# CROATIAN JOURNAL OF PHILOSOPHY

Vol. XX · No. 59 · 2020

*Croatian Journal of Philosophy*

1333-1108 (Print)
1847-6139 (Online)

*Co-published by*
"Kruzak d.o.o."
Naserov trg 6, 10020 Zagreb, Croatia
fax: + 385 1 65 90 416, e-mail: kruzak@kruzak.hr
www.kruzak.hr

*and*

Institute of Philosophy
Ulica grada Vukovara 54/IV, 10000 Zagreb, Croatia
fax: + 385 1 61 50 338, e-mail: filozof@ifzg.hr
www.ifzg.hr

Available online at http://www.ceeol.com and www.pdcnet.org

# CROATIAN JOURNAL OF PHILOSOPHY

Vol. XX · No. 59 · 2020

*Political Philosophy*

*Deliberative Democracy and Political Epistemology*

*Articles*

## *Book Discussion*

## *Book Reviews*

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *John Dunn Interview*

**Ivan Cerovac**: *You are considered one of the finest experts in John Locke's political though. What drove you to this research topic? What sparked your interest in political philosophy in general, and what attracted you to Locke's political writings?*

**John Dunn**: I was not initially drawn to Locke himself through any direct personal attraction. I was born in Britain in the second year of the Second World War and had become keenly interested in politics before I got to University because of idiosyncratic family experience in Germany, Iran and India. I have been deeply preoccupied with politics ever since because I knew already by then that the stakes in politics for everyone are always vast, the situation of most human beings then alive in the world hazardous and often painful, and that the chances of its improving seriously were, as they remain, largely at the mercy of politics. I recognized quite young that my parents' vision of politics was in many ways unreal and absurd, and since I loved them and admired many things about them, I wished with some intensity to learn to see politics for myself more clearly and steadily, without illusion and without self-serving. Through a series of whimsical pieces of good fortune I have spent my very privileged life in trying to learn how to.

I was drawn initially to philosophy because I hoped it would show me how to see everything which mattered to me clearly and because for someone arriving in Cambridge at that point, in the lingering aura of Wittgenstein, philosophy still held the imaginative glamour to make that fond hope almost plausible. I was drawn to political philosophy a year or two later, and as a student of History, because I hoped especially that it would show me how to see politics steadily in that way.

After I had completed my undergraduate degree in History I decided to go on to undertake doctoral research in the history of political thought hoping to do so under the supervision of the inspirational teacher who had introduced me in my final year to the thought of the Scottish Enlightenment and in particular to that of David Hume and Adam Smith. I intended the dissertation I hoped to write to be on Hume's understanding of political obligation, which I felt was woefully insufficient, despite the dazzling intelligence of his vision as a whole. I wanted to understand why he had come to see that understanding as sufficient: to rethink his thoughts at this point as deeply as I could in the hope of somehow seeing beneath them and through them how to understand the political bonds which hold (or fail to hold) a society together more clearly.

The teacher in question, Duncan Forbes, was a figure of arresting intuition, strong passions, but less capacity for calm and intellectual composure. He was wrestling with Hume's thinking himself at the time and working towards his important study of *Hume's Philosophical Politics* (1975). The last thing he wanted was an ignorant and overconfident graduate students stumbling around in his vicinity, so he passed me smartly on to Peter Laslett, a very different figure, flamboyant, charming, enthusiastic and very much at home in the world. Laslett had just published his path-breaking edition of Locke's *Two Treatises of Government*, which demonstrated that these were written not to vindicate a Revolution which had already occurred but to justify revolutionary resistance to the government of Charles II. Laslett wished me to trace the impact of the *Two Treatises* in Britain, France and North America over the century following its publication as a pathway towards two later Revolutions, in France and North America, and a third potential revolution in Britain itself, which did not in fact occur. I studied this for three years but found that impact over most of the century, except perhaps in one or two specific ways in the run up to American Independence, shallower than was widely alleged, and not worth systematic presentation as a book. Whilst doing the research I had, however, also seen something I thought was really important about Locke's own thinking and it was that to which I devoted my first book and my unsuccessful submission for a doctorate. I still think the central perception of that book was right, as have a variety of other scholars since. It was that Locke's overall vision of politics and its place in human life depended for him on a Christian *Weltanschauung* and that his main arguments, as he said explicitly himself whilst he was in a position to do so, do not hold good without it. A lot of the subsequent intellectual history of the west and much of the global political vulnerability of liberalism as a political approach today has followed from that fact.

**IC**: *Many hold that your work in the 1960s, along with that of Skinner and Pocock, fundamentally changed how political philosophers read some of the most important past political thinkers. What was wrong with the political philosophy in 1960s and what methodological prescriptions did you suggest in order to improve the reading of classics such as Locke?*

**JD**: I am afraid that I don't think it is true that the work of Pocock, Skinner and myself has changed anything much in how political philosophers read important past political thinkers, though one or two major figures like John Rawls have made polite concessions to the need for a measure of historical accuracy in understanding the views of past thinkers they continue to take seriously. I remain confident that it would still be better for the historical turn to have more of an impact in that respect.

It is unsurprising that my own work should have had little impact to that effect, but Skinner and Pocock are both scholars of extraordinary ability, immense depth of knowledge and each has produced since an oeuvre of enormous distinction and range which should extend the political imagination of any philosopher who chose to take politics seriously. I see no pressing need for any political philosophy which does not take politics seriously and consider it simply a misnomer.

Political philosophy in the Anglophone world was at a low ebb in the early 1960s, apart perhaps from Herbert Hart's philosophy of law, and university teaching on the history of political ideas or political philosophy as a whole was parochial, unrelentingly self-referential and rather smug.

Skinner and I were friends, and also at that time close intellectual companions. What united our view of the limitations of our elders (and Oxford contemporaries) was the sense that they were seeing and thinking within an extraordinarily narrow range and learning very little from the texts they happened to study. We thought they were doing so because they were failing to recognize the drastic existential sources of the works in question or take in what their authors were doing in bothering to write them at all, and hence often even to recognize what those authors intended to argue. We thought that this amply sustained habit was foolish and self-harming.

As already said, I still think that what I saw about Locke himself fifty years and more ago was accurate and important; but it is only fair to acknowledge that seeing it did not at that point improve the political discernment of my own reading of his political works. It has taken some time for me to recognize quite how deeply Locke thought into the fundamental elements of politics, the resources through which human communities can live relatively benignly together in face of its hazards, and the always limited reserves of patience and generosity in their feelings towards one another. It was not until I came to register the political insight of his insistence on the centrality of trust (and distrust) in human life, and the discomfiting strains on mutual tolerance inherent in the unease with which human beings experience one another that I really saw how far in advance in these respects he remains of any of today's leading philosophers of politics. Who, setting out from the text of Rawls's *A Theory of Justice* could begin to imagine a world distantly resembling the world in which all of us are now living? But that is the world which has been made by politics and the world in which human beings must continue to live and die. I am not an enemy of intellectual division of labour and I do not think that philosophers should turn themselves into historians. I simply think that any philosopher today who hopes to do political philosophy of real value needs both to open their imaginations fully to the realities of politics today and to call on the aid of historians when they try to learn from the great political philosophers of the past.

**IC**: *Political philosophers throughout history, as well as today, often construct political systems designed for citizens understood as rational and well-informed individuals. Nonetheless, empirical research suggests that citizens often lack the basic understanding not only of the political process, but also of their own interests. The rise of populist politicians and movements, often related to fake news and anti-science (or anti-experts) movements, evokes worries that political philosophy has little to say about real-world politics. What are your thoughts on this? Is there a way for political philosophy to address these modern trends and to help us change the world for the better?*

**JD**: All citizens are only intermittently rational and incompletely informed. Any political philosophy premised on assuming otherwise can scarcely hope to illuminate politics.

As already indicated, I believe that any serious political philosopher must focus on politics as it is and think with and for their fellow citizens or fellow human beings as these too are.

**IC**: *You have written extensively on the politics of socialism and the (quite dangerous) Marxist hope that new social, political and economic structures will end the exploitation and lead to a better future. Though still far from Marxism, we are witnessing the rise in support for some populist left-wing politicians and social movements that presume that they are acting in the interest of the majority of people. Should left-wing parties appeal to (weak or modest) Marxist argumentation when they criticize the existing inequalities and offer solutions, or should the left abandon Marxism and start anew?*

**JD:** Marx himself did see, and many Marxist political actors and thinkers in his wake have since seen, many aspects of collective human life quite realistically. The deep failing of Marxism as a political heuristic has been its absurd promise of a world beyond politics which History would somehow in the end deliver, the opportunistic reach for power vindicated by that claim, and the grotesque underestimate of the durable harm inflicted on any human society by decades of brutal oppression under the aegis of that claim. The left should keep from Marxism what is descriptively true and disavow completely what was always fantastical and is now brazenly mendacious.

**IC**: *In one of your recent books, Setting the People Free, you focus on the story of democracy and how the word changed its meaning from Ancient Greece to contemporary western societies. There, you make a careful distinction between democracy as an electoral instrument and the democratization process. Can you explain this distinction and explain why do you think it pushed democracy in the political mainstream?*

**JD**: Ostensibly free elections on the basis of universal suffrage have become the canonical form for establishing and sustaining legitimate government within a single historical sequence. They have done so largely because they provide a more compelling picture of how a population can authorize and de-authorize its government than any extant rival. The original experience of democracy as a political form provided for the relatively narrow ranks of free citizens a far more direct relation between governing and being governed than any modern state could replicate. For the citizens themselves, it lifted the burden of personal subjection from their lives. Democratization is a far vaguer process of lightening the burdens of subjection across a population which has proceeded to varying degrees across many different societies over the last few centuries and has sometimes been consciously and quite effectively steered through political action. It will never be complete, but it is a denser existential transformation than any modification of the process of government could possibly be.

**IC**: *Democracy is, if I understand your position well, a way to think politics together. Do you think that we are in a danger of losing that way of thinking temporarily, or even permanently? Is democracy something that can be forgotten and then recovered? What are the conditions under which people understand their living together politically in democratic terms?*

**JD**: Democratic politics in that sense is a historical creation and it has to be created through political action, though it of course relies throughout on many social and economic preconditions. I believe it to be a creation of great value but also of ineliminable vulnerability. At present it is being wounded, deliberately and pretty brutally and effectively, in many different settings. Making room for it requires high political strategy and luck, but democratic politics itself must consist in the actions of very much wider circles of a population. It must make some sense to them and it must on balance benefit, not harm, them. In the hands of the unscrupulous and malign it is very easy for high politics to take away what only it could make the space for in the first place. Above all it requires a people (a *demos*) with the will and capacity to live together in peace. As any resident of former Yugoslavia knows all too well, high politics can destroy that fast and thoroughly. It cannot make it either fast or thoroughly. You could say only History can make it, but it would be better simply to recognize that it has to make itself and do so in time.

**IC**: *There seems to be a rise of illiberal democracies in the world. Apart from China and Russia, more and more European countries (with Hungary and Poland as notable examples) reject the liberal political tradition and embrace simple majoritarianism reinforced by shared religious or ethnic identity. Why is this the case? How do you see the future of global democracy? Will it continue to be intertwined with liberalism or will it make an illiberal turn?*

**JD**: I don’t think liberalism is a clear political category, any more than democracy. In those societies which have had the historical opportunity to develop democratic politics and experience it for some time I think it has had a liberalizing effect on the society over time and has in practice done so to some degree by now across a very wide range of cultures across the world from Taiwan, South Korea and even Japan to Uruguay. Illiberal residues remain very large in all societies and in many it is obviously wrong to view as residual since they constitute a substantial majority of the population. Where they do the freest and fairest of elections will not hand power to liberals and the prospects for establishing democratic politics or sustaining it for any length of time are poor. There is good reason to describe Hungary as an illiberal democracy, above and beyond the fact that its present and frequently re-elected leader chooses to do so. There is less reason to describe any state in which the rulers simply authorize themselves as a democracy at all. I doubt if democracies which it is reasonable to call liberal, where they happen to exist, are in much danger of being supplanted wholesale by any other state form so far invented, though they might of course be subjugated militarily in some way or other. What may destroy them from within is the failure of their democratic politics, but that necessarily will have to be a failure of the citizens themselves. Democracy is not a providential form. It is a collective opportunity for citizens to use or squander.

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *Political Parties as Corruption Hazards.*
## *The Republican Case for Sortition*

OLIVER MILNE
*National University of Ireland Galway, Galway, Ireland*

*In this paper, I do several things. First, I present a definition of 'corruption' as 'abuse of power that builds or maintains the abuser's power', arguing that this definition is more generally applicable than other definitions offered in the literature and that it highlights a crucial property of corruption, namely its tendency to metastasise, presenting a more and more serious danger to society. To defend the emphasis I place on this tendency, I then argue that corruption (as commonly understood) frequently produces three mechanisms pushing it to reproduce: self-perpetuation by the corrupt actors to protect themselves, formation of networks between corrupt actors which ensnare new participants, and normalisation of specific kinds of abuse of power in the corrupt actors' social environments. From here, I turn to political parties, arguing that they present fertile soil for the mechanisms just described. In their stead, I argue for sortition—a system whereby legislators are randomly selected from the population at large. I make the case that each of the three metastatic mechanisms I have described would have much more difficulty taking root in a sortitional-democratic system than in an electoral-democratic one, before concluding by responding to a major potential objection to such a proposal's feasibility—namely, that sortitional juries would be less competent than elected legislatures—and presenting a sketch of a sortitional-democratic system setting out how it could discharge the government's executive functions, in addition to the legislative functions already covered. The paper as a whole, in addition to its explicit arguments, may be considered to make an implicit case for non-ideal over ideal theory, in that it attempts to show the importance of that quintessentially non-ideal factor, corruption, to the nature of any political order.*

**Keywords:** Corruption, political parties, non-ideal theory, democracy, sortition, political psychology.

What is corruption? We tend to think we know it when we see it, but how should we conceptualise it? In this paper, I argue that it is—as the classical Greeks and Romans thought—something dangerous, destructive, and insidious, that threatens the very foundations of democracy; that political parties are by their very nature structurally inclined to foster it; and that, as a result, democrats should replace electoral systems with sortitional ones—meaning systems in which both legislative power and the power to appoint and dismiss the executive are held by randomly-selected juries of ordinary citizens serving fixed terms.

The first thing I want to do to make this argument is to put forward a very simple definition of corruption that helps bring its inherent peril into focus. The definition I propose is this: Corruption is abuse of power that builds or maintains the power of the abuser, or of some group, entity, or cause on whose behalf they act. This isn't meant to be a redefinition, but a formalisation that captures the essence of the word's common usages in a way that highlights what's important about the phenomenon thereby identified.

