if God does not exist, would mean sacrificing a carefree life in this world, Pascal finally concludes this would not be the case, saying:

> You will be faithful, honest, humble, grateful, generous, a sincere friend, truthful. Certainly you will not have those poisonous pleasures, glory and luxury ... I will tell you will thereby gain in this life ... you will see so great certainly of gain, so much in what you risk.[38]

Everything that, according to Pascal, would devolve upon the person who accepts Christian customs and rules would also make him conflict-free and desirable in the community. According to Sextus, this would make his life carefree. In order to achieve a good life, it would be enough, in both cases, to adhere to religious customs and rules and to believe in God's existence, regardless of whether God exists or not.

Bibliography:

Aristotle. 1960. *Posterior Analytics. Topica*. Cambridge, MA: Harvard University Press.

Aristotle. 1968. *Physics*. Cambridge, MA: Harvard University Press.

Aristotle. 1999. introd. and transl. T. Irwin Nicomachean Ethics. Indianapolis: Hackett Publishing Company. 1095b 1–4.

Bett, R. 2015. God: M 9.13–194. in K. Algra and K. Ierodiakonou (ed.) *Sextus Empiricus and Ancient Physics.* Cambridge: Cambridge University Press.

Bett, R. 2021. Introduction. In: Bett, R. (select. transl. and introd.) *How to keep open Mind: An Ancient Guide to Thinking like a Skeptik* (Sextus Empiricus). Princeton and Oxford: Princeton University Press.

University Press. p. 54). A skeptic would think that he cannot be wrong because he claims nothing.

37 Based on this, we might say that Sextus is a dogmatist.

38 Pascal, B. 1910. & 233.

Cicero. 1967. *De natura deorum/Academica*. Cambridge, MA: Harvard University Press.

Diels, H./Kranz, W. 1985. *Die Fragmente der Vorsokratiker*, Zurich/Hildesheim: Weidmann.

Giussepe, N. (ed.). 1954. *Hecataei Milesii Fragmenta*. Firenze: La nouva Italia.

Glucker, J. 1995. Probabile, Veri simile, and Related Terms. in: ed. and introd. J. G. P. Powell. *Cicero the Philosopher: Twelve Papers*. Oxford: Clarendon Press.

Hereen, H. L. A. (ed.). 1742. *Ioannis Stobaeus Eclogarum Phisicarum et Ethicarum*. Gottingae: Vandenhoek et Ruprecht.

Lesher, J. H. 1991. "Xenophanes on Inquiry and Discovery: an Alternative to the 'Hymn to Progress' Reading of F. 18". *Ancient Philosophy* vol. 11. 229–248.

Liddell, George H., Robert Scott, and H.S. Jones. 1940. *A Greek and English Lexicon*, Oxford: Clarendon Press, 1940.

Pascal, B. 1910. *Thoughts*. transl. Trotter, F. W. New York: P. F. Collier & Son Corporation.

Plato. 1960. *Timaeus, Critias, Cleitophon, Menexenus, Epistles*. Cambridge, MA: Harvard University Press, 1960.

Plutarch. 1892. *Moralia* 4. Gregorius N. Bernardakis (ed.). Leipzig: Teubner.

Proclus. 2008. ed. and transl. T. D. Runia and M. Share. *Commentary on the Plato's Timaeus*. Cambridge: Cambridge Univesity Press.

Quintero, G. C. 2022. *Academic Scepticism in Hume and Kant: A Ciceronian Critique of Metaphysics*. Cham:Springer.

Roberts, R. W (ed.). 1902. *Demetrius On style /De elocutione/*. Cambridge: Cambridge University Press.

Sexstus Empiricus. 1976. *Adversus Mathematicos*. ed. G. R. Bury. Cambridge, Mass.: Harvard University Press.

Sextus Empiricus. 1976. *Outlines of Pyrrhonism /Pyrrhoniae Hypotyposes/*. Cambridge, MA: Harvard University Press.

Striker, G. 1996. *Assays on Hellenistic Epistemology and Ethics*, Cambridge: Cambridge Universuty Press.

Thorsrud, H. 2010. Arcesilaus and Carneades. in: Bett, R. ed. *Cambridge Companion to Ancient Scepticism*. Cambridge: Cambridge University Press.

Thorsrud, H. 2011. Sextus Empiricus on skeptical piety. ed. Machuca, E. D. *New Essays on Ancient Pyrrhonism*. Leiden: Brill.

Tulin, A. 1993. "Xenophanes Fr. 18 D.-K. and the Origins of the Idea of Progress". *Hermes* vol. 121. 129–138.

Dan Đaković*

JACQUES MARITAIN ON FREEDOM AND FREE WILL

Abstract: Freedom is one of the biggest and perhaps the main topic of (political) philosophy. On the other hand, one of the main tasks of philosophy is to give definitions and to make distinctions between things and concepts. In this sense, the author in this paper discusses the problem of freedom and the important difference between free will (freedom of choice) and freedom as such, but also the difference between natural and transnatural aspirations of human person. The author does that by presenting a mosaic of passages of Jacques Maritain (1882–1973) who was a French Christian philosopher and one of the leading Thomists, the main author of the Universal Declaration of Human Rights from 1948 and one of the most influential figures of the 20th century.

Keywords: Freedom, free will, freedom of choice, emancipation, autonomy, natural and supernatural aspirations, person, Jacques Maritain

It's always refreshing to hear a philosophical questions about freedom. What it means to be truly free? Is it possible to be free in prison? How we conquer the freedom? Are we born free or do we become free? Who is really a free man? Can a people (mankind) be free or is only an individual free? Political and economic freedom? Religious freedom? Spiritual freedom? Freedom *from* and freedom *for*?

So the topic of freedom is one of the biggest and perhaps even the main topic of (political) philosophy. On the other hand, one of the main tasks of philosophy is to give definitions and to make good distinctions between things and concepts. In this sense, I understand and presume here (as a very important) distinction between freedom and free will, but also between natural and transnatural aspirations of human person. I

* Faculty of Philosophy and Religious Studies, University of Zagreb, Sveučilište u Zagrebu, Fakultet filozofije i religijskih znanosti, dan_djakovic@ffrz.hr

want to say something about this topic by presenting a mosaic of passages of Jacques Maritain (1882–1973). He was a French Christian philosopher and the main author of the Universal Declaration of Human Rights from 1948; one of the leading Thomists (personalists) and one of the most influential figures of the 20th century.[1] I see his thoughts as an important help also in the current debate against *transhumanism* understood as self-deification.

When discussing freedom, Maritain notes that he is not primarily concerned with free will or freedom of choice. The existence and value of that kind of freedom, however, he implies in everything he says. The freedom he deals with is the freedom of independence and exaltation, which can also be called the freedom of autonomy, or even the freedom of expansion, of the human person.[2] It presupposes the existence of freedom of choice, but it is substantially different from it (see Maritain 1944: 13).[3]

A philosophical theory that falsifies the second operation of the mind by which it knows itself explicitly, can suppress and paralyze the first, primary and natural operation of spontaneous consciousness. As long as we are not victims of this accident, as Maritain claims, each of us knows very well that we have the freedom of choice. This means that all our actions are what they are only because we have included our personality in them and thus arranged them to be this rather than that. But each of us knows very little in what freedom of choice consists.

According to Maritain, this uncertainty of spontaneous consciousness, which is incapable of anything more than implicit knowledge about this problem, enables philosophers, but also those who philosophize without even realizing it, to often ask this question (see Maritain 1944: 14).

Intellectualism is a term used to mark a philosophical position that gives priority to the mind over volitional, intuitive or emotional factors. For example ethical intellectualism is the name for Socrates' position on the connection between knowledge and virtue, according to which a wise man knows what is right, so he will act wisely, i.e. good, which means that no one will actually do evil knowingly. Those who promote absolute intellectualism cannot understand the existence of free will, notes Maritain, because for them the intellect (mind) not only precedes the will, but precedes it as a separate divinity, which would influence the will without being at the same time under any influence of the will and without receiving from it any qualifying motion. Therefore, the domain of formal

1 About Maritain and this topic I also wrote in my dissertation. See: Đaković 2021.

2 In Paulinian, not Kantian, sense. Maritain writes about the human person in many places, but here I especially recommend this place: Maritain 1989: 42–49.

3 See also: Maritain 2011: 118–144.

or specifying determinations (the so-called *ordo specificationis*) can never itself depend intrinsically upon the domain of efficiency or existential effectuating (the so-called *ordo exercitii*), and the will is reduced to a function by which the mind will realize ideas that, based on of the mere object they represent, would appear the best to the subject. This was the position of the great metaphysicians of the classical era (see Maritain 1944:14).

As Maritain thinks, pure empiricists also cannot understand the existence of free will, because, recognizing only sequences knowable by the senses, the idea of causality, exercised by a spirit upon itself, has no meaning for them. When they express an opinion on a question which, like this one about free will, is essentially of the ontological order, they cannot fail to interpret the empirical results of the observational sciences in the framework of classical mechanism inherited from Spinoza. They cannot fail, without knowing what they are doing, surrendering to the most naive extrapolations.[4] In proportion to the scientific discoveries of dynamic elements that work in our psychical activity, they see in the very existence of these elements the proof that they operate deterministically, i.e. that our actions are predetermined, and this is precisely what remains to be proved (see Maritain 1940: 632).

It is probably Freud who offers this empiricist pseudo-metaphysics the greatest opportunities for illusion. At the same time, Maritain notes that a clear distinction must be made between the psychoanalytic method, which opened up new and very important paths for the research of the unconscious, and the philosophy that Freud, leaving the field of his competence and giving complete trust to his dreams, sought in extreme empiricism (see Maritain 1944: 15). The fact, discovered in psychoanalysis, that there are unconscious motivations that the subject follows without being aware of them, is in no way a sufficient argument against free will, because free will begins with rational reasoning and awareness.[5]

To the extent that these unconscious motivations prompt us to act automatically, it is not a matter of free will at all; and to the extent that they give or encourage conscious judgment, it is a question of knowing whether they in themselves shape or do not shape that judgment, or they are made decisively motivated by that judgment – and that means through free choice (see Maritain 1940: 632–633). A parallel can be drawn with an automatic rifle, which, although it has automatic mechanisms, ultimately

4 Extrapolation here means a type of induction, i.e. reasoning by which one concludes from a known order of reality to a higher one, i.e. making conclusions starting from a very limited number of experimental facts and extending the law established in a narrower area to a wider area.

5 See also: Maritain 2011: 147–153.

fires only at the will of the soldier. In other words, the question is whether unconscious motivations necessarily determine our actions or only direct them, and it is clear that the mere fact that they exist is not enough to make a final judgment on the matter.

Free will does not exclude, but rather presupposes the enormous and complex dynamism of instincts, tendencies, psychophysical dispositions, acquired habits and inherited weaknesses, and on the top of all that is the point where this complex dynamism merges with the spirit world where freedom of choice is exercised – to give or not to give a decisive role to the tendencies and drives of nature. From this it follows that freedom of choice, as well as responsibility, implies in us a multitude of degrees or layers of which the Author of being is the sole judge. It does not follow that freedom of choice does not exist – quite the opposite! Maritain concludes quite logically – if it implies layers and degrees, then it must exist (see Maritain 1944: 16).

At the same time, he points out that the efforts of some scientists to connect our natural belief in free will with the non-deterministic theories of modern physics can be very significant and thought-provoking and effective in eliminating many prejudices, but he thinks that the strict evidence which gives that belief an undeniably intelligible basis cannot be found in that direction. The direction to go, for Maritain, is, of course, the metaphysical one. He leads us towards formulations like those of Bergson: „Our motives are what we make of them"; or „Our motives determined us only at the moment when they became decisive, that is, at the moment when the act was almost done." But such formulations acquire both their full meaning and their demonstrative value, not by a philosophy of pure becoming, but only by a philosophy of being and intelligence, such as the philosophy of St. Thomas Aquinas (see Maritain 1940: 633–634).[6]

As he says further, spirit as such implies a kind of infinity; his ability to desire comes out of himself and goes towards the good that completely satisfies him, that means towards the good without limits, and we cannot desire anything but the desire for happiness. But as soon as reflection takes place, our mind, faced with goods that are not Good, and judging them as such, brings into actuality the radical indeterminacy that our desire for happiness possesses with regard to everything that is not happiness itself. The effective motivation of a intelligent being can only be practical judgment. And that judgment is effective only because of the will; it is the will which, acting according to its unpredictable initiative towards the good

6 In a strictly philosophical sense, for Maritain, in his formative years, it was decisive his departure from Bergson and the introduction and acceptance of the philosophy of Thomism.

presented to it by such a judgment, gives that judgment the power to effectively direct the will (see Maritain 1940: 634; also Maritain 2011: 151).

The free act, in which the mind and the will involve and embrace each other, is therefore like a momentary flash in which the active and dominant indeterminacy of the will is performed with respect to the very judgment that determines it. The will can do nothing without the judgment of the mind. It is the will that determines itself by the judgment of the mind, and it does it by this judgment rather than by any other. Far from being just a simple function of reason, by which he would understand ideas which by their mere object prove to be the best, the will is the original spiritual energy of unlimited capacity which has control over reason and its judgments with respect to practical choice, and it does what here and now proves to be best. What, for Maritain, is the real mystery of free will is that, although it essentially needs reasoning and specification of reason, the exercise of will takes precedence over this reasoning and specification and holds it under its active and dominating indeterminacy, because only will can give it existential efficacy (see Maritain 1944: 18).

For the entire personalistic tradition, the human person is a great metaphysical mystery. An essential mark of civilization, if it wants to be worthy of that name, is the feeling and respect for the dignity of the human person. There is even an imperative to be ready to make the ultimate sacrifice to defend the rights of the human person and to defend freedom. What are these values that deserve such a sacrifice, encompassed in the person of man? What exactly do we mean when we talk about a human person? – asks Maritain. When we say that a man is a person, we don't just mean that he is an individual, in the sense that an atom, a blade of grass, a fly or an elephant are individuals. Man is an individual who somehow holds himself in his hand.[7] He does not exist only in a physical way. He possesses spiritual super-existence through knowledge and love. He is, in a way, a universe within himself, a microcosm, in which the great cosmos in its entirety can be encompassed through knowledge (see Maritain 1944: 19).

Through love, he can give himself completely to beings who are, in a way, his second self. It is a relationship for which there is no equivalent in the physical world. The human person possesses, according to Maritain, these characteristics because, ultimately, man's body, which is animated and set in motion by divine fire, exist precisely because of the existence of his soul, which dominates over time and death. Spirit is the root and foundation of a person. The idea of a person thus includes the idea of wholeness and independence. This is a particularly important aspect in the context of the inalienability of human dignity, which is actually the

7 Of course, he does this by reason and free will.

central theme of personalism. No matter how humiliated it is, a person as such always remains a whole and exists independently of any human will. To say that man is a person means to say that in the depths of his being he is more whole than part and more independent than subjected (see Maritain 1940: 635).

When we say that a man is a person, we mean that he is a small fragment of matter that is at the same time the entire universe. Maritain expresses it poetically – man is a beggar who communicates with absolute being, mortal flesh whose value is eternal, a piece of straw into which heaven enters. This is the metaphysical mystery that religious thought points to when it says that a person is the image of God. The value of a person, his dignity and his rights belong to the order of naturally sacred things that bear the imprint of the Creator of being and that have the end and purpose of their existence in Him. *Spontaneity* is a very important word here. The freedom of spontaneity is not, like free will, the power of choice that transcends all necessity and all determinism. It does not imply the absence of necessity, but only the absence of restrictions and coercion. It is the power to act in accordance with one's inner inclination and without coercion imposed by any external factor (see Maritain 1944: 21 and Maritain 2011: 126).

According to Maritain, this type of freedom admits all kinds of degrees, from the spontaneity of an electron that revolves around the nucleus „freely", i.e. it is not diverted from its path by the intervention of a foreign particle, to the spontaneity of the grass in the field that grows „freely" or a bird that flies „freely", i.e. respecting only the internal necessities of its own nature. When the freedom of spontaneity crosses the threshold of the kingdom of the spirit, when it is spontaneity of a spiritual nature, it becomes, properly speaking, the freedom of independence.[8] To that extent, as Maritain says, it does not consist only in following natural inclinations, but rather in actively making someone the sufficient principle of his own action. In other words, it consists in possessing, perfecting and expressing oneself as an indivisible whole in the act performed by the subject. This is why the freedom of independence can only exist in beings that have free will, and it presupposes the (correct) use of free will in order to reach one's own end or purpose (see Maritain 1940: 636). Correct use of free will means to live according to truth and morality. Thus the most moral man is also the most free among us.

It should not be understood that an independent being must also be an *uncaused* being, i.e. a being without a cause. It is more precise to say

8 We can also recall here the well-known distinction between freedom *from* and freedom *for*.

that it is a being who is the master or manager of himself. If the true mark of a personality consists in the fact that it is free and that it is a whole, even if imperfect, it is clear that the person and the freedom of independence are in a relationship and are inseparable. On the scale of being they grow together. At the top of that scale is God – a person in pure act and freedom of independence in pure act (*actus purus*). God is so much a person that His existence is His knowledge and love. He is so independent that, although He is the cause of all things, He Himself is absolutely uncaused. (see Maritain 1940: 637). The term absolute actually means that which is independent, uncaused, unlimited, unbound, freed from any necessary relationship, and this all refers to God, i.e. the Being. The essence of His being is His very existence. His essence is to be. He himself is the Being.

In fact, as Maritain reminds us, in each man personality and freedom of independence grow together. Because man is a being in motion, in creation. If he does not improve and grow, he has nothing and even loses what he had; he must conquer his own being (existence). The whole history of his misery and his greatness is the history of his efforts to win and conquer the freedom of independence along with his personality. He is called to the conquest of freedom!

Maritain points out two fundamental truths here. The first is that a human being, although it is a person and therefore independent because it is spirit, by nature it is at the lowest level of personality and independence, because it is spirit that is one in substance with matter and is inexorably bound to bodily conditions. Another truth is that, regardless of how miserable, how poor, how captive or humiliated a person may be, the aspirations of a person remain inviolable; and as such they strive, in the life of each person as well as in the life of the human species, towards the conquest of freedom (see Maritain 1940: 637).

In the context of a person's dignity, we usually speak about *inviolability*, but maybe it is more accurate to talk about the indestructibility and inalienability of dignity, which can still be violated to the extent that it is not respected or is trampled on. So, no matter how much a person's dignity is denied or trampled on, it, although violated, in its essence can never be completely destroyed or alienated.

When talking about a person's aspirations, Maritain highlights two types. On the one hand, aspirations come from the human person as human, i.e. constituted in such a (human) species. We can say that they are then *innate* (connatural) to man and specifically human. On the other hand, aspirations come from the human person so far as he is a person, that is, so far as he participates in that transcendental perfection which is a personality and which is realized in God infinitely better than in us.

Maritain says that in that case they are supernatural (transnatural) and metaphysical (see Maritain 1940: 637).

The connatural aspirations strive for relative freedom compatible with the conditions here on earth, and the burden of material nature brings them serious defeat and failure from the very beginning, because no animal is born poorer and less free than man. If the transcendental dimension of man is ignored, then he is actually just a prematurely born or failed animal.

The struggle to win freedom for the purpose of social life, as Maritain says, aims precisely to correct the defeat and failures mentioned above. The metaphysical aspirations of the person in us strive for superhuman freedom, freedom pure and simple. And to whom does such a freedom belong by nature, if not to God himself, who is the freedom of independence itself, subsisting by itself. Man has no right to the freedom proper to God. When he strives for this freedom with a supernatural desire, he strives for it in an ineffective way, without even knowing what it consists of. Divine transcendence thus imposes from the very beginning the admission of a profound defeat of such metaphysical aspirations of the person within us. However, that defeat is not irreparable, at least if the victor descends to the aid of the vanquished. The movement to win freedom with the purpose of spiritual life aims precisely at repairing this defeat (see Maritain 1940: 638).[9]

Maritain warns that this point, reached here by considering freedom, is crucial for the human being. Even the smallest mistake costs a lot here. At this crucial point, capital errors, mortal for human society and the human soul, are mixed with capital truths to which the life of the soul and that of society are bound. We need to be extremely careful here in order to distinguish the truths from the errors. Namely, there is a false conquest of freedom which is illusory and homicidal. Fortunately, there is also the true conquest of freedom, which is truth and life for man. In order to briefly describe both types of the conquest of freedom, Maritain states that the misunderstanding of the conquest of freedom is rooted in a philosophy that, in technical language, he calls *univocalist* and *immanentist*. In such a philosophy, the concept of independence and freedom admits neither internal variety nor degrees; and, on the other hand, God is conceived as a physical agent raised to the infinity. So, either He is considered a *transcendent* being and His existence is denied, because a transcendent being would be a sort of heavenly Tyrant imposing restrictions and violence on everything other than Himself; or His existence is affirmed and His transcendence is denied – in other words, everything that exists is considered,

9 Cf. Maritain 1944: 19–20.

in the manner of Spinoza or Hegel, as modes or phases of His realization (see Maritain 1940: 638).

According to this view freedom or autonomy exists only if no rule or objective measure is received from any being other than oneself.. Here the human person claims divine freedom for himself, whether man takes, in atheistic forms of thought and culture, the place of the God he denies, or whether he tries to realize in act, in pantheistic forms, his identity of nature with the God he conceives. To these views, Maritain opposes the understanding of the conquest of freedom, which is based on the philosophy of the analogy of being[10] and the philosophy of divine transcendence. According to this philosophy, independence and freedom are realized, on the different levels of being, in ways which are essentially different. In God it is realized in an absolute way, and therefore He, transcending all things, is the supreme (or deepest) interiority, of which every existing being is a participation (see Maritain 1940: 639). In us it is realized in a relative way, and thanks to the privileges of the spirit which, regardless of the state of dependence to which it is subjected by the nature of things, makes itself independent by its own activity when it interiorizes within itself, by knowledge and love, the law which it respects and which it obeys. For this philosophy, divine transcendence does not impose violence and limitations upon creatures, but rather infuses them all with goodness and spontaneity and is more intimately present and closer to them than they are to themselves.[11]

According to Maritain, the autonomy of a rational being does not consist in its receiving no rules or objective measures from any other being than itself. Rather, it consists in its voluntarily conforming to such rules and measures, because it recognizes them as just and true, and because of a love for truth and justice. This is true human freedom to which the person tends as to a connatural perfection. And if a person also aspires to superhuman freedom, this thirst for transnatural perfection, whose satisfaction is not due to us, will be completely quenched only if the person receives more than he desires, and thanks to a transforming union with Uncreated Nature. God is free, from all eternity; more precisely, He is (subsisting) Freedom in and by Himself. Man is not born free, Maritain claims, except in the sense of the fundamental powers and basic potencies of his being. Man becomes free, fighting with himself (and against himself) and enduring many difficulties. Through the effort of the spirit and virtue, by exercising his free will (correctly!), he gradually wins his freedom, so that, in the end, a freedom he is given is better than the one

10 *Analogia entis*.

11 In the words of St. Augustine: *interior intimo meo et superior summo meo* (*Confessions* III, 6, 11).

he expected. From beginning to end, it is the truth which liberates him (see Maritain 1940: 640–641).[12]

In this spirit and in this sense we can consider political and economic emancipation as well as the deification of man. On the trail of this philosophy we can properly distinguish true from false emancipation, and even more importantly – true from false deification of man.

Fortunately there is a true deification of man. Maritain reminds us of a psalm: *Ego dixi: dii estis* (Ps 81,6). This is called eternal life, but it begins as in a mirror already here on earth. It is fatal to renounce perfect liberation as well as to try to reach it by wrong ways, that is, by human forces alone. Here the supernatural aspirations are fulfilled supernaturally and by a gift that surpasses anything we can imagine. What is grace if not a formal participation in the Divine Nature, in other words, a deifying life received from God? The mystery and paradox is that the supreme freedom and independence of man are won by the supreme spiritual realization of his dependence – of his dependence on the Being which, being Life itself, vivifies, and, being Freedom itself, liberates all that participate in His essence. This kind of dependence is not some eternal constraint, as in the case of a physical being with regard to another physical being. The more man realizes this dependence, the more he participates in the nature of the Absolute...

Men who have achieved some of that, participate in the freedom of the One who cannot be contained by anything. In losing themselves they have won a mysterious and disappropriated personality, which makes them act in virtue of that which they are eternally in the Uncreated Essence. Born of the spirit, they are free like it. Properly speaking, they didn't win anything, they have received everything! While they were fighting to win freedom, it gave itself to them. The true conquest of supreme and absolute freedom is to be made free, freely consenting to it, by Subsisting Freedom. The true deification of man consists in his opening himself to the gift which the Absolute makes of Himself, and to the descent of divine plenitude into the intelligent creature (see Maritain 1944: 39–40).

What Maritain wants to say is that it is all a work of love. The law[13] protects freedom and educates us to be free. When love follows the path of the law, it leads us through the law to emancipation from all slavery, even from servitude of the law.[14] The following fragment from the *Summa*

12 See also Maritain 1944: 21–23.

13 This refers only to the law that is just, i.e. in accordance with the natural law, and in the religious sense it, of course, refers to God's law. Cf. Maritain 1992: 90–98.

14 We can say that it's a triple jump – from slavery to evil and sin, through slavery to the law, to true freedom.

Contra Gentiles, where St. Thomas Aquinas comments St. Paul, is very famous, and Maritain considers it one of the greatest texts absolutely fundamental for the spiritual constitution of true humanity:

> „We must observe that the sons of God are led by the divine Spirit, not as slaves, but as free. For, since to be free is to be cause of one's own actions, we are said to do freely what we do of ourselves. This is what we do willingly: and what we do unwillingly, we do, not freely, but under compulsion. This compulsion can be absolute, when the cause is wholly extraneous, and the patient contributes nothing to the action, for example, when a man is compelled to move by force; or it may be partly voluntary, as when a man is willing to do or suffer that which is less opposed to his will, in order to avoid that which is more opposed thereto. The sanctifying Spirit inclines us to act, in such a way as to make us act willingly, inasmuch as He causes us to love God. Therefore, the sons of God are led by the Holy Ghost to act freely and for love, not slavishly and out of fear. That is why the Apostle says: *You have not received the spirit of bondage again to fear; but you have received the Spirit of adoption of sons* (Rom 8,15). The will is by its essence directed towards what is truly good: so that when, either through passion or bad habit or disposition, a man turns from what is truly good, he acts as a slave, in so far as he is led by something extraneous, if we consider the natural direction of the will. But, if we consider the act of the will, as inclined here and now towards an apparent good, he acts freely when he follows passion or an evil habit, but he acts as a slave if, while his will remains the same, he refrains from what he desires through fear of the law which forbids the fulfillment of his desire. Therefore, when the divine Spirit by love inclines the will to the true good to which it is naturally directed, He removes both – the servitude (heteronomy) whereby a man, the slave of passion and sin, acts against the order of the will, and the slavery whereby a man acts against the inclination of his will, and in obedience to the law, as the slave and not the friend of the law. Wherefore the Apostle says: *Where the Spirit of the Lord is, there is liberty* (2 Cor 3,17); and:
>
> *If you are led by the Spirit, you are not under the Law* (Gal, 5,18)" – (*Summa Contra Gentiles*, IV, 22).[15]

According to Maritain, there is a big difference between the imperfect liberation whereby the highest techniques of natural spirituality[16] oblige nature to satisfy in some way the transnatural aspirations of the human person, and the perfect freedom whereby the supernatural gift (!), that the Divine Personality makes of itself to created personality, far more than fulfills these aspirations and gives much more than what was asked for and expected. While leaving intact the distinction of natures, love, which

15 Quoted according to Maritain 1940: 648.

16 He probably thinks mostly on Eastern spiritualities and techniques, but not only on them.

at the end of spiritual growth creates this perfect freedom, truly makes man a god by participation. At the same time, far from enclosing itself in an altogether intellectual contemplation that would abolish action, the freedom in question lives on a contemplation which, since it proceeds from love, superabounds in action and penetrates into the most hidden structures of the world. The heroism it implies does not retreat into the sacred; it spills over into the profane and sanctifies it. Detached from perfection in perfection itself, because it thinks of loving more than of being itself without fault (!), it gradually awakens goodwill and brotherly love (see Maritain 1940: 648–649).[17]

Emphasizing the distinction between the political or socio-temporal and the spiritual, between the things that are Caesar's and the things that are God's, Maritain concludes that if the false deification of man results in the confusion of the temporal and the spiritual and the perverse adoration of the social and of temporal relativities elevated to the absolute, then the true deification of man, on the contrary, and precisely because it is perfected by the grace of the Incarnation and draws to itself all that is human, asks of divine things to descend into the most profound depths of the human. It asks that the social and political order, remaining essentially different from the spiritual, be pervaded and intrinsically superelevated by the sap which flows into souls from the Absolute. In the degree, small as it may be in fact, that things are this way, in that degree the historical march of civilization towards the conquest of *relative* freedom, which answers to the connatural aspirations of human personality, is in accord and mutual cooperation with the suprahistorical movement of the soul towards the conquest of *absolute* freedom, which corresponds, though transcending them divinely, to the transnatural aspirations of the person as person (see Maritain 1944: 42).

Finally, we can conclude that in order to understand Maritain's concept of freedom, we should first distinguish between freedom and free will, then between created and uncreated will, and finally, between the natural and supernatural aspirations of the person. These wills and aspirations can and should be in a synergistic and synchronous relationship. In other words, the only true freedom is won by finetuning created and uncreated free will, and natural and supernatural aspirations of human person. In the end, Maritain, remaining a philosopher – but taking into account the theological data of the Christian faith, thought and believed that uncreated free will was revealed to men through (by) nature, but also from above as a pure divine gift.

17 This may even be a key passage for understanding Maritain's concept of freedom. Cf. Maritain 2011: 171–211.

Bibliography:

Anshen, Ruth Nanda (ed.) (1940), *Freedom: Its Meaning*, New York: Harcourt, Brace and Co.

Augustinus, Aurelius, *Confessions*.

Đaković, Dan (2021), *Politika i religija u filozofiji Jacquesa Maritaina*, Doktorski rad. Zagreb, Zagreb: Sveučilište u Zagrebu, Fakultet filozofije i religijskih znanosti

Maritain, Jacques (1940), "The Conquest of Freedom", in Ruth Nanda Anshen, (ed.) *Freedom: Its Meaning*, New York: Harcourt, Brace and Co., pp. 631–649.

Maritain, Jacques (1944), *Principes d'une politique humaniste*. New York: Éditions de la maison Française.

Maritain, Jacques (1989), *Cjeloviti humanizam*. Zagreb: Kršćanska sadašnjost.

Maritain, Jacques (1992), *Čovjek i država*. Zagreb: Globus.

Maritain, Jacques (2011), *Scholasticism and Politics*. Indianapolis: Liberty Fund.

Višnja Knežević*

FILOZOFIJA U DOBA PANDEMIJE
JEDAN PRIMER IZ ANTIČKE ISTORIJE

Apstrakt: Strah od smrti jedna je od temeljnih ljudskih emocija, koja posebno dolazi do izražaja u kriznim situacijama kao što su pandemije i ratovi. Iskustvo sa zarazom virusom SARS-CoV-2 to nam je još jednom potvrdilo. Može li filozofija da bude od pomoći? I na koji način? Vraćajući se u antiku, pokušavamo da odgovorimo na to pitanje usredsređujući se na pandemiju kuge, koja je zahvatila Atinu za vreme Peloponeskog rata. Posebno izdvajamo Tukididov opis raspoloženja i ponašanja Atinjana u to vreme i nastojimo da ga protumačimo u kontekstu straha od smrti. Istovremeno, tome suprotstavljamo primer Sokratovog ponašanja, koje se ni u situaciji neposredne egistencijalne ugroženosti od zaraze i ratne pogibije – kao ni kasnije, u iščekivanju smrtne kazne – ni po čemu nije razlikovalo od njegovog uobičajenog držanja. Taj fenomen objašnjavamo Sokratovim sistemom verovanja u vezi sa smrću, za koji tvrdimo da je velikim delom posledica uloge filozofije u njegovom životu.

Ključne reči: strah od smrti, pandemija (kuga), Peloponeski rat, Sokrat, racionalnost verovanjā, filozofija.

Pandemija SARS-CoV-2 još uvek traje, ali se čini da smo tek sada u stanju da smirenije promislimo o okolnostima i zapitamo se šta nam je dalje činiti. Filozofija, u tom smislu, može da bude od odlučujuće pomoći. Da bismo to pokazali, osvrnućemo se na jedan drugi, istorijski čuveni slučaj pandemije i pokušati, na njegovom primeru, da pokažemo individualnu, socijalnu i egzistencijalnu dobrobit koju filozofsko promišljanje ima u sudaru sa apsurdom. Naime, iako imaju svoj naučno objašnjiv kauzalitet, u momentu kada se pojavljuju i kada još uvek za njih nema leka, pandemije smrtonosnih bolesti nas postavljaju u egzistencijalni apsurd utoliko što nas suočavaju sa mogućnošću smrti – bliske, a možda i bolne. Upravo u takvoj situaciji našli su se građani Atine u leto 430. godine stare ere.[1]

* Institut za filozofiju, Filozofski fakultet Univerziteta u Beogradu, *visnja.d.knezevic@gmail.com*.

1 Sve godine pomenute u radu odnose se na staru eru.

Nešto više od godinu dana po izbijanju Peloponeskog rata zadesila ih je tzv. kuga. To je bila bolest sa vrlo teškom kliničkom slikom[2], o čemu svedoči Tukidid (2.49–50 *ff.*), koja je usmrtila do 100.000 ljudi, to jest četvrtinu atinskog stanovništva (uključujući i pridošlice iz provincije – Littman 2009: 456). Pri tome, „nijedan organizam, bio on snažan ili slab, nije sam po sebi [bio] dovoljan da se odupre, nego je kuga sve odnosila, pa i one kojima je ukazivana sva moguća nega" (Thuc., 2.51).

Povrh opisa simptoma i toka bolesti koju je i sam preležao, Tukidid opisuje i njene posledice po mentalno stanje građana i po celokupno društvo. Kuga je obolele dovodila do emocionalnog distresa praćenog depresivnim afektom, a u Atini je zavladalo i „veliko bezakonje" (2.51, 53).[3] Da bi se razumeli razlozi u osnovi te socijalne situacije, navešćemo, a potom i analizovati, pasaž iz *Peloponeskog rata* koji to ilustruje:

> Pritisnuti nesrećom, ljudi nisu znali šta će se sa njima dogoditi, pa su postajali ravnodušni prema zakonoma, i profanim i svetim. [...] A i inače, sa ovom boleštinom počelo je u Atini i veliko bezakonje [...]. Ljudi su videli kako je sreća promenljiva [...]. I tako su se odlučili da potraže trenutna uživanja i da zadovolje svoje strasti, *smatrajući prolaznim i svoja tela i svoja imanja*. A unapred se mučiti za nešto što se smatralo ličnim ugledom, za to niko nije bio raspoložen, jer je

2 Pokušaji da se utvrdi koja bolest je tačno u pitanju traju već dugo. U poslednje vreme, primenom matematičkog modelovanja i uključivanjem epidemiološkog uz standardni, klinički model, moguća oboljenja su svedena na dve glavne grupe. U prvu spadaju tifusna oboljenja i kuga, a u drugu velike boginje. Primena ovog, kompleksnijeg modela eliminisala je male boginje i oboljenja izazvana bakterijama *Staphylococci*. Za više o tome, vid. Littman 2009.

3 Tukidid navodi (2.51) da je bolesnike hvatalo „očajanje", „čim bi osetili da su oboleli", da im je „duše [...] obuzimala beznadežnost, pa su se mnogo više predavali bolesti nego što su joj se odupirali". Od depresivnih poremećaja koje DSM-5TM prepoznaje (APA 2013: 155–188), a koji bi se mogli dovesti u vezu sa Tukididovim opisom, tzv. poremećaj prilagođavanja sa depresivnim afektom (*Adjustment disorder with depressed mood*) čini se najprikladnijim. Taj poremećaj je diferencijalna dijagnoza za tzv. veliki depresivni poremećaj i veliku depresivnu epizodu (*Major depressive disorder*, *major depressive episode*), (APA 2013: 168). Veliki depresivni poremećaj (MDD) dijagnostifikuje se pri pojavi diskretnih epizoda u trajanju od minimum dve nedelje, koje uključuju jasne promene afekata, kognicija i neurovegetativnih funkcija, kao i remisije između samih epizoda (APA 2013: 155). Na prvi pogled, čini se da Tukididov opis zadovoljava dijagnostički kriterijum za MDD – naravno, pod pretpostavkom da oboleli nisu umirali u roku kraćem od dve nedelje. Međutim, dijagnostikovanje MDD zahteva da depresivne epizode ne mogu biti uzrokovane drugim medicinskim stanjem, a to se u slučaju koji Tukidid opisuje ne može isključiti. Naprotiv, u poremećaju prilagođavanja sa depresivnim afektom depresivno raspoloženje se javlja kao direktan „odgovor" na činjenicu da je osoba obolelela od kuge, to jest kuga ovde funkcioniše kao psihosocijalni stresor, koji ima za posledicu depresivni afekt (opet u trajanju od minimum dve nedelje). Zahvaljujem se Mariji Kušić na ovoj primedbi.

> svako mislio da je neizvesno da li će, pre nego što se sve to postigne, umreti. A ono što je u svemu predstavljalo trenutno zadovoljstvo i što je vodilo uživanju, činilo im se lepim i korisnim. Nije ih mogao obuzdati ni strah od bogova, ni ljudski zakon; i *pošto su videli da svi ljudi podjednako propadaju, smatrali su da i pobožbost i bezbožnost izlaze na isto*; i niko se nije nadao da će toliko dugo živeti da bi za svoje pogreške mogao iskusiti kaznu, već je smatrao da mu *nad glavom visi mnogo veća i već presuđena kazna* i da je pravo da još malo uživa u životu pre nego što ga ta kazna stigne (*Peloponeski rat*, 2.52–53, prev. D. Obradović, podvukla V. K.).

U ovom pasažu se otkriva kompleksan sistem verovanja Atinjana u vreme pandemije, na osnovu kojeg Tukidid objašnjava njihovo ponašanje. Taj sistem sadrži implicitna verovanja o ljudskoj prirodi i utilitaristički orijentisanoj etici, a eksplicitno – odnos prema smrti. Ovde ćemo se usredsrediti na poslednje. Tvrdimo, naime, da je odnos prema smrti centralna i „neuralgička" oblast tog sistema verovanja, koja se pak konstituiše oko tri sledeća fundamentalna stava: 1) smrt je nešto loše – dolazi kao „već presuđena kazna", i to mnogo veća od ma koje zakonske kazne za neko krivično delo; 2) materijalistička metafizika – u doba pandemije, Atinjani su počeli da smatraju „prolaznim i svoja tela i svoja imanja"; 3) pobožnost i bezbožnost su izjedačene po svojim posledicama, odakle sledi logički pogrešan, etički zaključak da i jedna i druga „izađu na isto", te da je, dakle, 4) dozvoljeno, lepo, dobro i korisno činiti zločine.

Prvi stav nije obrazložen, ali se može pretpostaviti da je zasnovan na shvatanju smrti kao *kraja života* te na hipotezi da je kraj života nešto po sebi loše (*cf.* Deretić 2020: 38). Kada je reč o drugom stavu, izgleda da se verovanje u materijalističku metafiziku, ako već ne pre, pojavilo kao posledica uvida u visoku stopu smrtnosti od kuge. Inače, ta metafizika je implicitna u pominjanju prolaznosti tela i imanja, ali u nepominjanju duše. Treći stav, da smrt čeka i pobožne i nepobožne, ukazuje, međutim, na to da su Atinjani smatrali da je reč ne samo o smrti tela već o totalnoj smrti, odnosno da nema duše ili bar ne takve koja bi preživela smrt tela. Pobožnost i bezbožnost su, naime, svojstva duše, a ne tela. Oba pojma su u vezi sa pojmom pravde, i to u najvišem mogućem smislu – onoga što je *hósion*, to jest što je u skladu sa božanskim zakonom, čemu je podređen pojam onoga što je pravedno prema ljudskom zakonu (što je, dakle, *díkaion*). Drugim rečima, stavovi Atinjana o smrti mogli bi se izložiti na sledeći način: 1) tanatološki – smrt je kraj života, a kraj života je apsolutan, stradaju i telo i duša; kraj života je, uz to, kazna, odnosno nešto rđavo; 2) etički stavovi, posledica tanatoloških – pravednost i nepravednost su svojstva duše; pošto ista kazna (smrt, nešto rđavo) sustiže i pravedne i nepravedne a duša ne preživljava, isto je biti pravedan i nepravedan.