Now this is an umbrella definition—in fact, in the remainder of this paper I refer to it as such—and this is one of its major advantages. The two big questions that obviously follow in its wake—'what kinds of power are there?' and 'what constitutes their abuse?'—are huge, open fields of research and public contestation. Other conceptions of corruption, in particular fields such as political institutions, can very often be used (with minimal adaptation) as partial specifications of their answers. Emanuela Ceva's (2018) account of political corruption as the use of a publicly entrusted power of office (the relevant kind of power) for a publicly unaccountable reason (which constitutes its abuse), for example, can fit snugly under this framework in just this way. In fact, the restriction the umbrella definition places on the contents of the 'publicly unaccountable reason'—namely, that either it or the actor's use of their entrusted power must involve getting or keeping power of some sort—helps address the criticism that Ceva's conception is too broad: outside the umbrella, her definition encompasses every abuse of public office not demonstrably due to negligence.

Lawrence Lessig's (Lessig 2011) account of dependence corruption, wherein politicians become psychologically dependent on the generosity of benefactors and as a result are more open to their lobbying, can be similarly incorporated. The lobbyists abuse the wealth at their disposal to create a psychological dependence in the politicians that gives them (the lobbyists) greater power to influence public policy; the politicians abuse their power by using it only in ways that keep them getting the perks to which they've become accustomed. Whether this abuse is conscious or unconscious is beside the point: another advantage of the umbrella definition is that it's not so much concerned with actors' inner attitudes as with the impacts of their actions, which are generally more empirically observable.

Dennis Thompson's tripartite definition of institutional corruption sits even more easily within this framework. This is a special case of corruption as 'abuse of public office for private gain' in which

> (a) the gain an official receives is more institutional than personal, (b) the advantage the official provides takes the form of access more than action, and (c) the connection between the gain and the advantage manifests a tendency to subvert legitimate procedures of the institution, regardless of whether an improper motive is present. (Thompson 2018: 11–12)

The 'institutional gain' referred to here has to do with the operations of power: the core cases Thompson has in mind are U.S. congressional campaign contributions. Condition (c), meanwhile, is a definition of a certain kind of abuse of power.

So a lot of different accounts of corruption play nicely with the umbrella definition. But one immediate objection might be that, regardless of its intertheoretic merits, it doesn't seem to capture the most low-level, everyday kinds of corruption: if I'm an underpaid traffic cop, and I shake down motorists so I can pay my rent, what 'power' am I maintaining? Calling my ability to live under a roof 'power' seems *technically* accurate, but a little odd in its emphasis. But this is exactly why such a general umbrella definition is useful: it's a shift of emphasis—from petty corruption to grand corruption, but also from individual instances to the dynamics of corruption over time, and the ways petty and grand corruption can connect. It frames corruption as an investment of power that yields a return.

A second objection (for which I am obliged to Enes Kulenović) can be drawn out by use of a real-world example. In Croatia, ownership of property must be officially registered in order for the owner to sell the property, obtain a mortgage on it, and so forth. A number of years ago, officials in the Zagreb property-registration office cooked up a scheme for extracting bribes from people wishing to register property in a timely manner, by slowing down the registration process unless a 'fee' was paid for a 'legal memo' to speed it up. When this scheme was uncovered by the police, it emerged that some of the conspirators had avoided promotion within the office in order to keep taking bribes. Isn't this, the objection goes, a counterexample to the umbrella definition—a case of corrupt agents *avoiding* power for the sake of corruption?

This objection is instructive because it highlights the question of how we measure power. The typical post-Foucauldian approach, often taken in, for example, gender studies, is to consider power in terms of the damage it does. To put it crudely, on this approach, more damage equals greater power. This lends itself to what might be termed intersectional analysis, looking at how multiple different types of power act together on one object—a discourse, institution, or particular group of people—to provide a comprehensive understanding of the harms that object suffers. The approach implicit in the umbrella definition of corruption is different. Because the umbrella definition hinges on the corrupt agent, rather than the injured parties, the appropriate understanding of power has to begin with what that power enables that agent to do: from the agent's perspective, power is measured not by the damage it can do but by the degree to which it can help them achieve their ends. This is power in the sense of ability rather than power in the sense of domination. The kinds of analysis *this* approach works best for—of which this paper is an example—have to do with the motivations of, and pressures upon, power's wielders.

This way of measuring power allows us to answer Kulenović's objection. Higher positions in the bureaucratic hierarchy would doubtless have given the corrupt officials greater control over, and potential to dominate, their coworkers; but the accompanying effective pay cut would have reduced their ability to send their children to private schools, buy new cars, or whatever else they might have done with their ill-gotten gains. From the officials' perspective, a promotion would have tied their hands with regard to the things they actually cared about doing. The money, therefore, represented more power for them than the promotion.

A third potential criticism of this definition, from the opposite angle to the first two, is that it captures too much, specifically in the personal sphere. An abusive husband who emotionally manipulates his wife to prevent her from leaving him is abusing his power over her in order to keep it, but nobody would describe this as 'corruption'. Against this criticism, I take a rather different line: this is actually a useful and illuminating extension of the meaning of the term, because it helps highlight the links between power in the personal sphere and power in society at large.

Suppose, for example, our abusive husband is a gangster, or an investment banker of the old school. His wife's position at his side helps him retain the status he needs amongst his colleagues to be taken seriously, which in turn helps him to make deals, make money, and avoid getting screwed over and tossed to the wolves by his chums. That money and status, in turn, make it easier for him to prevent his wife from running away, both because they allow him to more lavishly gild her cage and because they would (he would like her to believe) enable him to hunt her down more easily. The two kinds of power, from the husband's perspective, are both simply tools in the same toolbox. This is not to say that every instance of spousal abuse enhances the abuser's power in other fields—the motives of domestic abusers are far beyond the scope of this article. The point is that it can sometimes be so used.

These are not corner cases, either. The entire practice of mediaeval-style political marriage—whether to secure inheritance of an empire or of an acre of cropland, as still happened in Ireland well into the 20th century—is this connection writ large. Less dramatic continuities between these kinds of power are still widespread today. As essayist Laurie Penny puts it:

> One of the ways men bond is by demonstrating collective power over women. This is why business deals are still done in strip clubs, even in Silicon Valley, and why tech conferences are famous for their "booth babes." It creates an atmosphere of complicity and privilege. It makes rich men partners in crime. (Penny 2018)

To extend our concept of corruption to incorporate these kinds of abuses of power, we may conclude, enriches our understanding of both the concept and the abuses.

To see the case against political parties is likewise a matter of following the umbrella definition's lead, and investigating the dynamics of the corrupt accumulation of power we thereby unearth. So let's consider another example.

If I'm a mayor and I make a chunk of money throwing city contracts to businesspeople who give me backhanders, I'm thereby building up power in at least two ways. First, and most obviously, I can spend all that money on, say, my re-election campaign. But alongside that, those contractors are now my cronies. We have a relationship. We each have reason to be confident that the other can keep a secret and may be interested in further underhanded dealings. And our shared secret gives us certain common interests—such as keeping prying eyes away from it—that encourage us to cooperate with each other.

What this second kind of power does is allow us to use one another's power to get things done. It acts as a force multiplier for every other kind of power: wealth, office, popularity, violence, further connections—all of these powers become available not merely to their holder but to their holder's friends and business partners, and the basis of this is the relationship of trust and mutual interest between them. (I say 'trust' here, but I should emphasise that this is a limited form of trust—if I'm a crooked mayor and you're a shady businessman, I might know very well that you'd stab me in the back to make a profit, but I can trust you're not going to call the cops if I offer you a mutually beneficial trade that just so happens to fall outside the law. And your attitude to me might be very similar.)

Talking about these different kinds of power also illustrates a reason it's worthwhile to think about multifarious varieties of corruption under the same heading: power is fungible. One kind of power can be used to gain another, which can be used to gain another still, and so on. If you're smart, ambitious, and unscrupulous, you'll use every tool in the box to advance your interests—and if we want to stop that from happening, or control how it can happen, we have to think about all those tools operating together as a system, rather than restricting our interest to some subset of them. Corruption as 'abuse of power that bolsters the abuser's power' is a conceptual frame within which to do that.

So these kinds of dynamics are all very familiar to us from popular media—I'm talking about things like *The Wire*, or *Game of Thrones*, or *Boardwalk Empire*, or the news. The point I want to make is that this frame captures what's distinctively, fascinatingly disturbing about it all: in each case, the players are building up their power, or trying to hold on to it. That's the ongoing transformation that propels the narrative arc and gives corruption its dramatic interest, and it's also what makes it distinctively dangerous in real life. An agent willing to engage in corrupt activities can thereby become capable of carrying off more audacious misdeeds—and, in fact, may have to, to avoid getting caught. That's one way that the little stuff turns into the big stuff: to stop his past shakedowns coming back to haunt him, perhaps the newly-promoted police captain will have to offer something to—or menace—a local journalist. And thus the corruption spirals—and the plot of your story about it gains momentum.

So already we can see two really important facets of corruption:

1. Corruption tends to perpetuate itself, not only because it builds power for the corrupt actor, but because it gives them a pressing reason to avoid losing that power—namely, to avoid suffering the consequences of their abuse of power.
2. Corruption tends to form networks: since most corrupt actors aren't wizards or superheroes, they mainly accumulate their power by forming connections that allow them to reliably make use of others' power—connections that often compromise the recipient in some way, inducing them to engage in corruption themselves.

There's one more significant facet of corruption I want to touch on, and that's its tendency to normalise itself. This is where an institution or milieu has a culture in which corrupt activities are commonplace, even celebrated, reducing the risk to corrupt actors and reducing their dependency on secretive trust relationships with other individual corrupt actors. So, for example, both our traffic cop shaking down motorists and our corrupt mayor have a pretty pressing interest in making sure their respective co-workers turn a blind eye to their activities. And if everyone's on the take, who's going to rat them out? So they foster a culture that tolerates it—through jokes, casual comments, mockery of 'holier-than-thou' attitudes, and so on. Non-corrupt actors participate in these, and it thereby becomes psychologically easier for them to ignore signs of actual corruption, and to conceive of and yield to the temptation to behave corruptly themselves. A problem endemic to the Croatian healthcare system, namely that everyone likes the bureaucrats who help their friends jump to the front of waiting lists, illustrates exactly this[1]. It also bears similarities to the way a culture of sexism can shield serial sexual predators—the formation of that kind of culture is for that very reason a form of corruption in the sense we're talking about here.

But this doesn't just apply to things like graft and sexual violence—typically thought of as bad things done by people who know, or could reasonably be expected to know, that they're bad things to do. Normali-

[1] My thanks to Ana Matan for this example.

sation of corruption also encompasses the generation of more ambitious political or ideological justifications for it—justifications that are supposed to persuade the public, or the members of the relevant institution, that the corrupt activities in question are in fact totally fine and nothing to be ashamed of.

Suppose, for example, you're a *sans-culotte* in the French Revolution, who has just chopped off the head of an 'enemy of the people'. It's super important for you, now, that this be the kind of thing it's OK to do, because you've just done it. You've just become heavily invested in ideologically justifying the shedding of your political opponents' blood 'for the good of the nation'—both for your own psychological wellbeing, and in order that nobody decides you yourself should go to the guillotine for it. But of course, if you have this ideological justification floating around the public sphere, not only is it going to make it easier for you to do the same thing again, but other people are going to be influenced by it, too. More heads are going to come off in 1790s Paris, just as in a police department developing a culture of venality, more cops are going to start taking bribes. And the disparity of these examples shows off another advantage of the definition of corruption as 'the abuse of power that builds or maintains power': it encompasses both the word's common uses—graft on the one hand, and the corruption of ideals and idealistic movements on the other.

So here's the bulletpoint version of the third aspect of corruption:

3. Corruption propagates itself from one actor to another, not only through the networks it forms, but through the normalisation or even ideological legitimisation of corrupt activities, both of which are important strategies corrupt actors can use to protect themselves.

Now none of these points is likely to be a huge revelation to anybody. But taken together, there's a clear conclusion to be drawn from them: corruption produces mechanisms that allow it to spread and worsen. Corruption metastasises. This is visible in all sorts of different historical and contemporary political situations, from the collapse of the democratic hopes of the French Revolution into the autocracy of the Directorate and then the Napoleonic Empire, to the situation in Hungary over the past couple of decades. This isn't a manageable chronic illness; it's a life-threatening condition that threatens the entire nature of a society.

This extraordinary peril inherent to corruption loomed large in the minds of the American Founding Fathers. Their thought was shaped by the conflict in Britain between the so-called 'Court' and 'Country' parties, in particular as expressed through Trenchard and Gordon's 'Cato's Letters'. Written under a pseudonym derived from a famously incorruptible conservative opponent of Caesar, these were described by historian Clinton Rossiter as 'the most popular, quotable, esteemed source of political ideas in the colonial period' (Rossiter 1953: 141). The political clash in which they intervened centred around the London-based political elite's—that is, the 'Court party's'—alleged use of patronage to build up its power and undermine Parliament to the benefit of the Prime Minister's office, threatening the liberties of the landowning public in England and Scotland, a.k.a. the 'Country party'. This is about as clear-cut a case as you can get of corruption as the abuse of power to build power, and Trenchard and Gordon's presentation of it was, for the revolutionaries, a vivid illustration of the connection between graft and tyranny. It was with a view to avoiding this kind of scenario that the leading lights of the American Revolution designed their constitution's separation of powers, and on those grounds that they publicly defended it, notably in the *Federalist Papers* (Hamilton et al. 1788). In other words, corruption's status as an existential threat to society's freedom has a long and distinguished pedigree.

Now one of the crucial factors in each of the three aspects of corruption I've highlighted is their dependence on relationships that persist over time. A corrupt actor has to keep hold of their power in order for it to grow. A network of relationships has to persist for it to be useful—if the relationships dissolve, or the people in them lose their other powers, the network is no longer effective. And an institutional culture has to propagate itself, enmeshing new recruits to the institution and keeping existing members behaving according to its patterns, in order to survive. So we can ask the question: what keeps these elements going over time? What mechanisms preserve them? What are the infection vectors in which they hide and grow? And how can these processes be disrupted?

Considered from this angle, one obvious answer jumps out at us: political parties. These are autonomous organisations in whose success their members are invested, making them less likely to blow the whistle on corrupt behaviour among their own ranks. Within parties, influential members can pursue careers lasting decades, so their power persists. The intake of new members is effectively vetted by their internal power structures, allowing them to filter out people who might pose a threat to corrupt activities. And, of course, successful parties have a great deal of power: they influence the opinions and actions of their members and supporters, and control access to high political office, whether that be through their reserves of electoral campaign funds and volunteers, or through a non-democratic hold on state institutions.

In principle, then, it looks like political parties ought to be hotbeds of corruption. And, in fact, the evidence bears that hypothesis out. Political parties and elected officials consistently come top in Transparency International's Global Corruption Barometer reports (Hardoon and Heinrich 2013; Pring 2017; Riaño et al. 2009). Now I should qualify this by saying those reports are based on survey data, and competing political parties do have an interest in making one another out to be corrupt, so we would expect them to be a little overrepresented. But if we look at contemporary cases of democratic backsliding, political parties are right at the heart of them, whether we're talking about Fidesz in Hungary, Law and Justice in Poland, the Republicans in the United States, or Mongolia's bipartisan decline into autocracy under President Khaltmaa Battulga (Tumurtogoo 2019). In each of these cases, the persistent power structure and culture of the party or parties involved have enabled their slide into corruption, simultaneously solidifying their hold on the levers of power and driving them to become progressively more unscrupulous and rapaciously venal.

A particularly striking demonstration of this tendency on the other end of the political spectrum can be found in Rojava, the autonomous territory in northern Syria. The official ideology of the ruling Democratic Union Party, or PYD (Partiya Yekîtiya Dîmokratik), is based on the work of green-anarchist theorist Murray Bookchin, and emphasises direct democracy and civil rights. But a 2016 report by Rana Khalaf for the thinktank Chatham House (Khalaf, 2016) claims that the PYD's efforts to consolidate power have put its actions at odds with its words: it restricts independent journalists and civil society organisations, and packs supposedly democratic councils and committees with its own placemen. It seems more likely than not that the party's need and desire for power will overwhelm their ideological goal of democracy, rendering that goal moot.

So political parties of all stripes are liable to corruption. What are we supposed to do with this shocking news? If there were no alternative, this analysis would be nothing but a reason for pessimism. But I want to make the case that there *is* a viable alternative, one that shares electoral democracy's merits while avoiding its vulnerability to corruption.

That alternative, I claim, is sortitional democracy—that is, government by randomly-selected juries. Under such a system, ordinary citizens would be selected by lot to serve as parliamentarians for terms of several years, and paid for their service. A similar system worked in classical Athens for more than two centuries, until it was curtailed by Macedonian imperialism (Raaflaub et al. 2007). More recently, proposals for sortitional systems of government have been put forward over the past decade by scholars including Alexander Guerrero and Terrill Bouricius (Bouricius 2013; Guerrero 2014), as well as appearing in public life in the form of citizens' assemblies, which feature prominently in the demands of the Extinction Rebellion climate protest movement, and one of which has actually been implemented as an official governmental advisory body in Ireland.

A sortitional system is resistant to all three of the aspects of corruption I've detailed. First and foremost, whereas under electoral democracy political parties can hold on to offices for decades, office-holders in a sortitional system are swept out of office every term and replaced with people who have—this is crucial—no prior connection to them.

The office-holders have neither the need nor the ability to try and hold on to power. They don't have any elections to win or patrons to appease. Even if their station goes to their heads, the public will be under no illusions about their special suitability for office—they were chosen literally at random. To try and maintain power based on an instance of governing jury service would be a Herculean task.

This disconnect between each successive cohort of office-holders also makes it harder for wrongdoers in office to get away with it. Elected members of political parties can call on their comrades in office to protect them and put them back in power later, even if they as individuals run up against term limits or otherwise lose their position in the formal state hierarchy. But jurors selected by lot are very unlikely to have any sense of obligation to protect corrupt strangers from the consequences of their own misdeeds. Indeed, the more corruptly the system behaves in one sitting, the more the random citizens selected to the next are likely to resent the corrupt actors.

This means that any corrupt political network in a sortitional system must re-corrupt the office-holders from scratch every four or five years. This is a huge ongoing risk for the network's existing members. Approaching an office-holder who's an unknown quantity and trying to embroil them in a network of corruption of whatever kind is inherently hazardous because you don't know how they're going to react. The exchange could easily blow up in your face, endangering you and potentially your allies, too. This is part of why corruption forms networks in the first place: you need trusted people you can deal with.

Now this particular consideration is clearly more applicable to corruption—defined, remember, as abuse of power that builds or maintains power—that isn't operating under an ideological shield. But even with that shield, it's much trickier for your would-be Lenins or Öcalans to enlist the cooperation of randomly-selected jurors in abuses of power than it is to get their loyal cadres to play along. It's also much more difficult for them to install those cadres in power in a sortitional system than it is under an electoral one, even if they don't play fair. The result of a rigged election generally looks at least vaguely like the result of a clean one, but the result of a rigged jury selection is immediately obvious to everyone: in a chamber of several hundred jurors, *any* disproportionate allocation of seats to supporters of one political faction is, statistically, such an unlikely outcome from a fair lottery that it's a sure sign the draw's been fiddled. The need to convince every new batch of office-holders of the total righteousness of your cause and the necessity of liquidating the kulaks (or what have you), and the risk that they won't buy it, is therefore a serious obstacle for a ruthless ideologue operating within a sortitional system. To get around it, they would need to achieve much higher level of public consensus around their ideology than they otherwise would, in order to ensure incoming jurors are sufficiently amenable to their advances.

So that's the case that sortitional democracy would, in principle, be much more resistant to corruption than its electoral cousin. Before I finish, though, I just want to preempt a couple of objections to the proposal's feasibility.

The first objection concerns the competence of the juries. How can such a system ensure a satisfactory minimum level of performance in administration, without the candidate vetting usually performed by parties in electoral democracies? This is too big a topic for me to cover comprehensively here, but there are a couple of things to be said in response.

First of all, a case can be made that sortitional democracy has certain advantages over electoral democracy when it comes to the quality of the decision-making process. First and foremost, the greater diversity of a sortitional chamber gives it an epistemic advantage over a chamber of elected politicians, who are, in most electoral democracies, drawn mainly from the ranks of an educated élite. The variety of perspectives and life experiences present in the room means the sortitional chamber has fewer blind spots than its elected counterpart, and is therefore better able to consider all of a proposal's impacts, all else being equal.

Secondly, the elephant in the elected chamber is the politicians' need to chase votes. Sometimes this imperative coincides with the public interest; frequently it does not. Elected officials' competence benefits the public very little when it is misdirected. The pressure to garner votes means the ignorant, information-poor choices made by voters reverberate through many different policy areas, as elected politicians do not what's best but what's most popular.

And this brings us to the central problem with the competence criticism. The claim is that under government by jury ordinary citizens will be making important decisions on things they're not competent to make important decisions about. But this also happens in electoral democracies, at every election. The difference between sortitional and electoral democracy is that the jurors—who, as a representative sample of the citizenry, have the same baseline level of competence as the voters in an electoral democracy—are not making their decisions from the average voter's (quite rational) position of ignorance, but are paid to consider the issues full-time, able to summon and consult relevant experts one-on-one, and engaged in a deliberative process aimed at producing the best decisions. This improved division of labour means sortitional democracy ought to be much *less* vulnerable to ignorant populism, and better able to make hard-but-necessary decisions, than its electoral cousin.

Over and above these arguments, there are structural measures that can be taken to improve the expected quality of a sortitional legislature's decisions. Rather than aping the general-purpose chambers common to electoral democracies, for instance, a sortitional democracy could have many different legislative chambers, each focusing on a specific issue, dramatically lessening the jurors' epistemic burdens. Both Guerrero's and Bouricius' proposed models of sortitional democracy are organised along these lines (Bouricius 2013; Guerrero 2014). A sortitional reconciling chamber could then exercise a veto over the specialist chambers' proposals, to act as quality control and hammer them into a coherent policy platform and budget, while being prohibited from making proposals itself.

Additional measures, such as providing a 'warm-up' period of several months between jury selection and the jurors' taking their seats, to allow them to get up to speed on their subject area, and having experts address the chambers at the start of each legislative session, have also been proposed to ameliorate the jurors' lack of prior training and vetting for competence. The Irish citizens' assembly boasts both of these latter features, and, over the three years since its inauguration, has successfully produced high-quality recommendations on a number of controversial and technical issues, including abortion, climate change, fixed-term parliaments, and the conduct of referenda.

The second major objection to the feasibility of sortition, which was pressed on me by John Dunn, concerns the executive functions of government. How is a sortitionally-based government supposed to carry these functions out, and avoid the onset of executive autocracy? I shall conclude this paper by providing a barebones sketch of a system that might handle this problem adequately.

One of the advantages of having specialist chambers is that each specialist chamber would be well placed to appoint and oversee the executive head of their particular department, for those areas where a department is required. These executive heads would serve at the pleasure of their respective chambers and be subject to term limits. The reconciling chamber could likewise appoint and oversee an executive chairperson, with the right to address any chamber, whose job it would be to coordinate between departments, and who could dismiss department heads with the prior approval of the reconciling chamber. On this model, policy direction as well as law would be devised by the specialist chambers, approved (or vetoed) by the reconciling chamber, and put into practice by the departments, with the executive chair's ability to dismiss department heads being the means by which the reconciling chamber would protect against failures of oversight by the specialist chambers, as well as preventing them from going rogue and unilaterally enforcing non-approved policies.

One more move that might be made would be for all these appointed officials to be politically restricted civil servants in the British mould, forbidden from publicly taking political stands—the point being to prevent them from taking advantage of the 'bully pulpit' to build personal public support, using their superior expertise and political savvy to undermine the public standing of the inexperienced sortitional jurors.

This is also why I strongly advocate against the executive being elected. Their power base—in particular, their legitimacy—needs to be kept in check. These two moves separate executive leadership from what might be called moral leadership of the public. But it's also possible the speech restriction could hamper these officials' effective execution of their duties, or prevent them from doing potentially-vital things like flagging up jurors' underperformance to the public. The question of which consideration is more important can only be answered empirically—a test to which I hope it will one day be put.

## *Bibliography*

Bouricius, T. G. 2013. "Democracy Through Multi-Body Sortition: Athenian Lessons for the Modern Day." *Journal of Public Deliberation* 9.

Ceva, E., 2018. "Political corruption as a relational injustice." *Social Philosophy and Policy* 35: 118–137.

Guerrero, A. A. 2014. "Against Elections: The Lottocratic Alternative." *Philosophy and Public Affairs* 42: 135–179.

Hamilton, A., Jay, J., and Madison, J. 1788. *The Federalist Papers*, 12/12/2011. ed. Project Gutenberg.

Hardoon, D. and Heinrich, F. 2013. *Global Corruption Barometer 2013, Global Corruption Barometer*. Transparency International.

Khalaf, R. 2016. "Governing Rojava: Layers of Legitimacy in Syria" (Research Paper). Chatham House, London.

Lessig, L. 2011. *Republic, Lost*. New York: Twelve.

Penny, L. 2018. "Four Days Trapped at Sea with Crypto's Nouveau Riche." *Breaker Mag*.

Pring, C., 2017. "People and Corruption: Citizens' Voices from Around the World." *Global Corruption Barometer*. Transparency International.

Raaflaub, K.A., Ober, J., and Wallace, R.W. 2007. *Origins of Democracy in Ancient Greece*. Berkeley: University of California Press.

Riaño, J., Hodess, R., and Evans, A. 2009. *Global Corruption Barometer 2009, Global Corruption Barometer*. Transparency International.

Rossiter, C. 1953. *Seedtime of the Republic: the origin of the American tradition of political liberty*. New York: Harcourt, Brace.

Thompson, D. F. 2018. "Theories of Institutional Corruption." *Annual Review of Political Science* 21: 495–513.

Tumurtogoo, A. 2019. "Mongolia's President Is Slicing Away Its Hard-Won Democracy." *Foreign Policy*.

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *How to Craft Economic Policy: Values in Economics*

HANA SAMARŽIJA
*University of Zagreb, Zagreb, Croatia*

*This article argues that all economic theory presupposes implicit political premises, and that these affect its scientific conclusions. More specifically, I will argue that neoclassical economics trades the epistemic values of predictive accuracy and explanatory strength for an image of the capitalist economy as sustainable, which renders it unequipped to analyze its crises. Echoing Anwar Shaikh's analysis, I will show that neoclassical economics, by constructing idealized settings and misleading metrics, obscures the inherent conflicts of capital accumulation. As this tendency leads to an incomplete understanding of the current system, I will argue that neoclassical economics cannot inform effective economic policy. To explain the difference between epistemic and non-epistemic values, I will begin with a brief historical overview of the role of values in science. I will then, by analyzing economic metrics and the basic assumption of perfect competition, proceed to show that neoclassical economics is both empirically and logically underdetermined. Once I have shown there is no epistemic argument in favor of neoclassical economics, I will argue that this choice of theoretical framework was mandated by underlying political concerns. I will end by discussing the relationship between engaged philosophy and public policy in times of crisis.*

**Keywords:** Social epistemology, political epistemology, philosophy of economics, philosophy of science, objectivity.

> This is, I believe, a serious problem throughout much of the contemporary world: erroneous policies based in erroneous theorizing are compounding the economic difficulties and exacerbating the social disruption and misery that result. (Harvey 2015: 10)

## 1. *Introduction*

As far as social sciences are concerned, economics is a unique case. Economics is everywhere. When economic theory goes awry, it fails to predict crises and proposes policies that impair millions of lives. When it ignores reality, economic theory overlooks the urgency of what is now dubbed a climate crisis in favor of corporate interests and snubs necessary institutional revisions as radical and unrealistic. If what is at stake is the everyday survival of millions and the future survival of the natural world, then the task of crafting economic policy demands the utmost caution. Given our current success at tackling both poverty and environmental devastation, with temperatures soaring above the recommended maximums, and inequality, in the United States alone, reaching rates unseen since the Great Depression (Zucman 2019), several questions seem central. Is neoclassical economics, with its models and idealizations, at all equipped to deal with these existential threats? Should policy-makers reconsider heterodox economic approaches, browsing their toolkits for responses to pressing issues? Do the issues of welfare economics and climate policy require a new attitude towards the ethics of policy making, and where, if anywhere, does that place the ethical foundations of economic theory? Before answering these questions, we might want to explore how neoclassical economics rose to become the present orthodoxy. We might wonder, for starters, whether the reasons behind this theory choice were strictly scientific. Did neoclassical economics offer shrewder predictions and a simpler explanatory framework than its competitors? Maybe its models painted a particularly precise image of real economic interactions? Did competing theoretical approaches, deprived of the neoclassical vocabulary, fail to reach basic economic conclusions? If the rationale behind choosing neoclassical economics was not epistemic, and I will proceed to show that it was not, we will need to find a way to explain it without resorting to empty talk of ideology.

My central claim is that all economic theory presupposes implicit political premises, and that these determine its scientific conclusions. Closer to the point, I will argue that neoclassical economics trades predictive accuracy and explanatory strength for an image of the capitalist economy as fundamentally sustainable, which renders it ill-equipped to analyze its crises. At the most basic level, all economic theory implies specific beliefs about proper state action concerning individual wellbeing. Higher up, it presupposes beliefs on what constitutes a dignified human life, and on whether the state should have anything to do with the makings of such a life. Economic theories presuppose and justify entire economic systems. What we focus on when phrasing our theory determines whether an economic system will seem sustainable. Science is about inquiry. It is about seeking answers to questions and constructing frameworks for making sense of those answers. As Elizabeth Anderson pointed out in 1995, even the most neutral theories answer particularly worded questions, make particular classifications, and opt for particular ways of managing brute data, and it is these choices that inform economic theory with most of its implicit premises (Anderson

1995). Impartial economic theory is merely theory whose assumptions have, through its prevalence in public discourse, briefly become invisible. They are, however, still there, and are, as much as ever, pliant to philosophical analysis.