Najpre treba razumeti da u osnovi straha koji se javlja u pandemiji kao što je kuga, ili COVID-19 danas, leži strah od smrti: ono čega se plašimo nije zaraza virusom *per se* nego visok stepen verovatnosti smrtnog ishoda. Pošto uverenja o prirodi smrti u bitnoj meri određuju i prirodu našeg *odnošenja* prema činjenici smrti, ta ista uverenja će određivati i to kako ćemo se odnositi prema zarazi smrtonosnom bolešću, ali i prema životu samom, odnosno njegovim vrednostima. Depresivna i manična ponašanja Atinjana u doba kuge, koje Tukidid opisuje, ali i na bilo koji drugi način ugroženo mentalno zdravlje obolelih ili onih koji veruju da će oboleti, u tom slučaju, nastaju kao rezultat straha od smrti neposredno, a posredno – naših verovanja o prirodi smrti. Kako, pak, strah od smrti utiče na odnošenje prema životu samom i njegovim vrednostima, dobro je ilustrovano logički pogrešnim, etičkim zaključkom koji su Atinjani izveli u doba pandemije – da su, s obzirom na činjenicu smrti kao neminovnog ishoda kuge, pravednost i nepravednost izjednačene.

Logički pogrešno rezonovanje Atinjana razložićemo u redovima koji slede. U prvom koraku, Atinjani su uočili da i pravedni i nepravedni građani umiru od kuge, što se formalno može predstaviti na sledeći način:

$$\exists x\,(Px \wedge Sx), \quad (1)$$
$$\exists x\,(\neg Px \wedge Sx),$$

gde je x promenljiva koja označava ma kojeg građanina, P predikat pravednosti, $\neg P$ predikat nepravednosti, a S predikat smrtnosti, koji u konkretnom slučaju označava činjenicu umiranja od pandemije. Te dve premise izražavaju istinite opažajne sudove Atinjana da *neki* pravedni i *neki* nepravedni ljudi umiru od kuge (ma o kako velikom broju da se radilo, to i dalje nisu *svi* pravedni i *svi* nepravedni građani Atine). Takođe, važno je istaći da se na osnovu pomenutih sudova ništa ne može zaključiti o odnosu predikata P i $\neg P$ ni o tome kakvu ulogu oni imaju u drugim kontekstima.

Međutim, Tukididovi Atinjani upravo ovo čine – oni, u drugom koraku, vrše pogrešnu generalizaciju da će *svi* pravedni i *svi* nepravedni ljudi umreti od pandemije, te da pravednost i nepravednost, *kao takve*, imaju istu posledicu ‑ smrt. Drugim rečima, Atinjani veruju da

$$\forall x(Px \rightarrow Sx) \wedge \forall x(\neg Px \rightarrow Sx). \quad (2)$$

Konačno, u trećem koraku, na osnovu prethodnog, dedukuju, opet logički pogrešno, identitet antecedenasa, to jest zaključuju da su, s obzirom na posledicu smrti, pravednost i nepravednost jednake:

$$(\forall x(Px \rightarrow Sx) \wedge \forall x(\neg Px \rightarrow Sx)) \rightarrow \forall x(Px \leftrightarrow \neg Px), \quad (3)$$

što je ne samo pogrešan zaključak, jer se na osnovu istog konsekvensa ne može zaključiti identitet antecedenasa, već je formalno i kontradikcija.

Štaviše, na osnovu Tukididovog teksta – „[...] i pošto su videli da svi ljudi podjednako propadaju, smatrali su da i pobožnost i bezbožnost izlaze na isto" – moguće je da su Atinjani implicitno izveli jedan, još jači zaključak, a to je da su, s obzirom na činjenicu smrti, pravednost i nepravednost iste *u svim svojim posledicama*, da imaju isto značenje u svim kontekstima. Odnosno, da važi

$$\forall x((Px \rightarrow Qx) \leftrightarrow (\neg Px \rightarrow Qx)), \qquad (3a)$$

gde je *Q* predikat koji označava proizvoljno svojstvo, na primer, „biti dobar vernik" (tj. poštovati bogove, kultove, religijske svečanosti, hramove, herme itd.), „biti dobar sin", „biti dobar roditelj", „biti dobar prijatelj", „biti dobar građanin" itd. Zaključak ne sledi jer, kao što je već rečeno, na osnovu (1) ne može se zaključiti ništa o tome kakvu ulogu *P* i ¬*P* imaju u drugim kontekstima, čak ni s obzirom na činjenicu smrti. (3a) je jači sud od (3), a razlika između ta dva suda ima svoje poreklo u mogućnosti dvojakog tumačenja reči „isto" u rečenici: „[...] i pošto su videli da svi ljudi podjednako propadaju, smatrali su da i pobožnost i bezbožnost izlaze na *isto*." U prvom slučaju, (3), *isto* se razume kao neizbežna smrt od kuge, a u drugom, (3a), *isto* podrazumeva sve posledice *Qx* koje slede iz *Px*, odnosno iz ¬*Px*, s obzirom na činjenicu neizbežne smrti. Oba rezultata su logički neodrživa.

Zašto su prosečno dobro obrazovani Atinjani rezonovali tako pogrešno? Na ovom mestu ne možemo ulaziti u *background* teorijske razloge njihovog izbora, koji se tiču Tukididovih uverenja o ljudskoj prirodi[4], niti su ti razlozi presudno značajni za naš argument. Za naše svrhe dovoljno je prihvatiti hipotezu da su mišljenje i ponašanje Atinjana u doba rata i pandemije u značajnoj meri bili determinisani strahom od smrti i dvoma uverenjima na kojima je taj strah počivao – prvo, temeljnim uverenjem da je smrt, *kao takva*, rđava i, drugo, induktivnim zaključkom, dovoljno potkrepljenim empirijskom evidencijom – da je s obzirom na prirodu i tok pandemije verovatno da će ih smrt ubrzo stići. Potonje je racionalno zasnovano uverenje, ali ono nije primarni motivator, mada jeste katalizator straha građana Atine u doba kuge. Primarni uzročnik je uverenje da smrt po sebi predstavlja zlo. Polazeći od tog uverenja, strah od smrti se posledično javlja kao najjači ekstralogički razlog, koji i najobrazovanije ljude motiviše da počine logičke greške koje im se, inače, verovatno ne bi tako lako „potkrale". Pa pošto su izveli pogrešan zaključak da je, s obzirom na ishod smrti, isto (u bilo kojem od dva ili u oba navedena smisla) biti pobožan i biti bezbožan, odnosno činiti pravdu i nepravdu, građani Atine su se predali nepravdi, koja im se činila dobrom i korisnom.

Dakle, i samo okretanje zločinu u tom slučaju treba sagledati u svetlu straha od smrti, a ne bolesti; to pak, na koncu, znači da ga treba sagledati

4 Više o tome, vid. Salins 2014; Jordović 2009.

u svetlu pogrešnog uverenja da je kraj života nešto rđavo. Atinjani su se, naime, okrenuli bezakonju u doba kuge zato što su zaključili da će, ako se razbole, verovatno umreti *i* zato što su verovali da je smrt neko zlo. Materijalistička metafizika implicitno, a strah od smrti, koju su pak izjednačili sa najvećim zlom, eksplicitno, naveli su Atinjane i na logičke pogreške i na to da bez stida čine zločine.[5]

Ostavljajući materijalističku metafiziku po strani, usredsređujemo se na verovanje da je smrt nešto po sebi loše. Nisu svi Atinjani tako mislili, odnosno znamo da bar jedan nije. Sokrat, naime, nije verovao da je smrt *per se* loša i smatrao je strah od smrti racionalno neopravdanim. Prema njegovom shvatanju, plašiti se smrti značilo je „ništa drugo do držati se mudrim, a ne biti mudar. To znači misliti da čovek zna ono što ne zna" (Pl. *Ap*., 29a6–8). Niko, naime, ne zna da li je smrt nešto dobro ili zlo pošto se niko nije vratio iz mrtvih da bi nam to kazao (*cf*. 29b5–6). Sokrat odbacuje strah od smrti kao na neznanju i na neistinitom verovanju zasnovan strah. Neistinito, naime, *verujemo* da je smrt loša jer to niti znamo niti možemo znati. Sokratov pristup strahu od smrti je kognitivistički i intelektualistički, on je smatrao ne samo da u osnovu emocije straha počiva kognicija, (pogrešno) verovanje o prirodi smrti nego, štaviše, da je to što inače nazivamo emocijom straha od smrti isto što i verovanje da je smrt nešto loše (Pl. *Prt*., 358d6–7; Deretić 2020: 25 *et pass*.; *cf*. Deretić & Smith 2021).

U tom smislu, moglo bi se reći da je Sokrat anticipirao prve hipoteze onoga što danas poznajemo kao kognitivno bihevioralni pristup u psihoterapiji.[6] Značajna odlika tog pristupa je usredsređivanje na sistem (klijentovih) verovanja, uz ispitivanje njihove opravdanosti i racionalne zasnovanosti. Verovanja koja ne izdrže tu vrstu provere uklanjaju se iz kognitivnog sistema ili, ukoliko to nije moguće (a često nije), nastoji se da se raskine kauzalna veza dotičnih verovanja i akcija (Dryden & David, 2008). Racionalno neosnovana verovanja, drugim rečima, prestaju da imaju ulogu mo-

5 Na ovom mestu uputno je setiti se Platonovog mita o Eru iz X knjige *Države*. Taj eshatološki mit trebalo je da pokaže da je pravednost, osim po sebi, dobra i po svojim posledicama. Mrtvim dušama, koje preživljavaju smrt tela, sudi se za pravednost, odnosno nepravednost. Duše koje su za života (u telu) bile nepravedne sustiže pravedna kazna, koja kod Platona ima korektivnu, a ne retributivnu funkciju. Slično je i u eshatološkom mitu o suđenju dušama iz dijaloga *Gorgija*. Jasno je da takva verovanja nisu postojala u dušama Atinjana u trenutku pandemije, sve i da su u njih (ili u neka slična) verovali pre pandemije.

6 Istina, osnivač REBT (*Rational emotive behaviour therapy*) A. Elis i osnivač CBT (*Cognitive behavioural therapy*) kao kišobrana za sve kognitivno bihejvioralne terapije A. Bek pozivali su se na stoicizam (posebno Epikteta) kao na filozofski osnov i preteču respektivnih pristupa terapiji (Robertson, 2016; Still & Dryden, 1999). Međutim, s obzirom na to da je stoička etika zapravo u biti sokratovska, nećemo pogrešiti ukoliko Sokrata proglasimo prvim koji je anticipirao osnovne hipoteze pomenutih savremenih psihoterapijskih pristupa; tim pre što je prvi eksplicitno zastupao kognitivističko razumevanje emocija.

tivatora ponašanja, ili se tome bar teži. Iako je ovo objašnjenje znatno pojednostavljeno, za naše svrhe je dovoljno: verovanja uzrokuju ponašanja, pa pogrešna ili iracionalna verovanja izazivaju neadekvatna, destruktivna ponašanja (kako po sebe, tako i po društvo) – ili, najjednostavnije rečeno, ljudsku patnju i rđava dela. Vraćajući se, dakle, Atinjanima u doba kuge, rekli bismo da su njihovo bezakonje i očajanje neodgovarajuća ponašanja, koja proističu iz jednog neistinitog i racionalno neutemeljenog verovanja – verovanja da je smrt neko zlo, kazna itd.

Budući da je Sokrat *istraživao* to pitanje, pa je nakon istraživanja zaključio da nije opravdano verovati da je smrt neko zlo, neadekvatna ponašanja su kod njega izostala. Na drugoj strani, tvrdio je da pouzdano *zna* (*oîda*) da je „rđavo i sramotno činiti nepravdu" (*Ap*., 29b6) i kršiti zakone (Brickhouse & Smith 1994: 144). Sokrat se, naime, smrti nije plašio, ali se plašio da slučajno ne počini nepravdu. Za to da je reč o znanju, a ne o istinitom verovanju, navodio je dva argumenta: 1) neposredno svedočanstvo, intenzivno apotreptičko iskustvo u saglasnosti sa racionalnim razlozima – tzv. daimonion (Deretić 2020: 43, 48); 2) stav da je vrlina ekvivalentna znanju, a porok neznanju. Ne možemo ulaziti u detalje sokratovske etike i tanatologije, koje su kompleksnije no što ih ovde predstavljamo, te načelno udaljene od materijalističke metafizike, a u pojedinim, značajnim elementima divergiraju i od kognitivno bihevioralne paradigme. Ipak, možemo reći da je Sokrat svoje znanje da je rđavo i sramno činti nepravdu zasnivao na argumentaciji koja je počivala ili na racionalnim razlozima ili je bila u saglasnosti sa njima, a isto važi i za njegovo verovanje da smrt nije po sebi rđava. U osnovi i jednog i drugog bili su istraživanje i promišljanje.[7]

Sokratova „filozofija života" bi glasila da je dobar život – život istraživanja. Naime, filozof je tvrdio da, za čoveka, neistražen život nije vredan življenja (*ho dè anexétastos bíōs ou biotṓs anthrṓpōi* – Pl. *Ap.*, 38a4), što se prevashodno odnosilo na odbacivanje nekritičkog odnosa prema stvarima. Otuda, kada je tvrdio da je strah od smrti lažno verovanje da je smrt neko zlo ili da pouzdano zna da je rđavo i sramno činiti nepravdu, Sokrat je to činio tek pošto je ta pitanja kritički preispitao. Živeći život posvećen istraživanju, *ne* zanemarujući pritom svoje dužnosti atinskog građanina[8], Sokrat se držao smireno i hrabro, što se posebno moglo primetiti u doba

7 Čak je i njegovo apotreptičko iskustvo uključivalo istraživanje. Naime, iako je bezuslovno verovao svom daimonionu, Sokrat je taj božanski glas naknadno reflektovao, to jest pitao se o razlozima zašto mu se daimonion javlja tada kada mu se javlja i zašto mu poručuje baš to što mu poručuje (*cf*. Pl. *Euth*., 272e3). Na taj način je utvrdio da je daimonion uvek u saglasnosti sa razumom, čak i ako ne potiče od razuma: daimonion nikada Sokrata nije odvratio od toga da učini nešto što bi inače bilo razumno i dobro.

8 Prema vlastitom svedočenju na sudu (*Ap*., 28e1–2), Sokrat je učestvovao u dve, za Atinu izuzetno politički značajne kampanje: na Potideju i na Delij i Amfipolis. Više o značaju Sokratovog političko ratnog angažmana, uz filozofski: *cf*. Anderson (2005).

rata i pandemije. Alkibijad svedoči (Pl. *Symp*., 219e3–220b5) da je za vreme kampanje na Potideju[9] Sokrat „prevazilazio u vojničkim naporima ne samo mene (*sc*. Alkibijada) nego i ostale", bio otporan na zimu i nedostatak hrane i nikada ga niko nije napio. Alkibijad, takođe, tvrdi da ga je Sokrat spasio od izvesne smrti na bojnom polju (220e1 *ff.*)[10], kao što je kasnije spasio ranjenog mladića Ksenofonta, prilikom povlačenja vojske iz Delija 424 . godine (221a1 *ff.*).

Za nas je ovde važnije, međutim, to što kada se, tokom kampanje na Potideju, mnogo atinskih vojnika zarazilo kugom, Sokrata nisu spopali kukavičluk, malodušnost i očajanje nego je nastavio da čini ono što je oduvek činio – da živi život istraživanja. Utonuo je, naime, u misli, po svoj prilici kritički ispitujući nešto:

> Zamislio se u nešto u ranu zoru i stajao je onde na istom mestu razmišljajući, i kad mu nije polazilo za rukom da to reši, nije s mesta odlazio, nego je i dalje stajao ispitujući. I već je bilo podne, i ljudi su to primećivali i u čudu jedan drugome kazivali da Sokrat od rane zore onde stoji i o nečemu razmišlja. Na posletku, kad je bilo već po večeri, neki Jonjani izneli su svoje prostirače napolje, jer je tada bilo leto: i jedni su spavali u hladovini, a drugi su pazili na njega hoće li i preko noći stajati. A on je stajao dok je zora svanula i sunce granulo, zatim se pomolio suncu i otišao (*Gozba*, 220c3–220d3, prev. M. N. Đurić).

Na osnovu ovog kratkog pasaža možemo videti da čak ni u situaciji neposredne egzistencijalne ugroženosti od združenih rata i kuge, Sokrat nije prestao da se bavi istraživanjem niti je na bilo koji način promenio svoje uobičajeno (ekscentrično) ponašanje. Tvrdimo da je takva smirenost i ovde posledica njegovog odnosa prema smrti, koji je usvojio posle refleksije, pošto je ustanovio da sud da je smrt po sebi loša nema racionalnog opravdanja. Za razliku od Tukididovih Atinjana, kod Sokrata se tokom

9 Kampanja na Potideju, odnosno opsada Potideje, trajala je tri godine (432–429). U kontekstu sukobā između Korkire i Korinta, bila je jedan od povoda za Peloponeski rat (Thuc., 1.24 *ff*., posebno 1.56 *ff.*). Bila je posebno surova jer su, ne uspevajući da je pokore, Atinjani stanovnike Potideje u drugoj godini opsade zatvorili u grad, gde su ih držali sve dok ovi nisu, dve godine kasnije, ostali potpuno bez hrane i čak pribegavali kanibalizmu. Tako je Potideja i kapitulirala (2.70). Sokratovo učešće u opsadi je nesporno, ali nije sasvim jasno kada se uključio i u kojim je bitkama tačno učestvovao. Naučnici danas smatraju ili da je Sokrat učestvovao u opsadi Potideje od samoga početka (Anderson 2005: 279, beleška 11) ili da je pak došao u okviru kontingenta koji su Perikleove kolege, stratezi Hagnon i Kleomp, poveli iz Atine na Trakiju i Potideju u leto 430. godine, kada je, zajedno sa njima, u vojsku stigla i kuga (Planeaux 1999; *cf*. Thuc. 2.58).

10 Bitka o kojoj govori Alkibijad ili nije bitka o kojoj Platonov Sokrat svedoči u *Harmidu* (153a1–d1), (Planeaux 1999: 75) ili nije čuvena Potidejska bitka, koja se odigrala u pozno leto 432. godine nego bitka u okviru kampanje na Potideju, a kod Spartola, koja se odigrala u maju 429. godine (Anderson 2005: 278–279, beleška 11).

pandemije nisu javljala depresivna raspoloženja, anksioznost ili bilo kakav mentalni poremećaj, a nema svedočanstva ni da je logički sled njegovih rasuđivanja na bilo koji način bio ugrožen. Naprotiv, sva svedočanstva tvrde suprotno: u situaciji u kojoj bi većina ljudi bila uzdrmana, Sokratu se to ne dešava. Verujemo da je razlog upravo u njegovoj opredeljenosti za život istraživanja. Sokrat, tako, svojim životom posvećenim istraživanju pokazuje dobrobit koju filozofija ima u kriznim situacijama. Ta dobrobit je, kako smo na početku sugerisali, individualna, egzistencijalna i socijalna. Za vreme pandemije, taj filozof nijednog momenta nije prestao da se ponaša kao politički *individuum*, odnosno kao građanin (*polítēs*), nije zanemario svoje dužnosti prema polisu i zajednici, što se ne bi moglo reći za njegove sugrađane. Bio je u stanju da reflektuje, potpuno smireno, činjenicu smrti, uprkos tome što je to ključna determinanta ljudske egzistencije. Već sama po sebi, činjenica smrti je fundamentalni uzrok ljudske anksioznosti. Istraživački život, odnosno bavljenje filozofijom, omogućio je Sokratu da napravi kritički otklon i od toga. Konačno, istraživanje i reflektovanje smrti omogućilo mu je da ne izgubi emociju koja ima suštinski značaj u vremenima teških kriza, a posebno takvih kao što su rat i pandemija – Sokrat, naime, ni za vreme kuge nije izgubio empatiju, kao što za svoje sugrađane nije prestao da brine čak ni kada je osuđen na smrt. Možemo se samo pitati koliko bi velika Atina dobila da su se i ostali njeni građani u većoj meri posvetili kritičkom istraživanju. Koliko je izgubila time što nisu – to već znamo.

Zahvalnica

Hvala Kanjevac Slobodanu na podsticaju za ovaj rad. Veliko hvala Kostić Jovani, Kušić Mariji i Nurkić Petru na konsultacijama, čitanjima i komentarima, koji su znatno unapredili njegov kvalitet. Sve eventualne greške su moje.

Literatura

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5.* 5th edition. American Psychiatric Publishing.

Anderson, M. 2005. Socrates as hoplite. *Ancient philosophy* 25: 273–289.

Brickhouse, Th. C. & Smith, N. 1994. *Plato's Socrates.* OUP.

Cooper, J. M. (ed.). 1997. *Plato. Complete works.* Hackett Publishing Company.

Deretić, I. & Smith, N. 2021. Socrates on why the belief that death is a bad thing is so ubiquitous and intractable. *The journal of ethics* 25 (1): 107–122.

Deretić, I. 2020. *Smrt i besmrtnost u Platonovoj filozfiji*. Neven.

Dryden, W., & David, D. 2008. Rational emotive behavior therapy: Current status. *Journal of cognitive psychotherapy* 22 (3): 195–209.

Đurić, M. N. (prir.). 2015. *Platon. Gozba.* Dereta.

Đurić, M. N. (prir.). 2020. *Platon. Odbrana Sokratova.* Dereta.

Jordović, I. 2009. Istorijski koreni učenja o pravu jačeg. *Anali Pravnog fakulteta u Beogradu* 57 (2): 212–228.

Littman, R. J. 2009. The plague of Athens: epidemiology and paleopathology. *Mount Sinai journal of medicine* 76 (5): 456–467.

Obradović, D. (prir.). 1999. *Tukidid. Peloponeski rat.* Beograd.

Planeaux, C. 1999. Socrates, Alcibiades, and Plato's ΤΑ ΠΟΤΕΙΔΕΑΤΙΚΑ. Does the *Charmides* have an historical setting? *Mnemosyne* 52 (1): 72–77.

Robertson, D. J. 2016. The Stoic influence on modern psychotherapy. In *The Routledge handbook of the Stoic tradition* (pp. 394–408). Routledge.

Salins, M. 2014. *Zapadnjačka iluzija o ljudskoj prirodi.* Anarhija – blok 45.

Still, A., & Dryden, W. 1999. The place of rationality in Stoicism and REBT. *Journal of rational-emotive and cognitive-behavior therapy* 17 (3): 143–164.

Višnja Knežević

Philosophy in times of pandemic
A case study from antiquity

Summary: Fear of death is one of the fundamental human emotions. Today's experience with the SARS-CoV-2 virus pandemic is probably the most novel confirmation of this plain and simple truth. Can philosophy be of help in such situations? If it can, in what way? I attempt to answer by providing a historical distance, i.e., analysing the critical situation of 430 BC when the plague struck Athens. I scrutinise the Athenians' mood and behaviour, as described by Thucydides, in the context of the fear of death, to which I then contrast Socrates's conduct at the time. Even in the situation of immediate existential threat from infection and death in war, Socrates did not act any different than he usually did, which – I argue – is to be explained by his belief system concerning death, but more importantly, is mainly due to the role philosophy had in his life.

Keywords: fear of death, pandemic (plague), Peloponnesian War, Socrates, the rationality of beliefs, philosophy

Vanja Subotić*

THE APPLIED ETHICS OF COLLEGIALITY: CORPORATE ATONEMENT AND THE ACCOUNTABILITY FOR COMPLIANCE IN THE WORLD WAR II

Abstract: Recently, I have proposed an extension of the framework of the ethics of collegiality (Berber & Subotić, *forthcoming*). By incorporating an *anti-individual perspective* and the notion of *epistemic competence*, this framework can reveal the epistemic *virtue/vice relativism*, which, in turn, charts the tension between being a good colleague and an efficient, loyal employee. In this paper, however, I want to sketch how the ethics of collegiality could be applied to practical domains, such as the *historical accountability* and *atonement* of corporations that participated in the anti-Semitic policies of the Third Reich and contributed to the Holocaust by using slave or forced labor. New studies suggest that corporations ought to engage in deeper historical reflection and ethical dialogue between Shoah survivors and top managers to address the issue of industrial compliance (Federman 2021), whereas most of the work on this topic traditionally focused on the issue of reparations litigation (Kelly 2016, Neuborne 2003). Through the notions of *collective institutional epistemic vice* and *institutional ethos* (Fricker 2021), the upshot is to assess whether it is feasible for corporations to be genuinely repentant regarding their role in the Holocaust thanks to the ethics of collegiality instead of merely offering compensation. I will argue that instead of emphasizing ethical leadership and the top-down approach to the (re-)implementation of values in corporate conduct, the spotlights should be on the *bottom-up approach* grounded in *collegial solidarity* among all employees.

Keywords: Corporate Accountability, Corporate Atonement, Corporate Ethos, Epistemic Vice, Ethics of Collegiality, Holocaust.

* Institute of Philosophy, Faculty of Philosophy, University of Belgrade, 1994, vanja.subotic@f.bg.ac.rs.

1. Introduction

Primo Levi, a twentysomething Italian chemist, who turned partisan, found himself in the notorious *Auschwitz Birkenau* concentration and extermination camp (*Konzentrationslager*) in February 1944. There he got a tattoo on his arm – he was no longer Primo, but *Häftlige* 174517 (Levi 1959: 22). Primo was among the few lucky ones who survived long enough to see the camp liberated by the Soviet Red Army in January 1945 – one of the 20 Italian Jews who were once part of the cohort of 650 living souls that were transported in cattle trucks. His expertise helped him to get around *Monowitz*, a labor camp (*Arbeitslager*) that was part of the deadly system of subcamps that constituted *Auschwitz*.

The peculiar thing about the *Monowitz* is that it was built and envisaged by the executives of *IG Farben* (*Interessengemeinschaft Farbenindustrie AG*), a German chemical and pharmaceutical company that was one of the largest conglomerates in the world back then.[1] *IG Farben* invested 700 million Reichsmark[2] in establishing a factory for the production of synthetic rubber in *Monowitz*, namely *Buna Werke*, for exploiting slave labor (Borkin 1978). Primo was among the 35,000 inmates who worked in *Buna*, thanks to his previous education as a chemist. He described the everyday routine in *Buna* in his memoir *If This is a Man*:

> *The hours of work vary with the season. All hours of light are working hours: so that from a minimum winter working day (8–12 a. m. and 12.30–4 p.m.) one rises to a maximum summer one (6.30–12 a.m. and 1–6 p.m.). Under no excuse are the Häftlinge allowed to be at work during the hours of darkness or when there is a thick fog, but they work regularly even if it rains or snows or (as occurs quite frequently) if the fierce wind of the Carpathians blows; the reason being that the darkness or fog might provide opportunities to escape.* (Levi 1959: 32)

The *SS* (*Schutzstaffel*) charged three Reichsmarks for unskilled workers, four Reichsmarks for skilled ones like Primo, and around one for children

1 *IG Farben* was formed in the 1920s when six chemical companies – *BASF*, *Bayer*, *Hoechst*, *Agfa*, *Chemische Fabrik Griesheim-Elektron*, and *Chemische Fabrik vorm* – decided to merge. Among industrial chemists employed in the company there were three Nobel laureates (Carl Bosch, Friedrich Bergius, and Gerhard Domagk). Interestingly enough, at first, the company had been denounced by the Nazi party as being capitalist and Jewish, but had come a long way to become the main government contractor and Nazi party donor during the World War II. The employees, chief scientists and physicians working for *Bayer*, even participated in medical experimentation on humans in *Auschwitz Birkenau* and *Mauthausen* by deliberately infecting inmates with diseases. Think about this next time you go to the nearest pharmacy and ask for *Bayer*'s aspirin.

2 To get a more vivid picture, that would be *more* than 3 billion (inflated) euros.

workers (Sofsky 1996: 175). The life expectancy was three to four months (Sofsky 1996: 182) – the inmates would die out of weariness, exhaustion, diseases such as scarlet fever, starvation, beatings from the assigned *Kapo*, or they would be gassed if deemed unfit for further exploitation. Around 10,000 people lost their lives in *Monowitz*, and the top management of *IG Farben* approved *any* method that would enforce inmates' productivity, as stated in reports sent from and to Frankfurt am Main, where the headquarters of the company were located.[3]

All major German companies, such as *BMW*, *Volkswagen*, *Siemens*, *AEG-Telefunken*, *Daimler-Benz*, *IG Farben*, *Deutsches Bank*, *Krupp*, and *Bosch*, were parts of the military economy and industry and relied on the slave labor force (Hayes 1995: 68). Specifically, *IG Farben* supplied the Third Reich with synthetic fuel and nitrile rubber thereby facilitating war efforts of Wehrmacht and was included in the production and distribution chain of Zyklon B – used for murdering more than million people, mostly European Jews, in gas chambers – along with another company, namely *Degussa* (*Deutsche Gold– und Silber– Scheideanstalt vormals Roessler*). This company acquired 25 Jewish firms and parcels of the real state (this was called the Aryanization of Jewish business), as well as rights to process gold and silver plundered from Jewish families sent to concentration and death camps (Rosenbloom & Althaus 2010: 185). They also used slave laborers to build new facilities. For some construction sites, slave laborers represented a horrific majority – or 76% of the total workforce (Rosenbloom & Althaus 2010: 186). In the 1920s, *Degussa* and *IG Farben* each acquired 42,5% of *Degesch* (*Deutsche Gesellschaft fur Schadlingsbekampfung*) shares. *Degussa* retained managerial control of the company during the war, while *IG Farben* had placed its directors as members of the *Degesch* executive board. *Degesch* developed the use of a pesticide that releases hydrogen cyanide in specific conditions and owned brand rights when it comes to the name Zyklon, whereas *Degussa* possessed the chemical formula (Hayes 2004: 275).

In September 2002, *Degussa AG* received a request for a bid to supply graffiti-resistant coating for the Memorial to the Murdered Jews of Europe in Germany's capital Berlin. The two main symbols of the Holocaust are Zyklon B and deportations via cattle trucks and trains. In this sense, the Holocaust legacy of *Degussa* and German corporate history play a significant role in ethical decision-making on the institutional and governmental levels. Similarly, in France, the role of the French National Railway

3 As stated in the educational material on the official page of *Auschwitz Birkenau Museum*: https://www.auschwitz.org/en/history/auschwitz-iii/living-conditions-and-number-of-victims/

(*Société Nationale de Chemins de Fer Français – SNCF*) in World War II came under scrutiny when a group of Shoah survivors in France and the United States requested that *SNCF* takes responsibility for organizing and participating in the transportations of 76,000 Jews to the German border, from where they were directly taken to concentration and extermination camps (Federman 2021: 410). This resulted in the Holocaust Rail Justice Act of 2013. The impetus for this was, again, bidding. *SNCF* aimed to sign lucrative commuter, regional, and high-speed rail contracts with the US government.

The questions that emerge are whether the top management and the employees of the corporations knew what was happening to the unfortunate and persecuted European Jewry during World War II and whether there was a way to remain profitable without being compliant with the Nazi regime and its nefarious policies. Moreover, do current CEOs and employees feel the burden of notorious corporate history? Every year, there are fewer and fewer Shoah survivors to tell their stories and to help us navigate the moral, social, and political waters so that Holocaust never happens again. It, thus, seems pertinent to discuss what constitutes ethical leadership and the genuine atonement of companies since commercial settlements and strategic re-branding do not seem to address the real issue – how was it possible for "ordinary men" to turn a blind eye to increase in Zyklon B turnover which went as high as 13.4 short tons[4] in 1943, or to horrific conditions in cattle cars where packed deportees died in significant number from asphyxiation, hypothermia, or thirst?

I will approach the issue from a somewhat unusual collectivistic perspective by analyzing corporate accountability and atonement through the intertwined frameworks of *the ethics of collegiality* and *vice epistemology*. I will start by introducing the said frameworks (**Sect. 2**), especially its point of intersection – the notion of *corporate ethos*. I will then present legal aspects of dealing with the *IG Farben*, *Degussa*, and *SNCF* cases (**Sect. 3**) in order to show that atonement must come from a deep reflection on corporate history and critical dialogue through which a company would realize what values constitute its ethical core (**Sect. 4**). In a nutshell, financial and legal accountability are hollow without moral, epistemic and historical accountability which would allow corporations to rebuild its deteriorated set of values, i.e., ethos, that should have a socially integrative role within work collectives forming corporations. Collegial solidarity should represent a solid ground for such efforts.

4 Keep in mind that 1 ton was sufficient for murdering around 300,000 people. In 1943, *Degesch* earned as much as 544,000 Reichsmark for selling Zyklon B to concentration and extermination camps (Hayes 2004).

2. From Anti-Individualism in the Ethics of Collegiality to Corporate (Counter-)Ethos

Normative ethics has seen a new development in the work of Monika Betzler and Jörg Löschke (2021), who proposed the inauguration of the ethics of collegiality as a new subfield between friendship and family ethics on the one side and business ethics on the other side. As authors rightly notice, given the extent to which collegial relations can impact our lives in terms of well-being and the sense of belonging to a particular work collective, it is quite odd that philosophers have not thought it through much earlier. Anyhow, let me start by introducing you to the framework.

Betzler & Löschke (2021) hold that collegial relationships should be regarded as intrinsically valuable iff two features are present – *collegial recognition* and *collegial solidarity*. This means that you will deem person *X* as a good colleague only if *X* is performing her job well, i.e., she is competent *and* if she is willing to help you. What matters here is the assumption that you and *X* are *peers,* and by virtue of being peers, you can assess each other's contribution to the company. To be labeled as peers, you and *X* should fulfill at least one of the following criteria: **(i)** You share the same domain of activity, **(ii)** You share the same affiliation, i.e., work for the same company or institution, **(iii)** You match when it comes to the work purpose, **(iv)** You have the same level of responsibility. To sum up, according to this framework, if you are bad at your job and you are systematically mistreating your co-workers, nobody will think of you as being a good colleague. Simple as that.

However, in a recent paper, my co-author and I proposed the extension of the initial framework so that it could be plausibly applied to real-world cases (Berber & Subotić, *forthcoming*). Although we think that Betzler & Löschke can rest on their laurels, given that they inaugurated a novel and important subfield in normative ethics, our view is that the natural next step for the ethics of collegiality is to turn towards anti-individualism. In other words, the current framework is focused on what it takes for an individual to be a good colleague, whereas we want to point out that the individual is always embedded in the work collective within a company, and the collectivistic perspective may dictate different norms for being considered a good colleague. Betzler & Löschke insinuated that the tension between being a good colleague and a loyal employee could easily be imagined: sometimes, corporate ethos will require that we owe our loyalties to the employer rather than our team or individual colleague. This is the borderline case that I will be examining through the prism of World War II-related atonement debates pertaining to the legacy of companies such as *IG Farben*, *Degussa*, and *SNCF*.

The first part of the framework extension has to do with the notion of *epistemic competence*. Not only do you (and should) care about the skills and corpus of knowledge of colleague *X*, but *X*'s epistemic character matters as well for performing the job. Her epistemic character is constituted by a mash of epistemic virtues and epistemic vices. Virtue and vice epistemology were predominantly individualistic at the beginning (see Zagzebski 1996 and Cassam 2016), as is the case with the ethics of collegiality now. Nonetheless, in the past decade, the anti-individualistic turn has changed virtue and vice epistemology to their core (see Smart 2018). The key point of anti-individualism in the virtue and vice epistemology is that groups can be considered as independent epistemic agents having an epistemic character in the same manner as individuals have it.

Miranda Fricker has recently proposed – *per analogiam* with the epistemic character of individuals – that institutions have an *ethos* that "(...) includes collective motivational dispositions and evaluative attitudes, whereas good or bad ends orientate the actions based on the ethos" (Fricker 2021: 91). In other words, we can dissect the values, virtues, and vices of institutional bodies thanks to their professed ethos: the absence or presence of particular values and virtues helps us understand and evaluate epistemic outcomes of such bodies.[5] This is similar to our idea that work collectives also have epistemic character *per analogiam* with the epistemic character of colleagues and co-workers constituting it; the difference is merely in the size of the chunk that is being analyzed. For the purpose of this paper, I will take Fricker's institutional ethos as a synonym for *corporate ethos*.

Collegial relations contribute to the corporate ethos and are an integral part of it as long as good colleagues are also loyal employees. As we have argued in Berber & Subotić (2021), the assessment of what makes one a good colleague heavily depends on one's contribution to the team or work collective: sometimes, individual epistemic vice may bring about a positive pattern of epistemic conduct of the collective and, conversely, individual epistemic virtue may hinder the positive epistemic outcome. In this sense, a good colleague need not be epistemically virtuous at all costs, but rather his position should be evaluated within a broader network of co-workers such as her team. This was a bottom-up approach. Here, however, I intend to use the top-down approach, i.e., I want to examine how companies, through the professed set of values, influence teams, and individuals.

5 In Sikimić et al. (2021), my co-authors and I offer an empirical in-depth analysis of the ethos of scientists comprising scientific institutional bodies with respect to the influence of their political attitudes on their epistemic atittutes.

Of course, the idea that companies can and do incorporate specific values is not new. Pruzan (2001), who even designed workshops in business ethics for CEOs in large multinational companies, argues that no company can be described as successful and visionary without the implementation of core values since legal and financial liability is hollow without accounting for the social and ethical aspects of corporate activities. Moreover, as he points out, there is a non-symmetric relationship between the decision-makers and decision-receivers – the management has a social responsibility that extends beyond maximizing profits, i.e., to *create* a set of values that should be *shared* among other employees and stakeholders. How to make sense of collectives and corporations sharing values, though?

You can choose to be a *summativist*, *non-summativist*, or the proponent of the *joint-commitment model* regarding group phenomena, be they belief, intentionality, conduct, or epistemic character. Summativists (e.g., Wray 2007) hold that groups cannot be endowed with epistemic character but rather an aggregation of individual employees' epistemic virtues or vices. If all co-workers in a team are intellectually humble, then the whole team *must* be intellectually humble. Non-summativists (e.g., Lahroodi 2007) claim something completely different – your co-workers may, in fact, lack intellectual humbleness, but that does not mean that the team cannot exhibit intellectual humbleness. Finally, in Margaret Gilbert's (2013) joint commitment model, groups are *plural subjects,* not mere aggregations or emergent entities. This means that groups are bonded by shared values, i.e., corporate ethos. If there is no such joint commitment, then corporate ethos deteriorates. Like Pruzan, Fricker (2021: 94) notices that, at first, there can be a mismatch between the newly committed values at the executive level and their implementation at the level of employees, but if such values are not professed among those who are in top positions we ought to doubt the viability of such ethos.

Arguably, only temporally and counter-factually stable values can be part of the ethos. This point is crucial for determining whether lapses of judgment were a one-time thing or whether the deterioration of values suggests that the ethos has crumbled. For instance, does the overt anti-Semitism and compliance with the Nazi regime's atrocities count as a one-time thing or serve as proof of crumbled corporate ethos during the time of crisis? As both Fricker and Gilbert argue, becoming a party to joint commitments, i.e., corporate ethos, has genuine *normative pressure*, which, sometimes, may not serve good ends. In this sense, one could claim that the employees of *SNCF*, *Degussa*, or *IG Farben* were jointly committed to the anti-Semitic policies endorsed by their managers and company directors. Their adherence to such policies would make them loyal employees, and conversely, were they opposed to it, they would be violating something to which they pledged, i.e., corporate ethos.

Moreover, we could go as far as to say that these companies had their own *counter-ethos* – collective motivational dispositions and evaluative attitudes that are easy to condemn from the contemporary perspective but that were gradually implemented in employees once the top managers and directors realized that it was the most rational way to be profitable during the war. After all, they all have to bring bread to the table, right? Social and ethical aspects of corporate actions could be further redeemed by pointing out how many families were sustained through the war. Turning a blind eye here and there, i.e., when *cheminot* working for *SNCF* notices that his former Jewish colleague's family is in a cattle car, but refrains from doing anything, only means that *cheminot* is loyal to the company in the kairotic moments when loyalty is a rare gem. Counter-ethos could be seen as a set of binding commitments which would allow some of the values to become suspended or vices endorsed and vindicated for the purpose of surviving in times of crisis, similarly as some national constitutions presume that in such times president takes charge, whereas other democratic institutions, e.g., people's assembly, are temporarily dismissed.

Would the endorsement of counter-ethos make corporations bulletproof when it comes to their culpability once the crisis has ended? In other words, would the conduct of corporations be irreproachable in that case? Legal accountability of corporations (and persons) is what remains stable across the periods of peacetime and wartime – this is what allows us to try those who have transgressed during the war. Thus, any fleeting lapse of judgment of corporations is still under the auspices of transitional justice.

3. Lex Paciferat: Redeeming Corporate Ethos through Trials and Compensations?

Recall the pertinent question posed in the **Introduction**: did the CEOs and employees of companies such as *IG Farben*, *Degussa*, and *SNCF* know what was going on with their Jewish neighbors, acquaintances, and co-citizens? If they did know, why did they not do something? As I have sketched in **Sect. 2**, adherence to the counter-ethos and loyalty to the company in times of crisis may be possible answers. Nonetheless, this does not strip one of the legal culpability. Thus, one more question can be added here: were the CEOs punished in any way for compliance with the regime that brought about the mass atrocity unheard of in modern European history?