Much like Anderson, I will argue that contextual social values are not a hindrance to objectivity; on the contrary, they are an essential element of scientific work, and should be handled with care. If we are to manage modern capitalism or to propose its corrections, we must first understand how it works. In this task, neoclassical economics fails us twice. Epistemically, it fails as a framework for understanding the dynamics of modern capitalism. Ethically, because its premise that capitalism is inherently stabilizing weakens its predictive accuracy, it fails to inform effective policy, which, in turn, damages millions of lives.

To prove this point, I will, echoing Anwar Shaikh's excellent analysis, show that neoclassical economics offers a distorted image of real economic practices (Shaikh 2016). In doing so, it obscures the internal conflicts of capital accumulation. By constructing idealized settings and misleading metrics, neoclassical economics portrays capitalism's cyclical products, such as economic stagnation, downward pressures on wages, unemployment, and financial crises, as its unfortunate aberrations. Predictable social maladies, sidelined by the constructs of neoclassical economics, become difficult to detect before they have gotten out of hand. Since it, as such, informs erroneous policy, neoclassical economics is not only epistemically dubious but ethically problematic. Once the choice of neoclassical economics emerges as epistemically unjustified, I will argue that the decision to embrace this theoretical framework was mandated by political concerns, interested in its ability to depict market capitalism as inherently sustainable. As the practical consequences of this choice will often be at odds with the ethical demands of policy-making—which must concern itself with poverty, housing, healthcare, and environmental preservation—this will lead us to our final topic, a discussion about philosophy and public policy in times of crisis.

Within the next twenty pages, we will be taking a detour from epistemological debates about the role of values in theory choice to recent discussions about the applied ethics of public policy. In the first section, I will show how Thomas Kuhn legitimized values as an aid in appraising rival theories, but limited his proposal to neutral epistemic values, such as predictive accuracy, coherence, and fruitfulness (Kuhn 1977). Continuing with Elizabeth Anderson's argument about value-laden inquiry, I will attempt to clarify the distinction between epistemic and contextual values. Why does this matter? Why should the difference between epistemic and non-epistemic values at all interest us? Because I will, in the second section, proceed to show that the decision to embrace neoclassical economics was epistemically unjustified and hence guided by another kind of motivation. Instead of tackling the whole of neoclassical economics, I will illustrate my argument by way of synecdoche, analyzing unemployment metrics, poverty limits, and the basic premise of perfect competition. I will close the article with a brief discussion about the relationship between philosophy and public policy.

If states want to craft effective economic policy, and we may assume they do, they must bring economic theory's implicit premises to the surface and assess their validity. At a time marked by rising inequality, precarious labor, insecure housing, and a looming climate crisis, there is little room for the pretense of impartial economics. Persisting with outdated poverty lines in the face of mounting disagreement is not an impartial decision. It is a claim about the relative weight of human hardship. Assessing economic health in terms of production and consumption while experts urge for circular economies is, rather than adherence with neutral scientific concepts, conscious insouciance to new research.

To sum up, my goal is to show that neoclassical economics presumes that capitalism is inherently stable and then builds its analyses upon this assumption, which makes it epistemically unfit to predict its crises. As long as clinging to the neoclassical toolkit continues hampering our efforts to resolve pressing issues, it will remain at odds with democratic standards and warrant an appropriate political response. If there is no decisive epistemic argument in favor of neoclassical economics, and I will show that there is none, we are invited to explore alternative approaches, those more apt at curbing inequality, restraining climate change and building a more fully just society for all. For the time being, we should do what we can. This point made, we can proceed to the first part of our discussion, a brief historical overview of values in science.

## 2. *Values in Science: From Epistemic Values to Implicit Premises*

The struggle to recognize the role of values in science was a lengthy endeavor. This reluctance was largely due to the rationalist legacy left by the logical positivists. Unlike contemporary philosophy of science, which places theory choice at the heart of scientific inquiry, logical positivism focused on work within a fixed research program. This confinement allowed it to reduce scientific work to induction from general laws, and to effectively purge science of subjectivity. According to positivist orthodoxy, the scientist's role was to infer scientific laws from individual observations. And the observations themselves were, in turn, treated as the unproblematic starting points of inquiry. Since scientists made these generalizations by applying shared skills and procedures, methods pliant to mutual accountability, scientific agency was portrayed as an inherently rule-governed business, and subjectivity was condensed to the necessary minimum. In Carnap's view, values were a thing of emotion and personal preference, and, as such, entirely foreign to the language of science (Carnap 1959).

Although this rationalist image of science was surprisingly durable, it had one fatal flaw: it bore little resemblance to the way science—a cooperative project encompassing thousands of fallible individuals—is actually practiced. Thus construed, logical positivism paid little heed to two crucial facts. First, observation in science is theory dependent. Observations do not automatically turn into propositions: we are the ones who, with the aid of a chosen vocabulary, must render them intelligible. An observed particle does not instantly manifest as an electron; we must first recognize it as such. It is solely by way of theory we can communicate our findings to others. Because we will interpret all observations in the language of our chosen theoretical framework, theory choice is not a provisional one-off affair but the starting point of all further inquiry. Second, scientific theories are underdetermined by the available evidence. In other words, there is no direct logical necessity between our observations and the chosen theory.

When faced with the problem of choosing one among competing theories, Carnap invoked the famous distinction between "internal" scientific questions, which can be answered within a given theoretical framework, and "external" questions, which concern the legitimacy of the framework itself (Carnap 1950). The internal questions of science were to be resolved, unsurprisingly, by logical induction from laws. However, when wondering whether a given research program suits our purposes, we could appeal to pragmatic criteria such as "fruitfulness" or "efficiency." These criteria were, of course, even if logical positivists did not yet recognize them as such, epistemic values. And theory-choice, which Carnap identified as "external" to science, was soon recognized as the most central of its activities: the choice of the framework which would inform the rest of our scientific agency.

It was Thomas Kuhn who, in his essay "Objectivity, Value Judgment, and Theory Choice," officially introduced values to science (Kuhn 1977). The question Kuhn sought to answer was how we choose between equally appealing theories that account for the same empirical data. Historically speaking, scientific theories are seldom singularly determined by evidence: empirical findings fit snugly in different explanatory frameworks, the same sets of facts give way to different readings, and rival scientists offer equally tempting interpretations. Thus, when faced with several equally viable theories, none of which is decided by brute evidence, we must, if we are to make a choice, resort to something other than the evidence at hand. This point is precisely where values come into play. According to Kuhn, we should then allow for a dose a subjectivity, evaluating the theories in line with a specific set of epistemic values and choosing those best suited to our respective research program. Kuhn's original scientific values were, as their name would have it, distinctly epistemic: they were meant to promote the epistemic quality of our scientific conclusions.

The upshot here is that scientific theories are often both logically and empirically underdetermined. Theory choice can thus seem like an arbitrary affair. Since we cannot fully justify our selection of either theory by referring to the available data, the fact it was more appealing than its competitor must lie in some external source of merit. Kuhn proposed five such epistemic values: predictive accuracy, internal coherence, external consistency, unifying power, and fertility (Kuhn 1977: 322). It is entirely clear why a physician researching vaccines might prioritize a more accurate theory over one that is, albeit greater in scope, more vulnerable to error. A theoretical physicist, whose field does not touch upon actual human lives, might, on the other hand, attribute greater weight to fertility, a theory's ability to overcome difficulties and stimulate further scientific research.

It is essential to note the extent to which Kuhn's values are already profoundly social. In employing different epistemic values, scientists must reflect upon the social configuration of their discipline and the social role of its scientific products. When evidence does not suffice, we fill it in with our metaphysical assumptions and practical interests. Is ours a branch that, as its results affect living human beings, must prevent errors and prioritize accuracy over loftier concerns? Are we dealing with a theoretical domain that profits from continuous disagreement and fruitful debate? If our scientific field partakes in policy-making, should it value correct predictions above thorough explanations? Simply put, when choosing our theory, we first ask what it is for. We ask what purposes it serves and what questions it is trying to answer. Inquiry is always driven by certain goals and interests. What Kuhn showed, albeit obliquely, was that theory choice *inherently* involves social factors, and that values cannot be purged from real scientific work. Furthermore, Kuhn saw that different scientists, guided by different practical interests, will attribute different weights to different epistemic values. Consider the following passage:

> The criteria of [theory] choice function not as rules, which determine choice, but as values which influence it. Two men deeply committed to the same values may nevertheless, in particular situations, make different choices, as in fact they do. (Kuhn 1977: 324)

However, back in the seventies, the realization that these values were social was not yet fully present. Even philosophers amicable of value-laden inquiry, such as Kuhn, tended to include a telling disclaimer: they would only speak of values in the natural sciences, where it was easier to portray them as strictly epistemic. In his eponymous essay on values in science, Ernan McMullin drew a sharp line between epistemic and non-epistemic values, rooting the difference in the very nature of science as a truth-seeking enterprise. Epistemic values seek to improve the epistemic quality of our theories and, ultimately, lead to truth. Non-epistemic values do not. What is more, McMullin envisioned for the correct usage of epistemic values to cleanse (natural) science of social and political influences, which can only detract from the final goal of our scientific efforts, objective truth (McMullin 1982: 20). The internal coherence of our theory, the fact it hangs well together, contributes to our quest for truth; its coherence with our political beliefs, on the other hand, does not. Our commitment to epistemic values will gradually lead to a better understanding of the world. The choice to indulge our ethical and political interests would only have us ignore all evidence inconsistent with a foreordained conclusion.

Our current topic owes far more to Elizabeth Anderson, who clarified the scientific role of values as we usually know them. Anderson's argument was not only that ethical and political values *can* play a decisive role in theory choice, but that they inevitably *do*. Our role, then, is to handle them with care. Unlike Kuhn and McMullin, Anderson did not limit her account to the natural sciences. This decision to include the social sciences, where the practical interests that inform theory choice are harder to distinguish from the content of the theory itself, enabled her to articulate a more faithful image of real scientific work. In defending the notion of feminist epistemology, she showed how contextual values could shape inquiry without falling into the trap of partial and irresponsible science. To do this, Anderson first had to dispel a common concern: that allowing moral values in science entails an immediate loss of objectivity to ideological pressures. In the eyes of rationalist philosophy of science, any defense of value-laden inquiry conjures images of Lysenko's biology, an infamous instance of totalitarian thought control uninterested in producing epistemically valuable results. Once we allow politics and morals to guide science, the argument goes, objective standards of excellence will quickly give way to a negligent scientific practice blind to facts that do not comply with the desired conclusion. Bad science will then degenerate into a muddle of rigged conclusions, and all scientific progress will, with mathematical certainty, come to a halt.

Anderson retorted that this is a misconstruction of the way science works. More importantly, she showed that the line separating epistemic and non-epistemic values is not as clear as might have seemed. Since theories are usually both logically and empirically underdetermined, underlying contextual values—metaphysical, ethical and political—jump in to take their place. Anderson illustrated this with a rich array of historical examples:

> Thus, Einstein initially appealed to thought experiments grounded in empiricist epistemological norms to argue for the superiority of the theory of relativity over classical Newtonian mechanics. (…) Functionalist explanation in sociology was discredited partly because it was incompatible with the non-teleological metaphysical framework of modern science: for those who accept this framework, merely pointing out that a social phenomenon promotes social stability does not provide a satisfactory explanation for why it exists. (…) In these cases, normative considerations about the conduct of inquiry, normative constraints on the form of acceptable data and of satis-

> factory explanations, and normative desiderata of calculative ease proved to be powerful arguments for theory choice. Where the data run out, values legitimately step in to take up the "slack" between observation and theory. (Anderson 1995: 29)

Don't these examples, though, still fall within Kuhn's epistemic values? Once we look beyond the content of epistemic values and focus on the reasons why we choose them, the distinction between epistemic and contextual values becomes blurry. Simply put, the choice of epistemic values is motivated by contextual social concerns. And our commitment to specific values will then proceed to shape our research program. The predictive accuracy of our theory is not only epistemically valuable but helps inform good public policy. Fertility, which motivates further research, looks to the future of our scientific community. As long as science continues taking place among real people, working in real scientific collectives and submitting their findings to real practical purposes, theory choice will inevitably hinge upon contextual values. Anderson thus proceeds with a series of examples where contextual values motivated theory choice to no epistemic detriment:

> Functional explanation in sociology was discredited not just because it did not offer a satisfactory scheme of explanation but because, by representing phenomena as functional for the social order, it underplayed the significance of social conflict and discouraged criticism of the status quo. A humanist interest in acknowledging and promoting the dignity and freedom of persons has influenced many social scientists. An emerging methodological norm among interpretive anthropologists is to show one's research to the subjects of study and respond to their criticisms. This norm serves the moral interest of respecting the dignity of those one studies. (Anderson 1995: 31)

If we conceive of the aims of science as at all broader than the bare accumulation of truths, we cannot maintain the pretense of disinterested research. Most disciplines have some practical application. The aim of medicine is to promote health; the aim of economics, likewise, is to prevent crises and to inform good economic policy. Anderson's most interesting theoretical innovation lies in where she located these implicit practical interests. She identified two places where contextual values enter science. First, when beginning our inquiry and wording our questions, we do it by considering the social purpose of our scientific discipline. The decision what will count as an answer will depend on our contextual values. Second, our values will also affect the way we classify our data. When separating relevant and irrelevant facts and molding our statistical categories, we will shape them into responses to our initial questions. Anderson's argument here rests on the concept of scientific significance. In other words, no theory can include all imaginable evidence: some facts will simply not count as significant. (What were the subjects wearing? What was the weather like?) To decide what evidence to feature in our theory, we will need certain criteria, and these criteria will depend on our practical interests.

To guide her point home, Anderson used an example that leads us to our main topic. She considered American unemployment rates, which exclude the category of discouraged workers, people who want paid work but have, conceding it is futile, stopped actively seeking it (Anderson 1995: 45). To count as an unemployed person, one needs to have sought work within the last four weeks. In this reduced form, unemployment rates are often used to assess macroeconomic health: lower unemployment rates are supposed to denote a flourishing and self-stabilizing economy. The noteworthy aspect of these statistics on unemployment is that they are, in the relevant epistemic sense, incomplete. To be sure, there would be nothing spurious in offering a second metric, one listing only those jobless people still actively looking for work and, hence, still exerting downward pressure on wages. Yet positing an incomplete figure as the sole statistic on unemployment fails at its main goal: the task of informing readers about the number of jobless individuals in a given economy. If the figure included everyone who said they wanted a job but could not find one they could have subsisted on, the number would be nearly double (Kudlyak 2007). In this deflated form, however, unemployment metrics take on a new rhetorical function. With artificially decreased rates of unemployment, states can depict their economic systems as more sustainable than they indeed are. The social malady of joblessness is thus successfully pushed aside until it escalates to the point it can no longer be ignored. Even if we assume no such deception is at play, lower unemployment rates demotivate policy-makers from focusing on joblessness. Incomplete theorizing thus leads to incomplete policy-making, and incomplete policy-making entrenches existing social problems.