Upon the liberation of *Auschwitz* and the capitulation of the Third Reich, The Allied forces organized the Nuremberg Trials between 1945 and 1949. Corporate entities did not face trials[6], but directors, board members, and CEOs of *IG Farben* and *Degesch*, besides several other companies that fueled the war efforts of Nazi Germany, were held accountable for committing crimes against humanity by using slave labor and supporting deportations to concentration camps, albeit nobody has served more than 8 years in prison (Federman 2021: 408). In fact, many of those who stood trial were acquitted and resumed their positions at *Bayer*. The director of *Degesch* was sentenced to only 5 years in prison, but neither the Nuremberg trials nor post-war investigations could disprove the testimony of *Degussa*'s leaders that they did not know that Zyklon B was used for the extermination of Jews in extermination camps (Rosenbloom & Althaus 2010: 187). *Degussa* initially dismissed members of the former Nazi party from management, board, and production in 1945. Unfortunately, however, the company rehired many of them in the years to come.

Interestingly enough, as Wiesen (2001) points out, the industrialists who stood trial wanted to deny compliance in war crimes and, simultaneously, to portray themselves as pragmatic and principled businesspeople who acted in accordance with corporate (counter-) ethos, which made them spend a pretty penny on newspaper articles, PR statements, pamphlets, and apologies. Maintaining profitability, caring for workers and their families, and sticking to high-quality manufacturing even amidst the war were the often pointed-out excuses by the industrialists. When faced with accusations of forced and slave labor, the answer was that corporations, in fact, saved the inmates from a much worse fate.[7] The industrials

6 According to historian Jonathan Wiesen (1999), the American prosecutors were careful to blame individuals rather than corporations so that they could vindicate the image of market economy. Moreover, as the Cold War heated the relationship between the Allies and Soviet Union, *ipso facto* between Western and Eastern Germany, the views of business complicity obtained an ideological shade: whereas Marxist and communist voices saw a link between nazism/fascism and market economy due to the role of German businesses in the World War II, the Western capitalist countries mostly lost interest in this issue the moment the reparations were ensued (Wiesen 1999: 4–5). The Cold War period was also cleverly used by the German industrialists who sought to wash their hands: they portrayed their pre-1933 role as saving the country from communism at all costs, which resonated with Western stakeholders (Wiesen 2001: 72).

7 The cynism of this line of argumentation could be refuted by relatively undemanding fact checking. Indeed, there were industrialists who actually *did* save inmates from much worse fate, and these efforts definitely did not include slave labor – take only a wildly popular example of Oskar Schindler. Moreover, historians Bernd Wagner and Piotr Setkiewicz found archival evidence that managers in *IG Farben* discussed labor conditions in *Auschwitz* and the ratio between *SS* and *Kapo* brutality and inmates

were, at the same time, victims of Nazism, virtuous Christians, devoted workers, loyal citizens, and apolitical patriots (Wiesen 2001: 70). Thus, the attorneys who worked for *IG Farben* argued the following about Karl Krauch, one of the directors:

> "*[I]nstead of being an ambitious and ruthless industrial magnate, Dr. Krauch is an honorable Christian, a simple man, a research-worker and scientist, conscious of his responsibilities, who never committed an offence but devoted his whole life to technical and scientific progress*" (cited in Wiesen 2001: 69).

Moreover, the industrialists went far to prove that they did not violate the intrinsic values constituting the corporate (counter-) ethos, such as adherence to Anti-semitism: many resisted the Aryanization and de-Judaization of companies from 1933 to 1938. For instance, *Degussa* didn't have any members of *NSDAP* (*Nationalsozialistische Deutsche Arbeiterpartei*) on the Board but several converts to Judaism were part of it (at least until 1938), and one of the last family members associated with the company, Walter Roessler, did not support the Aryanization (Rosenbloom & Althaus 2010: 185). However, the *Kristallnacht* in 1938 and further Anti-semitic policies such as the *First Ordinance on the Exclusion of Jews from German Economic Life* brought about the situation in which industrialists acted upon simple cost-benefit analysis – in the name of the company's overall interest, no resistance to governmental policies should be indicted (Wiesen 2001: 65–66). This is where the counter-ethos came to the scene – desperate times call for desperate measures such as the Aryanization of Jewish business and the usage of slave labor in order to save the ordinary German people struggling in the war-struck *Vaterland*.

The survivors' demands for reparations and compensation were met with varying success (Neuborne 2003, Kelly 2016). *IG Farben* failed to pay any money to Shoah survivors, and representatives blamed the legal disputes for not being able to put the company into liquidation (Borkin 1978). The company did, however, join the *German Companies Foundation Initiative: Remembrance, Responsibility, and the Future* in 2000, along with other 5000 corporate entities, including *Degussa*. Nonetheless, the fund struggled to obtain 10 billion *Deutschmark* for the ultimate compensation to former forced and slave laborers and came up with the total amount only after much international pressure (Wiesen 2001: 79, n. 2). *Degussa*'s sins were largely put to rest, and, recall, the company even produced anti-graffiti paint for the *Memorial to the Murdered Jews in Europe* in Berlin, although a considerable public controversy ensued when the

greater efficiency, whereas it is a widely known fact that *Siemens* used slave women labor in *Ravensbrück* concentration camp (Wiesen 2001).

company had submitted a bid (Rosenbloom & Althaus 2010: 183). In any case, the financial and legal aspects were taken as proof that the corporate ethos has been re-implemented in repentant German companies – the debt is paid, whereas moral responsibility and epistemic blameworthiness are now part of the history that should remain confined to archives. After all, *it* will *never* happen again, right?

SNCF, on the other hand, had a completely different historical trajectory being in the occupied zone rather than in the occupying country. An enterprise with a hybrid public-private identity was temporarily placed under German control in 1940, but the *SNCF* managed to handle its daily operations independently and billed Germans for all the provided services (transportation of soldiers, livestock, armaments, etc.). Moreover, French railway workers –*cheminots*– carried railroad sabotages during the war, which increased in intensity and number from 1943 onwards. Germans were never pleased with the lack of enthusiasm *SNCF* showed towards their requests (Federman 2017: 19). And yet, the employees and executives of *SNCF* organized the deportation of French Jewry to the border with Germany, from where they were taken to concentration and extermination camps. As Sarah Federman rightly notices, even if the employees were ignorant about the final destination of more than seventy convoys[8], the very conditions, and manner of deportation – witnessed by *cheminots* and bystanders alike – represented the violation of human rights:

> *"I saw a train pass by (...) Then, came the cattle cars packed. The skinny arms of children clinging to the bars. A hand outside flapping like a leaf in the storm. When the train stopped, voices cried 'Momma!'"* (cited in Federman 2017: 20).

Allegedly Germans ordered both deportations and conditions, whereas the task of the *SNCF* was to carry the orders. In post-war France, this episode was banished away from the collective memory due to the company's alleged role in the *Résistance*. Furthermore, in the 2000s, the company became an international player with worldwide revenue that is measured in billions of dollars (Federman 2017: 21). Holding such a company liable for its role in the Holocaust proved to be Sisyphus' job. First, in France, it was impossible to sue the state for policies imposed by the collaboration-

8 And, in any case, if the employees were ignorant about it, the "big shots" were not. In 1942, *SS-Obersturmbannführer* Adolf Eichmann, who planned and overlooked deportations, held a meeting in Berlin with those who were in charge of deportations in The Netherlands, Belgium, and France. The technicians from the *SNCF* as well as officials from the Vichy government were present and developed a deportation plan which was later passed on to *SNCF* general director, workers, local French prefectures, and police (Federman 2021: 415).

ist Vichy regime until the early 2000s. Second, survivors who decided to launch suits generally did not manage to obtain financial compensation, and the cases mostly outlived them (for a comprehensive list of lawsuits against the *SNCF,* see Federman 2017: 24–25). *SNCF* is a public company operating within private law, which essentially means that if there are no individual employees alive to be tried in criminal court (and, needless to say, there aren't any), then the company cannot bear any legal liability. The decisive ending of French Holocaust litigation came about in 2009. The ultimate result was *social* rather than legal or financial. *SNCF* started to take the Holocaust legacy on its shoulders: the company opened archives, had numerous exhibitions on deportations, took part in Holocaust commemorations, etc. (Wieviorka 2007).

However, in the USA, once the *SNCF* started bidding for rail contracts, especially in Maryland, the survivors engaged in lobbying and legal complaints, which brought about bad press. As opposed to France, the US public is sensitive to survivors' horrific experiences during the Shoah, and "a foreign, faceless, multi-national train company becomes all too easy to hate" (Federman 2017: 27). Moreover, rarely something conveys the symbolic of Shoah as trains since it would not be possible to proceed with methodical industrial killings without railroads and meticulous timetables. In 2014, after much pressure, France and the USA signed a settlement agreement to compensate the remaining survivors, and *SNCF* agreed to invest 5 million dollars in Holocaust research, commemorations, and similar educational projects (Federman 2021: 419). For instance, *SNCF* became one of the leading sponsors of the *Fondation pour la Mémoire de la Shoah* located in Paris and donated the land in Bobigny (the place from where most convoys departed) to the French Jewish community.

4. Applying the Ethics of Collegiality: Corporate Atonement through Ethical Leadership or Collegial Solidarity?

I have shown in the previous section how the legal and financial accountability of German corporations and French *SNCF* for their role in the Holocaust may be approached from the perspective of the shift from the crisis-induced counter-ethos to the re-establishment of ethos once the crisis has ended. The important moment here is whether the deterioration of intrinsic values and human rights, such as anti-racism and anti-discriminatory treatment of employees and stakeholders belonging to ethnic minorities, can be simply resumed and re-enacted through com-

pensations. For this reason, I will reiterate the issue of culpability in this **Section** so that I can propose a different and more effective type of atonement by applying the ethics of collegiality. My point will be that instead of focusing on a top-down approach, i.e., the culpability of top management and creating ethical leadership, one should turn to a bottom-up approach, i.e., the epistemic and ethical characters of employees constituting work collectives and collegial relations.

Recall Miranda Fricker's (2021) notion of institutional ethos that I used to account for corporate ethos and counter-ethos. She also argues that there are two distinct domains of potential culpability of institutions, namely the *inner ethos* (whether institutions are endowed with stable motives and values) and the *outer performance* (whether institutions achieved the ends of those motives). Thus, when assessing the culpability of institutions, one should pay attention to the violation and betrayal of intrinsic values or ends that should have been achieved through values. Specifically, when it comes to corporations, accountability refers to the amends a market actor must attempt in the aftermath of human rights violations (Federman 2017: 13). Regardless of the financial and legal accountability, Fricker's two types of culpability may be taken to show that corporations can be endowed with social and historical accountability as well. Take, first, the inner ethos – it is something that should survive the trials of the time and counterfactual situations. In this sense, crumbling inner ethos points out the historical accountability of specific companies: intrinsic values and core ideology of companies cannot be put on hold during the crisis. Once the values have been betrayed, there is no easy way back. The re-establishment of the inner ethos must be based on a deep and honest reflection on what went wrong in corporate history.

On the other hand, when it comes to outer performance, it is clear that the consequences of corporate conduct always have a bearing on the stakeholders. The post-war German industrialists were quite aware of it, so they spent a considerable amount of money to bleach the image and the brand of their companies so that they could regain the trust of both ordinary German folk and people who suffered heavily because of the Nazi regime. They sensed that the betrayal of inner ethos (regardless of the narratives that counter-ethos was indispensable for protecting German businesses and families) meant that the companies underperformed *despite* remaining profitable. Moreover, precisely because of the profitability, their outer performance was put under scrutiny.

Fricker (2021: 99) defines institutional epistemic vice as "a matter of culpable epistemic bad habits, where the culpable lapses might be in ethos or in implementation, or both," which allows for putting all pieces to-

gether. Corporations such as *IG Farben*, *Degussa*, *Degesch*, and *SNCF* were all guilty of culpable lapses in both inner ethos and outer performance, which made them epistemically vicious at the level of corporate epistemic character. Their betrayal of intrinsic values happened deliberately and consciously, which was further witnessed by their post-war conduct. The institutional or corporate epistemic vice in this regard forms the basis of their multifaceted accountability – legal, financial, social, and historical. Besides the usual charges of moral responsibility of corporations, here we can see that the notion of *epistemic blameworthiness* would be more useful. Both the employers and employees knew that they were taking part in something that went against the core values of their companies and offered different justifications to account for such lapses of good judgment. But, nonetheless, the employees were loyal to employers, and employers were committed to profit.

Take *SNCF,* for example. Instead of denying its participation in the process of organizing convoys, *SNCF* executives chose a strategy of victimization under German occupation: they emphasized plundering of assets, threats to employees and their families, Gestapo interrogations, etc.[9] Moreover, employers and employees "acted like the average French person; tired, afraid, and more concerned with their own survival than with the deportation of neighbors" (Federman 2021: 413). Can you blame anyone for being the average citizen amidst the occupation in war-torn France? Ludivine Broch (2014), the historian of the Vichy France, analyzed the relations within *SNCF* and suggested that they represented a complex web of advancement, hierarchy, and loyalty that resembled subservience: the cheminot were professionals, loyal to each other and to the company, who were ready to set aside any issues regarding human rights in order to perform their jobs competently. Only one (!) *cheminot* was honored as Righteous Amongst the Nations for rescuing his Jewish neighbors during the Shoah.

In the contemporary business ethics literature, the notions of ethical leadership and corporate social responsibility gained prominence in the 21st century (for an excellent review based on big data, see Liu et al. 2019). However, these analyses do not take into account the historical conduct of corporations but rather focus on mending the present consequences and forging trust with the idealized stakeholders who are living in the here and now. On the other hand, analyses such as Sarah Federman's (2021) do

9 Ironically, though, *SNCF*'s general director Robert Le Besnerais reported his own employees to Gestapo for the carried out and planned attempts of diversion (Federman 2021: 413). The employees were then deported to concentration camps.

take into account corporate history but still frame the issues around the same notions. According to this framework – let me label it as a *top-down approach* – the top management struggling with the Holocaust legacy of companies should provide their companies with ethical guidelines and communicate them to their employees instead of offering financial settlements, which ultimately amount to settlements of conscience.

However, the top-down approach does not seem to deal straightforwardly with the *ésprit de famille* that characterized work collectives in *SNCF*. The employees were true-blue patriots and genuinely cared for the company and its trains. Moreover, they themselves felt victimized by the German occupation. Communicating intrinsic values to such a collective would amount to endorsing counter-ethos through which top management would find excuses for complicity rather than fighting the institutional vice. In this sense, *SNCF* would not be any different from overtly anti-Semitic corporations such as *IG Farben*, *Degussa*, and *Degesch*. In other words, relying on the top-down approach shows only how values can be fickle instead of temporally and counterfactually stable – the big shots can make a sales pitch out of any kind of guidelines if the employees are only to be passive recipients. What needs to be implemented is a *participative process* (Pruzan 2001). The process would include developing a dialogue between the employees and management in such a way that values must serve a socially integrative function as opposed to discriminatory and racist policies that were justified by values constituting counter-ethos. In a nutshell, the corporations grappling with the Holocaust legacy need to, in fact, re-invent their ethos – specifically, their *ésprit de famille*. This could be done through a *bottom-up approach*.

Let me briefly remind you of the core features of the ethics of collegiality. The relation of collegiality is intrinsically valuable due to collegial recognition and collegial solidarity. Collegial recognition has to do with one's competence, including epistemic competence. Collectives may profit from an individual's unfavorable epistemic character in such a way that this brings positive epistemic output on the collective level. Conversely, an individual's favorable epistemic character may have adverse effects, i.e., negative epistemic output on the collective level. As we could witness in the case of *IG Farben*, *Degussa*, *Degesch*, and *SNCF*, individual's loyalties and dedication to producing high-quality goods or services had negative epistemic output on the collective level – the corporate ethos crumbled in times of crisis and distorted into counter-ethos due to individual's tunnel-view which resulted in the willingness to avoid facing the devastating consequences of counter-ethos endorsement such as deportations to concentration camps where people would be worked to their death.

This is, of course, not to say that loyalty should be a red flag. Rather, what went wrong is the omission of collegial solidarity from the equation. In a similar manner, as collegial recognition was extended in Berber & Subotić (*forthcoming*) to include epistemic competence as means of evaluating one's co-workers, here, the collegial solidarity should be extended to include the ones suffering the consequences of corporate conduct. Not only that one owes solidarity and empathy to colleagues and work collectives – after all, the most successful and visionary corporations harness *ésprit de famille* (cf. Pruzan 2001) – but one should discern institutional vice of inaction and indifference from intrinsic values and virtues that keep one from realizing the responsibility towards the end-users, consumers, or stakeholders. The values constituting corporate ethos must be actively shared within work collectives in such a manner that the corporate history serves as – pardon the *cliché* –a teacher of anti-discriminatory and anti-racist conduct. Without encouraging individual employees to express their solidarity and to be critical of the historical baggage of their companies, any donation to Shoah education or commemorations is hollow since the top management has not cultivated their own garden. Additionally, without the active participation of *all* employees in crafting the novel identity of the company, any litigation and financial compensation to survivors is more of a PR ruse than genuine atonement.

5. Conclusion

The upshot of this paper was to show one possible and important domain for the application of the ethics of collegiality, namely the historical and social accountability of companies *as* collective agents and means of their atonement. One of the darkest episodes of corporate history is the role of companies in the Holocaust. I have tackled the conduct of a negligent number of them – four (three German and one French) companies that were compliant with anti-Semitic policies of varying levels of human rights violation. Thus, *IG Farben* and *Degussa* were guilty of using slave labor in concentration and death camps; *Degesch* provided such camps with means to carry out mass atrocities, namely Zyklon B for gas chambers, whereas *SNCF* took part in the deportation of French Jewry to camps to meet their end there.

The similarity uniting these examples is the attempt to wash their hands in the post-war period by building idealized images of corporate conduct – German companies were trying to remain profitable to sustain ordinary German families, whereas the French company was itself the vic-

tim of German occupiers. As I have argued, such images can be taken to advance the argument that companies behaved in accordance with their counter-ethos, a set of values that emerge in times of crisis and are justified by such unfortunate and pitiful circumstances. I further argued that this is essentially a bad argument – values being temporally and counterfactually stable cannot simply be put on a halt due to both conceptual reasons and historical evidence that people knew that the deterioration of values is morally and epistemically reprehensible.

Contemporary business ethics has shown laudable interest in these issues albeit from the top-down perspective. The need for ethical leadership was emphasized at the expense of fine-grained analysis of regular employees' behavior and endorsed values. I proposed a bottom-up perspective through which one can apply the framework of the ethics of collegiality. The solidarity of colleagues constituting work collectives should be understood as extending beyond such collectives, and the values embodied in anti-discriminatory and anti-racist policies should be shared and co-created by all employees to ensure that never again one's whole being gets determined by the inscription *Arbeit Macht Frei*.

References

Berber, A., & Subotić, V. 2022. The Anti-Individualistic Turn in the Ethics of Collegiality: Can Good Colleagues be Epistemically Vicious? *The Journal of Value Inquiry*, online first at https://doi.org/10.1007/s10790-022-09922-5

Betzler, M., & Löschke, J. 2021. Collegial Relationships. *Ethical Theory and Moral Practice* 24: 213–229.

Borkin, J. 1978. *The Crime and Punishment of I.G. Farben*. The Free Press.

Broch, L. 2014. Professionalism in the Final Solution: French Railway Workers and the Jewish Deportations, 1942–4. *Contemporary European History* 23(3): 359–380.

Broch, L. 2015. French Railway Workers and the Question of Rescue During the Holocaust. *Diasporas* 25: 147–167.

Cassam, Q. 2016. Vice Epistemology. *The Monist* 99 (2): 159–180.

Federman, S. 2017. Genocide Studies and Corporate Social Responsibility: The Contemporary Case of the French National Railways (SNCF). *Genocide Studies and Prevention: An International Journal* 11(2): 5.

Federman, S. 2021. Corporate Leadership and Mass Atrocity. *Journal of Business Ethics* 172: 407–423.

Fricker, M. 2021. Institutional Epistemic Vices: The Case of Inferential Inertia. In: I. J. Kidd, H. Battaly, & Q. Cassam (Eds.), *Vice Epistemology* (pp. 89–106). Routledge.

Gilbert, M. 2013. *Joint Commitment: How We Make the Social World*. Oxford University Press.

Hayes, P. 1995. Profits and Persecution: Corporate Involvement in the Holocaust. In J. Pacy & A. Wertheimer (Eds.), *Perspectives on the Holocaust* (pp. 63–88). Westview Press.

Hayes, P. 2004. *From Cooperation to Complicity: Degussa and the Third Reich*. Cambridge University Press.

Kelly, M. J. 2016. *Prosecuting Corporations for Genocide*. Oxford University Press.

Levi, P. 1959. *If This is a Man* (transl. S. Wolf). The Orion Press.

Liu, Y., Feng, M., MacDonald, C. 2019. A Big-Data Approach to Understanding the Thematic Landscape of the Field of Business Ethics. *Journal of Business Ethics* 160(1): 121–150.

Neuborne, B. 2003. Holocaust Reparations Litigation: Lessons for the Slavery Reparations Movement. *NYU Annual Survey of American Law* 58: 615–622.

Pruzan, P. 2001. The Question of Organizational Consciousness: Can Organizations Have Values, Virtues, and Visions? *Journal of Business Ethics* 29: 271–284.

Rosenbloom, A., & Althaus, R. 2010. Degussa AG and its Holocaust Legacy. *Journal of Business Ethics* 92: 183–194.

Sikimić, V., Nikitović, T., Vasić, M., & Subotić, V. 2021. Do Political Attitudes Matter for Epistemic Decisions of Scientists?. *Review of Philosophy & Psychology* 12: 775–801.

Smart, P. R. 2018. Mandevillian Intelligence: From Individual Vice to Collective Virtue. In: A. J. Carter, A. Clark, J. Kallestrup, O. J. Palermos, & D. Pritchard (Eds.), *Socially Extended Epistemology* (pp. 253–274). Oxford University Press.

Sofsky, W. 1996. *The Order of Terror: The Concentration Camp* (transl. W. Templer). Princeton University Press.

Wiesen, S. J. 1999. German Industry and the Third Reich: Fifty Years of Forgetting and Remembering. *Dimensions: A Journal of Holocaust Studies* 13(2): 3–8.

Wiesen, S. J. 2001. Morality and Memory: Reflections on Business Ethics and National Socialism. *The Journal of Holocaust Education* 10(3): 60–82.

Wieviorka, A. 2007. La SNCF, la Shoah et le Juge. *L'Histoire* 316: 89–99.

Wray, K. B. 2007. Who Has Scientific Knowledge? *Social Epistemology* 21(3): 337–347.

Zagzebski, L. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press.

4. SCIENCE, FICTION, AND JUSTICE: A STUDY OF VIRTUES AND VICES IN TODAY'S WORLD

Ivan Mladenović*

THE PROBLEM OF POLITICAL POLARIZATION AND A WAY OUT OF IT

Abstract: Since political polarization significantly impacts contemporary politics and democracy, much of the research in the social sciences is dedicated to this topic. In recent times, philosophers joined the discussion related to the research on political polarization, primarily in the fields of political philosophy and political epistemology. The main aim of this paper is philosophical analysis of some dominant explanations of political polarization, but also to propose solutions for a way out of political polarization from the perspective of political philosophy. In a nutshell, to find solutions for a way out of political polarization, I will be looking in the direction of boosting epistemic rationality and fostering communication in conditions of tolerance and equality.

Key words: political polarization, motivated reasoning, rationality, public deliberation

Introduction

Since political polarization significantly impacts contemporary politics and democracy, much of the research in the social sciences is dedicated to this topic. In recent times, philosophers joined the discussion related to the research on political polarization, primarily in the fields of political philosophy and political epistemology. Robert Talisse's definition of political polarization will be used as a starting point: "Political polarization denotes a family of phenomena having to do with what might be called the *political distance* between political opponents and the consequent dissolution of common ground between them." (Talisse 2021: 209). The main aim of this paper is philosophical analysis of some dominant explanations of political polarization, but also to propose solutions for a way out

* Department of Philosophy, Faculty of Philosophy, University of Belgrade, ivan.mladenovic@f.bg.ac.rs

of political polarization from the perspective of political philosophy. In a nutshell, to find solutions for a way out of political polarization, I will be looking in the direction of boosting epistemic rationality and fostering communication in conditions of tolerance and equality.

1.

There is a scientific consensus among environmental scientists that climate change is happening and is caused by anthropogenic factors (Oreskes 2004, Cook et al 2016). Taking into account the scientific consensus, it may appear surprising that a significant number of people who do not engage in science reject the evidence on climate change.[1] The simplest explanation would be that people who reject the scientific consensus are not sufficiently informed about climate change. However, research has shown that among people who reject scientific knowledge about climate change there is a significant number of those who are well-informed and clearly understand the conclusions reached by science (Kahan et al 2012). This opens a question how could one understand a tendency to reject and deny scientific knowledge on climate change, given that the facts and the conclusions reached by science are accessible to those who reject them.

Rejection of scientific knowledge on climate change is just one example of a recent phenomenon of rejection on the part of the general public of some scientific theories or of science as a whole. This phenomenon is also a subject of numerous topical research, primarily in the fields of psychology and other social sciences. This research goes into two directions. One is identification of psychological mechanisms which can explain the phenomenon, and the other concerns identification of political factors impacting rejection of science. In fact, even though they represent two different strands of research, they are often viewed as complementary parts of a comprehensive explanation of the science denialism.

Recent review of the science denialism explanations in the field of psychology points to motivated cognition characterized by the following elements: reliance on heuristics, differential risk perception and a tenden-

1 The research on the rejection of environmental science usually refer to US data. These data show that while until 1970s there existed a consensus regarding environmental issues, polarization first occurred in the Congress, where by 1990s the gap has become apparent as Republican Party representatives started voting against environmental legislation. This political polarization was subsequently reflected in views of the general public so that a significant change in attitudes occurred in a relatively short time period – the difference between those supporting the Democrats and the Republicans regarding climate change increased in the period from 2006 to 2016 from 25% to 46%. For the aforementioned data see: Bayes and Druckman 2021: 27.

cy towards believing in conspiracy theories (Lewandowsky and Oberauer 2016). Contemporary psychology offered a lot of evidence that people often do not make decisions on rational grounds (in line with the assumptions of rational choice theory and decision theory) and that they do it in intuitive way, by using simple heuristics. Such heuristics can sometimes be useful and lead to good solutions. For that reason, it is considered that bounded rationality although may not lead to the best, may lead to sufficiently good decisions and solutions. Moreover, relying on heuristics can sometimes lead to better outcomes than decision-making in accordance with the rules that are characteristic for rational choice theory and decision theory (Gigerenzer 2007). However, relying on certain heuristics may also have adverse consequences, which is why it can be considered irrational, given the deviation from the cannons of rationality (Ariely 2008).[2]

In their study, Lewandowsky and Oberauer seem to have in mind the latter understanding of heuristics. Understood in this way, rejection of scientific knowledge on climate change suggests that a person does not take into account available evidence in a rational way. Motivated cognition implies that a person is inclined toward rejection of evidence and scientific knowledge if such evidence and knowledge is not in line with her prior attitudes and beliefs. So, when someone is rejecting climate science, that person is under the influence of mechanisms that protect her prior attitudes and beliefs from exposure to evidence that is not in line with it. This is precisely the point where the political dimension of the science denialism plays decisive role, given that the person engaging in such rejection usually attempts to protect her own *political* attitudes and beliefs. Motivation also affects a differential risk perception, so that those who due to their political attitudes and beliefs reject evidence on climate change usually underestimate its risks. In addition, Lewandowsky and his colleagues in their research payed particular attention to the third element of motivated cognition at work when rejecting science – the tendency towards believing in conspiracy theories (Lewandowsky et al 2013).

Concerning political dimension of the science denialism, Lewandowsky and Oberauer point out that although motivated cognition indicates individual irrationality, there are certain political and economic actors for which incentivizing such a relationship toward science is fully rational because it furthers realization of their political or economic goals.[3] They term this political aspect of the science denialism "institutionally or-

2 This can be explained by the fact that even though heuristics are adaptations, many of them are adaptations to ways of life in which people found themselves in distant past. On adaptive characteristics of heuristics see: Gigerenzer 2007: Chapter 4.

3 See also: Lewandowsky et al. 2018: 188–190.

ganized denial" and conclude that "although the rejection of science may be driven by a common set of cognitive processes, it is clear that political, ideological, and economic factors are paramount" and that "the communication of contested science is therefore inextricably caught up in political battles" (Lewandowsky and Oberauer 2016: 220).

In addition to predominantly psychological explanations, philosophical explanations of the rejection of science and scientific knowledge have recently also been formulated. Relying on psychological research, Neil Levy offered several additional explanations from the perspective of epistemology (Levy 2019). We have seen that psychological explanations are largely based on heuristics that may lead to rejection of evidence and scientific knowledge. Levy explains that people's inclination toward those heuristics, even though it may individually lead to a path of acquiring a wrong belief, can collectively be understood as a kind of adaptation for collective deliberation.[4] Namely, acquiring and firm adherence to various beliefs, even wrong beliefs, may actually contribute to the quality of collective deliberation because it brings in a larger number of perspectives, which necessitates arguing about which beliefs should be adopted and which ones rejected. In the case of inexistence of a multitude of perspectives and the consequent necessity to advocate one's own belief, it would be much easier for certain wrong beliefs to be adopted at the collective level.

Levy alternatively expresses this idea in terms of epistemic individualism (Levy 2019: 314). Epistemic individualism can be understood as people's inclination to give advantage to their own beliefs over the beliefs of other people.[5] The inclination toward epistemic individualism, even though it can lead to acquiring and adhering to wrong beliefs at the individual level can be understood as an adaptation for collective deliberation. Taking this into account, paradoxically, there are two tendencies that appear to be relevant for explaining the phenomenon of rejecting science and scientific knowledge. On the one hand, epistemic individualism explains why some people are inclined to reject results of collective deliberation arrived at, for example, by the scientific community. On the other hand, given that epistemic individualism is an adaptation for collective deliberation, people should also have an inclination to adhere to beliefs

4 In this regard, Levy relies on Mercier and Sperber's research on reasoning and argumentation: Mercier and Sperber 2011. I will discuss Mercier and Sperber's theory in more detail in the fourth section of this paper.

5 Michael Lynch points to a similar phenomenon that contributes to disagreement among people which he terms intellectual arrogance: "*Intellectual arrogance* is the psycho-social attitude that you have nothing to learn from anyone else about some subject or subjects because you know it all already. This is the arrogance of the know-it-all" (Lynch 2021: 252).

that are collectively accepted on the basis of evidence and argumentation. Why then all people do not believe in knowledge on climate change that is the result of collective deliberation within the scientific community?

To answer this question, Levy also thinks that a part of the explanation lies in political factors. However, he approaches the explanation from an epistemic perspective. In order to understand why there is a rejection of science and scientific knowledge, it should first be understood why there is acceptance of scientific results by the general public. That is, why people who are assumed to be epistemic individualists come to accept the results of collective deliberation. Levy suggests that the development of science from the 16th century onwards is not only a product of collective deliberation, but also of institutionalized collective deliberation (Levy 2019: 317). It means that a large portion of scientific success rests on institutional mechanisms which lead to reliable knowledge due to productive disagreement (for example, through anonymous reviews before academic findings are published, but also due to critical discussions within scientific community once they have been published).

For scientific knowledge to be accepted by the general public, not only the aspect of reliability but also the aspect of benevolence is important. In this regard, Levy maintains that when explaining rejection of scientific knowledge and testimony offered by scientific theories, political factors should also be considered. Namely, in the case when science itself is politicized, suspicion concerning its benevolence may lead to a rejection of its reliability. Thus, according to Levy, an important part of the explanation why a portion of the general public rejects scientific knowledge on climate change is that this knowledge is politicized, i.e., understood as expressing attitudes typical of a specific political viewpoint. So, those who reject it do not reject it as scientific knowledge, but as political views that are opposed to their political views.

In this section, I examined several explanations of the contemporary phenomenon of rejecting science and scientific knowledge. I illustrated this problem with a topical example of climate change, on which there is a scientific consensus, but which a portion of the general public nevertheless rejects. I have considered some dominant explanations both from the perspective of psychology and philosophy. Both types of explanations point to psychological mechanisms and political factors that impact rejection of scientific knowledge. In this section, I have tackled a specific example of rejecting scientific knowledge on climate change in the light of general types of explanation of this phenomenon from the perspective of psychology and philosophy. In the following two sections, I will tackle more general phenomenon of political polarization, focusing on specific mechanisms on which explanations of this phenomenon are based.

2.

An explanation for political polarization offered by social psychologist Jonathan Haidt is largely based on previously identified elements – motivated cognition and political factors (Haidt 2013). An integral part of the explanation is based on his empirical research concerning the foundations of morality (and in particular, moral foundations of politics). His comprehensive explanation of political polarization consists therefore of three parts. The first part of the explanation is characterized by Haidt's view that "*intuitions come first, strategic reasoning second*" (Haidt 2013: xiv). This view implies that people often make decisions by relying on intuition and heuristics. Haidt's view suggests not only that intuition in temporal sense precedes rational thinking and reasoning, but also that an exercise of rational faculties can only be seen as a justification of previously given intuitions a person has. In other words, people are mostly guided by their intuitions that do not have rational grounds, while using rational reasoning mostly in order to justify intuitions they already have. The second part of the explanation is characterized by the view that "*there's more to morality than harm and fairness*" (Haidt 2013: xv). Haidt argues that political polarization largely stems from differing foundations of morality people rely on in order to ground their political (primarily ideological) views. Finally, the third part of the explanation is based on the view that "*morality binds and blinds*", referring to political significance of identification with a specific group, for which Haidt finds sources within evolutionary theory, more specifically a (recent) theory of group selection (Haidt 2013: xvi).

Haidt places particular emphasis on two psychological mechanisms related to the view that "*intuitions come first, reasoning second*", which are particularly relevant for explaining political polarization. These psychological mechanisms are *biased confirmation* and *motivated reasoning*. In regard to biased confirmation, Haidt refers to well-known studies within psychology. For example, he points to a significance of the experiment carried out by Wason regarding "the 2–4–6 problem" (Haidt 2013: 92). In this experiment, the respondents were given a series of numbers "2–4–6" and they were asked to provide other number series to the experimenter in order to establish a pattern according to which the numbers had been ordered. The experiment showed that the respondents mostly cited series of numbers which confirmed the pattern they themselves assumed in advance, usually assuming that for any additional number one should add "+2" which is wrong because the experimenter had assumed a pattern of a series of numbers in which each successive number was greater than the previous one. This research has shown that people usually search for information that confirms their previous beliefs or assumptions rather than

the information that exposes their assumptions to being falsified. Given that in those cases people exclusively seek the information or evidence that confirms their previous beliefs and assumptions, this type of behavior has been termed *biased confirmation*.[6]

Motivated reasoning is a similar psychological mechanism which differs from biased confirmation because once it is engaged in reasoning, a person disregards or rejects the information not in line with her previous attitudes or beliefs. Haidt illustrates the distinction in the following way. In the case of biased confirmation, a person asks herself "Can I believe it?" and replies "Yes" if she finds evidence or pseudo-evidence that confirms her belief, while in the case of motivated reasoning, a person asks herself "Must I believe it?" and rejects the belief if she finds any sort of information that would undermine it (Haidt 2013: 98).

Ziva Kunda suggested, back in the 1990s, that the experimental evidence from various fields of psychological research points into the direction of the unique mechanism of motivated reasoning (Kunda 1990).[7] About the mechanism of motivated reasoning she says the following:

> "I propose that people motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer. They draw the desired conclusion only if they can muster up the evidence necessary to support it... In other words, they maintain an "illusion of objectivity"... The objectivity of this justification construction process is illusory because people do not realize that the process is biased by their goals, that they are accessing only a subset of their relevant knowledge, that they would probably access different beliefs and rules in the presence of different directional goals, and that they might even be capable of justifying opposite conclusions on different occasions." (Kunda 1990: 486)

The formulation that in the course of motivated reasoning people „attempt to be rational" should not be misunderstood. It merely means that people use their rational faculties to justify their desired conclusion which has already been determined in advance by their directional goals (previous attitudes and motivation). The reasoning process is therefore basically irrational despite the use of rational faculties. For that reason, motivated reasoning is one of the mechanisms which shows that, in Haidt's words, "*intuitions come first, strategic reasoning second*". The workings of the

6 For numerous other experiments concerning biased confirmation see: Nickerson 1998.

7 For the sake of precision, it should be noted that her paper strives to identify even more basic mechanisms found in the root of motivated reasoning that pertain to selective approach to memory and construction of beliefs. In retrospect, it seems that her paper has had a much larger influence regarding identification of the mechanism of motivated reasoning rather than these more basic mechanisms.

mechanism of motivated reasoning can be illustrated by the experiment carried out by Lord, Ross and Lepper regarding different views on capital punishment (Lord, Ross and Lepper 1979). In this experiment, the participants who opposed capital punishment as well as those in favor of capital punishment were given articles to read that contained arguments for and against capital punishment. The experiment showed that those who were in favor of capital punishment saw the articles as an additional confirmation of their prior views and vice versa.[8]

So, the first part of Haidt's explanation of political polarization is based on mechanisms of biased confirmation and motivated reasoning which show that the process of reasoning does not necessarily lead to rationally-based conclusions and beliefs. The second part of the explanation refers to different sources of moral intuitions people have. This part of the explanation is based on Haidt's research on foundations of morality and their political implications. Namely, Haidt and his colleagues have tried to identify modules behind various moral intuitions, building on insights from evolutionary psychology. On the basis of their research, they came to the conclusion that points to (at least) six sources of moral intuitions. These are the following foundations of morality: care, fairness, liberty, authority, loyalty and sanctity (Haidt 2013).

In what way are these foundations of morality relevant for understanding political polarization? The experiments conducted by Haidt and his colleagues show that research on foundations of morality has clear political implications. The conclusion they reached is that people who have liberal political views mostly ground their beliefs on three former foundations of morality (care, fairness, liberty), while those who have conservative political views ground their political beliefs on all six foundations of morality (although authority, loyalty and sanctity have crucial importance for them, while they interpret the former three in a way different from people with liberal views). The divergence of moral intuitions that have their political significance largely derives, in Haidt's opinion, from different moral foundations in which liberal and conservative views are grounded. Thus, it is largely because of the differences in foundations of morality of liberal and conservative views that people may end up in political polarization.[9]

However, the insight that "*there's more to morality than harm and fairness*", according to Haidt, is still not sufficient to explain political polarization. The third part of the explanation is also necessary, showing that

8 For numerous other experiments regarding motivated reasoning see: Kunda 1990.

9 Haidt notes that this ideological division is associated with the USA, where liberal orientation includes left-wing ideological views (Haidt 2013: xvii).

"*morality binds and blinds*". This part of the explanation is based on evolutionary biology, more specifically on multilevel selection. However, Haidt focuses on only two levels, the collective level and the individual level. He considers acceptable explanations of cooperation from the "selfish gene" perspective by means of a mechanism of reciprocal altruism, rejecting however the assumption that human cooperation can be explained only from such an individualistic perspective. He argues that the most recent findings on group selection also have to be taken into account. The point is that morality appeared largely a result of group selection, i.e., morality emerged due to the situations of conflict among groups. Given that cooperation has primarily developed within groups, it has led to parochial altruism which boosts cooperation within one's own group. Hostile feelings towards other groups can also be explained from this collective perspective. So, on Haidt's view, "*morality binds and blinds*" because people are bound primarily to members of their own group and beliefs they share.

This part of the explanation from the perspective of group selection is, according to Haidt, crucially important for understanding the phenomenon of political polarization. Haidt says that, "these tribal instincts are a kind of overlay, a set of groupish emotions and mental mechanisms laid down over our older and more selfish primate nature. It may sound depressing to think that our righteous minds are basically tribal minds, but consider the alternative. Our tribal minds make it easy to divide us, but without our long period of tribal living there'd be nothing to divide in the first place." (Haidt 2013: 246). So, political polarization does not emerge only because "righteous minds" are based on different foundations of morality, but also because people are inclined to side and identify with their own group, i.e., to consider views typical of the group they identify with correct and the views of opposing groups wrong. Finally, given that "*intuitions come first, reasoning second*", the sort of beliefs a person will accept or reject largely depends on whether they are in accordance with convictions of the group she identifies with.

Haidt thinks that realizing that there are foundations of morality on which different political (primarily ideological) convictions and beliefs are based may help not only explain, but also overcome political polarization. The route to a way out of political polarization would consist of better understanding of the reasons people have for different political views. But, aside from pointing out that better understanding of different foundations of morality may lead to better understanding of other people, subsequently leading to a realization of certain correct views in the political standpoint of an opponent, Haidt does not offer any institutional mechanism of political decision-making that would lead towards overcoming or at least

reducing political polarization. Moreover, people may gain a better understanding of why other people have attitudes and beliefs they do, and still continue to disagree with them. Better understanding of attitudes and beliefs of other people, does not necessarily lead toward overcoming or reducing political polarization. However, this is not to deny that it may be an important step in that direction.

There is also another problem with Haidt's explanation of political polarization. Namely, Haidt's emphasis on people being intuitive rather than rational beings, his emphasis of authority, loyalty and sanctity as foundations of morality, and on collective identity and collective values, overlaps with some of the basic tenets of conservative political views. This is problematic, because then research on political polarization between liberal and conservative views is largely based on prior acceptance of conservative views, which at the same time purport to be the object of analysis. Haidt himself admits that reading works of conservative political theorists has led him to realize this overlap with their views (Haidt 2013: 338). He also defends the conception of "Durkheimian utilitarianism" acceding greater correctness to such collectivistic conservative views inasmuch as they contribute to greater degree of happiness. The problem is that Haidt's views may lead to political polarization with respect to science, which is contrary to any rapprochement of political standpoints he allegedly advocates.

3.

The explanatory framework extrapolated in the previous section has recently been additionally specified inasmuch as *politically motivated reasoning* had been isolated as a basic psychological mechanism leading to political polarization. For that reason, methodology and specific experimental design that explore in what way politically motivated reasoning leads to political polarization have been laid out. As a part of this approach, issues of rationality and irrationality of politically motivated reasoning have been investigated, as well as whether it is a phenomenon typical of specific (mostly conservative) ideological view or whether there is a symmetry between different ideological orientations regarding political polarization.

Kahan and his colleagues conducted a series of experiments which demonstrate that *politically motivated reasoning* may be conceived as the main psychological mechanism behind political polarization (Kahan 2016a). In his research, Kahan starts from the assumption that contemporary political life is largely characterized by disagreement on factual mat-

ters, for example, whether climate change is happening. Given that factual issues are concerned, it is obvious that those who disagree do not do so on the basis of evidence, but on the basis of values.

However, in any explanation of political polarization, according to Kahan, a key role has to be played by specifically political and ideological values on which group identity rests. When engaged in politically motivated reasoning, people will be prone to reject evidence to the extent that it contradicts the view of the group they identify with. On Kahan's view, an important characteristic of politically motivated reasoning is precisely protection and defense of identity typical of the group with specific political and ideological values. Thus, politically motivated reasoning leads to "*identity-protective cognition*" (Kahan 2017: 1). Kahan summarizes his view on politically motivated reasoning in the following way:

> "Where positions on some policy-relevant fact have assumed widespread recognition as a badge of membership within identity-defining affinity groups, individuals can be expected to selectively credit all manner of information in patterns consistent with their respective groups' positions. The beliefs generated by this form of reasoning excite behavior that expresses individuals' group identities. Such behavior protects their connection to others with whom they share communal ties" (Kahan 2016a: 2)

And this leads, according to Kahan, to the following concequences:

> "When individuals apprehend – largely unconsciously – that holding one or another position is critical to conveying *who they are* and *whose side they are on,* they engage information in a manner geared to generating identity-consistent rather than factually accurate beliefs." (Kahan 2017: 6)

The explanation of political polarization on factual matters therefore lies in the way of reasoning individuals resort to in order to express and protect their political identity that essentially boils down to an identity of a specific political group or a group sharing common ideological convictions. Therefore, to the extent to which people reason in this way, they are prone to reject evidence which questions the values of the group they identify with. Kahan and his colleagues investigated this effect in the experiment which largely addresses polarization on climate change that was discussed in the first section of this paper (Kahan et al. 2011, Kahan 2016a).

In the experiment, respondents first read a short bio of a person they are told was an expert in the field of climate change. This basic information is such that on the basis of it, anyone could easily come to the conclusion that the person indeed was an expert in that field. However, after this initial piece of information, the respondents in the second part of the experiment are informed that the given person maintained that there was

a high risk (and alternatively a low risk) regarding climate change. We have seen in the first section that an important characteristic of motivated cognition regarding climate change was a differential risk perception. Relatedly, the experiment demonstrated that the additional information regarding climate change risk largely influenced the original assessment whether the person was an expert on environmental issues. Namely, the respondents of conservative ideological orientation were prone to believe that the person was not an expert if they received additional information about the person's belief in the high climate change risk, while persons of liberal political orientation regarded the same person as an expert in the light of the same piece of information. What this experiment demonstrates is that political and ideological factors may affect the judgement on whether someone was a climate change expert; on the basis of these factors, evidence is rejected if it does not accord with political and ideological views with which a person identifies, consequently leading to political polarization.

In the second experiment on political polarization, Kahan tested whether politically motivated reasoning can be considered rational or irrational and whether there is an asymmetry or symmetry in the inclination of people who have different ideological views to rely on this psychological mechanism (Kahan 2013). It is noteworthy that Kahan makes a difference between Bayesian rationality, as a typical model of rationality (where prior probability regarding an assumption or a hypothesis is adequately revised in the light of new evidence), biased confirmation (where prior probability regarding an assumption or a hypothesis directly determines the acceptability of evidence) and politically motivated reasoning (where prior probability regarding an assumption or a hypothesis is determined by political identity, which directly affects acceptance or rejection of evidence) (Kahan 2016a).

Kahan makes a difference among several approaches that generate different predictions regarding the role of motivated reasoning in political polarization. The first approach which is dominant within psychology of reasoning and rationality, termed the dual process theory, makes a difference between System 1, which is intuitive, fast, simple and primarily based on emotion, and System 2 which is reflexive, slow, requires analytical thinking and cognitive processes. According to Kahan, given the priority of intuitive system 1 when explaining motivated reasoning, the dual process theory approach presupposes decisive influence of that system for explanation of political polarization.[10] The second approach stresses

10 In the previous section, we have seen that Haidt's explanation of political polarization can also be understood in a similar way, because it is based on intuitive System 1.

asymmetry regarding different ideological views in relation to political polarization, assuming that people with right-wing and conservative ideological orientation are more prone to rely on intuition, and therefore more prone to use motivated reasoning leading to political polarization. Finally, Kahan advocates a third approach based on politically motivated reasoning. Quite contrary to previous approaches, this model envisages that System 2 has a greater effect on political polarization, but also that there is a symmetry between people with different ideological views regarding an inclination to politically motivated reasoning.

The design of the experiment is such that it consists of two parts. Within the first part of the experiment, the respondents take the Cognitive Reflection Test, a standard test on the basis of which it can be ascertained to what extent the respondents rely on System 1 and on System 2 (which usually shows predominant relying on System 1). The second part of the experiment consists of respondents being given a piece of information which informs them that those who achieved good scores on the Cognitive Reflection Test usually accept (or reject, respectively) evidence regarding climate change, on the basis of which they are expected to assess the validity of the test.

Kahan reports that the results of this experiment have shown that predictions from the perspective of the model of politically motivated reasoning are more accurate than predictions of alternative approaches. Recall that the dual process theory and ideological asymmetry theory predict that intuitive reasoning typical of system 1 was the primary factor for the explanation of political polarization. Quite the contrary, Kahan's experiment shows that persons scoring better at the Cognitive Reflection Test (which is one of the indications for greater reliance on System 2) are more inclined to rely on politically motivated assessment of the validity of the test on the basis of additional piece of information regarding acceptance (or rejection) of evidence on climate change. Furthermore, Kahan reports that this can equally be noticed among people who scored better at the Cognitive Reflection Test, both among those who displayed liberal views and those who displayed conservative views. In other words, the results of Kahan's experiment show not only that reliance on System 2 to a larger extent led towards political polarization, but also that an inclination to politically motivated reasoning was symmetrical in terms of different ideological standpoints.

Relying on results of the experiment, Kahan concluded that politically motivated reasoning can be considered an adequate explanation of political polarization, because in the light of additional information which directly referred to political and ideological identity, political polarization

was generated. In addition, Kahan drew two additional conclusions that politically motivated reasoning can be considered rational and that there was a symmetry in the inclination toward politically motivated reasoning regardless of ideological orientation. The first conclusion is somewhat surprising given the previous discussion in this paper. Namely, as we have seen, some dominant psychological explanations of political polarization and the rejection of scientific knowledge emphasize significance of motivated cognition, i.e., reliance on intuition and heuristics typical of System 1. In sharp contrast, Kahan emphasizes that "far from reflecting *too little rationality*, then, politically motivated reasoning reflects *too much*" (Kahan 2016b: 4). Kahan finds evidence for this conclusion in the fact that people who rely more on System 2 also have a greater inclination to politically motivated reasoning. He explains this in the following way:

> "Given the social meanings that factual positions on these issues convey, however, failing to adopt the stance that signals who she is – *whose side she is on* – could have devastating consequences for a person's standing with others whose support is vital to her well-being, emotional and material. Under these conditions, it is a perfectly rational thing for one to attend to information in a manner that promotes beliefs that express one's identity correctly, regardless whether such beliefs are factually correct... And if one is really good at conscious, effortful information processing, then it pays to *apply* that reasoning proficiency to give information exactly this effect." (Kahan 2016b: 4).
>
> "Far from evincing irrationality, this pattern of reasoning promotes the interests of individual members of the public, who have a bigger personal stake in fitting in with important affinity groups than in forming correct perceptions of scientific evidence." (Kahan 2017: 1)

However, there is an ambivalence in Kahan's specification of rationality. In order to see the problem, recall Kahan's initial differentiation between Bayesian rationality, biased confirmation and politically motivated reasoning. One of the key insights which Kahan reaches on the basis of his experiments is that unlike Bayesian rationality which is truth convergent, biased confirmation and politically motivated reasoning are not truth convergent. What distinguishes biased confirmation from politically motivated reasoning is that in relation to politically motivated reasoning it is possible to formulate specific predictions on the basis of ideological identity, which would not be possible with regard to biased confirmation. For example, if people do not have any previous knowledge on nanotechnologies, from the perspective of biased confirmation it is difficult to have any prediction what their views would be once they have been fed the information of such kind. However, if one assumes general disinclination towards new technologies as an important characteristic of an ideological view, from the perspective of politically motivated reasoning clear predic-

tions can be made regarding people's views once they have been fed the same type of information. Regardless of this difference, Kahan thinks that both types of reasoning clearly differ from Bayesian rationality inasmuch as evidence is not approached in a rational way.

We have seen that for Kahan politically motivated reasoning can be considered as a rational way of thinking. But how can politically motivated reasoning at the same time be rational and not be rational because it deviates from Bayesian rationality? It is obvious that there is ambivalence in Kahan's specification of rationality. Given that he does not make any further clarification, this ambivalence to a great extent limits the scope of his claim that relying on politically motivated reasoning is "perfectly rational". In other words, even if it is rational for a person to rely on politically motivated reasoning in order to promote her own interests, this cannot be rational in an epistemic sense of the term, because her way of reasoning deviates from adequate consideration of evidence and revision of degrees of belief.

Kahan's second conclusion suggests that there is a symmetry in the inclination toward politically motivated reasoning. Kahan thinks that his experiment only shows that the issue must remain unresolved and open for further research (Kahan 2016b: 5–6). In this regard, he actually compares results he had arrived at on the basis of his experiment with results of other experiments. However, when his symmetry thesis is viewed outside the lab context, it is obvious that the asymmetry thesis has a much larger evidential support that political science had mustered.

We have seen in the first section that climate change denial has its origin in conservative political beliefs and strategies adopted by the Republican Party for the sake of its own political agenda, which led to an emergence of public polarization on climate change (Bayes and Druckman 2021). Lewandowsky and Oberauer point out that this does not only pertain to scientific knowledge regarding climate change: "the rejection of specific scientific evidence across a range of issues, as well as generalized distrust in science, appears to be concentrated primarily among the political right" (Lewandowsky and Oberauer 2016: 218). However, our criticism of Kahan's symmetry thesis should not be misunderstood. It does not suggest that only conservatives are inclined toward politically motivated reasoning. The experiments clearly show that anyone regardless of ideological viewpoint can be subject to the politically motivated reasoning. The criticism merely suggests that in a situation when experimental evidence does not provide sufficient reasons to decide in favour of the symmetry thesis or the asymmetry thesis, additional evidence arrived at within political science about functioning of contemporary political life

can be relevant in this regard. And this evidence adds more weight to the asymmetry thesis than to Kahan's symmetry thesis.[11]

4.

So far I examined various explanations of political polarization. In this section, I turn to the question in what way political polarization can be overcome. At first glance, the very formulation of the question suggests that political polarization is something necessarily bad. For that reason, I would like to emphasize that one of the main tenets of democratic societies is the fact of disagreement. Taking into account that disagreement may not be something necessarily bad, I will make a distinction between *epistemically positive political polarization* and *epistemically negative political polarization*. What is epistemically positive political polarization? The dominant explanations of political polarization (including those we examined in the previous sections) view this phenomenon as the one in which both sides are equally under the influence of unconscious psychological mechanisms that lead them to a rejection of evidence which is contrary to their political identity. However, one of the sides may have correct beliefs that are based on evidence and the best scientific theories. In other words, rather than concluding that in the process of political polarization both sides are necessarily wrong due to the influence of psychological mechanisms such as politically motivated reasoning, one of the sides may actually have correct beliefs in the sense that the beliefs of that side are supported by evidence and formed in a rational way. The insistence on truth and correctness of belief *because it is based on evidence* is something that may lead to belief polarization, and even to political polarization. However, in that case, we have an epistemically positive political polarization, because it preserves knowledge and truth rather than political identity of a specific group.

The main problem regarding political polarization, at least in the form in which it emerges in contemporary societies is that it goes precisely in the opposite direction – in the direction of epistemically negative political polarization. Namely, a characteristic of epistemically negative political polarization inheres in the aspiration to disseminate wrong beliefs and incite irrational response towards available evidence, for the sake of achieving specific political and economic goals. The problem with epistemically negative political polarization is precisely that it aspires to align people into insular groups who share certain attitudes and beliefs and who view

11 On this point see also: Levitsky and Ziblatt 2019: Chapter 7.

other people who do not share their attitudes and beliefs as enemies rather than citizens who simply disagree with them. Such a kind of political polarization is harmful for democracy because it disrupts the ties of democratic citizenship that bind all citizens despite their different attitudes and beliefs. It leads towards aspiring to impose beliefs characteristic of one insular group on the entire society and guide society toward authoritarian forms of rule in which opposite beliefs and views are not tolerated and are moreover suppressed and considered undesirable. Therefore, the question regarding a way out of political polarization primarily refers to this type of epistemically negative political polarization that undermines democratic society and normalizes authoritarian forms of rule and behavior.

The solutions for a way out of political polarization (understood in the sense of epistemically negative political polarization) largely depend on a series of factors which pertain to a degree of polarization, level of development of democracy and democratic institutions, accessibility of scientific knowledge etc. To be sure, interdisciplinary research by different disciplines in social sciences and humanities can offer the best route to find solutions for a way out of political polarization. In the rest of the paper, I will suggest some routes from the perspective of political philosophy. In that regard, I will make a distinction between *individual* and *institutional* solutions for political polarization.

We have seen in the previous section that a large part of explaining political polarization refers to whether people approach evidence in a rational way. I have pointed out that one of the dominant explanations that adduces politically motivated reasoning is in fact ambivalent in that regard. This problem, in my view, is related to what I have termed individual solutions for a way out of political polarization. Glüer and Wikforss point out that in Kahan's view on politically motivated reasoning, no clear distinction has been made between epistemic and practical rationality (Glüer and Wikforss 2022: 38). In short, epistemic rationality refers to rational foundation of belief, while practical rationality refers to rationality of reasons for action. Their criticism also points out that even if Kahan's explanation of rationality can be understood in terms of practical rationality, it certainly cannot be understood in terms of epistemic rationality. Glüer and Wikforss suggest that for the phenomenon of "knowledge resistance" which has emerged in contemporary societies is characteristic "an irrational response to evidence" and that "it always includes irrationality" (Glüer and Wikforss 2022: 37, 43).

Clear understanding of aspects which pertain to rationality and irrationality is very important for individual solutions to find a way out of political polarization. Namely, if it is clear that one of the main problems

leading to political polarization is that individuals approach evidence in irrational way, then solutions should be sought in the direction of boosting epistemic rationality at the individual level. An objection could immediately be made that individual solutions are overly (and even hopelessly) optimistic if they expected individuals who in irrational way approach evidence to be ready to realize and, moreover, rectify this way of approaching evidence. However, this largely depends on various types of incentives which presently largely go in favor of epistemic irrationality. But, incentives in the direction of epistemic rationality originating from formal and informal education, as well as public policy, may be significant for individual solutions. The goal of this type of education and public policy is merely to make widely available knowledge about possible ways individuals have to overcome epistemic irrationality. In any case, how this knowledge will be used depends solely on individuals themselves. That these ideas are not overly (or hopelessly) optimistic is suggested by a psychological approach that testifies about positive effects of boosting rationality (Hertwig 2017).

I turn now to institutional solutions for political polarization. I already pointed out that institutional solutions pertaining to education and public policies may contribute to individual solutions regarding political polarization. However, the main institutional solution that I have in mind is public deliberation, that is, a sort of public discussion within which citizens in conditions of freedom and equality express their reasons for the views they advocate and listen to arguments by fellow citizens. This institutional solution and individual solutions are complementary in the sense that precisely a discussion with other people and new information acquired in that way may lead to boosting epistemic rationality. In the second section, I agreed with Haidt's view that confronting a contrary opinion and an inclination to understand why other people advocate contrary attitudes was an important step for overcoming political polarization, but I have also then pointed out that institutional mechanisms were needed for that purpose. Public deliberation is such a type of institutional mechanism because it fosters communication with other people in conditions that promote tolerance and equality. However, I emphasize that the purpose of public deliberation as *an institutional solution* is precisely in the provision of permanent institutional mechanism, not a one-off solution to the problem of political polarization.

In the context of this institutional solution, an objection can be made as well that it is an overly (and even hopelessly) optimistic expectation. In order to answer that objection, I turn to Mercier and Sperber's research on the function of reasoning. Their research shows that psychological mech-

anisms which lead to irrational way of reasoning at the individual level from an evolutionary perspective can best be described as adaptations for communication and collective reasoning (Mercier and Sperber 2011, Mercier and Sperber 2012). In their view, "reasoning has evolved and persisted mainly because it makes human communication more effective and advantageous" (Mercier and Sperber 2011: 60). Mercier and Sperber go on to explain that biased confirmation and motivated reasoning have important functions that are related to collective reasoning. Biased confirmation has an important function of protecting one's own standpoint in the course of discussion with other people and motivated reasoning has a function of not accepting lightly the views put forward by others. Thanks to those mechanisms, according to Mercier and Sperber, people only through discussion with other people arrive at the best solutions, that is, realization which reasons and arguments are the most convincing. They advocate the view that the function of reasoning is primarily associated with a collective plan of communication and that therefore reasoning functions well in that context.

Even though the main aim of Mercier and Sperber's research is explanation of reasoning in the light of evolutionary theory, Mercier and Landemore connected the results of this research with democratic theory pointing out its significance for understanding public deliberation and deliberative democracy (Mercier and Landemore 2012). They complement earlier insight that individual reasoning does not function well outside communication context with the view that it would not function well in conditions of a discussion between like-minded people. The psychological mechanisms such as biased confirmation and motivated reasoning in the context of a discussion among like-minded people lead precisely to a dynamic of group polarization. For that reason, they think that reasoning, in addition to functioning well at the collective level of communication, works best in conditions of mutual disagreement. And precisely communication with other people who initially disagree is a way to arrive at the best solution or the best decision. Mercier and Landemore conclude that „*fixing individual reasoning is not the solution*"; instead, to improve reasoning, „the changes should be made at the institutional rather than the individual level" (Mercier and Landemore 2012: 254). However, although I agree that institutional changes going in the direction of public deliberation would be important for a way out of political polarization, I do not fully agree with Mercier and Landemore's conclusion that solutions should not be sought at the individual level as well. Quite the contrary, I think that individual and institutional solutions are complementary and both important for a way out of political polarization.

Some recent experiments on public deliberation which directly or indirectly pertain to a possibility of reducing political polarization also show that the proposed institutional solution is not overly optimistic. One of the most important experiments regarding deliberative democracy is *deliberative poll*. In this experiment, randomly selected representative sample of citizens had an opportunity to discuss, in two days, working in smaller groups as well as in bigger plenary sessions (within which experts for a given area also took part), certain social and political issues. Before and after these two days, they obtained an identical questionnaire which pertained to the extent of their knowledge, as well as their preferences regarding the topic of discussion. Numerous experiments with deliberative polling have shown significant improvements in terms of the level of knowledge after only two days of participating in the experiment, as well as significant changes of preferences.

One of the criticisms of deliberative democracy has been that public discussion (especially among like-minded people) may lead to group polarization (Sunstein 2002). On the basis of an analysis of ten previously conducted experiments regarding deliberative polling, Fishkin and his colleagues have reached a conclusion that public deliberation may actually lead to depolarization (Luskin, Fishkin and Hahn 2007). These conclusions have been made on the basis of deliberative polling experiments, even though neither of experiments had been specifically designed to address the issue. Unlike these previous experiments, a recently conducted deliberative poll entitled *America in One Room* aimed precisely to establish to what extent public deliberation may contribute to a reduction of political polarization (Fishkin et al. 2021). The results show that owing to public deliberation, it is possible to arrive at a significant reduction of political polarization in two respects – regarding topics on which the citizens are most polarized, and regarding affective aspect of polarization.[12] In other words, owing to public deliberation it is possible to come to a rapprochement of attitudes, but also to a reduction of negative affects between political opponents. The results of the experiment also show that these effects do not merely occur among people who are moderately polarized, but also those who are extremely polarized. Moreover, the experiment has shown that when certain topics are concerned, two-way depolarization had oc-

12 Some recent experiments show that discussion can be particularly important for reducing affective polarization (Santoro and Broockman 2022). The results of these experiments have shown that affective depolarization can primarily be achieved in a discussion of opponents regarding topics that are not the object of polarization of attitudes, but also that effects of discussion in this regard are short-term (around three months) and that they may easily disappear if the discussion relates to topics that are the object of political polarization.

curred, but also a one-way depolarization (namely, rapprochement of attitudes due to greater changes among one of the polarized sides).

Recent research conducted by Mercier and Cladière, even though it does not tackle political polarization, indirectly concerns the institutional solution we proposed in this section (Mercier and Cladière 2021). These authors proceed from the hypotheses that public discussion would contribute to better knowledge and convergence of attitudes regarding factual issues. Even though the experiment directly addressed the „wisdom of crowds" when larger groups of people are concerned, it is indirectly relevant for proposed institutional solution, given that it explores the possibility of convergence of attitudes regarding factual issues on the basis of public discussion. The results of the experiment show that only 15 minutes of public discussion has made people give much more accurate answers to questions regarding factual issues compared to their initial individual responses.

Recently, an experiment has been conducted aiming to establish to what extent citizens' discussion within a smaller deliberative body or a mini-public about the facts relevant for enactment of policies may affect larger acceptance of evidence among the broader public (Már and Gastil 2020). This experiment is interesting in the present context because it aimed to establish to what extent a report of a deliberative body would affect motivated reasoning (given existing polarization regarding the topic of GMO regulation that was the object of public deliberation) and to what extent it would contribute to better realization of facts. The results of the experiment have shown that an information regarding deliberation on the given topic has actually among broader public led to a greater degree of knowledge about factual issues, rather than to a rejection of evidence on the basis of motivated reasoning. Moreover, the experiment has shown that acceptance of evidence and better knowledge regarding factual issues occurred even among people who had been most polarized on ideological grounds. Considering that in this paper I have mostly dealt with political polarization on factual issues, the results of aforementioned experiments provide some evidential support for the expectation that public deliberation may lead to depolarization on the questions of facts.

Conclusion

In this paper, I have focused on rejection of evidence as one of the main characteristics of political polarization. This does not mean that I consider the role of values less important for explanation of political polarization. On the contrary, as we have had a chance to see, values may

precisely be the sources for rejection of evidence, and among the drivers of epistemic irrationality. However, I focused on political polarization over evidence for two reasons. The first reason is that polarization over facts is surprising and requires additional explanation. If disagreement among people regarding values is something that is expected, disagreement over evidence certainly is not. The second reason is that the proposed solutions for a way out of political polarization may play a role precisely in this regard. Namely, the basic expectation from the proposed individual and institutional solutions is not convergence of value-related attitudes, but a possibility that people would approach evidence in a more rational way. So, the expectation is that the way out of political polarization may begin with the first step that pertains to a reduction of polarization over evidence and factual issues.

References

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York: HarperCollins Publishers.

Bayes, R. and Druckman, J. N. (2021). Motivated Reasoning and Climate Change. *Current Opinion in Behavioral Sciences*, 42: 27–35.

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., and Rice, K. (2016). Consensus on Consensus: A Synthesis of Consensus Estimates on Human-caused Global Warming. *Environmental Research Letters*, 11, 048002.

Fishkin, J., Siu, A., Diamond, L. and Bradburn, N. (2021). Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on "America in One Room". *American Political Science Review*, 115 (4): 1–18.

Gigerenzer, G. (2007). *Gut Feelings: The Intelligence of Unconscious*. London: Viking Penguin.

Glüer, K. and Wikforss, Å. (2022). What is Knowledge Resistance? In: Strömbäck, J., Wikforss, Å., Glüer, K., Lindholm, T. and Oscarsson, H. (eds). *Knowledge Resistance in High-Choice Information Environments*. London and New York: Routledge, pp. 29–48.

Haidt, J. (2013). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. London: Penguin Books.

Hertwig, R. (2017). When to Consider Boosting: Some Rules for Policy-makers. *Behavioural Public Policy*, 1 (2): 143–161.

Kahan, D. M. (2013). Ideology, Motivated Reasoning, and Cognitive Reflection. *Judgment and Decision Making*, 8: 407–424.

Kahan, D. M. (2016a). The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. In: Scott R. A. and Kosslyn, S. M. (eds). *Emerging Trends in the Social and Behavioral Sciences*, pp. 1–16.

Kahan, D. M. (2016b). The Politically Motivated Reasoning Paradigm, Part 2: Unanswered Questions. In: Scott R. A. and Kosslyn, S. M. (eds). *Emerging Trends in the Social and Behavioral Sciences*, pp. 1–15.

Kahan, D. M. (2017). Misconceptions, Misinformation, and the Logic of Identity-protective Cognition. The Cultural Cognition Project, Working Paper No. 164, Yale Law School.

Kahan, D. M., Jenkins-Smith, H., and Braman, D. (2011). Cultural Cognition of Scientific Consensus. *Journal of Risk Research*, 14: 147–174.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., and Mandel, G. (2012). The Polarizing Impact of Science Literacy and Numeracy on Perceived Climate Change Risks. *Nature Climate Change,* 2: 732–735.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108: 480–498.

Levitsky, S. and Ziblatt, D. (2019). *How Democracies Die*. London: Penguin Books.

Levy, N. (2019). Due Deference to Denialism: Explaining Ordinary People's Rejection of Established Scientific Findings. *Synthese*, 196: 313–327.

Lewandowsky, S. and Oberauer, K. (2016). Motivated Rejection of Science. *Current Directions in Psychological Science*, 25 (4): 217–222.

Lewandowsky, S., Cook, J. and Lloyd, E. (2016). The 'Alice in Wonderland' Mechanics of the Rejection of (Climate) Science: Simulating Coherence by Conspiracism. *Synthese*, 195: 175–196.

Lewandowsky, S., Gignac, G. E., and Oberauer, K. (2013). The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. *PLoS ONE*, 8 (10), e75637.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, 37 (11): 2098–2109.

Luskin, R. C., Fishkin, J. S. and Hahn, K. S. (2007). Consensus and Polarization in Small Group Deliberations. Paper presented at the Annual Meeting of the American Political Science Association, Chicago, IL, August 30-September 2, 2007.

Lynch, M. P. (2021). Political Disagreement, Arrogance, and the Pursuit of Truth. In: Edenberg, E. and Hannon, M. (eds). *Political Epistemology*. Oxford: Oxford University Press, pp. 244–258.

Már, K. and Gastil, J. (2020). Tracing the Boundaries of Motivated Reasoning: How Deliberative Minipublics Can Improve Voter Knowledge. *Political Psychology*, 41 (1): 107–127.

Mercier, H. and Cladière, N. (2021). Does Discussion Make Crowds Any Wiser? *Cognition*. https://doi.org/10.1016/j.cognition.2021.104912

Mercier, H. and Landemore, H. (2012). Reasoning is for Arguing: Understanding the Successes and Failures of Deliberation. *Political Psychology*, 33 (2): 243–258.

Mercier, H. and Sperber, D. (2011). Why do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences*, 34 (2): 57–74.

Mercier, H. and Sperber, D. (2012). Reasoning as a Social Competence. In: Landemore, H. and Elster J. (eds). *Collective Wisdom: Principles and Mechanisms*. Cambridge: Cambridge University Press, pp. 368–392.

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2: 175–220.

Oreskes, N. (2004). The Scientific Consensus on Climate Change. *Science*, 306: 1686.

Santoro, E. and Broockman, D. E. (2022). The Promise and Pitfalls of Cross-Partisan Conversations for Reducing Affective Polarization: Evidence from Randomized Experiments. *Science Advances*, 8, eabn5515.

Sunstein, C. R. (2002). The Law of Group Polarization. *Journal of Political Philosophy*, 10: 175–195.

Talisse, R. B. (2021). Problems of Polarization. In: Edenberg, E. and Hannon, M. (eds). *Political Epistemology*. Oxford: Oxford University Press, pp. 209–225.

Iris Vidmar Jovanović*

VIRTUES AND VICES OF FICTIONAL CHARACTERS: (WHY) DO THEY MATTER FOR SPECTATORS' MORAL SENSIBILITY?

Abstract: In this paper I explore the practical significance of the question of the moral impact of narrative works of art, mainly serialized fiction, on our moral agency. I address this question by analyzing the process of moral reasoning in the spectator's artistic experience, which may justify Plato's worries regarding the possibility of moral corruption through art. I counter those by suggesting some of the benefits that engaging with immoral characters can have for spectators. My ultimate aim however is to highlight the importance of conducting more research on the topic of the (im)moral impact of works of art.

1. To Watch or Not to Watch: The Challenge of Violent Fiction

Korean fiction serial *The Squid Game* became an overnight hit, but as reviews came pouring in praising the show's artistic elements – surprising narrative twists, outstanding performances, its overall production, to mention but few – educators, psychologists and teachers from all over the world issued severe warnings regarding the negative impact of the serial on children. Apparently, within a short period of time, children began imitating the serial's violent games and their behavior towards each other became noticeably more aggressive.[1] The old Platonic worries regarding the negative impact of certain works of art and entertainment have thus reverberated once again, warning us to be cautious over the kind of works our children, and possibly more mature spectators, are exposed to.

* Department of Philosophy, University of Rijeka, ividmar@ffri.uniri.hr

1 https://edition.cnn.com/2021/10/24/entertainment/squid-game-children-netflix-wellness-cec/index.html; https://7news.com.au/lifestyle/parenting/concerns-over-netflixs-squid-game-series-raised-by-childrens-mental-health-experts-c-4333696.

The *Squid Game* might have pushed the boundaries of the kind and extent of violence shown on television a bit further, but it is by no measures a unique or isolated phenomenon. From numerous 'revenge movies' of artistic masterminds such as Tarantino, to "extreme cinema" such as the *Saw* franchise or *Antichrist*, to television works such as *Breaking Bad (*AMC 2008 – 2013*)*, The *Sopranos* (HBO 1999 – 2007) or *Sons of Anarchy* (FX 2008 – 2014), viewers have been exposed to all sorts of violence, aggression, brutality and inhumanity for decades.[2] As several prominent philosophers and media scholars argued (mostly in reference to the widespread appearance of antiheroes, i.e. bad protagonists), the worry with these shows is not only their extensive depiction of violence. Rather, the concern is that such works encourage the spectators' sympathetic responses towards the protagonists who are morally blameworthy. Thus, it is not only that the viewers might, via repeated exposure to violent and morally dubious behavior, develop a kind of apathy with respect to such instances. The problem is, because they like and/or sympathize with morally troublesome characters, their moral sensibility may be marred. As Carl Plantinga wonders, it just might be the case that contemporary screen stories "create ethical confusion and moral dumbfounding" (2018, 276). And if that is the case, Plantinga argues, the worries regarding the impact of such works on our ethical sensibility are justified.

Debating the question of the potential impact of art on individuals is usually related to what Berys Gaut (2007) calls the "causal question": the one asking if exposure to works of art tends to affect us morally – morally to improve or morally to corrupt us". A lot has been written on this issue recently, with empirical psychologists joining forces with philosophers in exploring how precisely artistic fictional works impact our moral and cognitive sensibility; whether, in other words, our art engagements can boost our skills of reasoning, our knowledge, moral capacities, empathy and prosocial skills. Unfortunately, however, the question of art's impact on its consumers remains one of the hardest to crack, with research repeatedly showing contradictory results and scientists concluding that more research is needed before we can have a definitive answer to understanding the impact of art.[3]

Regardless of such inconsistency in results, and bracketing the skeptical voices over the very possibility of exploring the question empirically,[4] two views have recently emerged on this question. What I will call the 'no impact view' summarizes the theories of those scholars who dismiss art's

2 See Frey (2013) for an analysis of extreme cinema.

3 Maibom (2020); Winner (2019).

4 McGegor (2018).

causal impact, claiming that art does not make substantial contribution to our moral or cognitive agency. Lena Wimmer and her colleagues (2022) for example argue that the results of various empirical studies failed to reveal evidence for enhanced social or moral cognition with increasing lifetime exposure to narrative fiction. Psychologist Ellen Winner defends similar conclusions. While she does claim that "the arts offer a way of understanding unavailable in other disciplines" (2019, 187) she nevertheless dismisses the causal relation between enhanced academic performance and education in arts.[5] She is also rather dismissive of the capacity of art to enhance those of our skills related to empathic understanding of others. As she sees it, "People who read a great deal of fiction also have high cognitive empathy skills. But there is no reliable evidence yet to allow us to choose whether the causal arrow flows from fiction reading to empathy, from empathy to fiction reading, or in both directions. In terms of compassionate empathy, we can conclude that people who read prosocial stories and who get transported into them are likely to behave more prosocially immediately after reading, but thus far this has been shown only for behaviors that involve little cost or sacrifice" (2019, 201). This would imply that fiction does not make any significant, long-term contribution to how we reason or behave; whatever virtues or vices we have or do not have are barely affected by fictional characters' virtuous or vicious behavior. Different conclusion is voiced by those who argue in favor of 'impact view', i.e. who claim that art can impact our cognitive and ethical agency. Echoing this view is James Young: "The recent psychological literature provides empirical support for H, the hypothesis that reading literary fiction makes people more virtuous" (2019, 105).

My intention here is not to solve these opposite views, but to provide an argument as to why the question itself matters, and why research should continue until we come up with a more conclusive answer to it. Overall, I concur with Plantinga and my aim in this paper is to accept his invitation to continue the conversation on these matters. I do so by examining some of the ways in which our moral reasoning is affected by the narration. Regardless of some of the worries that this analysis triggers, in the second part I provide a reason for not dismissing morally blurry characters from our engagements. On the whole however, I do invite more research on the topic, primarily in light of some practical consequences of the question of art's causal effect. Understanding why the question of ethical impact of art matters practically, rather than solely theoretically, is important within the society which is sensitive to the moral character of

5 As she argues, while the research shows correlation between artistic education and high achievements, it is more probable that high achievers choose artistic education, than that artistic education casually impacts their high academic scores (see Winner 2019).

its citizens, and which holds art in high esteem.[6] I consider some such practical consequences in the next part.

2. Art, Moral Education and Censorship

One of the traditional arguments in favor of the unique value that art has in our civilization rests on its alleged cultural and educational value: art makes us cognitively and morally better, we learn more about the world and other people, we develop our moral sensibility and become better equipped to deal with morally blurry issues. The claim is that through narrative engagements, we develop empathy, and in that way a lot of social injustice can be set right: racism, homophobia, mistreatment of immigrants, and the like. One of the most famous examples used to support this view is *Uncle Tom's Cabin*, which, as many indicated, did wonders in establishing anti-slavery and changing the mindset of thousands of people who came to sympathize with the slaves and called for their liberation. *Uncle Tom's Cabin* is a telling example of the power of fictional work to bring about significant moral change: as numerous critics, philosophers and literary scholars argue, its impact on abolishment of slavery can hardly be overestimated. By depicting the slaves as loving and compassionate family members, Beecher Stowe managed to invite a reconsideration of the moral status of slaves in her contemporaries, and eventually to get them to form and act upon a moral judgment that was based on their recognition of the slave's humanity.[7] Of course, the novel was not the only social force trying to achieve such an impact, but its contribution to the development of a different socio-political climate, one which eventually led to liberation of slaves, should be acknowledged.

Now, if this view is correct and art indeed can have positive impact on us, we have good reasons to insist that certain works of art – those that are particularly well equipped to help us develop our moral sensibility – be 'forced upon' citizens so as to assist them in developing their moral

6 My claim here is meant to capture the fact that engagements with narrative works of art (high and low) which will exclusively interest me here, is a wide-spread form of artistic engagements for contemporary citizens. While most of my examples pertain to serialized fiction, I presuppose the philosophical debate on the moral impact of art that has been going on since Plato. I do not presuppose any particular definition of art or fiction. I am aware however that artistic properties of a work of art may have a significant role to play when it comes to the moral or cognitive impact of a work; that however is a problem for another occasion. For some of the thoughts on the role of artistic properties in this context see McGregor (2018) and Vidmar Jovanović (2020).

7 See Carroll (2014) and Wolterstorff (2015) for discussion on Beecher Stowe's narrative techniques.

sensibility. Such works could be included on the school syllabi, available on TV or staged in national theaters.[8] In other words, the view that art can have a positive impact on people can justify subjecting art to the purposes of moral education. Not everyone thinks this would be a proper way to treat artistic works, arguing that instrumentalizing art in that way diminishes its artistic value.[9]

The view that there is a moral impact that works of art can achieve cannot ignore the possibility of the morally corruptive impact, as suggested by the *Squid Game* example. If art can impact our moral sensibility, then those works which are in any way morally problematic could have precisely such an impact: make us worse off from the moral perspective than if we had not engaged with the work. Our moral sensibility would deteriorate, in light of a work inspiring us to, for example, become more violent or less tolerant. Margaret Mitchell's *Gone with the Wind* is certainly one of the most famous examples of a work which, while held in high regard by endless generations of readers, was primarily written with the intention of inducing people not to recognize the humanity of Black people. There is a sense in which moral sensibility of readers who accepted that perspective was corrupted by the reading experience. Therefore, if works of art can have such an impact, we might have a good reason to accept the strict norms of Plato's censorship.

Practical implications on how we understand the role and value of art within the society follow also from the no impact view. If art does not influence our moral sensibility, the problem of immoral art disappears; we can engage with various works that are morally problematic without the fear of becoming corrupted. Censorship is not necessary and people do not need protection, since there is nothing dangerous that might harm them. Furthermore, we can engage with art for its own sake, since instrumentalizing art for the purposes of moral education is no longer a viable option. However, if art does not impact our morality, we lose those of its benefits that were historically related to its positive contributions to our cognition and moral development at least since Aristotle (1989) set out to refute Plato's views. In that case, it might be hard to defend art's value and the role it

8 See McGregor (2018) for a proposal along these lines with respect to works which could potentially help reduce criminal inhumanity and eradicate social harm. For some of the worries regarding his proposal, see Vidmar Jovanović 2020.

9 For example, Peter Lamarque and Stein Olsen (1994) argue that the value of literature lies in its literary achievements, not in the potential benefits it can have for human cognitive or ethical sensibility. While I do not see this as a form of instrumentalization, on the condition that works used in the process of moral education are not appreciated solely for their educational capacities, the argument against instrumentalization should keep us alert to the value of art which is distinct from its capacity to educate.

has within our society, primarily if we are not satisfied with the claim that the ultimate value of art is its capacity to provide hedonic pleasure.

3. Violence on Screen

The question of the causal impact of art began with Plato's strict rules for censorship.[10] Plato called for rejection of all works of art except for those whose form and/or content could provide positive moral benefits. His account of censorship was complex, and two worries in particular interest us here: his belief that our artistic engagements leave us emotional (perhaps even irrational) and that we identify with the characters. Various ways in which we emotionally react to fictional characters can (as I discuss below) interfere with our moral judgment of their deeds, and this might in turn make us think that certain forms of immoral behavior are appropriate. This is all the more evident in cases in which we identify with characters: although more work is needed to properly understand the notion of identification, it is not an exaggeration to say that sometimes we find ourselves reflected in certain fictional protagonists (or vice versa), where such identification can advance some of our (im)moral acts.[11] So, if Plato is right and we react emotionally and /or identify with fictional characters, their moral virtues and vices might impact our moral development.