Consider another popular economic metric, the poverty line. Before his appearance at the World Economic Forum's meeting in Davos, Bill Gates lauded a graph which claimed that global poverty has, as a success of global neoliberalism, declined from 94% in 1820 to just 10% today. Similar statements, such as those made in Steven Pinker's *Enlightenment Now*, rest upon biased readings of economic data (Pinker 2018). These statistics are not untrue. Yet, thus presented, they do not offer enough background information for an adequate understanding of global poverty. In his retort to Gates' diagram, anthropologist Jason Hickel placed the data in the appropriate context (Hickel 2019). Instead of a vision of linear progress, he offered an image of enforced colonization and growing inequality, where masses of people trade rural living for a new place within the global proletariat. First, Hickel pointed out that data on poverty has only been collected since 1981, rendering any prior measurements either sketchy or meaningless. All that these numbers reveal, according to Hickel, is that people used to live in non-urban societies where very little actual money was required to survive. We have shifted from communities that subsisted by sharing abundant natural resources to a global market economy where millions of people, in changed circumstances, have to struggle on microscopic amounts of money. Hickel then considered the poverty line itself:

> But that's not all that's wrong here. The trend that the graph depicts is based on a poverty line of $1.90 (£1.44) per day, which is the equivalent of what $1.90 could buy in the US in 2011. It's obscenely low by any standard, and we now have piles of evidence that people living just above this line have terrible levels of malnutrition and mortality. Scholars have been calling for a more reasonable poverty line for many years. Most agree that people need a minimum of about $7.40 per day to achieve basic nutrition and normal human life expectancy, plus a decent chance of seeing their kids survive their fifth birthday. And many scholars, including Harvard economist Lant Pritchett, insist that the poverty line should be set even higher, at $10 to $15 per day. (Hickel 2019)

If we were to adjust the figures to the more conservative suggestion, shifting the poverty line to seven dollars (Woodward 2015), we would end up with an inverse image of global hardship: Hickel closed the article by showing that the number of people living on less than seven dollars a day has, rather than dropped, rocketed since the oldest data in 1981. So, although the initial facts were not strictly *untrue*, the way they were framed did not amount to an adequate understanding of our social reality. Closer to our point, it traded an accurate image of global inequality for an image of market capitalism as inherently stabilizing. Again, even if this was not a case of conscious dishonesty, such artificially soothing tales of sustainability may derail policy-makers from pressing social problems.

Our initial skepticism, then, was not entirely unfounded: implicit political premises can impede our quest for the whole truth. If we refuse to acknowledge evidence that disagrees with our preordained conclusion, science is sure to suffer as a result. How did Anderson resolve this challenge? Good science, she stressed, possesses internal mechanisms that guard against such miscarriages of objectivity (Anderson 1995: 32). Standardized practices such as blind reviews, regulated methods, and strict regimes of mutual accountability prevent science from deteriorating into a state where we opt for theories whose political implications we hold particularly dear. Impartiality does not require we ask our questions pretending to be clean of all contextual interests. It requires that we, once we have begun our inquiry, fairly assess all incoming evidence, including that which might disagree with the solution we might have hoped for. Neoclassical economics, then, does not err when presupposing that market capitalism is self-regulating, but when it clings to that assumption in the face of opposing evidence. If the specific products of neoclassical economics, such as unemployment statistics and poverty lines, seem to forfeit completeness for a more sustainable image of the current system, it might be interesting to explore whether the same holds for its underlying theoretical assumptions. And this is the topic of our next section.

## 3. *Perfect Competition and the Lacking Epistemic Case for Neoclassical Economics*

Pre-classical and classical political economics, as championed by Adam Smith, Karl Marx, and David Ricardo, analyzed capitalist practices by observing actual business behavior. Inequality, class struggle, and power differentials were crucial in understanding how the system works. And yet, this empirical approach could not be more different from present economic orthodoxy. Today, what we have is neoclassical economics, an approach focused on determining goods, outputs, and income distributions through laws of supply and demand. Modern neoclassical economics, which has dominated economic discourse since the 1980s, when the American economy, with its professed aid, recovered from the recession, derives its macroeconomic models from idealized accounts of individual behavior (Keen 2011: 35). Namely, it imagines individuals as fully rational and self-interested agents seeking to maximize their utility. Neoclassical models hinge on the assumption that, as markets are by definition self-stabilizing, crises can only emerge from excessive government interventions in the market, rather than from the market itself.

Why was classical political economics, a narrative discipline grounded in historicized empirical analyses, abandoned in favor of the neoclassical paradigm? What epistemic advantages did its models offer? According to heterodox economist Anwar Shaikh, neoclassical economics met the added political requirement of depicting capitalism as an ideal system (Shaikh 2016: 340). To maintain this image, neoclassical economics shunned the former focus on production, marred with unequal starting positions and differential access to capital, for a focus on exchange between abstract individuals. Exchange could be portrayed as a moment of equality: when exchanging goods and services, we encounter each other as free and equal agents who can opt out of the transaction. Shaikh illustrates the weaknesses of neoclassical economics by criticizing its assumption of perfect competition. While political economists saw trade between firms as a struggle for dominance, neoclassical economics refurbished it as a benevolent interaction wherein all agents emerge better off than when they began. This vision, according to Shaikh, would not have been picked up had it not been for the changing politics.

In Shaikh's recounting, this gilded depiction of capitalism as a system that satisfies the interests of all parties was a political response to the protracted economic crisis between 1873 and 1896, known as the Long Depression (Shaikh 2016: 341). As this period of entrenched economic pessimism required a theory that would reinstate trust in capitalism's functioning, Leon Walras articulated a mathematical vision of a perfectly competitive market (Walras 1874). What kind of theory, then, did these demands generate? Where classical political economy, concerned with actual business behavior, described aggressive companies that monitored each other's behavior and violently cut prices to achieve a market advantage, Walras offered a model of static equilibrium, now complete with the notions of perfect knowledge and passive price taking. How did Walrasian perfect competition work? In a complete departure from empirical data, perfect competition presupposed an infinite number of identical tiny firms who "operated as traders in specific auction markets managed by all-seeing auctioneers." Shaikh briefly describes Walras' model of perfect competition:

> Trading began with an announced market price that elicited buy or sell offers for quantities of individual commodities and labor power; this price being in accordance with the assumed utility-maximizing behavior of individual participants. If the resulting quantity demanded in the given market price was not equal to the offered supply, the price would be appropriately raised or lowered. The change in price would in turn elicit a fresh round of buy and sell offers, until each market "groped" its way to a balance at some particular price. (...) In the end, the only possible state of rest was one in which all markets were simultaneously in balance—general equilibrium. (Shaikh 2016: 342)

If we observed firms monitoring other firms to gather information about quality and pricing, it would be an aberration from perfect competition. If we observed firms lowering prices in response to the others' behavior, not to lose customers, we would, again, be dealing with imperfect competition, another departure from the modeled norm. The same scenario would occur if we witnessed firms struggling to automate or reducing wages to cut production costs. It would soon become evident that this model could not survive any empirical instance of trade. When faced with the temporal dimension of his theory, the fact that groping takes time, Walras assumed that individual firms would only act when the imagined auctioneer accepted their offer. The auctioneer himself would only accept offers when all markets were balanced, i.e., when supply was perfectly proportional to demand, and when all agents were guaranteed to have their wants satisfied. It is obvious how this model produced an image of market capitalism as inherently sustainable. Yet, if there is no imaginary auctioneer, who is setting the prices? Nobody. This is an acknowledged void in neoclassical economics: since firms are passively accepting market prices, they are not setting the price, and neither are the customers, who use prices to determine which product to purchase. This is not the only theoretical corollary of price taking. Because firms do not determine prices, neoclassical economics cannot explain conflicts between labor and companies that bring down wages to cut production costs.

The theory of perfect competition, nevertheless, held steadfastly. Although the model was soon condemned as empirically invalid (Kuenne 1954) and inoperable (Walker 1987) for analyzing actual capitalist economies, neoclassical economics continued using perfect competition as a methodological and pedagogical tool. Shaikh identifies eight telling similarities between Walras' early model of perfect competition and modern neoclassical economics (Shaikh 2016: 343). Both accounts (i) offer an idealized image of market capitalism as inherently sustainable, (ii) reduce economic phenomena to individual choices, and (iii) generalize the principle of scarcity from land and agriculture to all factors of production. They, moreover, (iv) transform the notion of "cost" to include a normal profit range, which was entirely foreign to classical political economics, where firms often emerged, as they do in reality, as complete losers. More pertinent to our point, they (v) envision economic dynamics as an equilibrium that is automatically reinstated as soon as it is disturbed and (vi) assume that economic activities only take place in a state of equilibrium, which is a glaring deviation from empirical reality. Finally, they (vii) presuppose that full-employment always obtains as a result of the market functioning at equilibrium and (viii) that all firms passively accept set market prices. These idealizations obscure the contradictions of modern capitalism—such as the tendency towards monopoly and the conflict between labor and capital—and make it difficult to analyze its real, imperfect dynamics.

How can we relate this to Anderson's notion of scientific significance? In purging competition of its empirical constituents, such as firms observing each other's behavior and cutting production costs, neoclassical economics trades an accurate image of real business behavior for a depiction of market capitalism as self-regulating. By positing empirical data as irrelevant and presuming its conclusion, neoclassical economics limits itself to preordained results. As an alternative to perfect competition, Shaikh, renewing the legacy of classical political economics, develops the theory of real competition: competition as warfare, with individual firms seeking to undermine each other by bringing down prices and the cost of production (Shaikh 2016: 259). This methodological choice gives him a clearer picture of the market forces which drive down wages, encourage automation, shape prices, and, in the end, produce monopolies. In comparison with Shaikh's approach, it is easy to see that the theory of perfect competition was empirically underdetermined. As the assumption of perfect knowledge, which would mean that each firm somehow knows what the others are doing, contradicts perfect competition, Shaikh proceeds to show it is also internally inconsistent (Shaikh 2016: 346). If there is no epistemic argument in favor of perfect competition, its choice must have been mandated by some external source of merit. In this case, it gratified contextual political values, the need to restore faith in capitalism's sustainability. It succeeded at this feat by containing implicit premises—namely, that capitalism is self-stabilizing—most apparent in the empirical behavior it chose to exclude.

In 2002, aiming to show that the very foundations of neoclassical economics are intellectually unsound, heterodox economist Steve Keen came out with a thorough retort, aptly titled *Debunking Economics*. By way of basic calculus and plain language, Keen argues that neoclassical economics cannot derive a coherent theory of consumer demand, that the theory of supply and demand is fundamentally flawed, and that its conception of the labor market cannot explain actual social dynamics. After the financial crash in 2008, the book reappeared for a second edition, now two hundred pages longer and complete with an urgent plea for a new economic paradigm (Keen 2011: 49). The way economy is taught at universities is, according to Keen, unacceptable: one's initiation into economics is more akin to indoctrination than to education, and students, as the basic premises of their discipline, learn disputed claims devoid of intellectual validity. In Keen's view, knowing neoclassical economics is not only useless but actively dangerous:

> The most important thing that the global financial crisis has done for economic theory is to show that neoclassical economics is not merely wrong, but dangerous. Neoclassical economics contributed directly to this crisis by promoting faith in the innate stability of a market economy, in a manner which in fact increased the tendency to instability of the financial system. With its false belief that all instability in the system can be traced to interventions in the market, rather than the market itself, it championed the deregulation of finance and a dramatic increase in income inequality. (Keen 2009)

The reasons why neoclassical economics has proven so durable despite its epistemic shortcomings are, according to Keen, twofold. First, neoclassical economics offers an idealized image of capitalism as a meritocratic and fundamentally sustainable system, and economists choose to believe it. As support for the neoclassical paradigm is often equated with support for capitalism itself, economists less eager to identify with the left wing of the political spectrum feel further reluctance to question its premises. Second, Keen argues that economic education stifles critical thinking and demotivates students from casting doubt on what they are taught. What, then, does Keen imply? Are all neoclassical economists just rampant ideologues, rigging the numbers in favor of an oligarchic status quo? Not at all. In fact, this dogmatism is not peculiar of economics. It is characteristic of inquiry within an established scientific paradigm, or within what Thomas Kuhn dubs "normal science" (Kuhn 1962).

In their study of scientific collectives, Margaret Gilbert and James Owen Weatherall show that a certain dose of dogmatism is not an aberration from normal scientific behavior (Gilbert and Weatherall 2016). On the contrary, it is essential in maintaining group cohesion. To stay on good terms with their colleagues, to advance their careers, and to prevent the corrosive incursion of cognitive dissonance, scientists will seldom look into the foundations of their discipline. Instead, assuming all is in order, they will follow what they have been instructed and apply the learned procedures. This obstinacy sometimes entails harmful epistemic consequences. The task of avoiding cognitive dissonance demands insouciance towards opposing evidence, and scientists are likely to dismiss criticisms coming from outside their group as threatening or irrelevant. Critical voices within the group are likely to be silenced, and more inquisitive scientists will shy away from confronting their colleagues on contested theoretical issues. Although this kind of behavior does not obstruct scientific progress, it can impede the transition to a new paradigm. It is hardly surprising that some of the most lucid criticisms of neoclassical economics had to come from geographers (Harvey 2007) and anthropologists (Graber 2014), scientists who, belonging to different disciplines, were not constricted by the premises of their particular branch. To sum up, I have attempted to argue that the neoclassical notion of perfect competition sacrifices completeness and empirical adequacy for an image of market capitalism as inherently sustainable. Because it is both empirically and logically underdetermined, its choice seems to have been mandated by contextual political values: namely, by the political task of depicting capitalism as fundamentally self-regulating. Hoping that this discussion has sufficed to show that neoclassical economics is neither the only nor the best approach to our economic reality, we can now explore the relationship between philosophy and policy-making.

## 4. *Conclusion: How to Craft Economic Policy*

We have seen that, by presuming that capitalism is inherently sustainable, neoclassical economics trades accuracy and completeness for a contrived image of the present system. This self-imposed methodological limitation renders it unfit to predict economic crises and hampers us in detecting social problems before they have gotten out of hand. Now is the time to answer our introductory question. How to craft economic policy in times of crisis and growing inequality, when a new economic paradigm is nothing but a moralistic pedagogical proposal, and the orthodox approach provides no tools for managing modern capitalism? Finally, what is the role of engaged philosophy in guiding and overseeing this process? Shaikh and Keen's critiques of the neoclassical paradigm take after Elizabeth Anderson's good science: mutually accountable researchers hold each other to high epistemic standards and scrutinize the other's outputs, detecting intellectual weaknesses and demanding they be resolved. In calling for changes to the way economics is taught, Keen goes a step further, looking to the future of economics as a scientific discipline. As an alternative to neoclassical economics and standard heterodox approaches, Shaikh offers us an empirically grounded revival of classical political economics, a theory sensitive to unequal starting positions and power differentials (Shaikh 2016: 4). Keen, working in another tradition, gives us a fruitful methodological framework which, taking account of time and disequilibrium, manages to predict crises and model depressions (Keen 2011: 426). Although both economists approach their task from explicit ethical standpoints, they do not sacrifice empiricism and scientific standards to some preordained conclusion.