Strong empirical support for this claim comes from L. Rowell Huesmann and his colleagues, who conducted numerous research concerning the impact of media violence on young spectators. As they argue, research consistently supports the claim that children who see violent film behave more aggressively toward each other. The underlying mechanism that explains this impact is mimesis – children imitate the actions they see, and if they are repeatedly exposed to aggression, they will start to imitate aggressive behavior – much like Plato feared they might. Huesmann repeatedly stresses this kind of causal contribution of violent media, claiming that „the evidence is already substantial that exposure to media violence is ... long-term predisposing and short-term precipitating factor" (2003, 201)[12] which contributes to development of aggressive behavior.

Elaborating further on his findings, Huesmann enlists several manners in which media violence can have negative impact on children. As he argues, short-term effects include a child's disposition to interpret a

10 Plato advances these arguments in his *Republic*.

11 Smith (1995) argues against identification and insists that moral evaluation of characters is central to our experience of art. For the relation between media and self-identity, see Cohen et.all. (2019).

12 See Huesmann et all (2003, 201); see also Bushman and Huesmann (2006).

provocation as more severe than it is because the emotional response stimulated by the previously observed violence is misattributed as being due to the provocation; consequently, child may respond more violently to an act s/he perceives as provocation, given that repeated exposure to violence may reduce the inhibiting mechanisms' capacity (such as normative beliefs about the wrongness of violence) to restrain aggression.

As Huesmann's findings indicate, violent representations do not only have short-term impact on behavior. Rather, they also impact children's cognitive economy in the long-run, in that through observational learning one acquires three structures related to social cognition: schemas about a hostile world, scripts for social problem solving that focus on aggression and normative beliefs that aggression is acceptable. Somewhat simplified, this means that a child who is repeatedly exposed to violence will come to believe that the world is a hostile place to begin with; that one can apply violence in order to fix problems in such a world and that doing so is morally acceptable. The reports on children's behavior, following their exposure to the *Squid Game*, offers reasons to believe that Huesmann's account correctly tracks the impact of art on children's reasoning. Notice that the underlying structure of how such impact takes place – i.e. how one comes to form certain beliefs and act upon them – is aligned with theoretical postulates of aesthetic cognitivism (the view that art imparts knowledge) dating as far back as Aristotle's *Poetics*, where the view was originally formulated. According to Aristotle, and as argued by contemporary advocates of aesthetic cognitivism, our engagements with fiction help us come up with certain perspectives on how the world is; they offer possible ways in which to handle such a world and they inspire normative beliefs in spectators. As Elisabeth Camp explains in her work on perspectives, this happens because perspectives structure one's way of thinking, so that the person interprets the events (in the real world) in accordance with a particular perspective. So, if a child repeatedly exposed to works such as *Sons of Anarchy* comes to believe that an individual has a right to ignore social norms, laws and rules of conduct for his own financial benefits, and is excused for killing his rivals in doing so, he or she may start to behave in such a way.[13]

Huesmann's research primarily concerns violent representations, but it is not an exaggeration to suggest that pornographic content might have the similar effect, in contributing to an individual's coming up with mis-

13 Consider, as an example of a work eliciting a certain perspective, Vaage's interpretation of antihero series as inviting a negative perspective on domestic sphere and femininity: „ Increasingly, the antihero series has emphasized how the male can live a transgressive and exciting life if only he breaks free from the home, restricted as it is by a traditionally feminine sphere portrayed as boring. (2017, 154).

representation of human sexuality. Numerous literary masterpieces were throughout the history banned or destroyed for fear of what they might cause their audiences to do.[14] Certain forms of entertainment could also have some corruptive impact: for example, the widespread presence of sympathetic and likeable, yet morally shady characters, may motivate spectators to form beliefs (and act upon those) about certain aspects of human sexuality that may be detrimental to their wellbeing. The examples I have in mind concern characters such as *Two and a Half Men*'s (CBS 2003 – 2015) Charlie Harper (Charlie Sheen), *How I Met Your Mother*'s (CBS 2005 – 2014) Barney Stinson (Neil Patrick Harris) or *Big Bang Theory*'s (CBS 2007 – 2019) Penny (Kaley Cuoco). With the repeated emphasis on their promiscuity (and, in some cases, proneness to alcohol and/or substance abuse), young spectators may get an impression that such a life style is desirable, given that these characters are loveable and their hedonic, visceral life style is for the most part presented as enjoyable. In light of such positive perspective on promiscuity, spectators might feel they are being actively encouraged to engage in such a behavior themselves.

4. Moral Reasoning in The Context of Fictional Engagement

Someone may object to the view expressed above and argue that a sort of misrepresentation of love and sexuality is the underlying premise of the comedic aspect of these shows, one which a spectator is supposed to understand if she is to enjoy the show, much like she has to be aware of the generic norms of horror story or murder mystery in order to properly respond to the movies pertaining to these genres.[15] In other words, when Barney explains cheating – "It's not cheating if you're not the one who's married. It's not cheating if her name had two adjacent vowels, and it's not cheating if she's from a different area code." – viewers laugh precisely because they recognize that such reasoning is wrong and would not accept it in nonfictional context. In other words, we know we are watching fiction and, given that we are familiar with the norms of fiction watching, we suspend our believes and make no inference from fictional to non-fictional world. We enjoy the show because we know it is fiction.

14 For an overview, see Ladenson (2006).

15 The underlying assumption here is that the spectators take a fictive stance (Lamarque and Olsen 1994) toward the content of representation and block inferences from the fictional world into nonfictional reality. Roughly, this implies that the spectator is aware, going into the story, that she should not take it as representing how things actually are.

This is an interesting proposal and it explains much of our behavior as we engage with fiction: we do not call the police when someone is killed and we do not shout to warn characters when they are in danger. However, it is not necessary so that, even if spectators acknowledge the fictional content and consider certain character traits and moral actions as pertaining to the generic norms (rather than to the over-arching moral lesson the work is to bestow upon them), that they do not incorporate beliefs based on such representations into their conceptual framework.[16] Someone indeed might come to believe that promiscuity is a desirable life style and that intimate, monogamous relationship deprives one of adventurous sexual experiences which can make life more enjoyable. In addition, considering that these shows are sitcoms, which the audience attends to with their 'moral radar' shut down, and in pursuit of entertainment, it is highly unlikely they give any serious thought to evaluating these characters' life choices. As Margrethe Bruun Vaage argues (2017), quoting a lot of empirical research, when it comes to fiction, viewers deliberately avoid making moral judgments in order to maximize their enjoyment, processes known as moral disengagement and fictional relief. Consequently, it might be the case that spectators fail to rebuke certain behaviors which they would rebuke in nonfictional circumstances. That in itself does not mean they would necessarily develop those same habits or imitate immoral life style. But more research is needed to understand the circumstances when the spectators would or would not accept certain kinds of behavior they laugh at in the fictional world as acceptable in nonfiction.

Such worries may be even more prominent with respect to the genre that Plantinga is worried about, drama series. Given the omnipresence of anti-heroes in popular culture (just consider Tony Soprano (James Gandolfini), Jax Teller (Charlie Hunnam), Dexter Morgan (Michael C. Hall) or Walter White (Bryan Cranston)), the worry about moral corruption induced by media may not be without foundation. As Vaage (2017) shows, various narrative strategies commonly employed by these shows make it more likely for the viewers to bracket their moral judgment of certain individuals and to side with them, cheer for them and consider them morally superior to others. What she has in mind are strategies like the one used in the case of Walter White: because the pilot portrays him as an underdog, unfulfilled genius repeatedly tormented by his wife Skyler (Anna Gunn),

16 Consider for example the research on the phenomenon of fictional persuasion, which suggest that viewers do not necessarily consider the reliability of the representation when forming beliefs about something (Steglich-Petersen, 2017; Sulliavn Bissett and Bortoloti (2017). For additional worries regarding reasoning to what is true in the fiction and what can viewers reasonably export from it see Matravers (2014), Abell (2020), Vidmar Jovanović (forthcoming).

ignored by his son, mocked by his students and dismissed as insignificant, viewers are quick to develop sympathy for him even before they learn of his terminal disease. Such sympathy excuses all of his crimes and misdemeanors, none of which would likely be excused in nonfictional context (or with the character who is less tormented by the other characters). Such moral blindness on the part of the viewers is evident not only in their repeated support for Walter (as evident in numerous fan sites), but, more worryingly, in the so called Internet hate-fest towards Skyler. As the actress herself states, not only did some of the fans openly express their desire to kill the character, seeing her as the "irredeemable bitch", but the actress received death threats due to the antipathy they felt for Skyler.[17]

Another narrative strategy invoked to maintain viewers' sympathy for the bad characters is partiality: as Vaage shows, because the protagonist is the focus of the story, viewers become familiar with him, and they even start considering him as a 'friend', as someone close to them. Furthermore, because the fictional world is organized around the protagonist, we know more about him than about other characters, some of whom may be, objectively, morally preferable to the protagonist (as Skyler is, compared to Walter). However, our focus on and interest in the protagonist makes it hard for us to make an impartial judgment: we resent Skyler for mistreating Walter, and we go on resenting her regardless of Walter's many transgressions. Similarly, we forgive Dexter for killing a serial child molester because we know the horrors he went through in his childhood, we know 'what made him the way he is', as he puts it. But we do not know, because the fictional world does not reveal it to us, what made his victims what they are; consequently, we cheer for him and we feel a relief when he takes down a killer or a rapist, often failing to consider the fact that he himself is no different. Even if we do acknowledge his moral failings, we are nevertheless happy when other characters take the fall for him. We sacrifice, in other words, morality of the character for the aesthetic enjoyment the show provides in light of that precise immorality.[18] In some cases of

17 See Vaage (2017, ch.6). Interestingly, Gunn attributed such reactions to widespread misogyny; Vaage on the other hand argues that antipathy for Gunn is motivated and sustained by the narration itself: „she is holding her husband back from what the audience perceive as enjoyable transgression" (151) and is repeatedly shown as obsessive, cynical, dull, hypocritical. For more interesting examples of hate directed at fictional characters, see https://screen-queens.com/2014/08/18/the-real-hated-housewives-of-tv/

18 Such reasoning is partly explained by our narrative (or work) desires: we want the work to go on and it can only go on if the protagonist continues to do what he does. This argument has a considerable force, but what matters here is the fact that not many viewers express moral regret when, for example, Doax is mistaken for the Bay Harbor Butcher, thus making it possible for Dexter to go on with his killings. Narrative desires also do not explain the hatred for Skyler. I'll come back to this below.

course the immorality can get a moral spin: when an antihero commits murders we think are justified, as we usually do in the case of Dexter's killings, spectator's satisfaction is not only aesthetic, it also has deep moral roots.[19] On the whole however, not every morally corrupted protagonist redeems himself through acts of kindness or benevolence and there are many who remain appealing to the audience even when they miss all chances of redemption (as I would suggest Don Draper (Jon Hamm) or Jax Teller do).

5. Now What?

There are of course additional factors to consider before we conclude that all violent fiction is potentially morally corruptive: factors relating to social status and viewing preferences of the viewers, social background of the child, circumstances in which one experiences a given representation, relation between artistic properties and moral dimension, and the like. As I argued elsewhere (2023), research into the moral impact of art often fails to consider cognitive and moral agency of spectators: perhaps the reason why people behave badly or aggressively is not only due to their viewing experiences and accumulation of violent screen stories, but also because they simply do not care about the requirements of morality, or lack the psychological strength to change their behavior, overcome prejudice or modify their behavior. In addition, it is often the case that immoral fiction contains precious epistemic lessons, in light of which the work can provide insightful moral education.[20] For reasons of space I cannot develop this argument here, but in many cases the morally murky characters themselves express their dissatisfaction with their life choices and can therefore serve as a sort of a cautionary tale for those who think of imitating them.

More importantly, fictional worlds are psychologically complex and that complexity enables the viewers to understand some of the causes of immoral behavior, as well as some of the downfalls of such actions – this is, after all, one of the reasons why the *Squid Game* represents a powerful criticism of contemporary society, making the show more than an exer-

19 See Vaage (2017) for the appeal of immoral characters.

20 While McGregor (2018) does not discuss antihero stories or immoral characters, his theory (narrative justice, as he calls it) is, I would suggest, an insightful example of how fictional work which depicts morally troublesome actions ends up delivering epistemic good: as he argues, criminological inquiry (as the one found in narratives) identifies the cause or causes of particular crime, and understanding these causes can contribute to reduction of crime.

cise in brutality. Another telling example is the character of the fictional detective Andy Sipowicz (Dennis Franz), whose positive moral qualities (dedication to his work, bravery, willingness to sacrifice himself for those in need, loyalty) barely trump his moral flaws. His drinking aside, Andy repeatedly goes out of line in his racist and homophobe outbursts, insulting and mistreating people of color and gay people. However, the show (*NYPD Blue,* NBC 1993 – 2005) does a wonderful job in putting an educational spin on such immoral actions: not only does it enable the viewers to recognize how racism manifests itself (where such behavior may not always be recognized as racist by those who act in such ways) and how it affects one's private and professional domain, but also in trying to modify such behavior. Because Andy is repeatedly challenged and reprimanded for his racial outburst by his colleagues, partners, and people of color, he manages to modify his behavior. Thus, for all of his immoral qualities, he sets a positive moral example. The interesting question then becomes: will a racist viewer modify her behavior in light of positive moral changes exemplified by Sipowicz, or will she come to resent Andy his moral conversion? And that of course is what this discussion is all about: the question of moral impact of art is never just a question of the moral character of a work or those who inhabit its fictional world. Many other factors need to be considered if we are to understand moral impact of art.

As the example of Anna Gunn mentioned above shows, many of the *Breaking Bad* spectators remain united in their hatred for her character, and their animosity was often expressed by fans on social media. Emily Nussbaum's (2014) account of bad fans – roughly, those who take pleasure in supporting morally problematic characters – shows that open support for morally shady characters is a widespread phenomenon which is, in itself, a source of pleasure. The phenomenon of fandom shows how strongly people unite in their aesthetic preferences; as numerous media studies reveal, fans often consider themselves as part of a wider community in light of their artistic choices, where such consideration is an important aspect of their identity and self-understanding. As Rebecca Williams (2015) argues, such identification, and commitments to both, a particular work and the wider community of fans, often functions as a kind of ontological stability which safeguards one from existential worries and uncertainties. If this is so – and numerous testimonies written by fans on various platforms of social media give us strong reason to think that it is – then, I suggest, we need to reconsider the implications of fans gathering around morally problematic shows, celebrating their immoral characters.

As Vaage points out, with respect to most antihero fictions, and most fans, at certain point in the narrative, it becomes obvious that the pro-

tagonist has gone too far in his moral transgressions, at which point the viewers' sympathies are no longer on his side. Such 'reality checks' as she explains, "remind [the viewer] of the consequences of the [protagonist's] actions, had they been actions in the real world, not just a fictional one." (2017, 25). It is at this point that the spectator snaps out of fictional relief and is no longer disregarding the immorality of the character: the narrative or work desire for the show to continue no longer comes at the price of disregarding the immorality of the protagonist. Interestingly, Vaage considers this to be a response of a properly engaged viewer, and believes such a viewer would not be writing hate messages to Anna Gunn, or be a bad fan. In other words, someone who appreciates the enjoyment of the antihero's immorality as pertaining only to the fictional world, but is fully aware of the unacceptability of such immorality in any other contexts, is a morally mature person who will not succumb to the appeal of immoral portrayals in nonfictional context. That certainly seems right. However, it does not explain all of our worries. Rather, it only generates more questions: does the phenomenon of the bad fan show that not everyone manages to resist the allure of immorality, just like in the Andy Sipowicz example, not everyone manages to pick up moral lessons? This seems right, but it brings us back to Plato's worries: people can be morally corrupted through fiction. What we need to do is find the relevant factors that explain when and why that happens. On the other hand, perhaps the phenomenon of bad fandom represents an aspect of human behavior that we have yet to properly characterize and understand. But to do so, more research is needed.

To conclude. To understand not only our moral and artistic engagements with works of narrative art but also the very nature of our moral reasoning, we need to see what makes a difference between those spectators who fail to respond to reality checks and those who do not, or those who respond to the moral lessons of a work and those who do not. Again, more research is needed to do so. I imagine such research should also address the relevance that our artistic engagements have for us, and should shed some light on the process of identification with certain characters, as well as the manner in which people balance moral assessment of works and the hedonic enjoyment from works. Furthermore, we need to explore the relation between ethical and aesthetic dimension of a work, and the phenomenology of the spectators' experience itself. My aim here was to show why such research is important.[21]

21 This work has been supported by the Croatian Science Foundation under the project UIP-2020–02–1309.

Bibliography

Abell C. (2020), *Fiction. A Philosophical Analysis,* Oxford: Oxford University Press

Aristotle (1989), "Poetics", in *A New Aristotle Reader*, ed. J. L. Ackrill, Princeton University Press

Bushman B. J. & L. Rowell Huesmann (2006), "Short-term and Long-term Effects of Violent Media on Aggression in Children and Adults", *Arch Pediatr ADolesc Med*, 160, 348–352

Camp E. (2013), "Slurring Perspectives", *Analytic Philosophy* 54, 3, 330–349

Carroll N. (2014), "Fiction, Film, and Family" in *Cine-Ethics: Ethical Dimension of Film Theory, Practice and Spectatorship*, eds. Jinhee Choi and Mattias Frey, Routledge, 43–56

Cohen J., Appel M. & M.D. Slater, "Media, Identity, and the Self", in *Media Effects: Advances in Theory and Research*, eds. Oliver, M.B., A.A. Raney & J. Bryant, Routledge, 179–194

Frey M. (2014), "The Ethics of Extreme Cinema", in *Cine-Ethics. Ethical Dimensions of Film Theory, Practice, and Spectatorship*, eds. J. Choi and M. Frey, Rouledge, 145–162

Gaut B. (2007), *Art, Emotion and Ethics*, Oxford, Oxford University Press

Huesmann L.R et all, (2003), "Longitudinal Relations Between Children's Exposure to TV Violence and Their Aggressive and Violent Behavior in Young Adulthood: 1977–1992", *Developmental Psychology*, 39, 2, 201–221

Ladenson E. (2007), *Dirt for Art's Sake, Books on Trial from "Madame Bovary" to "Lolita"*, Cornell University Press

Lamarque P. & S.H. Olsen (1994), *Truth, Fiction and Literature: A Philosophical Perspective*, Oxford University Press

Maibom H. (2020), *Empathy*, Routledge

Matravers, D. (2014), *Fiction and Narrative*, Oxford University Press

McGregor R. (2018), Narrative Justice, Rowman and Littlefield

Nussbaum E. (2014), "The Great Divide: Norman Lear, Archie Bunker, and the Rise of the Bad Fan", *The New Yorker*, April

Plantinga C. (2018), *Screen Stories: Emotion and the Ethics of Engagement,* Oxford University Press

Plato (1997), *Complete Works*, ed. John M. Cooper, Cambridge: Hackett Publishing Company

Smith M. (1995), *Engaging Characters. Fiction, Emotion, and the Cinema*, Oxford University Press

Steglich-Petersen A. (2017), "Fictional Persuasion and the Nature of Belief", in *Art and Belief*, eds. Sullivan-Bissett E., Bradley H. and P. Noordhof, Oxford University Press, 174–193

Sullivan-Bissett E. & L. Bortolotti (2017), "Fictional Persuasion, Transparency and the Aim of Belief", in *Art and Belief*, eds. Sullivan-Bissett E., Bradley H. and P. Noordhof, Oxford University Press, 153–173

Vaage M. (2016), *The Antihero in American Television*, Rutledge

Vidmar Jovanović I. (forthcoming), "Perspectivism, cognitivism and the ethical evaluation of art", *The Journal of Aesthetic Education*

Vidmar Jovanović, I. (2020), "Becoming Sensible: Thoughts on Rafe McGregor's Narrative Justice", *The Journal of Aesthetic Education*, 54, 4; 48–61

Vidmar Jovanović, I. (2023) "Art and Moral Motivation: Why Art Fails to Move Us", *Journal of Aesthetic Education,* 57 (1): 19–35

Williams R. (2015), *Post-object Fandom. Television, Identity and Self-Narrative*, Bloomsburry

Wimmer L., Currie G., Friend S. & F.J. Ferguson (2022), "The Effects of Reading Narrative Fiction on Social and Moral Cognition: Two Experiments Following a Multi-Method Approach", *Scientific Study of Literature*, 11,2, 223–265

Winner, E. (2019), *How Art Works. A Psychological Exploration*. Oxford University Press

Wolterstorff N. (2015), *Art Rethought. The Social Practices of Art*, Oxford University Press

Young, J. (2019), "Literary Fiction and the Cultivation of Value", in *Narrative Art, Knowledge and Ethics*, edited by Vidmar Jovanović I., 87–107. Faculty of Humanities and Social Sciences Rijeka

Miljana Milojević

ONE HEALTH, EXTENDED HEALTH, AND COVID-19[1]

Abstract: The aim of this paper is to critically assess the One Health approach in medical sciences and to contrast it with the alternative Extended Health approach. The mentioned assessment is going to be conducted in line with different methodological and ontological criteria, in order to evaluate whether the One Health approach and which of its variants, can be proven as a fruitful and productive paradigm in medical sciences. After the proposed evaluation, an argument for the reconciliation of the particular kind of non-radical One Health approach with the Extended Health one will be offered. To illustrate the utility of the defended position, an analysis of the case of the COVID-19 pandemic will be given and it will be shown how its management could have been greatly improved by joining these alternative new approaches to medicine.

Keywords: One Health, Extended Health, COVID-19, Anti-Individualism

1. One Health

The general approach to health and disease in the 21st century is starting to seriously take into account various factors external to primary affected organisms in order to identify, treat, and prevent various illnesses. External factors were, of course, trivially always considered when it comes to various diseases, especially those caused by pathogens. Viruses, bacteria, fungi, and parasites, but also chemical and mechanical injuries, constitute the majority of causes of bodily harm and disease and they certainly originate outside of the affected organism. Except for genetic and

* Department of Philosophy, Faculty of Philosophy, University of Belgrade, mrmiloje@f.bg.ac.rs

1 This research was financially supported by the Ministry of science, technological development and innovation of the Republic of Serbia as part of the funding of scientific research at the University of Belgrade – Faculty of Philosophy (contract number 451-03–47/2023–01/ 200163).

immune disorders, diseases are most often caused by environmental and biological factors that are external to the patient. Thus, when we say that the 21st century medical approach started to seriously look into the factors external to the subject of disease we do not mean that medicine is starting to look at the mere external immediate causes of disease. Rather, we are pointing to a certain methodological shift towards a more holistic approach in medical sciences that tries to meaningfully identify and track complex dynamics of the human and animal populations together with their changing environments, and with respect to the emergence of new pathogens, paths of transmission and similar. This holistic paradigm is now known under the name "One Health".

One Health is advocated in some form by all major health organizations today, but its main tenets were recognized a long time ago. The origin of the insight that human and animal health are complexly intertwined with their environment can be tracked to ancient times, and holistic methodology in medical sciences started to be explicitly advocated in the 1800s. It was Rudolf Virchow, a German pathologist, who was inspired by the works of Louis Pasteur and Robert Koch and his own investigations into *Trichinella spiralis*, that found deeper connections between human and animal health and coined the term "zoonosis" to designate an infectious disease passed between humans and animals (Virchow 1859). The discovery of these tight linkages between animal and human health prompted subsequent considerations of the connections between veterinary and human medicine. Nevertheless, it was not before Calvin Schwabe and his 1964 *Veterinary Medicine and Human Health*, that this paradigm of uniting the treatment of human and animal health in a single medical science got its own name and program. Namely, Schwabe coined the term "One Medicine" and called for a unified medicine and a collaborative approach between practitioners of veterinary and human medicine and epidemiology in order to effectively prevent and treat zoonoses.

Although the concept of One Medicine existed in a fairly developed form since the 1960s several decades needed to pass in order to make its main principles recommended as an overall approach for understanding health and, in particular, as providing general framework for controlling infectious diseases. In 2008, One Medicine, now termed "One Health", became a global reality when International Ministerial Conference on Avian and Pandemic Influenza in Sharm el-Sheikh, Egypt was held, after which the One Health strategy was released in a form of a document named "Contributing to One World, One Health™ – A Strategic Framework for

Reducing Risks of Infectious Diseases at the Animal-Human-Ecosystems Interface". The document was drafted by the experts of FAO (The Food and Agriculture Organization), WHO (The World Health Organisation), and WOAH (The World Organization for Animal Health) in collaboration with UNICEF, the World Band, and UNSIC (United Nations System Influenza Coordination), and marks an important point in time in which multiple global health agencies pledge to dedicate themselves to a unifying holistic approach in preventing and fighting disease.

We can find the newest definition of "One Health" in the publication of the One Health High-Level Expert Panel (Adisasmito, Almuhairi, Behravesh, Bilivogui, Bukachi, et al. 2022), in which they claim:

> "One Health is an integrated, unifying approach that aims to sustainably balance and optimize the health of people, animals and ecosystems. It recognizes that the health of humans, domestic and wild animals, plants, and the wider environment (including ecosystems) are closely linked and interdependent.
>
> The approach mobilizes multiple sectors, disciplines and communities at varying levels of society to work together to foster well-being and tackle threats to health and ecosystems, while addressing the collective need for clean water, energy and air, safe and nutritious food, taking action on climate changes and contributing to sustainable development."

A similar determination of One Health can be found on the CDC website, where it is said that:

> "One Health is a collaborative, multisectoral, and transdisciplinary approach — working at the local, regional, national, and global levels — with the goal of achieving optimal health outcomes recognizing the interconnection between people, animals, plants, and their shared environment." (CDC website)

On the WHO website, we find a time-relevant remark about the importance of One Health approach for the management of the COVID-19 pandemic:

> "'One Health' is an integrated, unifying approach to balance and optimize the health of people, animals and the environment. It is particularly important to prevent, predict, detect, and respond to global health threats such as the COVID-19 pandemic." (WHO website, https://www.who.int/news-room/questions-and-answers/item/one-health)

Putting forth the One Health approach in the occurrence of a global pandemic is no surprise, especially if we have in mind the plausible zoonotic nature of COVID-19. Given that the discovery of zoonosis was one of the main originators of the One Health approach, most of the efforts of

One Health advocates are aimed at preventing and controlling diseases of zoonotic origin[2]. Nevertheless, special attention to zoonoses of One Health supporters is not only to be found in its past and its origins but also in recognizing the fact that the majority of emerging diseases in the past 50 years were of zoonotic origin. Some of them are HIV (AIDS), Hemorrhagic fever from Hantavirus, Lassa fever, Marburg fever, Lyme disease, Rift Valley fever, Ebola, Nipah disease, West Nile virus, Spongiform bovine encephalopathy, Avian influenza, Zika Gastroenteritis, and Monkeypox. According to the One Health approach, the emergence of new zoonotic diseases in turn asks for a holistic approach in which anything from climate changes, habitat changes, social circumstances and changing ways of living, availability of clean water, etc. has to be taken into account to manage – prevent and control, the emergence of new zoonoses.

2. Philosophy and One Health

Given the rising importance of the One Health approach in medical sciences and health management it is not surprising that a number of philosophical articles appeared that analyze and critically assess this new approach. In this paper, we are going to overview several arguments against radical versions of the One Health approach and try to distill the lessons they teach us. We will also try to make our own distinctions which should hopefully further the discussion.

Sironi et al. (2022) differentiate between two main versions of the One Health approach – the Prudential One Health Approach (POHA) and the Radical One Health Approach (ROHA). The difference between these two approaches is spelled out in terms of their subjects of attention. Namely, POHA is centered on human well-being, while ROHA "considers the overall balance of the living eco-system and the environment from a broader perspective than the human one" (Sironi et al. 2022). Thus, POHA is instrumental to human health, so we are not healing the planet or es-

2 Here it is said "zoonotic origin" as many of infectious diseases start as zoonoses which are transmissible inter-species, but afterwards a pathogen mutates and adapts in such a way that it affects only one species after the mutations. Example of such cause of disease is, for instance, HIV. On the other hand, we have "full-fledged" zoonoses such as rabies, West Nile virus, etc. As for COVID-19 as a disease caused by one of the emerging coronaviruses (including SARS and MERS), there is much supporting evidence that it is a case of zoonosis, though some authors demand that COVID-19 should be classified as an "emerging infectious disease (EID) of probable animal origin" (Haider, Rothman-Ostrow et al. 2020).

tablishing a balance between species for its own sake, but to improve the quality of life of our own species and to eradicate or manage illnesses that affect humankind. On the other hand, ROHA can leave humans behind to perish if it turns out that larger ecosystems suffer from our presence. They can both be seen as valid standpoints, but with very different ethical assumptions behind them, and very different practical and epistemological implications in front of them.

If we look at the definitions of the One Health approach, we can easily observe that they do not incorporate the explicit focus on the health of one group or the other, thus they are usually compatible with both POHA and ROHA interpretation of the One Health approach. If One Health is seen as an approach that originated and is still a part of human medicine, then it can easily be read as POHA. On the other hand, by not singling out human species in its manifesto and major definitions One Health can be also interpreted as ROHA. The authors of the paper rightly emphasize the importance of this distinction as it has vast implications on both scientific approaches in health management as well as on policy making. Also, the two versions of the One Health approach have gravely different ethical and epistemological consequences – while the application of POHA would certainly improve human condition, and it would be easier to implement, ROHA requires a radical change in our ethics and faces great epistemological challenges and ethical dilemmas.

If we look at the current practices that are put under the name of One Health, like the management of the COVID-19 health crisis, we will find that they are almost exclusively in line with POHA. Namely, in a holistic treatment of the pandemic we can see that different environmental factors are identified, but they are put into an equation with the distinct anthropocentric perspective. Climate change as an environmental factor is seen as having a causative effect on the changing habitat of some animal species, in this instance bats. Nevertheless, this causative effect, namely the migration of bats, is identified only because the bats were moving closer to human habitats and the growing urbanization is destroying existing buffer zones between the habitats of these two species. Thus, POHA "calls for an important "broadening" of the factors considered without, however, a real change in perspective, method and purpose of knowledge relating to health, such as to configure a radical epistemological shift" (Sironi et al. 2022). On the other hand, ROHA should not focus only on those effects that have a direct bearing on us as a species but should change and broaden the very perspective on how we see health and interconnections between species and their environment. For instance, an extinction of a

species, any species, even human, due to a cyclical event such as climate change can be seen as a natural phenomenon, one which is not to be interfered with.

The difference between POHA and ROHA can also be seen as a difference in an adopted ontology – the ontology of health and relations between species. Namely, adoption of ROHA could be a consequence of adopting a global instead of individualistic ontology of health. It is a question without a clear answer whether we should commit ourselves to a single entity such as the health of a global ecosystem, or we should still focus on individual entities that constitute this ecosystem. The non-existence of the clear answer to this question relies on the normative nature of the notion of "health". Namely, there is an underlying ethical question about what should be healthy, or what should survive – is it a cell, an animal, a species, human race, or an Earth's ecosystem? Also, the interdependence of different organisms and species blurs the boundaries of the proper subject of health, and it is questionable can we even speak of the health of an individual without making a reference to other individuals, including those from other species. Taking the interdependence of human organisms and their microbiota into consideration, for instance, illustrates this point, as it is not clear should we talk about the health of a human or a health of a human+microbiota system.

Thus, evaluation of POHA vs. ROHA can be conducted according to various criteria: ethical, epistemological, methodological, ontological, etc. Nevertheless, in evaluating POHA and ROHA versions of the One Health approach Sironi et. al take a practical stance. The main criterium for deciding between anthropocentric individualistic POHA and non-anthropocentric holistic ROHA is realistic policy and decision making "in an attempt to preserve the planet" (Sironi et. al 2022). Thus, although the more eco-centric view is to be ethically preferred in their opinion, where moral agency is attributed to other species and anthropocentrism is mitigated, POHA and its inter-connectionism that presupposes the individuality of entities seems as a more realistic option. The shift in perspective, which does not have to go all the way to the extreme ends of the ROHA, is welcomed and a number of authors call for the environmental health which is not merely instrumental to the human health (Lysaght et al. 2017). It is pointed out that our efforts must take into account inextricable connections between animals and humans and that One Health has to take an ethical stance which strives to improve the health of humans, animals, and whole ecosystems (Capps and Lederman 2015). It is clear that such an approach prefers practicality and ethics to ontology or metaphysics.

2.1 Assessing the ontological implications of One Health

In their 2018 paper Morar and Skorburg focus on ontological issues surrounding the One Health approach. Sironi et al. (2022) notice that POHA and ROHA would have different ontological implications (we would say conversely that different ontologies can imply different One Health approaches), but do not analyze them in detail. On the other hand, Morar and Skorburg take One Health and similar approaches to be challenging the dominant view that an individual organism is a bearer of health and disease states. We should make a remark here. While Sironi et al. make a difference between POHA and ROHA, by which they undoubtably read at least one version of One Health as endorsing individualistic ontology of health, namely POHA, which advocates the health of human individuals as a function of broader systems; Morar and Skorburg seem to read One Health as adopting anti-individualistic ontology in general when they say that it "proposes a holistic conception of health and disease that extends beyond the traditional individual in order to account for the intricate links between humans, wildlife, and environmental health" (Morar and Skorburg 2018: 351). Thus, Morar and Skorburg's reading of ontological commitments of One Health can be seen as too strong. It can be said that they are more in line with commitments of what was called ROHA, but it is also important to notice that even ROHA was not defined through its ontological commitments but through its priorities and preferences of methodological and ethical kind. In other words, ROHA would follow from adopting global anti-individualistic ontology of health, as we have already noticed, but the converse is not true. Namely, the radicalism of ROHA is in shifting the priority which was on human health to health of other/all species, and not in its radical holistic ontology. Thus, ROHA can still adopt the individualistic ontology just with a shift of interests and ethics to different kind of individuals. Nevertheless, there is, certainly, a space for anti-individualistic ROHA, especially if we take into account all those requests to take care of "the health of the planet", and to introduce balance between species. Thus, we can call this approach OROHA or Ontologically Radical One Health Approach, one which asks for anti-individualistic ontology which transcends the health of individuals.

The ontological issues at hand and OROHA as an approach have to be critically assessed. First, it is clear that One Health as a single approach does not clearly specify its ontological commitments. Second, we can define versions of One Health which have specific ontological commitments, like OROHA that employ extremely extended ontology of health. Third, it is prudent to assume that at least some advocates of One Health implicitly endorse something like OROHA when defending claims about the planet's

health and the overall mutually beneficial connections between species. Thus, even if only some of the One Health variants accept that health can be ascribed to large, weakly connected collectives, then the plausibility of such implications needs to be analyzed.

Morar and Skorburg start from naturalistic assumptions advocated by Boorse (1975, 1997) according to whom statements about diseases are value-neutral, and a matter of natural science. Without challenging the naturalism about health and disease they continue to evaluate the statements and implications about bearers of these states in different approaches to health and medicine. They conclude that in the naturalist camp predominant view is that the relevant individual that bears the medically relevant properties in question is a singular organism.

By briefly examining the assumptions of the opposite side – those of normativist approaches to health and disease – they come to similar conclusions. The normativists unlike naturalists do not focus only on the physiological states, but instead emphasize that for making judgments about health and illness we have to take into account "the ways in which a patient experiences (or not) this condition as something to be avoided, as an illness, along with the social norms that carve out her lived experience" (Morar and Skorburg 2018, cf. Ereshefsky 2009). If certain physiological dysfunctions do not lead to unfavorable subjective, nor social, conditions then it is at least unclear whether such a state should be treated as a disease or an illness. Also, there are the opposite cases where there is no existing physiological dysfunction, meaning all bodily systems are performing their proper functions, but a certain state or behavior is treated as a disease based on social perception and norms (one such case is homosexuality which was treated as a disease until recently, or until the social norms changed). It is an interesting question, but not the one to be dealt in detail here, is how starting from a normative, usually social, identification of an illness, still often asks for a naturalistic explanation of its origin. Thus, seeing homosexualism as an illness, or advocating the inferiority of certain ethnic groups, was almost always followed by series of experiments intending to "prove" such claims empirically by identifying biological defficiencies or malfunctions. So, we can say that even if the normativist approach identifies and defines disease and illness from an externalist social and normative dimension it still seeks explanations on the physiological and biological level. Also, it should be kept clear that we don't see normativism as a platform for advocating racism or bigotry, as it has also clearly helped to see different physiological deficiencies as being "normal", healthy or non-diseased. Nevertheless, it should be noted that normativist views on health are as "good" as our social norms are. As the health terms

usually bear normative connotations in natural language, and we associate need for avoidance, disgust, fear, etc. with "disease", "illness", "sick", etc. we have to be careful in ascribing those attributes.

To get back to the main topic of this article – despite the clear inclusion of external factors, such as social norms, into the individuation process of a disease, illness, or health, normativists still adopt the individualistic organism-centered stance toward the bearers of health and disease states. To corroborate their claims Morar and Skorburg cite Engelhardt (1975) and Goosens (1980) who purport that the appropriate subjects of health and disease ascriptions are individuals whenever such subjects exist.

After surveying the standard individualistic approaches to bearers of health and disease ascriptions, the authors turn to possible anti-individualistic hypotheses. First, they review the importance of the discovery of the symbiotic relationship found in humans and their microbiota, as well as "transactive goal dynamics" framework in psychology advocated by Gráinne Fitzsimmons and colleagues[3], to show cases in which individualistic health models are challenged. After analyzing these potential cases of extended or collective bearers of health and disease ascriptions, Morar and Skorburg turn to the One Health approach as a possible framework for an anti-individualistic treatment of certain health/disease states.

It is safe to say that Morar and Skorburg read the One Health approach in the OROHA way. They sharply contrast the One Health ontology with the individualistic ontology of more conservative approaches. But as we have seen it is only the most radical OROHA, and not POHA, or even ROHA without further specifications, which asks us for a radical ontological revision. For simplicity we can call the latter two approaches "OPOHA" or "Ontologically Prudential One Health Approach", or the One Health approach which adopts individualistic ontology. In context of the debate about the 4E approaches in cognitive science, OROHA can be connected by analogy to the Extended approaches to cognition which see the bodily and environmental factors as constitutive of cognitive processes, while OPOHA, can be connected by analogy to the Embedded approaches which see environmental factors as highly influential on cognitive processes and crucial in understanding their nature and dynamics, but not

3 The framework offered by Fitzsimmons et al. is built to deal with mechanisms of "self-regulation" in a novel way. It is well known that a number of health issues directly depend on the ability to self-regulate such as obesity or addiction, but this ability was usually conceptualized as an individual's capability connected to the ability to delay gratification. On the other hand Fitzsimmons et al. investigate how close relationships influence goals and achivement of those goals of the partners and claim that "self-regulatory systems become inextricably linked, part of a complex and messy web of interdependence" (Fitzsimmons et al. 2015).

at the same time as constitutive of them. In that way, OPOHA can also be seen as sympathetic to normative approaches to health and disease in which both physical and sociological environment are seen as necessary for individuating the health/disease states, but on the other hand not as a constitutive of those states. The imbalances in nature, for instance, could be seen as descriptors of diseased states in certain species, but not at the same time as (partial) bearers of those states themselves. OROHA on the other hand asks us to ascribe the relevant health states to ecosystems and potentially to all living beings making them, as a collective, a suitable bearer of health/disease states.

2.2 Problems for OROHA

Morar and Skoburg frame the debate in terms of the narrowness of individualistic approaches to health and the excessive wideness of the One Health approach. Interpreting One Health as OROHA which asks for entities that span biological, sociological and economical domain they argue that such an approach would be far too permissive. They too make a comparison between the OROHA and the Extended Cognition approach and offer argument against the former analogous to the one found in the literature on Extended Cognition (ExCog). Namely, they refer to the arguments for coupling-constitution fallacy and cognitive bloat.

We should briefly get acquainted with what is claimed by ExCog. Clark and Chalmers now famously write in their 1998 paper "The Extended Mind" that:

> "if, as we confront some task, a part of the world, functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head!" (1998: 8)

By claiming this they, in fact, argue in favor of a non-chauvinistic extended ontology of cognitive processes, according to which there are no *a priori* reasons to believe or claim that cognition must take place solely within the boundaries of the skull or in a neural matter of a certain subject. The argument is based on functional assumptions about the mental and the cognitive. If a cognitive process or a mental state is identified by suitable functional roles there are no conceptual or theoretical limitations on the realizers or the location of such states and processes. Thus, at least hypothetically, extended, or partially externally realized cognitive processes and mental states, can exist. Examples of extended mental states could be, for instance, dispositional beliefs stored in a notebook of an Alzhei-

mer patient; and those of extended cognitive processes could be various manipulations of external structures in order to facilitate epistemic tasks, such as the spatial manipulation of shapes in a Tetris game by physical manipulation of appropriate buttons in order to offload the task of mental rotation of shapes onto the console itself.