However, while academic economics can play for time with its transition to a new scientific paradigm, testing different approaches and schools of thought, policy-makers do not enjoy this privilege. Similarly, unlike the lofty realm of epistemology, which benefits from unremitting debate, policy-makers cannot resolve the problem of dissenting experts by suspending their judgment and waiting for some calmer moment within economic discourse (Sosa 2010). What should we do? How can we, as comparative laypeople, choose the economic theory best suited to our social reality? According to Alvin Goldman, laypeople cannot discriminate between competing experts by evaluating the esoteric content of their claims (Goldman 2001). In other words, we lack both the knowledge and the time needed to study the internal propositions of some scientific discipline, which renders us unequipped to assess the expert's status within his branch. Policy-makers are just as unlikely to trudge through the margins of economic theory. What we can do, Goldman argues, is refer to the expert's track record of successfully solving problems. Translated into the language of economic policy, we should favor those economic approaches which have managed to foresee crises and have shown a commitment to human welfare. After the financial crash in 2008, Dutch economist Dirk Bezemer compiled a list of economists who, using heterodox methodological tools, predicted the supposedly unpredictable economic crisis (Bezemer 2009). What the twelve cataloged economists, Steve Keen included, had in common, was an empirical approach to the economy, a concern with debt, and a regard for the relationship between the financial and the real sector. As Keen points out, these features stand in stark opposition to neoclassical economics, which barely accounts for finance and which, due to its idealized assumptions, lacks the tools to model depressions (Keen 2011: 47).

In a recent article, Jonathan Wolff drew up the distinction between applied and engaged philosophy, the latter of which seeks issues of ethical interest and endeavors to resolve them through public policy (Wolff 2019). As philosophers, we have been granted the privilege of a life spent working through arguments, managing abstract concepts, justifying theories, and comparing information garnered from diverse sources. This fortunate position obliges us to put our tools to good use, applying them to unearth the implicit assumptions of modern society and to assess their validity. At a time marked by myriad social ailments—record rates of inequality, environmental degradation, racism, sexism, nationalism, and imperialism—we are obliged to understand and counter the forces that reproduce them. Engaged philosophy is much like Elizabeth Anderson's good science: conscious of its social role and willing to disclose its values, it addresses evidence and arguments with an open mind, browsing through historical lessons and studying policies in search of the most effective solution. Correctly understood, this engagement presupposes an interest in the economy, the material basis of all social life. Our shift towards a world where each person will be able to pursue their goals and fulfill their potential demands a stable economic footing; to build it, we will need a theory that can grasp economic reality as it is. Neoclassical economics, in assuming that the current system is sustainable, presumes its foregone conclusion, hampering our efforts to shape a just world. I have attempted to show that, in heterodox economic approaches, there are viable alternatives at hand. Transitioning to a new economic paradigm will surely be a formidable task. Yet this is the task ahead of us.

## *Bibliography*

Anderson, E. 1995. "Knowledge, Human Interest, and Objectivity in Feminist Epistemology." *Philosophical Topics* 23 (2): 27–58.

Bezemer, D. 2009. *'No One Saw This Coming:' Understanding Financial Crisis Through Accounting Models*. Groningen: Faculty of Economics, University of Groningen.

Carnap, R. 1950. "Empiricism, Semantics and Ontology." *Revue internationale de Philosophie* 4: 20–40.

Carnap, R. 1959. "The Elimination of Metaphysics through Logical Analysis of Language." In A. J. Ayer (ed.). *Logical Positivism*. Glencoe: Free Press: 60–81.

Gilbert, M, and Weatherall, J. O. 2016. "Collective Belief, Kuhn, and the String Theory Community." In M. Brady and M. Fricker (eds.). *The Epistemic Life of Groups*. Oxford: Oxford University Press: 191–218.

Goldman, A. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63 (1): 85–110.

Graeber, D. 2014. *Debt: The First 5000 Years*. Brooklyn: Melville House Publishing.

Harvey, D. 2007. *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

Harvey, D. 2015. *Seventeen Contradictions and the End of Capitalism*. Oxford: Oxford University Press.

Hickel, J. 2019. "Bill Gates says poverty is decreasing. He couldn't be more wrong." *The Guardian*.

Keen, S. 2009. "Neoclassical Economics: mad, bad, and dangerous to know." *Real World Economics Review* 49.

Keen, S. 2011. *Debunking Economics: The Naked Emperor Dethroned*. London: Zed Books.

Kudlyak, M. 2007. "Measuring Labor Utilization: The Non-Employment Index." *Federal Bank Reserve of San Francisco Economic Letter*.

Kuenne, R. 1954. "'Walras, Leontief, and the Interdependence of Economic Activities." *Quarterly Journal of Economics* 68 (3): 323–354.

Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuhn, T. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press: 320–339.

McMullin, E. 1982. "Values in Science." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2: 3–28.

Pinker, S. 2018. *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. London: Penguin Books.

Shaikh, A. 2016. *Capitalism: Competition, Conflict, Crises*. New York: Oxford University Press.

Sosa, E. 2010. "The Epistemology of Disagreement." In R. Feldman and T. Warfield (eds.). *Disagreement*. New York: Oxford University Press.

Walker, D. 1987. "'Walras, Leon." In J. Eatwell, M. Milgate, and P. Newman (eds.) *New Palgrave: A Dictionary of Economics*. London: Macmillan: 852–863.

Walras, L. 1874. *Eléments d'économie politique pure; ou théorie de la richesse sociale*. Paris: Imprimerie L. Corbaz.

Wolff, J. 2019. "Method in philosophy and public policy: Applied versus engaged philosophy." In A. Lever and A. Poama (eds.). *The Routledge Handbook of Ethics and Public Policy*. New York: Routledge: 13–25.

Woodward, D. 2015. "Incrementum ad Absurdum: Global Growth, Inequality and Poverty Eradication in a Carbon-Constrained World." *World Social and Economic Review* 4.

Zucman, G. 2019. "Global Wealth Inequality." *Annual Review of Economics* 11 (1).

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *Democracy, Truth, and Epistemic Proceduralism*

IVAN MLADENOVIĆ
*University of Belgrade, Belgrade, Serbia*

*The usual justifications of democracy attach central importance to fair decision-making procedures. However, it is being increasingly emphasized that it is necessary to address epistemic considerations to justify democracy and democratic authority. In her book* Democracy and Truth: The Conflict between Political and Epistemic Virtues*, Prijić-Samaržija defends the view which places emphasis on the necessity of epistemic justification of democracy. In this paper, I will discuss her criticism of epistemic proceduralism, which can be considered major contemporary framework for epistemic justification of democracy. Within the framework of epistemic proceduralism, for justifying democracy and democratic authority it is necessary to take into account both political and epistemic values. Nevertheless, Prijić-Samaržija thinks that epistemic proceduralism is not sufficiently epistemic and that it reduces epistemic to political values. I shall argue that epistemic proceduralism can be defended from this kind of criticism.*

**Keywords:** Democracy, truth, correctness, legitimacy, democratic authority, epistemic proceduralism.

## 1. *Introduction*

The usual justifications of democracy attach central importance to fair decision-making procedures. However, it is being increasingly emphasized that it is necessary to address epistemic considerations to justify democracy and democratic authority. In her book *Democracy and Truth: The Conflict between Political and Epistemic Virtues*, Prijić-Samaržija defends the view which places emphasis on the necessity of epistemic justification of democracy. In this paper, I will discuss her criticism of epistemic proceduralism, which can be considered major contemporary framework for epistemic justification of democracy. Within the framework of epistemic proceduralism, for justifying democracy and democratic authority it is necessary to take into account both political and epistemic values. Nevertheless, Prijić-Samaržija thinks that epistemic proceduralism is not sufficiently epistemic and that it reduces epistemic to political values. I shall argue that epistemic proceduralism can be defended from this kind of criticism.

The structure of this paper is as follows. In the second section, the distinction between proceduralist and epistemic justification of democracy is introduced. I also take into consideration certain distinctions within epistemic justification of democracy and present reasons underpinning Prijić-Samaržija's criticism of epistemic proceduralism. The third section explores the distinction made by Prijić-Samaržija between pure and moderate epistemic proceduralism. In this section I discuss her arguments against pure epistemic proceduralism. The fourth section of the paper examines her criticism of moderate epistemic proceduralism. In this regard, the role of truth in the framework of moderate epistemic proceduralism is particularly scrutinized. Section five concludes.

## *2.*

I will start my analysis by introducing the distinction between political and epistemic values (Prijić-Samaržija interchangeably uses terms political and epistemic virtues). Basic political values include principles of fairness, primarily freedom and equality. It is less clear what should be included among basic epistemic values relevant for the political domain. In any case, Prijić-Samaržija conceives of epistemic values as a broad set of values that include truth, correctness, problem-solving, epistemic responsibility (Prijić-Samaržija 2018: 117). In her book *Democracy and Truth: The Conflict between Political and Epistemic Virtues*, Prijić-Samaržija examines the significance of political and epistemic values for justification of democracy. We usually refer to political values of freedom and equality when we want to answer the question what makes a political decision-making procedure fair. Obviously, in that respect democracy has advantages over non-democratic decision-making procedures since it treats all participants in a fair way.

In sharp contrast to this, epistemic values in the political domain do not necessarily have to favor democracy. Ever since Plato it has been claimed that epistemic considerations suggest that the most desirable way of political decision-making is the rule of those who know best, which implies the rule of a few. However, this form of rule can be rejected because it does not treat all members of society fairly when deciding about political issues (Estlund 2008: 35–36). So, it can be concluded that democracy is the most desirable way of decision-making. However, that raises an additional question whether democracy itself adds to the quality or correctness of political decision-making. Once importance of epistemic values in the political domain is recognized, it becomes necessary to answer the question whether a democratic way of decision-making matters not only due to procedural fairness, but also due to certain epistemic values. If it would prove impossible to furnish such an answer, epistemic values could always be evoked when someone wants to criticize democracy. The most recent justifications of democracy therefore consider it necessary to demonstrate that democracy encapsulates both political and epistemic values.

In addition to the distinction between political and epistemic values, there is a related distinction between proceduralist and epistemic justification of democracy. Proponents of proceduralist justification of democracy maintain that freedom and equality should be understood as purely procedural values. If they are conceived of as procedure-independent values, then such justifications of democracy may favor setting limits on democratic decision-making procedures (Dahl 1989: 169–170). The obvious problem for proceduralist justification of democracy is that it does not provide any criterion for distinguishing between good and bad outcomes of democratic decision-making. In this view, legitimacy of democratic decision-making is guaranteed by the very fairness of the procedure. However, bad decisions can be brought as an outcome of fair procedures. Proponents of epistemic justification of democracy therefore maintain that epistemic criteria must be taken into account in order to assess outcomes of democratic decision-making. Those espousing epistemic justification of democracy however disagree among themselves whether such epistemic criteria should be a part of the procedure of democratic decision-making or should be understood as standards of correctness that are independent of the procedure. In any case, proponents of epistemic justification think that in addition to political values, justification of democracy must necessarily include some epistemic standards, regardless whether they are understood as inherent to the procedure of democratic decision-making or independent of it (which of course does not preclude the possibility of making both types of standards a part of justification).

Prijić-Samaržija propounds epistemic justification of democracy that takes into account both political and epistemic values. She calls this a hybrid justification of democracy, because it strives to balance both kinds of values. Obviously, aforementioned types of epistemic justification of democracy, which include types of epistemic proceduralism, can also be considered hybrid because they strive to balance political and epistemic values. However, despite that, Prijić-Samaržija maintains that in various types of epistemic proceduralism epistemic values are reduced to political values. She says “that they failed to offer a hybrid stance *at all* because epistemic justification was immediately either dismissed as secondary or downright sacrificed in favor of the political and ethical” (Prijić-Samaržija 2018: 145). By contrast, she argues that epistemic values should not be reduced to political values (Prijić-Samaržija 2018: 14). Instead, one should strive to find a model of justifying democracy which to the greatest possible extent will lead to true beliefs and correct decisions. Prijić-Samaržija therefore holds that since it insufficiently takes into account the epistemic value of truth, neither type of epistemic proceduralism is adequate enough for epistemic justification of democracy. In the following two sections, I will examine more closely her criticism of epistemic proceduralism.

## *3.*

Prijić-Samaržija makes a distinction between two types of epistemic proceduralism which she terms pure epistemic proceduralism and moderate epistemic proceduralism (Prijić-Samaržija 2018: 122). Even though her analysis takes into consideration pure epistemic proceduralism first, it should be pointed out that this position, defended by Fabienne Peter, had actually emerged as a criticism of the standard version of epistemic proceduralism espoused by David Estlund (and which Prijić-Samaržija qualifies as moderate epistemic proceduralism). Namely, Estlund has offered arguments in favor of epistemic proceduralism as the most adequate theoretical framework for justification of democracy and democratic authority. According to Estlund's conception of epistemic proceduralism, this type of normative justification has an advantage over alternative proceduralist and epistemic ways of justifying democracy. Unlike fair proceduralism, epistemic proceduralism takes into account procedure-independent standards of correctness. Estlund holds that this is necessary in order to be able to make any kind of difference between better and worse outcomes (Estlund 1997: 179). There must be some standards of correctness on the basis of which outcomes of decision-making procedure can be assessed. However, Estlund claims that epistemic proceduralism offers more adequate justification of democracy and its authority than classical epistemic justifications that he calls correctness theories (Estlund 2008: 102). According to correctness theory, a classical proponent of which was Rousseau, not only that a procedure-independent standard of correctness should be taken into account, but the decision-making procedure must be a fully reliable device for its realization. Rousseau therefore considered majority rule one such device so that those who find themselves in a minority after voting, have an obligation to act in accordance with the voting outcome, since it has been shown that their standpoint was wrong.

Estlund argues that correctness theory is too demanding for the purpose of justifying democracy and especially for justifying democratic authority (Estlund 2008: 104). He thinks that the standpoint of epistemic proceduralism offers a better alternative, because it can provide justification of democracy and its authority without recourse to requirements that are so demanding. Namely, if the fair procedure has a general tendency to lead to correct outcomes, this can be sufficient for justifying democratic authority. Therefore, to justify democracy and its authority, it is no longer necessary that the procedure should be a fully reliable device for realizing or advancing procedure-independent standards of correctness; instead, it should be reliable enough to generally have a tendency to lead to their realization. This is what sets epistemic proceduralism apart from correctness theories, even though both theories recognize the significance of procedure-independent standards. Fabienne Peter criticized Estlund's version of epistemic proceduralism because she thinks that procedure-independent criteria are not necessary for epistemic justification of democracy (Peter 2007: 343). Namely, according to the standpoint of pure epistemic proceduralism, epistemic quality can ensue from very decision-making procedures that treats all participants fairly. Unlike fair proceduralism, the significance of epistemic values is recognized, but unlike Estlund's epistemic proceduralism, this conception drops procedure-independent standards of correctness. In the rest of this section, I will take into consideration arguments which Prijić-Samaržija furnishes against pure epistemic proceduralism. In the following section, I will discuss her criticism of moderate epistemic proceduralism.

According to the standpoint of pure epistemic proceduralism, fair access to the process of democratic decision-making can lead to correct outcomes due to inclusiveness and diversity. Although prejudices and wrong convictions people hold might find their way into the process of democratic decision-making, Peter thinks that they can be filtered through the process of discussion with other people. Obviously, pure epistemic proceduralism would require a procedure of public deliberation as a necessary condition in order to arrive at correct decisions. The assumption is that in the process of public deliberation, wrong beliefs could be rectified and many prejudices and biases exposed. Peter therefore holds that in a fairly organized public deliberation, some obviously incorrect proposals would not be able to hold their ground and go through (Peter 2007: 346). The basic idea is that due to inclusiveness of a fair procedure, such attitudes would encounter justified criticism. Fair procedures, according to the standpoint of pure epistemic proceduralism, can lead to realization of the difference between correct and incorrect outcomes. Precisely because of that, fair procedures can lead to outcomes that are correct.

Two main strands of criticism of pure epistemic proceduralism offered by Prijić-Samaržija are the following. First, she claims that pure epistemic proceduralism in effect reduces epistemic values to political values. She holds that epistemic values, even though their significance is recognized, are derived from political values. Prijić-Samaržija says that "pure epistemic proceduralism is not sustainable because it leaves the realm of epistemic assessments and reduces the epistemic justification of democracy to the political" (Prijić-Samaržija 2018: 131). According to Prijić-Samaržija, the role of procedure-independent epistemic values is necessary in order to provide epistemic justification of democracy. Given that pure epistemic proceduralism does not recognize any procedure-independent epistemic value, Prijić-Samaržija concludes that the main problem with this standpoint is that it is not epistemic enough.

Second, she argues that the expectation that fair procedure of public deliberation will lead to correct outcome is overly optimistic. She draws attention to well-known facts regarding voter ignorance and lack of motivation to be informed about political issues, insisting that this should be taken into account when assessing epistemic contribution of public deliberation.[1] When in addition to these facts, the evidence about the difficulties in disseminating knowledge of more informed persons within deliberative groups are also taken into account, Prijić-Samaržija arrives to the conclusion that it is more appropriate to hold pessimistic expectations regarding the possibility that public deliberation would lead towards correct outcomes (Prijić-Samaržija 2018: 132–133). A related issue is that pure epistemic proceduralism does not offer any threshold for ascertaining whether public deliberation possesses an epistemic quality (Prijić-Samaržija 2018: 133). Prijić-Samaržija claims that this further corroborates her conclusion that pure epistemic proceduralism is not epistemic enough.

Regarding the second argument against pure epistemic proceduralism, it can be pointed out that findings about voter ignorance mostly pertain to existing democratic societies that do not function according to the principles of deliberative democracy, but primarily according to the majority rule.[2] The fact that this particular model of democracy does not motivate voters to become more informed, does not necessarily mean that they would remain equally uninformed had they had a possibility to engage in public deliberation to a greater extent. So, it seems that broadening the domain for public discussion within existing democracies could contribute to being more informed and thus to greater epistemic quality of the democratic process. The facts about voter ignorance thus do not necessarily lead towards a pessimistic conclusion about epistemic expectations from deliberative democracy.

If the first criticism that pure epistemic proceduralism reduces epistemic to political values is right, then the standpoint of pure epistemic proceduralism must be reduced to the standpoint of fair proceduralism. It is clear that fair proceduralism, which is based exclusively on political values, is not the same as pure epistemic proceduralism that is primarily interested in epistemic quality of fair procedures. Therefore, argument that leads to the conclusion that the standpoint of pure epistemic proceduralism is not epistemic enough does not necessarily show that in pure epistemic proceduralism epistemic values have been reduced to political values. The fact that epistemic values of fair pro-

[1] On this point, see also Ahlstrom-Vij 2019.

[2] Obviously, public discussion is not excluded, but seems insufficient from the perspective of deliberative democracy which emphasizes crucial importance of public deliberation for democratic legitimacy (Cohen 1997).

cedures are examined underlines that they hold significance for epistemic justification of democracy, which indicates that pure epistemic proceduralism should be distinguished from fair proceduralism. Pure epistemic proceduralism might not be epistemic enough, but in any case, unlike fair proceduralism, it recognizes the significance of epistemic values for justification of democracy.

## *4.*

Having discussed pure epistemic proceduralism, I now turn to Prijić-Samaržija's criticism of moderate epistemic proceduralism. As we have already seen, Estlund's version of epistemic proceduralism is referred to in her work as moderate epistemic proceduralism. Prijić-Samaržija argues that despite the fact that moderate epistemic proceduralism has certain advantages over pure epistemic proceduralism, this standpoint is still not epistemic enough. That in contrast to pure epistemic proceduralism, procedure-independent epistemic values are taken into account when justifying democracy, in her view, constitutes an obvious advantage of moderate epistemic proceduralism. It is worth reiterating that for Estlund's version of epistemic proceduralism both procedure-independent standards of correctness and epistemic properties of fair democratic decision-making procedures are important for justification of democracy and democratic authority.

However, Prijić-Samaržija holds that there is ambiguity concerning whether moderate epistemic proceduralism should be viewed as something which is proximate to correctness theories or whether it is a kind of a dualism of independent and purely procedural epistemic standards (Prijić-Samaržija 2018: 140). Prijić-Samaržija argues that in the light of significance attached to procedure-independent epistemic values, moderate epistemic proceduralism can be said to resemble correctness theories. But, given that equal importance is attached to fair procedures of democratic decision-making with some inherent epistemic characteristics, moderate epistemic proceduralism, according to Prijić-Samaržija, is rather akin to a kind of dualism of independent and purely procedural epistemic standards. If the first interpretation which reduces epistemic proceduralism to correctness theories is rejected (given that Estlund explicitly distances his position from correctness theories), only the interpretation of epistemic proceduralism as a dualistic position remains. The problem, according to Prijić-Samaržija, is that such a position is unstable since it cannot be seen how procedure-independent and procedural values can be balanced, since epistemic proceduralism, unlike correctness theories, does not maintain that correctness of outcomes is a necessary and sufficient condition for legitimacy of democratic decision-making.

Her main argument against moderate epistemic proceduralism starts from the assumption that "the result of a good democratic procedure will be epistemically legitimate even if it is incorrect and a decision made in a democratic debate will have epistemic value even if it is untrue" (Prijić-Samaržija 2018: 140). The basic point is therefore the following. If moderate epistemic proceduralism allows legitimacy of democratic decision-making even in the case of incorrect decisions, then correctness theories that do not allow this, from an epistemic point of view, are more adequate for epistemic justification of democracy. Furthermore, according to Prijić-Samaržija, moderate epistemic proceduralism proves to be an unstable position because in a case when an incorrect decision should be obeyed, the source thereof would lie solely in a decision-making procedure. Thus, moderate epistemic proceduralism is an unstable dualistic position that in justification of democracy adduces either procedure-independent standards of correctness or fair decision-making procedures. Moreover, considering importance attached to fair procedures, Prijić-Samaržija concludes that despite the starting premises of moderate epistemic proceduralism, epistemic values also become reduced to political values (Prijić-Samaržija 2018: 145).

First, it should be noticed that when the possibility of incorrect outcomes is allowed within epistemic proceduralism, it does not mean that such possibility should be seen as a benchmark for the epistemic significance of procedures. Epistemically relevant benchmark for assessing decision-making procedures within epistemic proceduralism is that they generally have a tendency to lead towards correct outcomes. If this is the case, such procedures have epistemic value despite the fact that in some of their instantiations outcomes might not be correct. Therefore, the possibility of an incorrect outcome is not a relevant benchmark for assessing the standpoint of epistemic proceduralism; rather, it is that procedures have a tendency to lead towards correct outcomes.

Second, as I have already pointed out, unlike correctness theories which require fully reliable procedures, epistemic proceduralism requires that procedures should be reliable enough. Unlike pure epistemic proceduralism which does not furnish any threshold for epistemic values, Estlund points out that this threshold is that outcomes of democratic procedures should be better than random. Estlund's argument substantiating this proposal is based on his criticism of fair proceduralism. Namely, it refers to the flipping a coin argument (Estlund 2008: 6). Voting is one fair procedure, but flipping a coin is one too. If we consider making decisions in a democratic way more significant than making decisions by flipping a coin, than it means that in order to justify procedures of democratic decision-making, they must be better than random.

Finally, Estlund thinks that epistemic proceduralism is more adequate than correctness theories for justifying democratic authority because in order to create political obligations, it is not necessary that a decision-making procedure lead to correct outcome in every single instance. It is sufficient that a decision-making procedure should have a tendency to lead towards correct outcomes. Therefore, for the creation of political obligations, it is not required, as in correctness theories, that those who have found themselves in a minority should consider their decision wrong. Unlike correctness theories, epistemic proceduralism envisages that people would accept obligations that derive from a decision-making procedure which has a general tendency to lead towards correct outcomes, even when they disagree with a particular decision and consider it wrong in the given instance. So, it can be concluded that epistemic proceduralism provides better foundation for justifying democracy and its authority than correctness theories which on epistemic grounds do not allow for a possibility of disagreement and any wrong decisions. In that respect, epistemic proceduralism is indeed a moderate epistemic position, because it drops overly demanding requirements of correctness theories. This can be considered its advantage rather than its disadvantage, at least when epistemic justification of democracy and its authority are concerned.

These are the reasons why (moderate) epistemic proceduralism differs from correctness theories and why it cannot be considered to harbour a kind of dualism of epistemic values. Prijić-Samaržija correctly notes that it is not easy to see what the role of truth might be in the framework of epistemic proceduralism (Prijić-Samaržija 2018: 141). In this regard, she says that "Eslund's dualism of epistemic and political values is problematic because he seems to claim that the epistemic value of democratic procedure simultaneously is and isn't related to truth" (Prijić-Samaržija 2018: 142). So, it is necessary to determine more precisely the role of truth in the framework of Estlund's epistemic proceduralism.

First, one should bear in mind that Estlund proposes a process of justifying democracy in two steps.[3] In the first step, from the perspective of reasonable persons or qualified points of view, non-democratic forms of decision-making are rejected because they do not pass the test of reasonable acceptance. This includes epistocracy or the rule of experts. Given that only democratic procedure remain in the game, in the second step of justifying democracy and democratic authority, the question is raised which democratic decision-making procedure is the most adequate in epistemic terms, that is, which democratic decision-making procedure is better than random and reliable enough to realize or advance certain procedure-independent standards of correctness. Estlund also thinks that public deliberation to a greater extent than other decision-making procedures, can be expected to satisfy these requirements.

Second, it should be noted that procedure-independent standards of correctness to which Estlund refers are not necessarily considered to have to be true. Namely, some criteria that can be an object of rea-

[3] This process is even more complex since it also includes the device of normative consent. For the purpose of explaining Estlund's epistemic proceduralism, it is however sufficient to stick to the first two steps.

sonable acceptance should not be required to be true, because such a requirement would be too demanding. From the perspective of reasonable persons or qualified points of view, such criteria can be acceptable despite the fact that they do not embody the whole truth as it is seen from the perspective of reasonable comprehensive doctrines. Insisting on entire truth as seen from the perspective of reasonable comprehensive doctrines would preclude reaching any kind of agreement. Therefore, it is more adequate to say that standards of correctness should be acceptable to all reasonable persons, rather than they should be true. This, however, does not mean that some procedure-independent standards cannot be considered true, at least in the minimal sense. Estlund says that "a statement "*x* is F" is true in at least minimal sense if and only if *x* is indeed F" (Estlund 2008: 25).

It is reasonable to suppose that what Estlund terms "primary bads" such as war, famine, economic collapse, genocide, belong to this class of procedure-independent standards (Estlund 2008: 163). Regardless of various reasonable comprehensive doctrines that they espouse, reasonable persons could consider it true that they are bads that should be avoided. I point out that all of this pertains to the first step of justification of democracy. In the second step, the question is raised which democratic decision-making procedure can be reliable enough (i.e. better than random) in order to realize procedure-independent standards of correctness or avoid primary bads (if they are taken as procedure-independent standards). The only reason why it could be said that democratic decision-making procedure in Estlund's version of epistemic proceduralism is unrelated to truth is that he does not claim that procedure-independent standards necessarily have to be true. However, this does not preclude the possibility that in non-controversial cases, reasonable citizens could consider them to be true (as in the case of primary bads). In both cases, however, in the second step, a democratic decision-making procedure is related to a procedure-independent standard of correctness, regardless whether it is considered true or acceptable to reasonable persons.

Third, it seems that Prijić-Samaržija's criticism presupposes that truth for Estlund is a fundamental epistemic value or procedure-independent standard. However, as we have seen, epistemic proceduralism has a much broader view on procedure-independent standards of correctness. Whatever criteria might be the case in point, what is relevant for justification of democratic decision-making is not to arrive to true beliefs, but to outcomes that realize or advance some procedure-independent standards of correctness (or avoid them if these standards are primary bads). Obviously, some political values will usually be considered procedure-independent standards of correctness. But this does not entail the reduction of epistemic to political values, because these values are considered in such a way as to yield an epistemic significance which is independent of the democratic decision-making procedure.

Here it would be helpful to draw attention to a distinction between theoretical and practical authorities. Theoretical authorities give us reasons for belief, while practical authorities give us reasons for action (Raz 1986: 29). When justification of democracy and especially democratic authority is concerned, what is important is not that a democratic decision-making procedure should lead to outcome which give us a reason for belief, but a reason for action. Consequently, for justification of democracy and democratic authority, it is not necessary that truth be the only relevant epistemic standard. It is more reasonable to assume that for justification of democracy, some other procedure-independent standards should have their epistemic significance and that democratic decision-making given its inherently epistemic characteristics should provide reasons for action or reasons to comply (Estlund 2008: 106).[4] Precisely because these procedural and procedure-independent standards have an epistemic significance, it cannot be said that in the framework of epistemic proceduralism, epistemic values are reduced to political values.

## *5. Conclusion*

Prijić-Samaržija's main objection to epistemic proceduralism is that this position is too procedural for the purpose of epistemic justification of democracy. We have seen that pure epistemic proceduralism does not require any epistemic threshold regarding democratic decision-making procedure. By contrast, moderate epistemic proceduralism sets a threshold that the procedure should be better than random. Obviously, Prijić-Samaržija thinks that this is also not sufficient and that justification of democracy requires fully reliable procedures that will lead to truth. But it is doubtful whether such a demanding epistemic criteria are necessary for justification of democracy and democratic authority. These criteria seem too demanding if we take into account that conception of epistemic proceduralism can be sufficient for justifying democracy and its authority.