Adams and Aizawa (2001; 2008) accuse proponents of ExCog of making a "coupling-constitution fallacy". They analyze a number of examples of extended processes found in the literature and claim that advocates of such examples make "a long description of the causal connections between the brain and environment followed by *the move* to the view that these causal loops constitute part of the cognitive process" (2008: 96). And *the move* is constituted by observation of causal dependencies between the environment and the cognizer and the conclusion that they jointly constitute extended cognitive processes when it is only shown that they are only causally connected and not coupled in a way that would constitute a new extended entity (Adams and Aizawa 2008: 91). One such example is found in Wilson (2004):

> "We solve the problem by continually looking back to the board and trying to figure out sequences of moves that will get us closer to our goal, all the time exploiting the structure of the environment through continual interaction with it. We look, we think, we move. But the thinking, the cognitive part of solving the problem, is not squirreled away inside us, wedged between the looking and the moving, but developed and made possible through these interactions with the board." (Wilson 2004: 194; Adams & Aizawa 2008: 93)

Accusations of coupling-constitution fallacy lead to the accusations of producing cognitive bloat. Namely, by postulating constitutional claims based on only causal dependencies advocates of ExCog risk to overextend the constitutional base of cognition rendering the very notion of cognition meaningless. The critics ask what stops us from considering our phones, our laptops, or even the internet as parts of our cognitive systems that at least partially constitute our cognitive processes. It could be said that information retrieved from the internet plays appropriate functional roles in the causal web of my mental states and my behavior, that I regularly rely on web searches in doing my research, and that I am causally connected to the internet through the physical manipulation of my laptop. So, if this is all that is needed for extension then it seems that our minds span over almost all information-bearing artifacts in our environment which seems implausible and detrimental to the research program of cognitive science.

In a similar vein, we could see OROHA as making the same mistake, at least this is what Morar and Skorburg claim.

> "The analogous objection would be this. If we permit constitutive connections among agents and their biological, social, political, economic, and ecological contexts as One Health specifies, where do we draw the line? If everything matters for health, then there is a sense in which nothing really matters." (2018:19)

Also,

> "The analogous concern is that by extending health and disease too far, we will then be forced to admit all sorts of entities into the domain of medical practice that are clearly outside the scope of realistic, effective medical interventions." (Ibid.)

Thus, these authors point out, in a similar fashion to Adams and Aizawa in the case of ExCog, that having overextended cognitive/medical processes or bloated cognitive/medical ontologies would dilute the basic notions of cognitive and medical sciences respectively. Medicine that deals with overextended entities would become clinically inert.

But here we would like to assess some differences between ExCog and OROHA in order to evaluate these arguments against them. First of all, proponents of ExCog do not advocate overextension. Their claims are limited to highly integrated body+artifact systems and they do not wish to claim that their, my, or your mind extends to every book in our respective libraries or similar. Critics then try to show that their arguments for these smaller, tighter systems stretch to the cases of overextension too, so they ask for more criteria – in the case of Adams and Aizawa, they ask for the "mark of the cognitive", or better specification of what makes a thing cognitive in the first place, and what separates these tighter systems from the overextended ones. On the other hand, proponents of OROHA (if they exist, because in the debate about One Health the ontological commitments are not fully spelled out, so there are no authors who explicitly advocate this position although we might say that it is implied by some of the definitions of One Health) would have to start from the claim that overextended systems exist. This would be the foundational claim of OROHA. So in this sense, OROHA proponents do not have to answer the cognitive bloat argument because answering it would mean abandoning their position, but they have to better establish their foundational claims and give us reasons for considering these overextended systems as appropriate bearers of health states. In that manner, the usual response of ExCog proponents against this attack which is spelling out additional conditions of extension such as, for instance, the existence of feedback causal loops between the parts of the system (these conditions should stop the overextension), would not and could not work in case of OROHA proponents, because in that case, they would advocate a completely different approach.

This brings us to the second point, the issue of the practicality of these overextensions and critics' claims that medicine would become clinically inert and meaningless, just like cognitive science will lose its proper subject of investigation. In the case of cognitive overextension, we were starting from individual subjects, and we were trying to find their physical boundaries or boundaries of their minds. It is undoubtedly impractical and meaningless to extend our minds to every possible available information source. Clinical psychology would have no use for it, nor cognitive science which searches for correct cognitive architectures which are responsible for the occurrence of cognitive abilities, and which could be potentially recreated in creating strong AI. But does this apply all the same to OROHA? We think not. Because OROHA is a different program from ExCog. OROHA is not starting from an individual human and her health and then trying to define it in terms of global systems. OROHA if it is at all advocated must be a claim about the health of a global system. Thus, such a program is not a program of human medicine. In that sense, Morar and Skoburg are right that *human* medicine would become clinically inert if it adopts such an ontology, but they miss to notice what Sironi et al. have recognized, and that is that ROHA in general asks for a radical change of perspective and ethics. OROHA medicine would be something completely different from our human medicine, and it also wouldn't be just a sum of botany, ecology, veterinary and medical sciences. It would be a new science of global health.

But is there a viable ontology that could support such a discipline? Perhaps a hypothesis about Gaia (Lovelock 1972, 1979) could be one contender for providing such ontology. According to Gaia hypothesis living beings and inanimate environments like climate systems are co-regulating, and the habitability of the planet, thus life itself, is depending on successful feedback loops that connect all living beings and the inorganic environment. Proponents of Gaia also advocate that there is a single entity, a life itself, that is comprised of all living beings having the same single ancestor. Thus, if the Gaia hypothesis or a similar one can be successfully defended then a medical science of a global entity can be established.

What we learned so far is that the One Health approach needs to be more carefully defined if promoted as a new medical paradigm. Sironi et al. focused on the primary goals of healthcare in this approach and identified POHA and ROHA as two basic forms which differ with respect to the main subject of medical interest – humans or all living species equally. They concluded that ROHA would ask for a too radical change which is not achievable without a complete shift in perspective and ethics. For these reasons, they advocate for anthropocentric POHA and additionally

welcome a slight shift towards planetocentric ROHA. On the other hand, Morar and Skorburg focused on the ontology of health and posited One Health on the extreme end of what we called OROHA. As we saw ROHA itself can keep the individualistic ontology by treating the health of all the species equally but can also take a more radical stance which is in line with the credo "heal the planet" which postulates super-extended entities on a global level to which we can apply medical predicates. Similarly to Sironi et al., Morar and Skorburg argued for the practicality of dispensing with OROHA and its ontology, but this time not because of the feasibility but because of the meaninglessness of medical concepts in this overextended context.

In both camps, we can recognize the call for practicality and understanding the methodologically plausible One Health approach as human oriented. This is not surprising if we recognize that the traditional subject of medical sciences is humankind and it's well-being. One Health in that context is a move for recognizing that human health cannot exist in isolation, but it is intricately connected to the health of other species and the balance of environmental conditions. Thus, it would be prudent to frame the One Health agenda in POHA individualistic terms, where reference to the health of other species, and that of a planet is in the function of human health. That way One Health is staying in a traditional framework of medicine, which is now broadened and takes into account the most diverse external factors. This means that One Health should be seen as closer in methodology and commitments to Embedded approaches to cognition, than to Extended ones. More radical versions of One Health would constitute a different discipline, one which is different from "human" medicine, and which is closer to Earth sciences.

Nevertheless, we can still ask wether the ontology of traditional individualistic medicine is still the proper one even if we reject the OROHA as a contender. This brings us to our last section.

3. Extended Health

In section 2. ExCog was described as an analogous position in cognitive science to One Health in medicine. Towards the end of section 2.2, we have made a remark about the difference between these approaches which is not based on their subject matter, but rather on their starting methodological assumptions. Namely, it was argued that ExCog starts from the traditional subject of cognitive science and then asks about its physical boundaries in order to show that this subject is in fact sometimes

extended, while OROHA starts from the "health of the planet" and thus abandons the traditional subject of medicine and begins with the postulation of a new global entity. We argued that as such OROHA should be better seen as a program of a new discipline different from medicine, and not as a corrective of an old program. Although ExCog is perhaps not the best analogous approach to OROHA, it can still give inspiration for proposing a new ontology for medical sciences.

Morar and Skorburg present Extended Health as an alternative anti-individualistic ontology for medicine apart from OROHA. While the One Health approach brings to focus the question of the complexity of environmental and inter-species relations in understanding health and illness and their management, as we saw previously this approach does not have clear ontological commitments. It recognizes different dependencies between biological subjects and various external factors, but it treats those dependencies equally. Thus, the two options for its ontology are traditional individualism or, as we have seen, an ontologically radical solution which recognizes overextended globally stretched entities or one single global entity which is the true bearer of health states. In other words, those complex dependencies or relations can either be seen as non-constitutive, or they are all constitutive. Extended Health is a middle ground between those two options. It makes a difference between constitutive and non-constitutive relations. The importance of this demarcation can be seen in various examples, some of which are mentioned in this article. For instance, the symbiosis between a human organism and its microbiota is of such a kind that makes it impossible to talk about human health without making a reference to the state of its microbiota. This could be seen as a reason enough to explore a kind of anti-individualistic ontology for medicine, one which is not over extensive and which can have useful implications for clinical practice. The question before us is how to carve nature at its joints and how to separate constitutive from non-constitutive causal connections between the individual organism and entities external to its biological boundaries.

Answers to this question can be sourced in the literature about ExCog, and this is the route that Morar and Skorburg also explore. Milojevic (2020) argues that arguments for ExCog must separately give criteria for the integration and identification of appropriate extended processes or states (for instance, functionalist criteria for extended cognitive processes). Fortunately, the literature on ExCog is bountiful with offers of such criteria, although they are not always called this way. The problem of integration in this literature can be found in analyses of different notions such as: "non-trivial causal spread", "dynamical coupling", "distributed func-

tional decomposition", "continuous reciprocal causation", "glue and trust conditions", etc. (Clark 2008). These notions should facilitate understanding why some interactions lead to creating integrated cognitive systems which can be seen as single entities. One way of explaining integration is by endorsing functionalism and employing the concept of distributed functional decomposition.

> "Distributed functional decomposition is a way of understanding the capacities of supersized mechanisms (ones created by the interactions of biological brains with bodies and aspects of the local environment) in terms of the flow and transformation of energy, information, control, and where applicable, representations. The use of the term functional in distributed functional decomposition is meant to remind us that even in these larger systems, it is the roles played by various elements, and not the specific ways those elements are realized, that do the explanatory work." (Clark 2008: 13–14)

Also, the causal connections between the parts of such systems are explained by concepts borrowed from dynamical systems theory. Thus, the components of these systems will interact in an ongoing, reciprocal way, creating feedback loops, where a change in one part of the system produces change in the other part of the system which in turn affects the first part. Such connections cannot be observed in the overextended "systems" that we mentioned before. Although the information from the internet can and does affect me, that information itself was neither affected by me prior to this effect, nor does it make me affect it later. In the case of a notebook and its user suffering from Alzheimer's disease, the flow of information between the human and the notebook is reciprocal. The information in the notebook is stored by the human, it in turn affects the behavior of its body and perception, which in turn affects the content of the notebook.

The integration of a cognitive system will also depend on a number of dimensions, such as:

> "the kind and intensity of information flow between agent and scaffold, the accessibility of the scaffold, the durability of the coupling between agent and scaffold, the amount of trust a user puts into the information the scaffold provides, the degree of transparency-in-use, the ease with which the information can be interpreted, amount of personalization, and the amount of cognitive transformation." (Heersmink 2017: 433–4)

We can notice that described criteria of integration are tailor-made for cognitive systems. Nevertheless, even if they make reference to notions like "flow of information", "representation", "cognitive transformation", etc. which are higher-level processes and entities found at the cognitive level, they also make use of lower-level processes and events like causal cou-

pling, feedback loop, reciprocal causation, etc. in describing integration. We can expect that we have to find these lower-level criteria satisfied in every kind of extended physical system, including ones that instantiate extended medical states and processes. On the other hand, those systems will have to have additional higher-level criteria satisfied, like joined metabolic processes, exchange of organic and inorganic matter, control of vital processes, etc. If we look at the example of human microbiome, we find the appropriate feedback loops and reciprocal causation. For instance, the microbiota is responsible for synthesizing certain vitamins, like vitamin B, and amino acids, breaking indigestible fibers, polysaccharides, and polyphenols (for which humans lack appropriate enzymes), producing short chain fatty acids, regulating fat deposits, etc. which in turn enable the host's survival, which in turn provide nutrition for microbiota (Rowland et al. 2018). These are only some of the many biological functions that sustain the life and health of a human+microbiota system. Thus, we can say that at least human+microbiota create an extended system to which appropriate medical terms can be applied.

4. Medicine of Extended Health

We saw that different authors conclude that for practical reasons we should prefer less radical and more prudent versions of the One Health approach. This means that the feasible One Health approach will still be focused on humans as their primary subject and it will not include any new ontology, instead, it will keep the traditional individualistic approach according to which the bearers of relevant medical states are individual organisms. As mentioned before that sort of approach is best compared with Embedded approaches to cognition according to which cognitive systems are deeply embedded in their environment and a proper understanding of cognitive abilities and their development calls for investigating and taking into consideration a variety of different external factors. Embedded approaches can be seen as calls for soft revolutions in science, which do not change the subject matter of a specific discipline or its primary methods and postulates, but point out the need to diversify and put its primary subject in a global context. What separates POHA from old-fashioned medicine is then the call for multidisciplinarity and greater collaboration between botany, veterinary sciences, ecology, and human medicine. With this widened interest One Health can certainly help in identifying the sources of certain diseases, mitigating their prevalence, and devising appropriate treatments.

On the other hand, Extended Health calls for a shift in perspective concerning the subjects of medical science. Unlike POHA which looks at usually causally far-removed external factors such as, for instance, the origins of zoonoses, Extended Health focuses on immediate causal factors that have a strong influence on an individual organism. Furthermore, those immediate causal factors create feedback loops with the primary system such that they become parts of thusly created new extended systems. Insisting on including the Extended Health perspective in medical sciences is motivated by the need to put back the subject of health states into the focus of clinical practice. While One Health approaches have the merit of extending the scope of medical research onto environmental factors and, thus, enable a better understanding of the origins, development, and spread of disease, they simultaneously affect medical practices in a way that can be detrimental to primary subjects of health and disease. Namely, by focusing on the disease One Health is losing perspective on the diseased. This can be seen in the management of the COVID-19 pandemic and the employment of specific strategies. In the remainder of the paper, we will briefly analyze a couple of examples that focus on a particular widely used strategy in fighting the COVID-19 pandemic – social distancing, and how it fits with One Health approaches on one side, and how Extended approaches might influence it.

While quarantining is a widely used and effective strategy in fighting infectious disease, by isolating infectious or exposed individuals, one of the main strategies in fighting the COVID-19 pandemic was isolating healthy individuals, or at least non-infectious ones. Isolation and social distancing of individuals was widespread and advocated on a global level by national health institutions and international institutions like WHO. It was carried out by closing public places of interest such as theaters and cinemas, transferring school courses to online platforms, leaving all but essential workers to work from home, employing a 1.5 or 2 m physical distancing rule in public spaces, not allowing seniors to leave their homes for several months, not allowing for larger gatherings at the privacy of a person's home, etc. This strategy was effective in slowing down the spread of the SARS-CoV-2 virus among humans, but it also affected the health of so isolated people in ways that could have been predicted, especially if the Extended Health approach was employed. Namely, we can distinguish at least three kinds of effects of the pandemic on the health of human individuals: 1) direct effect by infection with SARS-CoV-2, 2) indirect effect on mental and physical health induced by fears of infection, fear of our own death or death of close people, anxiety and stress of being infected or spreading the infection, and 3) indirect effect on physical and men-

tal health induced by strategies employed to "fight" the pandemic such as isolation and social distancing. In the past couple of years, researchers have identified a number of physical disorders induced by psychogenic factors such as weight gain (Zachary et al. 2020), and numerous psychological disorders such as depression, anxiety, insomnia, and PTSD. Causes of these disorders can be found also in group 2) of pandemic effects, nevertheless, we are now interested in the effects of strategies for reducing the spread of the coronavirus. This group of effects is identified by (Cenat et al. 2022), which also add fears of losing a job, the anxiety produced by financial insecurities, distress caused by media reports, etc.

Because the One Health approach focuses on the disease itself and its paths of transmission, its interspecies trajectory, and ways to stop its global spread, it is not at the same time too concerned of the effects its strategies have on the individuals. On the other hand, Extended Health does just the opposite. In the concrete case of the isolation of individuals during the COVID-19 pandemic, we can identify several ways in which individuals' health has been hindered. We'll briefly discuss Lyre's shared intentionality (2018), and Kosslyn's (2006) hypothesis of social minds (both examples are discussed in more detail in (Milojevic 2021)).

Lyre (2018) notices that cognitive extension does not have to be only into the body of the cognitive subject or into her physical environment, but it can also happen in a way that extends the original subject onto its "informational", and sometimes onto its social environment. As a case of social extension Lyre analyzes "shared intentionality". It is important to note that Lyre is writing about the strong constitutive kind of cognitive extension even in the cases of social extension, the same kind which Clark and Chalmers defended (1998), and thus he is not defending a claim about shared intentionality in group or collective minds but adopts and modifies the individualistic model of Bratman (1993). According to this model, an individual can have her own intention which is partially constituted of intentions and plans of another individual, though that primary intention cannot be ascribed to them together or jointly, nor to that other individual. In such a case, the nervous system of a second individual is becoming a constitutive part of the first's individual cognitive system. Such coupling can occur when both individuals intend to jointly accomplish a given action in a cooperative way, and in the process both become extended on the neural resources of another.

On the other hand, Kosslyn (2006) talks about social prosthetics and uses arguments independent from Clark and Chalmers (1998). Kosslyn sees social extensions of human cognitive systems as a natural consequence of their limited neural resources. Humans evolve in highly structured en-

vironments which reciprocally influence cognitive processes which are primarily realized in a neurally plastic flexible brain. According to Kosslyn the structure of our brains is not fully determined by our genes, but it is also strongly influenced by the environment. Thus, because of limited neural resources humans build a great number of tools that can be seen as cognitive prosthetics. Some of these tools are language, different kinds of notations, and some classical tools for navigation like compasses, etc. Nevertheless, Kosslyn's main point is that we do not build only artifacts to extend our resources, nor do we only structure our physical environment, but we also deeply rely on other people in performing cognitive and emotional tasks to the point that our personal identity depends on the people from our immediate surroundings. Others help us make decisions, form intentions, judge options, etc. Cases in which humans borrow parts of their cognitive resources from another human being in a transient or more permanent arrangement Kosslyn calls Social Prosthetic Systems.

Taking into account these hypotheses about a cognitive extension by social connections we can easily draw implications of social isolation of individuals in the period of COVID-19 pandemic. Despite the frequent claims that strategies for fighting COVID-19 pandemics were not employing social, but physical distancing and isolation, we claim that many social interactions have a physical component and that keeping virtual communication while removing the physical connections between people is sufficient for breaking many different social bonds. Cooperative endeavors frequently assume shared physical space and environment which allows for joint manipulation of task space. In that sense introducing physical distancing directly affects our methods of problem-solving, and with acceptance of Socially Extended Cognition our cognitive processing, and even our personal identity if we accept Kosslyn's claim about Social Prosthetic Systems. Thus, strategies of isolation can be seen as influencing the mental health of individuals not only by introducing negative external factors but also by influencing the integrity of the subjects themselves.

5. Concluding remarks

In this paper, we analyzed different methodological and ontological commitments of the One Health approach and we have concluded that they are not yet fully transparent in the offered formulations of this view. Nevertheless, we can differentiate between different versions of the One Health approach by explicitly employing the distinction between POHA and ROHA as two methodologically different approaches with two dif-

ferent aims, and OROHA and OPOHA as two ontologically different approaches that employ either radical anti-individualistic ontology or prudential traditional individualistic ontology of medical sciences. We have judged that OROHA is ontologically opposed to both traditional medical approaches and to Extended Health approach. Morar and Skorburg argued that Extended Health should be seen as an alternative anti-individualistic approach to the One Health one, but with more careful implementation of the introduced distinctions it is clear that this is true only for its radical ontological version. We have also argued that such a radical alternative would ask for the constitution of a different scientific and health discipline distinct from human medicine. Instead, we argued for cooperation between the prudential approaches of One Health which have practical, methodological, and ontological advantages, and Extended Health approaches. This amalgam of approaches should improve current practices on two fronts simultaneously: a) expanding the field of study and introducing multidisciplinarity which is needed for understanding the origins, spread and development of diseases that cross species barriers or occur because of environmental imbalances, and b) keeping the focus on an individual and her health by carefully examining the boundaries of systems to which we should apply health properties. The case of strategies employed to fight the COVID-19 pandemic showed that not introducing both approaches at the same time can lead to detrimental practices which safeguard the health of individuals on one front but negatively influence it on the other.

References

Adams, F. R., & Aizawa, K. (2001). "The bounds of cognition". *Philosophical Psychology* 14 (1), 43–64.

Adams, F. R., & Aizawa, K. (2008). *The Bounds of Cognition*. Oxford: Wiley-Blackwell.

Boorse, C. (1975). "On the distinction between disease and illness", *Philosophy & Public Affairs* 5(1):49– 68.

Boorse, C. (1997). "Rebuttal on Health". In J.M. Humber et al. (eds.) *What is Disease?*, 1–134. New York: Springer.

Bratman, M. (1993). „Shared intention". *Ethics* 104 (1):97–113.

Capps, B., Lederman, Z. (2015). "One health, vaccines and ebola: the opportunities for shared benefits". *Journal of Agricultural and Environmental Ethics* 28(6):1011–32.

CDC (2022). "One Health Basics". https://www.cdc.gov/onehealth/basics/index.html

Cénat, J. M., Blais-Rochette, C., Kokou-Kpolou, C. K., Noorishad, P.-G., Mukunzi, J. N., McIntee, S.-E., Dalexis, R. D., Goulet, M.-A., Labelle, P. R. (2021). „Prevalence of symptoms of depression, anxiety, insomnia, posttraumatic

stress disorder, and psychological distress among populations affected by the COVID-19 pandemic: A systematic review and meta-analysis". *Psychiatry Research* 295

Clark, A. & Chalmers, D. (1998). "The Extended Mind". *Analysis* 58(1):7–19

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Engelhardt Jr, H.T. (1976). "Ideology and etiology". *The Journal of Medicine and Philosophy* 1(3):256–268.

Ereshefsky, M. (2009). "Defining 'health' and 'disease'". *Studies in History and Philosophy of Biological and Biomedical Sciences* 40:221–227.

FAO et al. (2008). "Contributing to One World, One Health™ – A Strategic Framework for Reducing Risks of Infectious Diseases at the Animal-Human-Ecosystems Interface". https://web.archive.org/web/20150606122557/http://www.oie.int/doc/ged/D5720.PDF

Fitzsimons, G. M., Finkel, E. J., & van Dellen, M. R. (2015). "Transactive goal dynamics". *Psychological Review* 122(4):648–673.

Goosens, W. (1980). "Values, Health and Medicine". *Philosophy of Science* 47: 100–115.

Haider, N., Rothman-Ostrow, P., Osman, A.Y., Arruda, L.B., Macfarlane-Berry, L., Elton, L., Thomason, M.J., Yeboah-Manu, D., Ansumana, R., Kapata, N., Mboera, L., Rushton, J., McHugh, T.D., Heymann, D.L., Zumla, A. and Kock R.A. (2020). "COVID-19—Zoonosis or Emerging Infectious Disease?". *Front. Public Health* 8:596944. doi: 10.3389/fpubh.2020.596944

Heersmink, R. (2017). "Distributed cognition and distributed morality: Agency, artifacts and systems". *Science and Engineering Ethics* 23 (2): 431–448.

Kosslyn, S. M. (2006). "On the Evolution of Human Motivation: The Role of Social Prosthetic Systems". In S. M. Platek, T. K. Shackelford & J. P. Keenan (Eds.), *Evolutionary cognitive neuroscience* (Cambridge, MA: MIT Press): 541–554

Lovelock, J. (1972). "Gaia as seen through the atmosphere". *Atmospheric Environment* 6 (8): 579–580. doi:10.1016/0004–6981(72)90076–5.

Lovelock, J. (1979). *Gaia:a new look at life on earth*. Oxford UP.

Lyre, H. (2018). „Socially Extended Cognition and Shared Intentionality". *Frontiers in Psychology* 9.

Lysaght, T., Capps, B., Bailey, M., Bickford, D., Coker, R., Lederman, Z., et al. (2017). "Justice is the missing link in one health: results of a mixed methods study in an urban city state". *PLoS One* 12(1):e0170967.

Milojevic, M. (2020). "Extended mind, functionalism and personal identity". *Synthese* 197: 2143–2170. https://doi.org/10.1007/s11229–018–1797–5

Milojević, M. (2021). "Socially extended cognition and covid-19 pandemic". In Nenad Cekić (ed.), *Етика и истина у доба кризе*. Belgrade: University of Belgrade – Faculty of Philosophy. pp. 235–253.

Morar, N., Skorburg, J.A. (2018). "Bioethics and the Hypothesis of Extended Health". *Kennedy Inst Ethics J.* 28(3):341–376. doi: 10.1353/ken.2018.0020. PMID: 30369509.

One Health High-Level Expert Panel (OHHLEP), Adisasmito WB, Almuhairi S, Behravesh CB, Bilivogui P, Bukachi SA, et al. (2022). "One Health: A new definition for a sustainable and healthy future". *PLoS Pathog* 18(6): e1010537. https://doi.org/10.1371/journal.ppat.1010537

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., Tuohy, K. (2018). "Gut microbiota functions: metabolism of nutrients and other food components". *Eur J Nutr*. 57(1):1–24. doi: 10.1007/s00394–017–1445–8.

Schwabe, C.W. (1964). *Veterinary Medicine and Human Health.* Williams & Wilkins

Sironi, V.A., Inglese, S. & Lavazza, A. (2022). "The "One Health" approach in the face of Covid-19: how radical should it be?". *Philos Ethics Humanit Med* 17, 3. https://doi.org/10.1186/s13010–022–00116–2

Virchow, R. L. K. (1859/1978). *Cellular pathology*. London: John Churchill

WHO (2022). "One Health". https://www.who.int/news-room/questions-and-answers/item/one-health

Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. New York: Cambridge University Press.

Zachary, Z., Brianna, F., Brianna, L., Garrett, P., Jade, W., Alyssa, D., & Mikayla, K. (2020). „Self-quarantine and weight gain related risk factors during the COVID-19 pandemic". *Obesity research & clinical practice* 14(3): 210–216.

Jelena Pavličić*

THE VALUE LADENNESS OF SCIENTIFIC PRACTICE: "COVIDIZATION" OF RESEARCH AND TRUST IN SCIENCE

Abstract: In recent years, philosophers of science, including social epistemologists, have increasingly begun to focus on the role of value judgments in research activities and their consequences on the epistemic integrity of scientific inquiry. These considerations initiated a series of new practical and theoretical challenges, and "revived" old descriptive and prescriptive disagreements over the form of the relationship between values and scientific practice. In this article, we will attempt to frame the way in which values in science are discussed today, point to concrete examples that serve to illustrate the pervasiveness of value judgments in the scientific endeavour, and consider the question of how it is possible to ensure credibility in science and protect its epistemic integrity in the light of a value-laden framework.

Keywords: science, value-neutrality, the credibility of scientific results, public knowledge

The claim that scientific practice is deeply influenced by values is widely held and defended by philosophers of science today (Douglas, 2009: 15, Steel, 2015: 2; Elliott 2017: 8; Goldenberg, 2021: 100; Oreskes, 2019: 147–159 et al.).[1] However, this viewpoint has not always been accepted. For a time, it was believed that even though science has obvious political, moral, and socio-economic repercussions, it can and should be an enterprise that does not involve value judgments. This ideal of value-

* Institute for Philosophy, Faculty of Philosophy, University of Belgrade, jelena.pavlicic@f.bg.ac.rs

1 This paper has been written as part of the scientific research project *Humans and Society in Times of Crisis*, which was financed by the Faculty of Philosophy – University of Belgrade.

free science began to gain its dominance at the end of the 1950s. Although at the time there were theorists who were ready to admit that such an ideal is neither advisable nor realistic and warned that it is not possible to understand the practice of the sciences without considering the specific value judgments that are formed within it (Rudner 1953:6; Frank 1954: 143), discussions on values in science would almost completely fall silent during the 1960s (see: Douglas, 2009: 50, 62–5).[2] From the 1980s onwards, there has been a growing body of literature that seeks to show that when making decisions and providing answers to a number of questions that fall under their domain, scientists make extensive use of value judgments. But even today, when almost no one would be ready to deny that research practice is strongly permeated with values, attitudes that value-neutrality should be the goal of science are still present (Shrader-Frechette, 1991: 44; Ruphy 2006: 192; Koertge, 2000: 53). Such a belief is the result of recent conflicting and complicated proposals on how we should understand the nature, domain, and role of value in scientific work, as well as of an old concern that more transparent attempts to articulate the idea of value-laden science would damage the public's trust in the reliability of its results (Du Bois, 1912, 1935; Compton 1936; Merton, 1938). Having that in mind, significant literature in recent years has been developed with an aim to provide a satisfactory analysis of scientific practice that will take into account its strong interwovenness with values, but without unacceptable consequences regarding its epistemic integrity and the reliability of scientific results (Kitcher 2001; Douglas 2009; Elliot 2017; De Melo-Martin & Intemann, 2018, etc.). In this article, we will present a framework for understanding the way in which values in science are discussed today, focus on some concrete examples that highlight the range of ways in which value judgments influence scientific work, and consider the question of how it is possible to preserve trust in science and protect its integrity in the light of a value-laden scientific framework.

II

One way to approach the consideration of the relationship between values and science is to point to situations in which value judgments can enter research practice and interfere with scientific reasoning. Here, we will offer four different contexts in light of which it is possible to identify

2 More on the specific historical episodes that preceded and accompanied the ideal of value-free science, as well as on sporadic, marginalized and short-term deviations from it, see: (Douglas 2009: §3).

the relevance of value judgments to the researchers' decisions.[3] A closer look at them should contribute to the clarification of the manners in which values can influence the shaping of epistemic and organizational processes of science, as well as to a clearer understanding of the sources of conflict around the standards for assessing values' epistemic desirability in different stages of the scientific endeavor. In addition, the following examples will serve to summarize the main features of our viewpoint regarding the role of values in scientific practice which, in turn, will pave the way for addressing the concerns about public skepticism toward issues of scientific expertise and its resistance to scientific claims.

i) Selection of research problems

Since there are far more lines of inquiry that can be implemented in real-time, the selection of research problems and their corresponding goals tend to be influenced by the value preferences of the members of the scientific community. The extent to which normative-value frameworks can influence research directions could be effectively captured by the recent emergence of the *covidization of research* (Pai, 2020). This phenomenon is exemplified by the fact that in the period from January 1, 2020. to August 1, 2021, more than one in six active members of the scientific community decided to adapt or redirect their research activities to include the study of various aspects of the coronavirus pandemic-induced crisis (Ioannidis, et al., 2021). As a result of such a shift in research priorities, the number of works related to the study of COVID-19 (210,863) reached 3.7% of the total number of scientific works (5,728,015) that were published and indexed in Scopus (*Ibid*). Such a turn was, among other things, supported by the financial considerations of researchers.[4] But regardless of the extent to which researchers' choice sets were motivated by ethical, theoretical or financial reasons, they were manifestly not immune to social influences and were based on specific value judgments. In a similar manner as in the example of *covidization of research*, value judgments play an influential and important role in deciding which research topics we (as

3 In the literature, there are several different illustrations and classifications, see: (Machamer & Wolters 2004; Dorato 2004; Ward 2021; Elliot 2017) which include examples of how value-based choices and compromises affect model tuning (for example: De Melo-Martin & Intemann, 2018: 121) or dissemination of research findings (Elliot 2017); examples that we have not highlighted here.

4 Data show that by the end of June 2021, 14 billion dollars have been allocated for research activities related to the coronavirus pandemic, often at the cost of canceling or postponing the opening of regular invitations for research funding (Ioannidis, et al. 2021; Pai, 2020).

individuals) want to pursue, which areas of research are the most significant or promising for us (as a society), and which directions of research, given the limited resources for funding of science, should be prioritized.

ii) Establishing standards for performing responsible research

In addition to guiding our choice of research programs, value judgments significantly influence the steering of research projects as well as the decisions on whether they will be implemented at all. For example, the implementation of programs that include methodological approaches not in accordance with the informed consent of research subjects, those which violate the confidentiality of information about research participants, or those which propose experiments that would result in their physical (or psychological) harm (that is, violation of the principle of *primum non nocere*), cannot be allowed for obvious legal and moral reasons. In this sense, value judgments that include social, moral, and legal considerations, limit the range of means by which particular problems will be studied and shape the standards for the responsible conduct of research.

iii) Epistemic risk and loss function

As in choosing the subject of research and establishing standards for its responsible performance, value judgments play a significant role in determining how, within the framework of their statistical procedures – accepting or rejecting a statistical hypothesis – scientists deal with the risk of making a mistake in their decisions. Empirical knowledge achieved by scientists is beset with a variety of epistemic risks and in their procedures of arriving at it, scientists always face the risk of making two types of errors: accepting the wrong (type 2 error) or rejecting the true hypothesis (type 1 error). Since, taking a study design as given, these two errors are supplementary – the probability of committing one can generally only be reduced at the expense of increasing the probability of the other – the choice of how to manage or balance between those errors can be described as a loss function. It is important to underline that there is no firm methodological rule stating what the loss function should be i.e. what the acceptable balance between the risks of committing the two types of error is. These decisions are typically made in light of the interests and values which determine how grave the consequences of going wrong in either direction are. The quick sketch of an admittedly idealized context of medical research can illustrate how the loss function is decided by invoking value judgments. Let's consider a group of scientists developing a drug for an already well-managed disease that is somewhat superior to the existing treatments in terms of

efficacy. In this situation, rejecting the null hypothesis would require extremely strong evidence (i.e., a very low probability of type 1 error), so the process would be repeated many times in different populations under different circumstances, which would be followed for a long time to ensure with high probability that the new drug is superior to the existing treatment. Suppose, on the other hand, humanity is facing a progressive disease whose early clinical symptoms indicate certain fatal outcomes, against which no existing treatment is at all effective. Under these circumstances, a lower level of evidentiary support would be required to implement the drug and, in this admittedly simplistic case, researchers' relative tolerance for committing type 1 and type 2 errors would shift significantly towards the former. In other words, they would be more tolerant of providing a drug that may turn out to be insufficiently effective (type 1 error), than risk discarding one which may be effective (type 2 error).

Or, consider an example in economics. A central bank researcher is trying to predict whether or not the coming year will be an inflationary episode, and consequently whether interest rates should be raised. This, in turn, has political consequences – for the distribution of wealth between borrowers and savers, for example. Both the choice of the loss function (which side to err on) and choice of a statistical model to use (usually more than one is acceptable) come from values, and those have to do with the view of the researcher of what the consequences of going wrong in either direction might be. In other words, they risk having the end in mind – an economist who strongly believes in the power of markets to self-correct, for example, might tend to require stronger evidentiary support that inflation will occur than one who believes in government intervention to reduce inflation, or she might select a statistical model which is less likely to predict inflation next year. Drawing on examples like this, over the past several years, philosophers of science have been increasingly exploring not only how values influence the way scientists judge the output of their statistical test but also how value-laden determinations of loss function shape the statistical choices the researchers make while designing and directing their research programs.[5]

iv) Selection and definitions of variables

Formulating a statistical research problem inevitably involves a simplification of the world in the sense that we are choosing to focus on a few variables whose impact we want to measure. In macroeconomics, these

5 See: Zollman K., Values, Objectivity & Data Science – Philosophy of Data Science, Link: https://www.youtube.com/watch?v=9USkWtX-ydc

variables are typically aggregated. Economists often formulate questions in terms of the impact of a phenomenon on GDP, which is just the sum of all income generated by domestic households and firms. But the decision to focus on the sum is also a decision not to focus on the distribution of income (between rich and poor workers, or between workers and capital owners, etc.), which is a whole set of value judgments. Moreover, selecting GDP as the outcome of choice is a value judgment that income is what is *good* in an economy, but there are other indicators of economic and social progress. A discussion of these can be found in *Beyond GDP: Measuring What Counts for Economic and Social Performance* (2018), a book where Joseph Stiglitz and others discuss the limitations of GDP and propose a range of complementary measures of economic well-being, which include measures of economic insecurity, wealth and income inequality, social and environmental sustainability, trust in institutions and quality of life. Including any of these indicators in a research proposal represents a value judgment as to its importance.

A similar point emerges if we consider analyses of the concept of mental disorder. Let's restrict our attention to the definition of mental disorders as "harmful dysfunction". While "dysfunction" can be understood as the inability of an internal mental mechanism to perform a specific function for which it was predestined by evolution, the question of whether the dysfunction will have a detrimental effect on a person's well-being will depend on the social values (Wakefield, 1992: 385). For example, brain dysfunctions that can interfere with reading would not be considered "harmful" within the preliterate communities, while today children and adults who have difficulties in reading are diagnosed with dyslexia or decoding difficulty which is, along with dyscalculia and disorders of written expression, classified as a learning disorder (Üstün, Chatterji, & Andrews, 2002: 31; Snowling & Hulme 2012: 594). In a similar vein, it is suggested that many other classifications of diseases incorporate value influences and that on the questions of whether something is a medical disease or how it should be 'correctly' defined for purposes of research and diagnosis, there is often no value-free way to provide answers (Kukla, 2019).

So, from the decision to engage in science and that certain research projects are worth pursuing, to the evaluation of the output of statistical tests, to selecting and defining variables, moral, social, political, etc., values play a significant role in research activities and are inherent in scientific practice (cf. Rudner, 1953: 2; Douglas 2009: 112; Kitcher 2001: 63–82; Elliott 2017: 15, 166). In this regard, it should be emphasized that the study of the influence of values on the selection of research programs

and the formation of standards for their responsible performance (at least until recently) was not a subject of special interest to philosophers of science, since it was considered that value-laden decisions made in these so-called "external" stages of scientific endeavors do not threaten the reliability of its epistemic procedures (more on this: Dorato 2004: 57; Machamer & Wolters 2004: 1–4; Douglas, 2009: 45, 98; Kitcher 2011: §1; De Melo-Martin & Intemann, 2018:119). The polarization of opinion in the literature on science and values mainly refers to the determination of the kinds of values and the degree to which they should influence the development of the "internal" phases of science, which include data characterization, assessment of available evidence, acceptance of hypotheses, model development, etc. The first line of debate has argued that only those values that are "epistemic" in nature ("scientific", "internal" or "cognitive") such as the degree of evidentiary support, consistency, predictive and explanatory power, etc., can play a legitimate role in these processes, while the influence of "non-epistemic" ("non-scientific", "external" or "social") values that include moral, legal, political and socio-economic considerations must be eliminated or at least minimized (Shrader-Frechette 1994: 53). At the other end of the spectrum is the understanding that "non-epistemic" values should *in some form* "enter" all stages of scientific work since they are necessary to provide guidance for scientist when making judgments (Douglas, 2009: 112; De Melo-Martin & Intemann 2018: 119–22, Steel, 2015: 2), and that scientific communities which take that into account will be more successful, both in achieving their epistemic goals and in establishing a constructive relationship with the general public (De Melo-Martin & Intemann 2018: 119; Goldenberg, 2021: 125; Elliott, 2017: 166; Longino, 2004: 137).

Although launching into this discussion would go beyond the scope of this paper, it should be noted that philosophers have started to use the terms "epistemic" (scientific, internal, or cognitive) and "non-epistemic" (non-scientific, external, or social) in a very confusing way, which has led to the displacement and blurring of demarcation lines between these two camps. Thus, while some authors point to the fragility of separation between epistemic and non-epistemic values (Machamer & Osbeck 2004), others attempted to formulate a clearer demarcation criterion, trying to work out the exact meaning of the terms. As a result, once coextensive terms "epistemic" and "cognitive" (Lacey 1999: 221) began to diverge (Laudan, 2004: 19; Douglas, 2009: §5), as is the case with the meanings of the terms "non-epistemic", "external", "non-scientific" or "social" (Dorato, 2004: 53). Subsequently, others drew attention to examples in which non-epistemic values can be interpreted or seen as epistemic (Douglas 2009:

90; Wilholt 2006: 80). Still others indicated that there are values for which it is not clear how they should be classified and that the lines of demarcation are not easy to draw at all (Machamer & Wolters 2004: 3). As a result, there is a growing body of literature that points out that the principled difference between epistemic and non-epistemic values is not viable (see: Longino 2004: 128; Douglas 2009: 90).

Indeed, if this is the case and we take into account the failure of previous attempts to single out one class of values that could play a normatively acceptable role in scientific reasoning and on which the "protection" of science from problematic value influences could be based, the question arises: how to approach the fact that science is strongly imbued with values and that many of its activities take place precisely on their background? Following the example of some recent proposals, we can suggest that traditional attempts to classify values be replaced by approaches that focus on more detailed considerations of the question: how, when, and in what situations they "enter" scientific practice (Douglas 2009: 87) and those that point to the necessity of precise articulation of their role in different research activities (Elliott, 2017: 73; De Melo-Martin & Intemann 2018: §9). That the active mapping of values and their more precise articulation (focusing on individual local contexts, examining individual examples, and analyzing their details) could play a significant role in attempts to identify their potential adverse impact on science and the reliability of its results, will be illustrated in the light of the aforementioned emergence of *covidization of research* in the following section.

III

During the aforementioned covidization of research there has been a tendency in parts of the scientific community to focus excessive attention on efforts to understand the emergence of the coronavirus pandemic at the expense of dealing with questions in scientists' primary area of expertise (Pai, 2020). As some authors emphasize, while it is encouraging to see the extent to which the scientific community can be motivated and united in order to respond to existing social challenges, the question is whether its response – the amount of resources and energy spent – is proportional to the size of the existing crisis and what the real advantages of such hyper production of works are (Ioannidis, et al. 2021: §4). Unfortunately, existing analyzes suggest that much of the growing literature on the coronavirus pandemic crisis is of poor quality (Khatter, et al. 2021; Bagdasarian, Cross, and Fisher, 2020; Ioannidis, et al., 2021, 2022). Having

that in mind, the question arises as to whether some theorists are correct to claim that the values which influence the choice of research problems play only a sporadic role in achieving the epistemic goals of science. Namely, as already indicated, until recently the study of the influence of values in this phase of the scientific process was not of immediate importance in the context of the discussion about values in science because it was considered that they do not have an epistemically relevant character and therefore no immediate effect on the reliability of scientific results. And yet, it seems that the example of the covidization of research suggests that value judgments should not be assigned with privileged status in the "external" phase of the scientific process. Without pretending to go into consideration what motives contributed to the covidization of research, it seems quite reasonable to say that any decision of the members of the scientific community – regardless of the stage or phase in which it was made – based on the values that unjustifiably favor unidirectional research activities or impede the acquisition of appropriate evidence, may have an adverse effect on the epistemic engagement of science. In this regard, relying on strategies and approaches that propose constant monitoring and, if possible, critical reviews of the role of values in different research domains and from the perspectives of different stakeholders can contribute to preventing, where possible, the future neglect of equally important research projects as well as the lowering of epistemic standards in those ongoing. In other words, transparent and active discussions regarding the determination of facts and the way in which values pervade scientific procedures could be a useful set of tools for offering a more complete representation of their epistemic consequences and determining more precisely which value judgments (given a theoretical, social, technological, organizational, etc. context) can be assigned with normatively acceptable roles.

However, another question related to the previous considerations arises: would a transparent discussion of value-laden science lead to an erosion of public trust in science and have a negative impact on the public's motivation to comply with recommendations based on scientific judgments? Although even a partial review of numerous recent studies on public trust in scientific claims would go beyond the scope of this paper, what can be emphasized here is that they strongly indicate that the political and ideological orientations of individuals are a significant factor in establishing and maintaining trust in scientific evidence (De Melo-Martin & Intemann, 2018: 123; Elliot 2017:9).[6] These findings correspond to recent viewpoints that the problem of mistrust in science should primar-

6 See: (Pavličić, 2020 Pavličić, Petrović and Smajević Roljić, 2022) for a disscusion on the issue of public mistrust of scientific authorities.

ily be understood as a consequence of individuals' beliefs that scientific findings somehow threaten their values, religious convictions, political-ideological orientations or economic interests (Oreskes, 2019: 147, Kitcher 2011: §1). Such views are often accompanied by insights that a skeptical public "is better understood as a rejection of the values underlying the scientific consensus" (Goldenberg, 2019: 22) rather than as a consequence of the fact that the consensus includes values.

Does that mean that public trust in scientific results depends entirely on whether scientists adopt the values that society set for research? Although at the moment it is not possible to give a precise answer to this question, it is worth noting that there is an increasing number of examples that indicate that the failure of scientists to be transparent and honest about the assumptions underlying their research activities has contributed to the public's concern that certain political interests were prioritized over the search for scientific truth in their reports (see: De Melo-Martin & Intemann 2018: 115). Indeed, if that is the case, some authors are quite right to claim that scientists' further resistance to speak openly about value judgments would only worsen the situation by creating the impression that their values are somehow problematic and endangering their knowledge-seeking engagement (Oreskes, 2019: 153). Therefore, the scientific community should establish a more transparent and active dialogue on values-guided decisions between itself and the public, and implement complementary strategies which promote values that are inclusive and representative of the interests of different stakeholders. While the inclusion of the broader public in the scientific enterprise would contribute to determining research priorities and establishing more realistic expectations from science, open and critical discussions would help to form informed and reflective judgments in the light of which scientists themselves could identify the damaging impact of their values on the reliability and the significance of their studies. Although it is certainly necessary to conduct significant experimental research in order to determine what concrete strategies and tactics would enable an effective, acceptable, and quality institutional involvement of public opinion in science, it is quite resonable to say that a fair relationship between science and the public (their mutual understanding, cooperation and maintaining trust in the scientific community) requires a socially responsible science that strives to preserve its epistemic integrity and is transparent about its goals.

Literature:

Bagdasarian, N. Cross, G. B. Fisher, D. (2020) "Rapid publications risk the integrity of science in the era of COVID-19", BMC Med. 18: 192

Compton, K. T. (1936) "Science advisory service to the government", *Scientific Monthly* 42: 30–39

De Melo-Martin, I. & Intemann, K. (2018), *The Fight Against Doubt: How to Bridge the Gap Between Scientists and the Public*, Oxford: Oxford University Press.

Dorato, M. (2004). "Epistemic and non epistemic values in science", *Science Values and Objectivity*: 52–77.

Douglas, H. (2009), *Science, Policy, and the Value-Free Ideal*, University of Pittsburgh Press

Du Bois, W. E. B. (1912). The rural south. Publications of the American Statistical Association, 13(97), 80–84.

Du Bois, W. E. B. (1935). Black reconstruction in America. New York: The Free Press

Elliott, K. C. (2017), *A Tapestry of Values: An Introduction to Values in Science*, New York: Oxford University Press.

Frank, P. G. (1954) "The variety of reasons for the acceptance of scientific theories", The Scientific Monthly, Vol. 79, No. 3: 139–145

Goldenberg, M., (2021), *Vaccine Hesitancy: Public Trust, Expertise, and the War on Science*, Pittsburgh: University of Pittsburgh Press.

Ioannidis, J. P. A. Bendavida, E. Salholz-Hillelf, M. Boyackg, K. W. and Baas, J. "Massive covidization of research citations and the citation elite" Edited by Kenneth Wachter, University of California, Berkeley, CA; received March 7, 2022: 1–8.

Ioannidis, J. P. A. Salholz-Hillel, M. Boyack, K. W. Baas, J. (2021) "The rapid, massive growth of COVID-19 authors in the scientific literature", R. Soc. Open Sci. 8, 210389.

Khatter, A. Naughton, M., Dambha-Miller, H., Redmond, P. (2021) "Is rapid scientific publication also high quality? Bibliometric analysis of highly disseminated COVID-19 research papers". Learn. Publ. 34, 568–577.

Kitcher, P. (2001). *Science, Truth, and Democracy*. New York: Oxford University Press.

Kitcher, P. (2011). *Science in a democratic society*. New York: Prometheus Books

Koertge, N. (2000). Science, values, and the values of science. Philosophy of Science (Supplement) 67: 45–57.

Kukla, R. (2019). "Infertility, epistemic risk, and disease definitions". Synthese 196 (11): 4409–4428.

Lacey, H. (1999). *Is science value-free? Values and scientific understanding*. New-York: Routledge.

Laudan, L., (2004) "The Epistemic, the Cognitive, and the Social", *Science Values and Objectivity:* 14–24

Longino, H. (2004) "How Values Can Be Good for Science", *Science Values and Objectivity*: 52–77.

Machamer P. & Wolters, G. (2004) "Introduction", *Science, Values and Objectivity*, University of Pittsburgh Press: 1:14.

Machamer P., & Osbeck, L., (2004) "The Social in the Epistemic", *Science, Values and Objectivity*, University of Pittsburgh Press: 78–90

Merton, R. (1938) "Science and the Social Order", *Philosophy of Science, 5*(3): 321–337.

Nagel, Ernest. (1961) *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace & World.

Oreskes, N., (2019) *Why Trust Science?*, Princeton: Princeton University Press.

Pai, M. (2020) "Covidization of research: what are the risks?", *Nat Med* 26, 1159

Pavličić J. (2020) ''Konsenzus i poverenje u nauku: uvidi u oblasti socijalne epistemologije primenjeni u analizi socijalnog epistemičkog ponašanja u doba krize izazvane virusom korona (SARS-COV-2)", Бањалучки новембарски сусрети 2020: 311–324

Pavličić J., Petrović M., Smajević Roljić M. (2022), "The Relevance of Philosophy in Times of the Coronavirus Crisis", Philosophy and Society 33 (1): 233–246.

Rudner, R. (1953). "The scientist qua scientist makes value judgments". *Philosophy of Science* 20:1–6

Ruphy, S. (2006). "Empiricism all the way down: a defense of the value-neutrality of science in response to Helen Longino's contextual empiricism". *Perspectives on Science*, 14 (2): 189–214.

Shrader-Frechette, K., (1991). *Risk and rationality*. Berkeley and Los Angeles: University of California Press

Snowling, M. J. and Hulme, C. (2012), Annual Research Review: The nature and classification of reading disorders – a commentary on proposals for DSM-5. Journal of Child Psychology and Psychiatry, 53: 593–607

Steel, D. (2015) "Acceptance, values, and probability", *Studies in History and Philosophy of Science* Part A, Vol. 53: 81–88

Stiglitz, J., J. Fitoussi and M. Durand (2018), *Beyond GDP: Measuring What Counts for Economic and Social Performance*, OECD Publishing, Paris, https://doi.org/10.1787/9789264307292-en.

Üstün, T. B., Chatterji, S., & Andrews, G. (2002), International classifications and the diagnosis of mental disorders: Strengths, limitations and future perspectives. In M. Maj, W. Gaebel, J. J. López-Ibor, & N. Sartorius (Eds.), *Psychiatric diagnosis and classification* John Wiley & Sons Inc: 25–46.

Wakefield JC. (1992) "The concept of mental disorder. On the boundary between biological facts and social values", Am Psychol. Mar; 47(3): 373–88.

Wilholt, T. (2006) "Design Rules: Industrial Research and Epistemic Merit". *Philosophy of Science*, *73*(1): 66–89.

Zollman K., Values, Objectivity & Data Science – Philosophy of Data Science, Link: https://www.youtube.com/watch?v=9USkWtX-ydc

Jelena Pavličić

Vrednosno opterećena naučna praksa: „Kovidizacija" istraživanja i poverenje u nauku.

Apstrakt: U poslednjih nekoliko godina, radovi iz filozofije nauke i socijalne epistemologije nauke su sve više počeli da se fokusiraju na pitanja kakvi su status i uloga vrednosnih sudova u sprovođenju naučnih aktivnosti i kakve posledice njihova prisutnost može imati u pogledu epistemičkog integriteta naučnih istraživanja. Ova razmatranja su inicirala niz novih teorijskih i praktičnih nedoumica i„oživela" stara kako deskriptivna, tako i preskriptivna neslaganja u pogledu poimanja odnosa između vrednosti i naučnoistraživačke prakse. U ovom članku tematizovaćemo način na koji se danas diskutuje o vrednostima u nauci, ukazati na primere koji jasno svedoče o uplivu vrednosti u aktuelnu naučnoistraživačku praksu te razmotriti pitanje kako je moguće u svetlu pristupa koji uzima u obzir vrednosnu opterećenost naučnoistraživačkog rada očuvati poverenje u nauku i zaštiti njen epistemički integritet.

Ključne reči: nauka, vrednosna neutralnost, kredibilitet naučnih rezultata, javno mnjenje

Janko Nešić*

AFFORDING AUTISTIC PERSONS EPISTEMIC JUSTICE

Abstract: Autism is a psychopathological condition around which there is still much prejudice and stigma. The discrepancy between third-person and first-person accounts of autistic behavior creates a chasm between autistic and neurotypical (non-autistic) people. Epistemic injustice suffered by these individuals is great, and a fruitful strategy out of this predicament is much needed. I will propose that through the appropriation and implementation of methods and concepts from phenomenology and ecological-enactive cognitive science, we can acquire powerful tools to work towards greater epistemic justice for autistic individuals. I will use the resources found in the skilled intentionality framework, integrated with various phenomenological theories. From these approaches, we can view autistic impairments and disability relationally and how epistemic enablement and disablement form. Phenomenology and its methods help us learn more about the perceptual and social experiences of autistic individuals. The voices of the autistics themselves will be of the greatest importance here. I will show that, through restructuring our landscape of affordances and with a greater phenomenological understanding of the autistic inner world, we can devise new strategies that afford greater epistemic enablement and epistemic justice.

Keywords: autism spectrum disorder, epistemic injustice, phenomenology, enactivism, ecological psychology, landscape of affordances

I stim, therefore I am
Melanie Yergeau

1. Introduction

Autistic people face many injustices. A recent horrifying event that has befallen an autistic person testifies to the profound lack of people`s understanding of autistics. On May 20, 2020, a 32-year-old autistic Pales-

* Janko Nešić, Institute of Social Sciences, Belgrade, jnesic@idn.org.rs

tinian man, Eyad al-Hallaq, after being mistaken for a terrorist, was shot and killed by the Border Police on his way to a special needs school that he attended in Jerusalem.[1]

Even when prejudices towards autistic people are not that extreme, there seems to be a common belief that autistic people are inherently asocial (lacking sociability). Autistics have raised their voices against such qualifications (or prejudices). They have put forward the idea of *the double-empathy problem*, claiming that the difficulties in social interaction and communication between autistic and non-autistic (or *neurotypical*, as some autistics call non-autistics) people are a two-way issue (Milton, 2012). These stem from autistic phenomenology. The novel ideas about autism that come from autistic people themselves are integral to the neurodiversity movement that has played a crucial part in changing the perception of ASD in recent times. This raises the problem that autism is misrepresented and shows a lack of autistic personal voices being heard both in autism research and by the general public. These individuals are thus victims of *epistemic injustice*, and their epistemic agency is being neglected or thwarted.

In the current iteration of The Diagnostic and Statistical Manual of Mental Disorders (DSM-5), autism (autism spectrum disorder or ASD) is understood as a neurodevelopmental disorder which is characterized by deficits in social interaction and social communication (i.e., deficits in social-emotional reciprocity, nonverbal communicative behaviors) and repetitive patterns of behavior, restricted interests and activities (i.e., stereotyped or repetitive motor movements, insistence on sameness, highly restricted, fixated interests, hyper– or hyporeactivity to sensory input) (APA, 2013, p. 50). A common criticism of the DSM heard in modern times seems to be especially appropriate in the case of autism spectrum disorder. There is little or no mention of the first-person phenomenology of autistic persons. The philosophical and psychiatric understanding of ASD has changed since Kanner`s and Asperger`s time.[2]

Cognitivist models, like central coherence (agents give more attention to details than to global information, Happé, 1999; Happé & Frith, 1996) and mindblindness (autistic individuals fail to develop the capacity to mind-read or "mentalize", it is claimed, and lack the ability to understand mental states, hence mindblind; Baron-Cohen, 1995; Frith, 2003)

1 More about this case study can be found in Bader & Fuchs (2022).

2 It was recently revealed (Sher & Gibson, 2021) that the Soviet-Russian psychiatrist Grunya Efimovna Sukhareva gave the first clinical account of autistic children long before Kanner and Asperger. Her descriptions of autistic traits in six boys (between 2 and 14 years of age) from the 'hospital-school' at the Psychoneurological Department for Children in Moscow were published in a German journal in 1926.

were the first. Enactive and embodied accounts revolutionized how we understand cognition and autism (De Jaegher, 2013; Maiese, 2021; Krueger, 2021; Krueger & Maiese, 2018). Contemporary phenomenological accounts have emphasised that differences in autistic perception and interaction are to be sought on the pre-reflective level (Zahavi & Parnas, 2003; Bizzari, 2018; León, 2019). Along the way, and in synthesis with enactive approaches, predictive coding/processing explanations have also been put forward (Van de Cruys et al., 2014; Schilbach, 2016; the dialectical misattunement hypothesis, Bolis et al., 2017; Constant et al., 2018).

I will proceed in the following way. Section 2 will define epistemic injustice and how different kinds of inequities are inflicted upon autistic persons. In the same section, I thematize Catala et al.`s (2021) relational account of epistemic agency based on enactivism (Section 3). I take this approach as a starting point and extend it with the ecological perspective to arrive at an account of epistemic injustice in ASD within the ecological-enactive framework (Section 4). In Section 5, I discuss how to employ this integrative framework, together with phenomenology, to study autistic experience and get a better understanding of the autistic style of interaction and norms. In the end, these strategies could help fight epistemic injustice in autism, I will argue.

2. Epistemic injustice in autism

The kind of injustice that is markedly epistemic in nature, *epistemic injustice* consists of "a wrong done to someone specifically in their capacity as a knower" (Fricker, 2007, p. 1). These can refer to various mistreatments "that relate to issues of knowledge, understanding, and participation in communicative practices" (Kidd, Medina & Pohlhaus, 2017, p. 1). These unjust treatments of knowers can take the forms of exclusion, invisibility, misrepresentation, being instrumentalized and marginalized, and distrusted, to name just a few (Kidd, Medina & Pohlhaus, 2017). Miranda Fricker has distinguished two kinds of epistemic injustice in her work: testimonial injustice and hermeneutical injustice (Fricker, 2003, 2007). Testimonial injustice is inflicted when a hearer, due to prejudice and bias, reduces the credibility of the speaker`s testimony. Hermeneutical injustice comes on a more collective level (and at a prior stage) than testimonial injustice concerning participation in the process of production of knowledge. When there are "gaps in collective hermeneutical resources", one is disadvantaged in that her social experience will be hard to communicate because of those gaps in the collective/mainstream hermeneutical resources (Fricker, 2007, p. 1; Dinishak, 2021, p. 2).

Dinishak argues that there is a distinct form of hermeneutical injustice at work in the case of autism, one that concerns knowledge production, i.e. autistic autobiographies. It appears that their own first-personal accounts of autistic experience are being neglected in the formation of concepts about such experience. Using Hacking's work, she starts with considerations of the difficulties both autistics and neurotypicals face when they try to understand the behavior and experiences of one another. Hacking (2009) calls it the lack of "Köhler's phenomena"[3] in the two-way interaction and mutual understanding of autistics and neurotypicals. That is, both groups lack non-inferential, unmediated access to the mental states of the other (concept of *direct perception* in modern debates, Krueger & Overgaard, 2012). The behavior of the autistic seems completely "alien" to the observing neurotypical (and the same stands for autistic people, for example, Temple Grandin calls herself an anthropologist on Mars).

Now, the neurotypical is an age-old language used to describe their experiences, and the same could not be said about autistics; the language that will adequately describe their experiences, helping autistics themselves understand their own experience and communicate these experiences to neurotypicals, is still missing. That kind of language is now in the making, and one way to contribute to this language creation is through autistic autobiography. This is the crucial point at which autistics suffer hermeneutical injustice and hermeneutical marginalization, as Dinishak argues. Autistic people`s contributions to language and concept formation that describe their own experiences are still being neglected (Dinishak, 2021, p. 9). They are retooling and improving everyday language and "expert" language used to explain autistic behavior. Autistic biographies could help neurotypical people gain some insight into autistic experiences. This way, glimpses into the social life of autistics and "neurodivergent intersubjectivity" (Heasman & Gillespie, 2019) could be achieved.[4]

Now, focusing solely on autistic persons that are verbal and able to express their experiences would exclude a wide population of non-verbal autistics (many autistic children), which is something Dinishiak is aware of and highlights in her paper (2021, p. 12). Falling to include autistic individuals with whom autism researchers do not "share a common verbal mode of communication" and those who are nonspeaking would also be a kind of epistemic injustice (Hens, Robeyns & Schaubroeck, 2018). Testimonial and hermeneutical injustices again rear their ugly heads in these

3 Comes from Gestalt psychologist Wolfgang Köhler who pointed out expressive movements and practical behavior are, most of the time, "a good picture" of people`s inner life (Köhler, 1929, p. 250).

4 Victoria McGeer has pointed out that autistic testimonies are too often dismissed as unreliable (Boldsen, 2022; McGeer, 2005).

cases, particularly testimonial injustice. Now, the problem is how to get insight into the autistic experience when it comes to those who only rely on nonverbal modes of communication.

Lucienne Spencer, in a recent paper (Spencer, 2022), builds a case that the current definition of testimonial (in)justice should be expanded to include other forms of communication, both verbal and nonverbal. Spoken and written language difficulties are characteristic of neurocognitive disorders – intellectual disabilities, according to DSM-5, such as autism and late-stage dementia. Spencer argues that such individuals are subject to epistemic harm in the form of testimonial injustice, although they communicate non-verbally. She names this *non-verbal testimonial injustice* and uses dementia as a case study. Spencer adds autistic people (at least those that are non-verbal) as a population vulnerable to this kind of epistemic injustice (Spences, 2022, p. 6). Any ways of non-verbal communication are usually overlooked and disregarded when it comes to autistic behavior, and only close family members or carers see and understand such attempts to communicate. An autistic child's peculiar movements and gestures could be trying to convey an emotion or a desire, but only the parents would perhaps understand its meaning. Spencer employs a phenomenological framework drawn from Merleau-Ponty`s (2012) work to argue that non-verbal expressions (embodied gestures) are a meaningful form of communication. She broadens Miranda Fricker's idea of "testimonial sensibility" to "communicative sensibility" to include our ability to register other people's gestures as "epistemically loaded" (2022, p. 5).

Catala et al. warn that epistemic injustices to autistics are based on neuronormativity and neurotypical ignorance, and from this comes a specific kind of oppression. They focus on epistemic injustice that autistic people suffer from neuronormative/neurotypical biases about autistic sociability. Persuasively they argue for connections between testimonial and hermeneutical injustices and how they produce one another. Who appears as a credible knower affects who will be involved in the meaning-making. Who appears to be intelligible will influence who is viewed as credible, and so on. To argue for this, they show that there is conceptual and expressive hermeneutical injustice and find five types of epistemic injustices in this regard: "systematic testimonial injustice; preemptive testimonial injustice or quieting; testimonial smothering; contributory hermeneutical injustice; and expressive hermeneutical injustice" (Catala, Faucher, & Poirier, 2021, p. 9017). There are no adequate conceptual tools and proper terms to capture the experience of a certain group in the mainstream hermeneutical resources, in this case, autistics, and their experience is unintelligible; they cannot be understood. When conceptual and terminological developments have been made by a certain group (autistics have developed a new

language and concepts suitable for their experience), but their contribution is neglected, they are subject to contributory hermeneutical injustice (Dotson, 2012; Catala, Faucher, & Poirier, 2021, p. 9020).

In order to understand how epistemic injustice comes about and how to deal with it, Catala et al. introduce an important idea (which comes from an examination of autistic testimonials) that epistemic agency is a "fundamentally dynamical and relational process" (2021, p. 9022) as opposed to the internalist picture. This process involves not just the individual but other agents and the sociomaterial environment. According to them, epistemic injustice comes from neuronormativity and neurotypical ignorance. These types of identification force them to understand epistemic agency in this relational way. But the relational account of epistemic agency can also help us find ways to achieve greater epistemic justice.

While tracing the historical origins of the idea that agency can be dependent on the environment, authors eventually come to Varela, Thompson and Rosch's enactivism (*The Embodied Mind,* Varela et al., 1991). It is no surprise that Catala et al. turn to a different understanding of cognition to support their idea of epistemic agency as relational. In the end, they defend an enactive theory of epistemic agency that is in line with autistic experiences (Catala, Faucher, & Poirier, 2021, p. 9025). Let us unpack what this means.

3. Enactive solution

Enactivism, as a research programme, came about under the influence of ideas from biology, cognitive science and phenomenology of Merleau-Ponty (Thompson, 2007). Integrating these perspectives was the goal from the beginning (e.g., neurophenomenology, Varela, 1996). As opposed to the doctrine of cognitivism (mind/consciousness operates much like a computer with representations, and there is a clear divide between the inner and the outer world), enactivism understands cognition as embodied action that is not enclosed in the brain (or the organism that has it). The organism and the environment are dynamically coupled, making up a dynamical system. There is a "brain-body-environment" system to be accounted for, and the organism produces meaning in the world through the process of sense-making. Every live organism has consciousness, according to enactivists (*life-mind continuity thesis*; Di Paolo, 2009; Thompson, 2007). The organism's environment is meaningful; it is its ecological niche (nem. *Umwelt*; von Uexküll, 1909). One of the main tenets of enactivism is that sensory and motor processes are indivisible, entangled, perception and action in a circle.

Every type of cognition can be viewed through the enactivist lens, not just perception but intersubjectivity or social cognition (Di Paolo & De Jaegher, 2007), affectivity, and language, cognition of both the lower and higher forms. In addition, enactivism has been applied to psychiatry and psychopathology, de Haan, 2020; Maiese, 2016; for autism De Jaegher, 2013; Klin, Jones, Schultz, & Volkmar, 2003; for schizophrenia Kyselo, 2016).[5] The work of Saneke de Haan (2020) is of particular importance, given she expounds the most detailed and worked-out form of an enactive approach to psychiatry and understanding of psychiatric disorders.

In their enactive account, Catala et al. include ideas and concepts from the work of Rietveld and Kiverstein – the notions of the *landscape and field of affordances* (Bruineberg & Rietveld, 2014), as well as that of *mental institutions* (Krueger & Maiese, 2018). Both of these ideas heavily rely on the concept of affordance from the ecological psychology of Gibson (1979). Since the epistemic agency is relational, it would lead us to understand autism not as an individualistic condition but as one that comes about in the relationship between autistic people (and their norms) and neurotypical people (and their norms). Catala et al. here draw on Gallagher`s and Krueger and Maiese`s notion of neurotypical mental institutions. Authors argue that such a mental institution with its neurotypical "norm-governed practices, artefacts and traditions" (Krueger & Maiese, 2018, p. 10) sets up its own affordance landscape that is different from the affordance landscape of autistic people (the one they skillfully engage in). Now, the problem comes from the mismatch between neurotypical and autistic landscapes (and corresponding "institutions"). Autistics do not attune to neurotypical norms and the *epistemic disablement* of autistic people comes from this, as Catala et al. (2021, p. 9026) argue.

4. Ecological-enactive remedy

Since the problem of epistemic injustice and disablement comes down to the differences between neurotypical and autistic people that involve the ecological aspect (affordance landscape), it seems only natural that the enactivist account should be expanded with ecology (the famous fifth E in 4E approaches). Therefore, I think the best framework to understand epistemic injustice (and ASD more generally) is the *ecological-enactive*

5 I review the philosophical literature on enactive approaches to psychiatry and its combinations with ecological psychology in a different paper, Nešić (2022).

framework. I will use a particular EE framework – the *skilled intentionality framework* or the SIF (Rietveld, Denys, & van Westen, 2018). SIF combines embodied, enactive and ecological research programs and views cognition as skilled engagement with affordances (possibilities for action) in the sociomaterial environment, and this is how an individual tends toward the optimal grip. Part of SIF is an ecological-enactive interpretation of the free energy principle and predictive processing (Bruineberg & Rietveld, 2014).

According to SIF, members of the same species are situated within the same ecological niche, e.g., the human ecological niche. It is a rich *landscape of affordances*. These affordances correspond to the abilities available in a particular *form of life*.[6] Skilled intentionality is responsiveness to a landscape of affordances (which are relational). The landscape contains all the affordances that are available to a form of life in general (humans). These include social affordances. On the other hand, the *field of affordances* "reflects the multiplicity of inviting possibilities for action for an individual in a concrete situation" (Rietveld, Denys, & van Westen, 2018, p. 52; de Haan et al., 2013). A field of affordances is an individual "subset" of the whole landscape of affordances.

I find this delineation of totally separate landscapes of affordances of neurotypicals and autistics troublesome. This will depend on how we understand landscapes, but if we follow the ecological-enactive theory of Kiverstein and Rietveld, I think it would be wrong to posit several landscapes of affordances – there is one landscape of the human species. That said, the field of affordances of the autistic is different. Mental institutions are a useful concept and could be located somewhere between the landscape and the field of affordances. Gallagher (2018) himself acknowledges that his affordance space falls between field and landscape. Perhaps, when these authors say there are different affordance landscapes, they are just being imprecise.

Staying true to this distinction (field-landscape) and the claim that there is one landscape, and following the ecological-enactive approach, I find it useful to view autistic persons as having a different field of relevant affordances. Given that there are three dimensions to the field of affordances: width ("broadness of the scope of affordances"), depth (in terms of temporality), height (salience or "intensity of the relevance") (de Haan et al., 2013; de Haan, 2020), it can be said that autistic people have narrow

6 Rietveld and Kiverstein (2014) follow the Wittgensteinian (1953) notion of affordances. With the form of life they refer both to the kind of an animal (with an ecological niche) and to the sociocultural practices.

fields, with shallow temporal depth, and with great affective salience of the affordances that solicit them in the field.[7]

Similarly to Gallagher's notion of "disaffordances", Catala et al. introduce concepts of epistemic enablement and disablement. Epistemic disablement comes from the interactionist model of disability, saying that the disability stems from the discrepancy between the capacities of the individual and environmental conditions that can be resources or obstacles (Catala, Faucher, & Poirier, 2021, p. 9029). So factors or elements of the environment can hinder or enable certain capacities and, thus, be enabling or disabling (e.g. cultural norms). In their words, "epistemic disablement, a process that effectively removes the possibility for an individual or a group of individuals to engage in fair epistemic interactions and to successfully make fruitful epistemic contributions" (2021, p. 9031).

This is in line with the enactivist approach, but since they want to understand environmental influences, adding the ecological aspect to the enactive perspective would make more sense, that is to view the problem from the ecological-enactive approach.[8] The SIF defenders also propose an ecological-enactive model of disability (Toro et al., 2020), which emphasizes the role of a pragmatically structured sociomaterial environment in constraining and enabling behavior. Unlike the medical and social models, this model focuses on the experience of the lived body of the disabled person.

Catala et al. view enactivism as an epistemic enabler. Other ways of enabling include participatory research. The ecological-enactive approach that builds on enactivism and phenomenology would bring even more epistemic enablement. Epistemic enablement is needed to get to greater epistemic justice. Catala et al. note that enactivism enables getting to the cause of epistemic disablement, enabling greater epistemic injustice. I think that a better fit is the ecological-enactive perspective since it explicitly and in a detailed manner considers the environmental aspects. So, I defend an ecological-enactive account of epistemic injustice in ASD. From the ecological-enactive framework, we have a better perspective on what can be epistemically enabling for autistic individuals. Since the ecological-enactive framework is integrative and connects enactivism, ecological psychology and phenomenology, findings and strategies from all these disciplines can be of help.

---

7 I have developed an ecological-enactive account of autism in Nešić (2023).

8 Integrating these two approaches to cognition is not an easy endeavour. While the enactivist have criticised Gibson`s ecological theory of perception as one-sided (on the side of the environment), the ecologists pointed out that fot the enactivist environment has no meaning. See more about this in (Toro et al., 2020, p. 2).

5. Towards epistemic justice

In this section, I would like to discuss some strategies that can lead to greater epistemic enablement and justice for autistic people, given all that has been discussed so far. Spencer, in her work, has shown how phenomenology can be used to get a better understanding of non-verbal forms of communication and how they might appear in disabled people. Phenomenology can contribute to the debate surrounding epistemic injustices by exploring autistics' first-person and second-person experiences. So, in order to arrive at some strategies for greater epistemic justice, phenomenology seems like an invaluable tool. Boldsen (2021) uses a phenomenological framework based on Merleau-Ponty, which encompasses material objects and surroundings to analyze autistic social experiences and the specificity of autistic intersubjectivity. These approaches are further nicely aligned with ecological and enactive perspectives on autism. Boldsen shows that in autism, we find a different kind of intersubjectivity in which interactions include material spaces as well as bodies.

Catala et al. note similarly that in the case of non-verbal autistic persons and children and those with other intellectual disabilities, the enactivist approach can contribute to a deeper understanding of the movements and expressions of those individuals and so to the illumination of their experience (2021, p. 9034). I agree with this, and this is what participatory research built on enactivist and phenomenological foundations has been able to achieve.

For example, the psychiatric term "stereotypy" in DSM-5 designates those repetitive motor movements, like hand-flapping, finger flicking, and whole-body rocking movements (also called "self-stimulatory behaviors"). These behaviors are deemed problematic and are up for suppression and possible elimination in therapy. Now, autistics themselves have been outspokenly critical of how these types of behavior are seen and understood. They use terms like "stims"/"stimming" and "loud hands" to describe such behavior (Bascom, 2012; Kapp et al., 2019). Neurodiversity activists and autistics oppose eliminating these types of non-harmful behavior and point out that they can be seen in some instances as non-verbal means of communication. Different ways of stimming can be expressive and communicative (Bascom, 2012).

The DSM-5 has brought with it the collapse of Asperger's syndrome and autism spectrum disorder[9], and this had a negative impact on all people who identified as "Aspies" and caused a ripple in the community sur-

9 DSM-5 (APA, 2013) diverges from the fourth iteration in that it integrates previously separate categories of autistic disorder, Asperger syndrome, pervasive developmental

rounding the diagnosis (Scrutton, 2017; Giles, 2014). The patients themselves (autistic people) have not been involved or participated in defining their experience and their condition, so the diagnosis, once again, came from a third-person perspective. First-person accounts have been neglected in this discriminatory distribution of epistemic credibility.

As Catala et al. (2021) note, participatory research furnishes epistemic enablement. Others (Leary and Donnellan, 2012, p. 51) have argued that stims could be "effective ways of managing incoming sensory flows". Autistic habits of mind like self-stims have a "norm-governed character" (Krueger, 2021). De Jaegher has pointed out that there is evidence that activities like repetitive behaviors ("autistic sensorimotor and affective particularities") are connected to pleasure and well-being, though they may be seen as socially unacceptable. They are "beloved activities apparently associated with great positive valence" (Klin et al., 2007, p. 97; cited in De Jaegher, 2013, p. 10). This can be witnessed in the qualitative interviews conducted by Mercier et al. (2000; cited in De Jaegher, 2013). Such activities can have salience and relevance for autistic persons, which should be considered when dealing with the behavior – there is a possibility of "converting them into acceptable activities" rather than trying to extinguish them altogether (Krueger & Maiese, 2018, p. 27; Boyd, McDonough, & Bodfish, 2012).

Certainly, there are methodological problems with how to conduct interviews with autistics. This would seem almost impossible in the case of autistic children, who are often non-verbal. Methodological advances from 4E cognitive science and phenomenology can be epistemically enabling. Participatory research and phenomenological, semi-structured interviews provide for second-person methodologies and approaches to autistic experience that can directly include autistic individuals in the process of knowledge collection and production. Involving autistic persons in interviews, the most direct and precise tools for phenomenological data collection (see Henriksen et al., 2021), proves to be particularly difficult. For autistics to properly engage with the interviewer and supply fruitful feedback, the interview has to be set up to be conducted in special ecological and dialogical circumstances.

Consider the work of Sofie Boldsen (2022). To investigate disturbances of social experience and social interaction in autism, she uses empirical, phenomenological methods of the interview and participatory observation, working in groups with high-functioning autistics. These methods

---

disorder – not otherwise specified, and childhood disintegrative disorder into one consolidated umbrella diagnosis of autism spectrum disorder.

presuppose the use of the second-person perspective. Approaching the social experiences of autistics in this way and engaging with such experience head-on in group interactions is an invaluable way to work towards greater epistemic justice for autistic people. Similarly, participatory research conducted by enactivist such as Thomas Fuchs and Hanne de Jaegher (De Jaegher et al., 2017), in practical or empirical phenomenology through the PRISMA method ("the systematic unfolding of interactive experience"), enables us to get a better understanding of interactive experience in autism. Taking into account autistic first-person and second-person experience (as well as the second-person experience of those who engage with autistic persons) is a good remedy for epistemic injustice.

Phenomenological and enactivist accounts have stressed that we need to understand autism as a relational, "two-way phenomenon" (Krueger & Maiese, 2018), that it is not just an individual`s disorder but unfolds dialectically between a person and her sociomaterial environment (Boldsen, 2022, p. 204). Among the phenomenological strategies which help in the fight against epistemic injustice, we can now add those that come from the ecological understanding of autism. From the perspective of the skilled intentionality framework, which I find to be the most encompassing and useful one, greater epistemic justice for autistic people can be brought about through inclusive, relational changes in the landscape of affordances. Restructuring the landscape to include more appropriate affordances for autistics would allow them to feel less disabled and be able to search for and develop new skills. Since many problems for autistics come from sensory overload in the environment, for example, we (the neurotypicals) can make changes to the affordances in order to accommodate their field. This way, autistic people would be in a position to attune better to our norms and practices phenomenally.

We can get to greater enablement by designing more attractive landscapes of affordances that could promote actions from autistic people (e.g., with the arrangement of "place-affordances"). We can predict and reorder the available affordances of a particular place (as if in an art installation) to generate behavioral change in autistic subjects (see about the usefulness of the notion of *field of promoted actions,* Reed & Bril, 1996; Bruineberg et al., 2021, pp. 12834–36). The mismatch in norms between autistics and non-autistics (neurotypicals) can lead to epistemic disablement, as Catala et al. warn. However, since the disorder on the whole (and the disablement that comes with it) is constituted relationally from our side, we can work to make the sociomaterial environment more open and flexible for attunement to autistic norms.

6. Conclusion

Too often, autistic persons (especially children on the spectrum) are prejudicially discounted by neurotypicals and characterized as "not knowing anything", lacking any skills, and not being able to fit in the community. In this paper, I tried to hint at possible strategies that would be helpful in fighting epistemic injustice in autism. I endeavoured to do this by building on a recent account of epistemic injustice, that of Catala et al. (2021). They develop a relational account of epistemic agency in enactivist terms. In their account of epistemic agency, the epistemic injustice comes from neuronormativity and neurotypical ignorance, but they tried to show how enactive ways of epistemic enablement can be achieved. I aimed to argue how this framework for understanding epistemic agency and (in)justice can and should be extended by considering the ecological aspect of cognition. The appropriate encompassing framework for the task is an ecological-enactive one, the skilled intentionality, as I contended. With the ecological dimension of cognition added to the enactive one, and through notions of the field and landscape of affordances, we could see how disability (and epistemic disablement) can arise and be in a better position to find new ways to support epistemic enablement. I then argued that phenomenology, with its concepts and methods (interview and participatory observation) and as an integral part of the ecological-enactive framework, can be helpful in bringing epistemic justice to autistic people. Both phenomenological and participatory research on autism could contribute. These are all valuable strategies through which neurotypicals can eliminate prejudice against autistic people and bring greater epistemic justice for these individuals.

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of disorders (4th ed.)*. Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: American Psychiatric Association.

Bader, O., Fuchs, T. (2022). Gestalt Perception and the Experience of the Social Space in Autism: A Case Study. *Psychopathology*, 55, 211–218. DOI: 10.1159/000524562.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Bascom, J. (ed.) (2012). *Loud Hands: Autistic People, Speaking. Washington*, D.C.: Autistic Press.

Bizzari, V. (2018). Like in a Shell: Interaffectivity and Social Cognition in Asperger's Syndrome. *Thaumàzein*, 6, 158–179.

Boldsen, S. (2021). Social interaction style in autism: An inquiry into phenomenological methodology. *Journal of Phenomenological Psychology*, 52(2),157–192. https://doi.org/10.1163/15691624–12341389.

Boldsen, S. (2022). Material Encounters: A Phenomenological Account of Social Interaction in Autism. *Philosophy, Psychiatry, & Psychology,* 29(3), 191–208. doi:10.1353/ppp.2022.0039.

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C., Schilbach, L. (2017). Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity. *Psychopathology,* 50(6), 355–72. https://doi.org/10.1159/000484353.

Boyd, B. A., McDonough, S. G., Bodfish, J. W. (2012). Evidence-Based Behavioral Interventions for Repetitive Behaviors in Autism. *Journal of Autism and Developmental Disorders*, 42(6), 1236–1248.

Bruineberg, J., Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, 8, 1– 14.

Bruineberg, J., Seifert, L., Rietveld, E., Kiverstein, J. (2021). Metastable attunement and real-life skilled behavior. *Synthese*, 1–24. https://doi.org/10.1007/s11229–021-03355–6.

Catala, A., Faucher, L., Poirier, P. (2021). Autism, epistemic injustice, and epistemic disablement: a relational account of epistemic agency. *Synthese,* 199, 9013–9039, https://doi.org/10.1007/s11229–021-03192–7.

Constant, A., Bervoets, J., Hens, K. et al. (2018). Precise Worlds for Certain Minds: An Ecological Perspective on the Relational Self in Autism. *Topoi,* 39, 611–622 https://doi.org/10.1007/s11245–018–9546–4.

de Haan, S. (2020). *Enactive Psychiatry*. CUP

de Haan, S., Rietveld, E., Stokhof, M., Denys, D. (2013). The phenomenology of deep brain stimulation-induced changes in OCD: an enactive affordance-based model. *Frontiers in Human Neuroscience*, 7(653), 1– 14.

De Jaegher, H. (2013). Embodiment and sense-making in autism. *Frontiers in Integrative Neuroscience*, 7.

De Jaegher, H., Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6 (4), 485–507.

De Jaegher, H., Froese, T. (2009). On the role of social interaction in individual agency. *Adaptive Behavior,* 17(5), 444–460.

De Jaegher, H., Pieper, B., Clénin, D. et al. (2017). Grasping intersubjectivity: an invitation to embody social interaction research. *Phenom Cogn Sci* 16, 491–523. https://doi.org/10.1007/s11097–016–9469–8.

Di Paolo, E. (2009). Extended life. *Topoi*, 28 (1), 9–21.

Dinishak, J. (2021). Autistic autobiography and hermeneutical injustice. *Metaphilosophy* 00, 1–14. https://doi.org/10.1111/meta.12514.

Dotson, K. (2012). A cautionary tale: On limiting epistemic oppression. *Frontiers*, 33(1), 24–47

Fricker, M. (2003). Epistemic Injustice and a Role for Virtue in the Politics of Knowing. *Metaphilosophy* 34, 1–2, 54–73.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.

Frith, U. (2003). *Autism: Explaining the Enigma*. Hoboken, NJ: Wiley-Blackwell.

Fuchs, T. (2015). Pathologies of Intersubjectivity in Autism and Schizophrenia. *Journal of Consciousness Studies,* 22(1–2), 191–214.

Fuchs, T. (2019). The Interactive Phenomenal Field and the Life Space: A Sketch of an Ecological Concept of Psychotherapy. *Psychopathology,* 52, 67–74. DOI: 10.1159/000502098.

Gallagher S. (2018). The therapeutic reconstruction of affordances. *Res Philosophica*, 95(4), 719–736.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin

Giles, D. (2014). 'DSM-V is taking away our identity': The reaction of the online community to the proposed changes in the diagnosis of Asperger's disorder, *Health,* 18.2, 179–195.

Happé, F. (1999). *Autism*. London, UCL Press.

Happé, F., Frith, U. (1996). The neuropsychology of autism, *Brain,* 119, 1377–1400.

Heasman, B., Gillespie, A.. (2019). Neurodivergent Intersubjectivity: Distinctive Features of How Autistic People Create Shared Understanding. *Autism,* 23, no. 4, 910–21.

Henriksen M.G., Englander M., Nordgaard J. (2021). Methods of data collection in psychopathology: the role of semi-structured, phenomenological interviews. *Phenom Cogn Sci*., https://doi.org/10.1007/s11097–021-09776–5.

Hens, K., Robeyns. I., Schaubroeck, K. (2018). The Ethics of Autism. *Philosophy Compass,* 14, no. 2:E12559.

Kapp, S. K., Steward, R., Crane, L., Elliott, D., Elphick, C., Pellicano, E., Ginny R. (2019). 'People Should Be Allowed to Do What They Like': Autistic Adults' Views and Experiences of Stimming. *Autism,* 23, no. 7, 1782–92.

Kidd, I., Medina, J., Pohlhaus, G. Jr., eds. (2017). *The Routledge Handbook of Epistemic Injustice*. London: Routledge.

Klin, A., Danovitch, J. H., Merz, A. B., and Volkmar, F. R. (2007). Circumscribed interests in higher functioning individuals with autism spectrum disorders: an exploratory study. *Res. Pract. Pers. Sev. Disabil*., 32, 89–100.

Klin, A., Jones, W., Schultz, R., Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358 (1430), 345–360.

Köhler, W. (1929). *Gestalt Psychology*. New York: Horace Liveright.

Krueger, J. (2021). Enactivism, Other Minds, and Mental Disorders. *Synthese,* 198, 365–389.

Krueger, J., Maiese, M. (2018). Mental institutions, habits of mind, and an extended approach to autism. *Thaumàzein,* 6, 10–4.

Krueger, J., Overgaard, S. (2012). Seeing Subjectivity: Defending a Perceptual Account of Other Minds. *ProtoSociology,* 47, 239–62.

Kyselo, M. (2016). The enactive approach and disorders of the self. The case of schizophrenia. *Phenomenology and the Cognitive Sciences*, 15 (4), 591–616.

Leary, M. R., Donnellan, A. M. (2012). *Autism: Sensory-movement Differences and Diversity*. Cambridge: Cambridge Book Review Press.

León, F. (2019). Autism, social connectedness, and minimal social acts. *Adaptive Behavior*, 27(1),75–89.

Maiese, M. (2016). *Embodied selves and divided minds*. Oxford: Oxford University Press.

Maiese, M. (2021). Autism as disordered sense-making. *Constructivist Foundations,* 17(1): 056–058. https://constructivist.info/17/1/056.

McGeer, V. (2005). Out of the mouths of autistics: Subjective report and its role in cognitive theorizing. In A. Brook & K. Akins (Eds.), *Cognition and the brain: The philosophy and neuroscience movement*. Cambridge: Cambridge University Press.

Mercier, C., Mottron, L., Belleville, S. (2000). A psychosocial study on restricted interests in high functioning persons with pervasive developmental disorders. *Autism,* 4, 406–425.

Merleau-Ponty, M. (1945/2012). *Phenomenology of Perception*. Translated by D. A. Landes. Abingdon, Oxon: Routledge.

Milton, Damian E.M. (2012). On the ontological status of autism: the 'double empathy problem'. *Disability & Society*, 27:6, 883–887. DOI: 10.1080/09687599.2012.710008.

Nešić, J. (2022). Enaktivizam kao okvir za psihijatrijske poremećaje. *Engrami,* Vol. 44, (1). https://doi.org/10.5937/engrami44–40298.

Nešić, J. (2023). Ecological-enactive account of autism spectrum disorder. Synthese 201, 67. https://doi.org/10.1007/s11229-023-04073-x

Newen, A., de Bruin, L., Gallagher, S. (Eds.) (2018). *Oxford handbook of cognition: Embodied, enactive, embedded and extended*. Oxford: Oxford University Press.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Reed, E., Bril, B. (1996). The primacy of action in development. A commentary of N. Bernstein. In M. L. Latash (Ed.), *Dexterity and its development* (pp. 431–451). Erlbaum.

Rietveld E., Kiverstein J. (2014). A rich landscape of affordances. *Ecol Psychol,* 26(4), 325–352.

Rietveld, E., Denys, D., van Westen, M. (2018). Ecological-ecological-enactive cognition as engaging with a field of relevant affordances: The Skilled Intentionality Framework (SIF). In A. Newen, L. de Bruin, & S. Gallagher

(Eds.), *Oxford handbook of cognition: Embodied, enactive, embedded and extended*. Oxford: Oxford University Press.

Schilbach, L. (2016). Towards a second-person neuropsychiatry. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 371(1686),20150081. https://doi.org/10.1098/rstb.2015.0081.

Scrutton, A. P. (2017). Epistemic Injustice and Mental Illness. In Ian James Kidd, José Medina, and Gaile Pohlhaus Jr. (Eds.), *The Routledge Handbook of Epistemic Injustice*, (pp. 347–55). London: Routledge.

Spencer, L. (2022). Epistemic Injustice in Late-Stage Dementia: A Case for Non-Verbal Testimonial Injustice. *Social Epistemology*. DOI: 10.1080/02691728.2022.2103474.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge. MA: Harvard University Press.

Toro, J., Kiverstein, J., Rietveld, E. (2020). The Ecological-Enactive Model of Disability: Why Disability Does Not Entail Pathological Embodiment. *Frontiers in Psychology,* 11: 1162.

Van De Cruys, S., Eveers, K., Van der Hallen, R., Van Eylen, L., Boets, B., De-Witt, L., Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review,* 121(4), 649–675.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies,* 3(4), 330–49.

Varela, F., Thompson, E., Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

von Uexküll, J. (1909). *Umwelt und Innenwelt der Tiere*. Berlin: Springer.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

Zahavi, D., Parnas, J. (2003). Conceptual Problems in Infantile Autism Research: Why Cognitive Science Needs Phenomenology. *Journal of Consciousness Studies*, 10(9–10), 53–71.

Ivan Umeljić*

Petar Nurkić**

WHAT ARE WE TALKING ABOUT WHEN WE TALK ABOUT SCIENTIFIC OBJECTIVITY?

Abstract: Philosophers of science often suggest that the key feature of scientific research is striving for objectivity and that we should evaluate scientific practice by whether it is objective or not. In this paper, we will analyze several definitions of scientific objectivity to illustrate the complex meaning of this term and examine its role in evaluating scientific practice. First, we will introduce Lorraine Daston and Peter Galison's standpoint concerning the historical connection between the genesis and development of scientific objectivity and the practices of visual representation in the research practice of the 19th and 20th centuries. We will accomplish that by outlining the process of establishing scientific objectivity as an epistemic virtue and a vital feature of the "scientific self". Subsequently, using Heather Douglas and Marianne Janack's conceptual analysis of scientific objectivity, we will show that scientific objectivity is characterized by an "irreducibility of meaning" and an "endemic instability" caused by the overuse of metaphors in defining this concept. In the final section, in light of contemporary problems such as the crisis of reproducibility, we examine to what extent philosophical definitions help test the objectivity of scientific practice and point to an intriguing attempt to define "objectivity for the research worker" using the model proposed by Noah van Dongen and Michał Sikorski.

Keywords: scientific objectivity, scientific self, conceptual analysis, scientific research, reproducibility.

1. Introduction

Scientists are not isolated from society. In this sense, like all other citizens, they should respect the rights and property of other people, not

* Center for the Promotion of Science, 1977, iumeljic@cpn.rs

** Institute for Philosophy, Faculty of Philosophy University of Belgrade, 1991, petar.nurkic@f.bg.ac.rs

harm them, but aid them and comply with the law. As experts from a particular field, in addition to civic duties, scientists have specific obligations because of their distinct roles in the broader community, which consequently rewards them with a certain authority and autonomy. The broader inventory includes a whole range of norms that should be fulfilled in different domains of professional, scientific practice: impartiality, honesty, objectivity, openness, recognition of colleagues, respect for intellectual property, respect for colleagues, competence, legitimacy, social responsibility, efficient and responsible use of resources, verifiability, coherence, empirical support, precision, economy, etc. (Resnik, 2006: 36).

As we can notice, some of the mentioned norms are moral in nature, others come from law, and others are epistemic. Most frequently, objectivity stands out as the most important of all norms and is also the primary feature of a scientific enterprise with a dual, moral, and epistemic character (Daston & Galison, 2007: 42; Resnik, 2006: 45).

Objectivity is a trait of scientific conclusions, methods, and results that excludes personal social, economic, and political biases from the procedures of experimental design, testing, analysis, review, and publication (Reiss & Sprenger, 2017; Resnik, 2006: 35). Moreover, other moral and epistemic norms, such as honesty, openness, empirical support, verifiability, and precision, can be founded on different approaches to objectivity (Resnik, 2006: 52). Furthermore, if we glance at things from a broader perspective, objectivity is a normative ideal, like justice, virtue, or piety (*Ibid*).

In this article, we will analyze from different angles the connection between the epistemic authority of science and the concept of objectivity. First, we will consider some noteworthy philosophical definitions and then examine whether philosophers have provided a satisfactory conceptualization of objectivity that might be of practical use to researchers today. Given that objectivity is not only an incredibly vast topic but also a somewhat controversial one, it should be stated that we have covered it only partially. We have tried to provide a balanced and concise display of some of the relevant viewpoints and discussions in the paper, but understandably we have yet to address many issues.

We begin our articulation with the central ideas introduced in the most significant study of objectivity in this century, the book *Objectivity* by Lorraine Daston and Peter Galison, in which the authors present the history of this epistemological concept.[1] Genre-wise, some authors

1 "This is a book for meditation and loaning to friends. It's a book prize for the best undergrad in the class. The bounty of information, the charm of anecdote, the care with which each sentence is composed, the elegance of illustration, the power of the

assign this book to the history of science, others to historical epistemology, and still others, such as Ian Hacking, believe that it is a work from the field of meta-epistemology (Hacking, 2015: 19).[2] In any case, this comprehensive study of the concept of objectivity, its use over the last two centuries, and the associated practices is an indispensable reference for any discussion of this topic. For that reason, in the following section, we will take a closer look at the most significant insights and arguments related to it.

2. A brief history of scientific objectivity

Briefly, the central thesis that Daston and Galison advocate in their book *Objectivity* is that the connotation of terms associated with "objectivity" has varied considerably over the past two centuries (Daston & Galison, 2007: 35). Since the 19th century, "objectivity has had its prophets, philosophers, and preachers," (Daston & Galison, 2007: 17) but its distinctiveness, as Daston and Galison point out, was most evident in the example of a regular scientific practice – the production of images. For this reason, the two renowned historians of science have chosen to portray the history of objectivity through images from scientific atlases, or rather, through selected assemblages of images that served to identify the major research subjects in particular disciplines. Their comprehensive analysis of naturalistic illustrations from the eighteenth century onward reveals, as they state, three distinctive forms of objectivity– "truth to nature", "mechanical objectivity" and "trained judgement" (*Ibid*).

Their analysis commences with the recognition that throughout most of the eighteenth and nineteenth centuries, the process of constructing scientific knowledge was in many ways analogous to the approach of creating a work of art (Daston & Galison, 2007: 35; Ambrosio, 2015: 354). Namely, as scientists of the time sought to "capture nature in its ideal form", they

analysis, and the formidable structure of the whole". With these words, Ian Hacking began his talk at the *Objectivity from a Historical Perspective* roundtable dedicated to this book. Hacking was followed by talks of Peter Dear, Matthew L. Jones, and authors Lorraine Daston and Peter Galison (Dear, Hacking, Jones, et al. 2012: 17).

2 *Historical epistemology* is a collective term for several diverse approaches to studying the history of epistemic concepts such as objectivity, observation, experimentation or probability, as well as the historical trajectories of research subjects such as the electron, DNA, or the phlogiston. This term also refers to the primary research direction of the Max Planck Institute for the History of Science in Berlin, founded in 1994 and led by Lorraine Daston, who has contributed significantly to the popularity of this approach among historians of science (Feest & Sturm, 2011: 286).

paid attention to how they could depict a particular individual plant or animal as a "representation" of that ideal when preparing illustrations for atlases (Daston & Galison, 2007: 66). Objectivity, in this case, is defined by its affinity to realism-what Daston and Galison anoint a "truth to nature" perspective.

Pictures served the ideal of truth-and oftentimes, the ideal of beauty, along with the truth. "Truth to nature" requires a thorough knowledge of diversity and deviations in nature in order to "perfect" the individual phenomena found around us (Daston & Galison, 2007: 104). The scientists of the 18th and the first half of the 19th century had the "duty to correct nature for the sake of truth" (Ambrosio, 2015: 354): Their illustrations show that for them, representation was inseparable from the act of discernment, which meant visualizing, not individual natural phenomena, but their ideal manifestations.

The advent of photography brought about a radical change. Since 1839, when the first daguerreotype was displayed at the French Academy of Sciences, the status of photography has been the subject of heated debate. Initially, scientists believed it was the ultimate tool for achieving accurate observation and measurement. Its mechanical and reproducible character was the rationale for believing that the camera functioned as a kind of "artificial retina", devoid of subjective perspective (Daston & Galison, 2007: 187; Ambrosio, 2015: 358). By the end of the 19th century, photography was also being utilized to observe phenomena that were otherwise considered imperceptible. It also found its use to measure and obtain experimental records. Daston and Galison associated the emergence of the contemporary concept of scientific objectivity with photography's advent.

Although the concept of "mechanical objectivity" extended to a broader range of scientific instruments, they singled out photography as the principal reason that led scientists to adopt a non-interventionist attitude toward the subject of their research. Mechanical reproducibility contrasted sharply with the ideal of "truth to nature", in which the willful intervention of the researcher lends credibility and scientific status to the pictures. In contrast, "mechanical objectivity" mandates the researcher to adopt an *ascetic attitude* toward the object of scientific inquiry. Human intervention is substituted by the procedural use of technologies that ensure that the scientist's judgment is truncated in the visualization process. This form of objectivity went hand in hand with the increasing reliance of scientists on recording and measuring instruments, which, like the camera, promised to be able to thoroughly eliminate the human factor (Daston & Galison, 2007: 122; Ambrosio, 2015: 358; Christin, 2016: 27).

The increasingly frequent reliance on technologies in scientific practice has brought with it new moral and epistemic acuities.[3] As Daston and Galison noted, the virtues attributed to machines were stressed as models for humans to emulate, the most important being associated with diligence and a dedicated and focused work ethic. In addition, machines had the unique advantage of not comprehending theories and being unable to think about them, which held them from the inevitable bias characteristic of humans. Daston and Galison distinguish this widespread belief in machines' superior objectivity and quality as a paradigm of "mechanical objectivity" (Daston & Galison, 2007: 123).[4]

Over time, however, researchers realized that adherence to "mechanical objectivity" had its price: the machines registered only a small part of the natural phenomena the scientists wanted to record or left their imprints on objects that were not there. It turned out that the machine's photograph, imprint, or X-ray often required clarification and misled researchers, mainly because it contained too much information which was largely irrelevant or implausible.[5] In the late nineteenth and early twentieth centuries, a new paradigm emerged, "trained reasoning", in which scientists again started to rely on their expertise and experience to add their interpretation to the data provided by the machine, for example, by adding complex color schemes or combining different images to obtain composite vistas. For example, solar magnetogram require a trained expert to "extract" the correct signal from the data registered by the instruments. Daston and Galison cite this example as an illustration of "trained judgement" (Daston & Galison, 2007: 21).

3 For more on the ethical and epistemic challenges and complexities of relying on automation and mechanization in scientific practice, see Kušić & Nurkić (2019).

4 Stanford sociologist Angèle Christin sees a contemporary version of the ideal of mechanical objectivity in viewpoints that describe algorithms as value-neutral tools of rationalization and objectivity, as opposed to human individuals whose thoughts are shaped by various biases rendered by class, race, gender, or political attitudes (Christin, 2016: 28). Christin points out that *Big Data* analysis, which has fundamentally transformed practice in numerous scientific fields, is increasingly described "as the cure for 'broken' systems shaped by long histories of bias, inefficiency, and discrimination" (*Ibid*). Christin aside, Galison himself has criticized the assertions that algorithms managing artificial intelligence are more objective than human experts in procedural, methodological, and value terms (Galison, 2019).

5 In her study, in which she converses about how the views of artistic photographers influenced practices of visual representation in science, Chiara Ambrosio points out how pictorialists looked with scorn at the widespread attitude among scientists about the objective nature of photographs (Ambrosio, 2015: 359). A very frank polemic often had sarcastic undertones, as in a 1903 brief article entitled "Ye-Fakers," in which the pictorialist photographer Edward Jean Steichen explicitly ridiculed the asceticism preached by advocates of mechanical objectivity (Steichen, 1903: 48).

We conclude this brief review of the history of visual representation in scientific atlases by noting that objectivity has permanently moved along two tracks – one involving the development of the epistemology of scientific practices and the other leading to the adoption of the distinctive moral virtues that Daston and Galison refer to as the "scientific self" (Daston & Galison, 2007: 229).[6]

Having sketched the unusual historical development of scientific objectivity, we discuss below several influential viewpoints that reveal the tribulations we may encounter in attempting to define this concept. After depicting the views of Heather Douglas, Marianne Janack, and Ian Hacking, in the final part we will explore whether researchers can rely on and apply an appropriate conceptualization of objectivity in practice.

3. Endemic instability

It is unnecessary to emphasize that some of the most important questions within the philosophy of science have to do with objectivity in one way or another. We will only enumerate a few here: the problem of induction; the criteria for preferring a theory; the realism/anti-realism debate; scientific explanation; experimentation; quantification; application of statistics; the role of values in science; feminism. For instance, when articulating "epistemic risks", objectivity is viewed through the prism of the problem of induction, the notion of "procedural objectivity" is associated with experimentation, and "statistical objectivity" with the application of statistics, and so on (Harding, 2015; Biddle & Kukla, 2017; Douglas, 2004; Freese & Peterson, 2018). For a broader understanding of discussions of objectivity in the philosophy of science, an overview of a range of additional issues is necessary beyond our article's scope.

6 To illustrate, here is the opening paragraph of the first chapter of the book *Objectivity*, entitled "The Epistemologies of the Eye". "Scientific objectivity has a history. Objectivity has not always defined science. Nor is objectivity the same as truth or certainty, and it is younger than both. Objectivity preserves the artifact or variation that would have been erased in the name of truth; it scruples to filter out the noise that undermines certainty. To be objective is to aspire to knowledge that bears no trace of the knower — knowledge unmarked by prejudice or skill, fantasy or judgment, wishing or striving. Objectivity is blind sight, seeing without inference, interpretation, or intelligence. Only in the mid-nineteenth century did scientists begin to yearn for this blind sight, the "objective view" that embraces accidents and asymmetries, Arthur Worthington's shattered splash-coronet. This book is about how and why objectivity emerged as a new way of studying nature, and of being a scientist." (Daston & Galison, 2007: 17).

Before turning to some contemporary philosophical critiques of the concept of objectivity, let us briefly consider the interpretations of some of the most prominent philosophers of science of the 20th century. Underlying the viewpoint advocated by the leading proponents of logical empiricism is the conviction that facts are "out there somewhere" in the external world and that it is the scientist's mission to uncover, analyze and systematize them. Objectivity is the measure of whether they have been triumphant in this endeavor. In this sense, science is objective to the extent that it succeeds in discovering and generalizing facts and abstracting them from the subjective perspective of the individual scientist (Reiss & Sprenger, 2017).[7]

Robert Nozick equates objectivity with *invariance* and utilizes "objectivity" as a modifier for "truth" and "fact." Invariance, according to Nozick, is what remains when one abstracts from other properties of objectivity – accessibility from different perspectives, possibilities of intersubjective agreement, and the independence of a given truth or fact *p* from human "beliefs, desires, hopes, and observations or measurements that *p* is" (Nozick, 1998: 21). In contrast to Nozick, who focuses on the interactions between man and the world in his account of objectivity as invariance, Thomas Nagel refers to individual thought processes in his account of objectivity as *aperspectivism*, i.e., a "view from nowhere", while Bernard Williams refers to objectivity as an *absolute concept* (Nagel, 1986; Williams, 1985).

Recently, one of the critical topics of debate on scientific objectivity is the proliferation of meanings of this term (John, 2021: 4). Namely, its semantic richness is reflected in the multitude of possible categorizations and subdivisions, as Heather Douglas and Marianne Janack point out.

Although objectivity is one of the most prevalent concepts in the philosophy of science and epistemology, Heather Douglas believes that we are dealing with one of the most ill-defined terms. Douglas points out that every time we reach for objectivity, we appeal to its rhetorical power and say, "I endorse this and you should too" (Douglas, 2004: 453). In other words, and a milder form, we should trust the outcome of the process that objectivity produces. In exploring whether objectivity hauls with it something else besides this persuasive power and call to trust, Douglas was able to articulate eight distinct, operationally accessible meanings. Unlike many of her predecessors whose views we have mentioned, she concluded

7 In his book *The Advancement of Science: Science Without Legend, Objectivity Without Illusions* (Oxford University Press, 1993), Philip Kitcher criticizes this view and ironically calls it the "Legend" of how successive generations of scientists have written the entire true history of the world (Kitcher, 1993: 3).

that none of these eight meanings could be strictly reduced to one another, making objectivity an *irreducibly complex concept* (Douglas, 2004: 465).

Marianne Janack went a step further than Douglas. She noted the striking tendency in philosophic attempts to define objectivity – relying on the ideal of perspective in explaining something that is the opposite of perspective (Janack, 2002: 274). Janack's critique is directed at the overuse of metaphors in the conceptual analysis of the notion of objectivity. Without denying the importance of metaphors as a heuristic tool for our understanding of the world, she contends that in the case of defining objectivity, the problem is that perspectival metaphor is both a "cognitive frame for the concept" and an "explanation of the concept" (*Ibid*). "The 'frame'", as she states, "undermines the 'target' of the metaphor" because "we use the idea of perspective to explicate the ideal of perspectivelessness" (Janack, 2002: 275) – and so we get paradoxical definitions such as "a view from nowhere" that is a perspective that is not a perspective at all and the like, which is consistent with Lorraine Daston's supposition that the historical rise of scientific objectivity began precisely with the "escape from a perspective" (Daston, 1992: 598).

Janack also assumes that metaphorical determinations of objectivity are characterized by an endemic conceptual instability that she thinks is inevitable.[8] To reinforce this, she itemizes no fewer than 13 diverse meanings she has encountered in her inquiry of the relevant literature (Janack, 2002: 275):[9]

1. Objectivity as value neutrality;
2. Objectivity as lack of bias, with bias understood as including:
   a) personal attachment;
   b) political aims;
   c) ideological commitments;
   d) preferences;
   e) desires;
   f) interests;
   g) emotion.
3. Objectivity as scientific method;
4. Objectivity as rationality;
5. Objectivity as an attitude of "psychological distance";

8 For more information on other contexts of conceptual instability, see Nurkić (2022).

9 Janack verbatim states at one point, "philosophers and scientists writing on objectivity seem to abandon themselves to this 'drive to metaphorize' with nary a blink" (Janack, 2002: 274).

6. Objectivity as "world-directedness";
7. Objectivity as impersonality;
7. Objectivity as impartiality;
8. Objectivity as having to do with facts;
9. Objectivity, as having to do with things as they are in themselves; objectivity as universality;;
10. Objectivity as disinterestedness;
11. Objectivity as commensurability;
12. Objectivity as intersubjective agreement.

When we try to apprehend what is actually denoted by objectivity, we undergo, as Janack suggests, "a dizzying array of different kinds of virtues, ideals, metaphysical positions and psychological states" (Janack, 2002: 276), emphasizing that science is no exception in this regard. Not only is the internal use of this term no more uniform in the domain of science than in other fields, but her research has revealed that scientific back-and-forths draw on all of the above meanings of objectivity. To make matters more ominous, among the subcategories clustered around the second connotation ("objectivity as lack of bias") from Marianne Janak's inventory are terms from the domains of law and politics, which are often cited as epistemic ideals in scientific discussions. In the following section, we will return to this issue, considering the usefulness of the philosophical conceptualization of objectivity for researchers. Before proceeding to the analysis from the outlook of scientific practice, we will also mention another engaging critique that starts from the meaning of the concept of objectivity, put forward by Ian Hacking.

In one of his seminal books, *Social construction of what?,* Hacking notes that words such as "fact", "truth", "reality", or "knowledge" often operate at a different level than words used to denote ideas or objects, as he refers to them as "elevator words" (Hacking, 1999: 22). Support for this is found in Willard van Orman Quine's analysis of the terms mentioned above, according to which they serve "semantic ascent". Hacking argues that "facts, truths, reality, and even knowledge are not objects in the world, like periods of time or little children, fidgety behavior, or loving-kindness." (*Ibid*). "These terms are on a higher plane," Hacking acknowledges. He considers "objectivity" as one of them, which he asserts is not a virtue but instead accentuates the absence of vice. Such notions, he points out, lead to grandiose controversies that sound important but are empty (Hacking, 2015: 24).[10]

10 Hacking illustrates this with the following question, "Whose research in climate science meets the standards of scientific objectivity?" (Hacking, 2015: 20).

4. Objectivity in scientific reasoning

One of the strongest arguments in favor of scientific realism is that the only satisfactory explanation for the success of scientific theories is that they are true (or approximately accurate or proper in those respects that account for their success). This point of view is sometimes called the "ultimate argument for scientific realism" (Musgrave, 1988: 229). Without going further into the quarrels between realists and relativists in the philosophy of science, we would like to state at the beginning of the final chapter that the success of science is indisputable and that this is undoubtedly one of the main reasons why science is ascribed objective character and epistemic authority. We will also explore the extent to which current philosophical conceptualizations can be helpful to researchers and their practice, and draw attention to an interesting step in this direction. In particular, we will present a recent attempt to make the concept of objectivity advantageous for solving annoyances related to the crisis of reproducibility of the results of scientific theories (van Dongen & Sikorski, 2021: 2).

In recent years, as we know, numerous concerns in the scientific community have increasingly come to light, often labeled as unethical behavior, albeit for various reasons. Some involve overt fraud (such as fabrication and plagiarism), while others are somewhat more subtle but generally much more present and detrimental to the broader scientific community (Ioannidis, 2005; Open Science Collaboration, 2015). The collective term for this overall group of problems is the "crisis of reproducibility", which can be interpreted as being caused by a lack of scientific objectivity (van Dongen & Sikorski, 2021: 2). Existing philosophical theories of objectivity do not equip scientists with an appropriate conceptual framework to apply and improve their practice and eradicate (or at least reduce the likelihood of) the occurrence of this nuisance. One of the first and more substantial steps in this direction was recently undertaken by Noah van Dongen and Michal Sikorski to supply researchers with an empirically and methodologically sound inventory of facets that undermine scientific practice in their various domains (van Dongen & Sikorski, 2021: 8). They emphasize the conceptual framework that highlights the methodological quality of the research and the results obtained.

Van Dongen and Sikorski stress that their approach focuses on scientific problems that result from concrete decisions and practical actions by researchers. What exactly does this imply? Primarily the exclusion of several factors that are not under the immediate control of scientists and that have often been mentioned in eclectic definitions of objectivity. For

example, Dongen and Sikorski ruled out issues of an ethical, financial, and political nature, but also some specific external factors, such as limited access to samples, instruments, or the policy to which most scientific journals are committed, namely to publish articles conveying experiments with positive results (van Dongen & Sikorski, 2021: 9).[11] The problems mentioned are beyond the control of individual scientists. They concern the position of science and the scientific community in the broader societal context rather than the verifiable practical procedures and decisions of scientific workers.

Van Dongen and Sikorski focused on the specific decisions and actions of the researcher before, during, and after the research. Namely, before the research, the scientist can make *a priori* decisions about the design of the experiment and the method of data collection, which can reduce/increase the likelihood of an outcome and thus open the door to bias (van Dongen & Sikorski, 2021: 8). After the research, a similar approach can be taken to process and analyze the data by straining all combinations until the desired (positive) result is achieved (van Dongen & Sikorski, 2021: 9). Their presumption of *objectivity for the research worker*, which we briefly conveyed, implies a verifiable conceptualization that would prevent the emergence of intricate practices during research. Dongen and Sikorski have furnished a model of this conceptualization that they hope will soon grow into tangible protocols for verifying the objectivity of research in various scientific fields (van Dongen & Sikorski, 2021: 19–22).

5. Conclusion

Objectivity is an epistemic virtue or norm that invokes moral values on the one hand and pragmatic efficiency in ensuring the acquisition and verification of knowledge on the other. As Daston and Galison put it, "epistemic virtues earn their right to be called virtues by molding the self, and the ways they do so parallel and overlap with the ways epistemology is translated into science." (Daston & Galison, 2007: 41). In the previous part of our paper, we attempted to provide three possible answers to how epistemology is translated into science. First, we approached the question of what we are talking about when we speak of scientific objectivity from a historical perspective, then from the angle of conceptual analysis, and finally from the position of scientific practice. From there, we have drawn several valuable conclusions.

11 This last type of bias can shut the door on authors describing experiments with negative results, influencing the skyrocketing publication rate of articles with false positives.

Regarding the historical side of objectivity, we can conclude that everyone engaged in science evaluated their work to the extent that it fit the distinctive kind of "scientific self" they cultivated. Conceptual analysis has revealed that one of the key features of objectivity is conceptual instability due to the fact that philosophers often resort to metaphors when trying to define it. Finally, as far as scientific practice is concerned, it has been ascertained that objectivity is not so manageable to verify and evaluate but that there are exciting attempts in this direction that could contribute to the solution of some accumulated tribulations that have burdened scientists and the scientific community in recent years (van Dongen & Sikorski, 2021: 19–22). Finally, we would like to reiterate that objectivity is quite an extensive and controversial topic. Although we have done our best to make our analysis and the selection of topics we confer relevant and congruous, it is understandable that we still need to address some issues.

References

Ambrosio, C. (2014). Objectivity and Visual Practices in Science and Art. In Galavotti, M.C., Dieks D., Gonzalez W.J., et al. (eds.). *New Directions in the Philosophy of Science.* The Philosophy of Science in a European Perspective 5, Springer.

Biddle, J.B. & Kukla, R. (2017). The Geography of Epistemic Risk. in Elliott, K.C. & Richards, T. (eds), *Exploring Inductive Risk: Case Studies of Values in Science*. Oxford Academic.

Christin, A. (2016). From daguerreotypes to algorithms: machines, expertise, and three forms of objectivity. *ACM SIGCAS Computers and Society*, 46 (1), 27–32.

Daston, L. (1992). Objectivity and the Escape from Perspective. *Social Studies of Science*, 22(4), 597–618.

Daston, L. & Galison, P. (2007). *Objectivity*. Zone Books.

Dear, P., Hacking, I., Jones, M.L. et al. (2012). Objectivity in historical perspective. *Metascience* 21, 11–39.

Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese* 138 (3), 453–473.

Feest, U. & Sturm, T (2011). What (Good) Is Historical Epistemology? Editors' Introduction. *Erkenntnis*. (75), 285–302.

Freese, J. & Peterson, D. (2018). The Emergence of Statistical Objectivity: Changing Ideas of Epistemic Vice and Virtue in Science. *Sociological Theory* 36(3), 289–313.

Galison, P.L. (2019). Algorists Dream of Objectivity. u: Brockman, J (ed) *Possible Minds: 25 Ways of Looking at AI*. Penguin Publishing Group.

Hacking, I. (1999). *The Social Construction of What?*. Harvard University Press.

Hacking, I. (2015). Let's Not Talk About Objectivity. In Tsou, J.Y., Richardson, A. & Padovani F. (eds.) *Objectivity in Science*. Springer Verlag.

Harding, S. (2015). Objectivity for Sciences from Below. In Tsou, J.Y., Richardson, A. & Padovani F. (eds.) *Objectivity in Science*. Springer Verlag.

Ioannidis, J.P.A. (2005). Why Most Published Research Findings Are False. *PLoS Med* 2(8): e124

Janack, M. (2002). Dilemmas of objectivity, *Social Epistemology*, 16(3), 267–281.

John, S. (2021). *Objectivity in Science*. Cambridge University Press.

Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.

Kušić, M., & Nurkić, P. (2019). Artificial morality: Making of the artificial moral agents. *Belgrade Philosophical Annual*, (32), 27–49.

Musgrave, A. (1988). The Ultimate Argument for Scientific Realism. In: Nola, R. (eds) *Relativism and Realism in Science*. Australasian Studies in History and Philosophy of Science, vol 6. Springer, Dordrecht.

Nagel, T. (1986). *The View From Nowhere*. Oxford University Press.

Nozick, R. (1998). Invariance and objectivity. In *Proceedings and addresses of the American Philosophical Association* (Vol. 72, No. 2, pp. 21–48). American Philosophical Association.

Nurkić, P. (2022). Retorika državne nestabilnosti. *Međunarodne studije*, *22*(1), 97–113.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*. 349(6251):aac4716.

Reiss, J., & Sprenger, J. (2017). Scientific objectivity. In E N Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Winter 2017. Metaphysics Research Lab, Stanford University.

Resnik, D.B. (2006). *The Price of Truth: How Money Affects the Norms of Science*. Oxford University Press.

Steichen, E. (1903). Ye Fakers. *Camera Work*. (1), 48.

van Dongen, N. & Sikorski, M. (2021). Objectivity for the research worker. *European Journal for Philosophy of Science*. 11(3):93.

Williams, B. (1985). *Ethics and the Limits of Philosophy*. Fontana

REFEREES OF THE PAPERS

1. Slobodan Perović (Department of Philosophy, Faculty of Philosophy, University of Belgrade)
2. Voin Milevski (Department of Philosophy, Faculty of Philosophy, University of Belgrade)
3. Aleksandar Dobrijević (Department of Philosophy, Faculty of Philosophy, University of Belgrade)
4. Aleksandra Zorić (Department of Philosophy, Faculty of Philosophy, University of Belgrade)
5. Jovana Kostić (Department of Philosophy, Faculty of Philosophy, University of Belgrade)
6. Vladimir Cvetković (Faculty of Security Studies, University of Belgrade)
7. Stefan Jerotić (Clinic for Psychiatry, University Clinical Centre of Serbia)
8. Nena Vasojević (Institute of Social Sciences, Belgrade)
9. Miroslav Galić (Faculty of Philosophy, University of Banja Luka)
10. Marija Petrović (Institute of Psychology, Faculty of Philosophy, University of Belgrade)
11. Marija Kušić (Institute of Psychology, Faculty of Philosophy, University of Belgrade)
12. Vuk Samčević (Chair for Byzantine Studies, Faculty of Philosophy University of Belgrade)
13. Aleksandar Puškaš (Chair for History of Yugoslavia, Faculty of Philosophy, University of Belgrade)

CIP – Каталогизација у публикацији –
Народна библиотека Србије, Београд

17(082)
165.21(082)
17.022.1:165.21(082)
17: 32(082)

VIRTUES and vices – between ethics and epistemology : edited volume / Nenad Cekić (editor). – Belgrade : University, Faculty of Philosophy, 2023 (Beograd : Službeni glasnik). – 375 str. ; 24 cm. – (Edition Humans and society in times of crisis)

Radovi na engl. i srp. jeziku. – "This collection of papers was created as part of the scientific research project 'Humans and society in times of crisis' ..." --> kolofon. - Tiraž 200. – Str. 7–9: Introduction / editor. - Napomene i bibliografske reference uz radove. - Bibliografija uz svaki rad.

ISBN 978-86-6427-257-5

а) Етика -- Зборници
б) Епистемологија -- Зборници
в) Врлине -- Епистемологија -- Зборници
г) Филозофија политике -- Зборници

COBISS.SR-ID 115290633

Collection of papers titled *Virtues and Vices – Between Ethics and Epistemology* resonates not only with recent normative debates in ethics and epistemology but also shows how it is possible to engage with other philosophical disciplines. The authors in this collection explore the theme of virtue and vice in the fields of ethics, political philosophy, epistemology, philosophy of mind, philosophy of science, and aesthetics. At the same time, it should be noted that the discussion is rounded out by papers that explore the theme of "virtue and vice" from a philosophical-historical perspective, including the genesis of the debate and various possibilities for interpretation and reading.

Prof. dr Snježana Prijić Samaržija

This collection of essays, *Virtues and Vices – Between Ethics and Epistemogy*, successfully brings together discussions on the concepts of virtue and vice in ethics and epistemology, on the one hand, with various debates in the history of philosophy, ontology, phenomenology, moral psychology, philosophy of science, philosophy of medicine, political philosophy, aesthetics, philosophy of religion, and philosophy of education, on the other. As a result, this collection can inform the contemporary discourse on a range of issues that we face in the realm of the individual or, more broadly, at the social level, and it is reasonable to expect it to be of interest not just to academic circles but to professionals and various institutions alike.

Dr Dejan Šimković

The thematic focus of the collection *Virtues and Vices - Between Ethics and Epistemology* successfully summarizes recent debates on the relationship between virtues and vices. Whether from the perspective of ethics or epistemology, this collection of papers represents a significant contribution to both academic and non-academic communities by providing answers to questions that occupy us daily but for which we never seem to have enough time. A stimulating journey through the diverse articles in this collection inspires reflection on common human experiences and encourages us to strive to transcend the bundle of everyday practices and habits.

Prof. dr Vojislav Božičković

By relying on their shared normative context, which becomes evident after a deeper philosophical analysis, this collection of papers sheds new light on the relationship between ethics and epistemology. This connection arises through the consideration and discovery of valid normative standards in moral judgment, political decision-making, and the pursuit of true knowledge. This collection will be of great benefit to students in the fields of epistemology, moral and political philosophy, as well as other scholars working in the social and humanistic sciences. *Virtues and Vices - Between Ethics and Epistemology* represents a significant philosophical and scientific contribution in the fields it addresses, particularly in our academic community.

Prof. dr Milorad Stupar

ISBN 978-86-6427-257-5

UNIVERSITY OF BELGRADE
FACULTY OF PHILOSOPHY