## *References*

Ahlstrom-Vij, K. 2019. "The Epistemic Benefits of Democracy: A Critical Assessment." In M. Fricker et al. (eds.). *The Routledge Handbook of Social Epistemology*. New York and London: Routledge: 406–414.

Cohen, J. 1997. "Deliberation and Democratic Legitimacy." In J. Bohman and W. Rehg (eds.). *Deliberative Democracy*: *Essays on Reason and Politics*. Cambridge: MIT Press: 67–92.

Dahl, R. A. 1989. *Democracy and Its Critics*. New Haven and London: Yale University Press.

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

[4] Of course, within certain limits. My aim here however is not to discuss the limits of political obligation, but its sources.

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

Peter, F. 2007. “Democratic Legitimacy and Proceduralist Social Epistemology.” *Politics, Philosophy, and Economics* 6 (3): 329–353.

Prijić-Samaržija, S. 2018. *Democracy and Truth: The Conflict between Political and Epistemic Virtues*. Milan: Mimesis International.

Raz, J. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.

*Croatian Journal of Philosophy*
Vol. XX, No. 58, 2020

# *The Epistemic Justification of Democracy*

SNJEŽANA PRIJIĆ SAMARŽIJA
*University in Rijeka, Rijeka, Croatia*

*In the article, I am concerned with the epistemic justification of democracy: what does the epistemic justification of democracy consist of, and how can we assure that democracy indeed generates decisions of the highest epistemic quality? However, since it is impossible to speak about the epistemic justification of democracy without considering its relation to political justification, and their tension, this article will also question the relationship between epistemic and political justification. I endorse a position called the hybrid stance, not only because I think that, when justifying democracy, we need to consider both the political value of fairness and the epistemic values of truth-sensitivity and truth-conduciveness, but because I believe we should appropriately harmonize them. While the advocates of epistemic proceduralism hold that it best harmonizes the political and epistemic values of democracy, I argue that they do not separate epistemic values as intrinsically different from the political. On the other hand, even if we accept that epistemic justification is tied to intrinsically truth-respecting practices, the question remains which decision-making processes best satisfy this demand. In simpler terms, we must inquire how to divide epistemic labor between citizens and experts. I will try to show that the optimal model needs to preserve both the epistemic potential of the diversity present in the collective intelligence of citizens, and the epistemic potential of the factual knowledge embodied by the individual intelligence of experts.*

**Keywords**: Epistemic justification of democracy, epistemic proceduralism, division of epistemic labor, experts, epistemic diversity.

## 1. *Introduction*

In my book, *Democracy and Truth*, my central claim was that democracy's legitimacy stems not only from its political but from its epistemic justification. In simpler terms, democracy is as good and as desirable a political system as it is fair—meaning, as much as it supports the freedom and equality of every citizen, but also as much as it generates decisions of high epistemic quality (that is, decisions that solve the citizen's problems, evidence-based decisions, decisions that are the consequence of epistemically responsible conduct or, more succinct, decisions that are truth-conducive or truth-sensitive).

It is not entirely simple to offer a specific and concise explanation of what it means for democracy—a system of procedures, practices, and institutions—to be epistemically justified or to generate epistemically valuable decisions. Truth has traditionally been the foundational epistemic value. However, since truth is something we attribute to propositions, it is not entirely appropriate for assessing (collective) epistemic agency. If we are looking to offer an epistemic appraisal of systems, practices, and procedures, we will need to utilize new standards of epistemic evaluation. In this sense, I have chosen to speak about the epistemic features of a system whose decisions solve its citizens' problems, or, more precisely, about epistemically virtuous practices, procedures, and institutions that generate truth-conductive or truth-sensitive, evidence-based beliefs/stances/decisions. If democracy, as a system, manages to solve its citizens' problems and creates decisions that are—because they are made through scientific methods such as critical thinking, deliberation, argumentation, epistemically virtuous disagreement resolution or like—correct and accurate, then we can say democracy possesses the epistemic feature of being a truth-conducive and truth-sensitive system.

Democracy is undoubtedly the best existing system when it comes to the political value of fairness. Still, it seems that, even though it satisfies the demand for fairness, it frequently produces decisions of low epistemic quality. These decisions are not only inadequate for citizens because they do not answer their problems. They also inspire disagreement through social divisions, exclusion, radicalism, and terrorism; they generate decisions detrimental to the quality of life, health, education, and survival of their citizens and make the citizens' lives miserable by deepening social inequalities. It could be claimed, of course, that only an unfair system that just appears democratic will generate decisions of low epistemic quality. Apart from the option that an only ostensibly democratic system could be both politically and epistemically unjustified, research has shown that even fair democratic procedures—such as majority voting, representative democracies that include debates between party representatives, referendums, and public forums—often generate decisions and beliefs of low epistemic quality. It is always possible for some fair practice to produce a low-quality decision owing to particular circumstances or the quality of the citizens' contribution. However, recent social phenomena imply there is no reason to think that some procedure—a procedure that is politically justified because it includes all citizens as free and equal—will, as if through an invisible hand, generate decisions of high epistemic quality.

There is nothing in the very nature of fair procedures, no relevant epistemic feature of such methods by themselves, that would guarantee the truth-conduciveness of its decisions (Goldman 2010).

In my book, I have attempted to show there is an intrinsic tension between the political value of fairness on the one hand and the epistemic value of truth-sensitivity on the other. Even Ancient thinkers, most prominently Plato, have emphasized that including all citizens in the decision-making process will not generate decisions of the highest epistemic quality. Plato's *kallipolis* is nothing else but a proposal that, to preserve the epistemic quality of decisions, we should sacrifice the political values of equality and freedom: simpler, he pleaded for the epistemic virtues of epistocracy. The tenacity of this conflict is also visible in other practices, such as the different forms of epistemic paternalism and programs of affirmative action I write about in my book. Most political philosophers, aiming to avoid the thorny question of the non-democratic character of the practices that best promote epistemic quality, have purposely omitted the epistemic status of democracy, centering only on its political justification. Contrary to epistocracy and all kinds of elitism, I assume that a desirable political system must surely be politically justified and fair. Nevertheless, it also needs to be epistemically justified.

In this article, I deal specifically with the epistemic justification of democracy: what does the epistemic justification of democracy consist of, and how do we assure that democracy indeed generates decisions of the highest epistemic quality? However, since it is impossible to speak about the epistemic justification of democracy without considering its relation to political justification and their tension, this article will also question the relationship between epistemic and political justification. I endorse a position called the hybrid stance, not only because I think that, when justifying democracy, we need to consider both the political value of fairness and the epistemic values of truth-sensitivity and truth-conduciveness, but because I believe we should *appropriately* harmonize them. In that sense, political instrumentalism, the stance that we should prioritize political values, while epistemic values are only secondary or derived from the political, is as unacceptable as epistemic instrumentalism, the view we should prioritize epistemic values while sacrificing and forgoing the political.

***

In my discussion about the epistemic justification of democracy, I will rely on the arguments authored by Ivan Mladenović and Nenad Miščević, both of whose articles analyze the nature of the epistemic in democracy. Ivan Mladenović raises the question of what demands a democratic system must satisfy, apart from its political rationale, to be considered epistemically justified. Nenad Miščević, assuming that epistemic justification is tied to truth-respecting practices, wonders which decision-making processes best satisfy this demand. Even though their articles abound in arguments and solutions that entail a value independent from their debate about my book, I will focus on their central claims related to the epistemic justification of democracy.

## 2. *Epistemic value: is it intrinsic or derived from political value?*

In his article Ivan Mladenović (Mladenović 2020) endorses the stance of epistemic proceduralism, which, as he claims, best harmonizes epistemic and political justification. Epistemic proceduralism, as developed by David Estlund and Fabienne Peter, is a standpoint deserving of our maximum attention, primarily because it was these philosophers who first tackled the epistemic justification of democracy (Estlund 2008a, 2008b, Peter 2008, 2013). In my book, I have tried to show that, despite their pioneering efforts, their positions do not acknowledge epistemic justification. Instead, they either derive it from or reduce it to political justification, which is why I describe them as political instrumentalists.

I claim that Peter's pure epistemic proceduralism, the position that there are no procedure-independent standards of epistemic justification, ultimately reduces epistemic justification to the political, assuming that fair and inclusive procedures will by themselves generate epistemic quality. Moreover, Peter claims that, in the context of the epistemic justification of democracy, we can only understand epistemic quality as the outcome of fair procedures. Estlund is somewhat more moderate, which is why I call his version moderate epistemic proceduralism. Although he does concede there are procedure-independent standards, he assumes the final stance that fair democratic procedures tend to generate correct or epistemically valuable decisions—which is why independent standards appear only as a welcome supplement for particular situations. In any case, Estlund does not have to establish truth-conducive standards separate from fair procedures because he believes that democratic processes inherently contain enough reliability to produce correct outcomes.

We can interpret his ambiguous attitude towards independent epistemic norms as a worry that independent epistemic criteria could imperil the fairness of democratic procedures. Estlund is probably one of the best contemporary critics of epistocracy, and we can understand his final position about epistemic justification as the stance that setting an independent truth-conducive standard could give rise to epistocracy. I believe that Estlund is right to choose this tactic. Namely, establishing separate epistemic standards indeed raises the question of the conflict and the balance between political and epistemic values—where an independent epistemic perspective forms new demands for democratic procedures. However, if we genuinely want to endorse the epistemic justification of democracy, we will need to tackle the question of resolving the tension between political and epistemic criteria. The tactic of minimizing those epistemic standards for evaluating democratic decision-making processes that are independent of political values is not appropriate for dedicated advocates of the epistemic justification of democracy.

Contrary to my stance, Mladenović holds that Estlund's epistemic proceduralism best harmonizes the political and epistemic values of democracy, and that epistemic proceduralism is an example of a hybrid perspective. As I have already said, while it is vital to support the inclusion of epistemic justification alongside the political, it does not satisfy my understanding of a hybrid approach. Any proper hybrid approach assumes that each value bears an equal independent weight in the evaluation, and strives for an optimum balance of intrinsically different values that occasionally come into conflict. First, both in Peter and in Estlund, epistemic quality, or a reliable record of generating correct decisions, is derived from fair procedures. Second, there is no conflict between these values because any divergence is eliminated by the assumption that epistemic value can be inferred from the political.

Both for Peter and Estlund, there is no *objective* epistemic value (the epistemic feature of truth-conduciveness) that would be the necessary presumption of any intrinsically epistemic justification. For both epistemic proceduralists, epistemic value is a consequence of politically appropriately organized and conducted procedures. While Peter endorses the stance that fair processes will certainly generate epistemic quality, Estlund holds that fair procedures only tend to produce correct decisions. Although it is not entirely clear what "correct" is supposed to mean, the concept resembles Rousseau's correctness theory, where a "correct" decision is one that is supported by the general will. Therefore, a correct attitude is the one that is supported by the majority through fair democratic procedures and is unrelated to objective epistemic value. Fair procedures alone, independent from the informed and non-egoistical stance of those participating in the discussion, possess no epistemic feature that would guarantee epistemic quality. When Estlund and Mladenović claim that adequate democratic procedures tend to generate correct decisions, they are both telling us that they do not perceive epistemic values as separate and intrinsically different from the political. And this is my main point of conflict with epistemic proceduralism: I hold that epistemic value is objective and inherently different from how it is defined by various forms of social constructivism—which upturns the concept of truth by relativizing and annulling its objective value, or by considering it a product of (good) political processes.

Although he does not endorse Peter's pure epistemic proceduralism, Mladenović notes that I erroneously equate her pure epistemic proceduralism with mere fair proceduralism, as Peter advocates precisely for epistemic proceduralism. In contrast, real fair proceduralism is utterly insensitive to epistemic justification. Likewise, he holds that I wrongly neglect the fact Estlund does understand the importance of procedure-independent standards. I repeat that both Estlund and Peter have done a pioneering job in tackling the question of epistemic justification. Before them, political philosophers have ignored, purposely disregarded, or explicitly refused the issue of the epistemic value of democracy. However, once Estlund and Peter have put this question on the table and opened this significant philosophical debate, we must ask whether they correctly understood the nature of epistemic justification. Pure epistemic proceduralism, precisely because of its utter reduction of epistemic justification to the political, is essentially no different from that proceduralism that deals only with the political values of democratic procedures. Fair proceduralism and pure epistemic proceduralism are not entirely equal because Peter emphasizes the significance of epistemic justification, but their final evaluations are very much alike: everything comes down to assessing and improving the political value of procedures. To make myself clear, while I do believe that the political value of processes is significant, it is not sufficient for justifying democracy.

In his version of (moderate) epistemic proceduralism, Estlund claims that epistemic justification must possess independent standards, only to end with the conclusion that it is acceptable for democratic procedures not to generate truth in some particular case as long as they, in general, produce correct decisions. Mladenović, echoing Estlund, claims that establishing truth as a procedure-independent standard that must always be met is not necessary and that my criterion of epistemic justification is both redundant and overly ambitious. "It is reasonable to suppose that what Estlund terms 'primary bads' such as war, famine, economic collapse, genocide, belong to this class of procedure-independent standards," he writes, and continues "Consequently, for justification of democracy and democratic authority, truth doesn't need to be the only relevant epistemic standard. It is more reasonable to assume that for justification of democracy, some other procedure-independent standards should have their epistemic significance and that democratic decision-making given its inherently epistemic characteristics should provide reasons for action or reasons to comply" (Mladenović 2020). It is evident that the critical difference between epistemic proceduralism and my stance, which I call reliability democracy, is in how I understand the nature of epistemic justification. While I see epistemic justification as an intrinsic feature independent from political justification and use this definition in establishing procedure-independent epistemic standards, Estlund and Mladenović do not consider this necessary. While I believe that the epistemic justification of democracy must rest on objective epistemic value or quality, they seek epistemic value in the constructs of fair political procedures or, in particular cases, in some political/ethical values, such as "the elimination of primary bads," independent from what processes can generate.

Mladenović's concept of reasonable agreement or acceptance, which he proposes as a suitable replacement for my outmoded insistence on truth, is undoubtedly a significant political value that ensures society's basic functioning in times of disagreement. However, for epistemic justification, what we need is truth-conducive agreement. We can also attain reasonable agreement around beliefs that are not true, and in procedures that are not (sufficiently) truth-conducive. There is nothing in the nature of fair processes, the general will, and reasonable agreement, that would imply they necessarily possess the epistemic feature of truth-conduciveness. In ideal epistemic circumstances of informed and epistemically responsible citizens who do not dogmatically hold on to their original stance, reasonable agreement could enjoy the feature of truth-conduciveness. However, real epistemic cases are sub-ideal. People are often both inadequately informed and unmotivated to form true beliefs and are subject to biases and prejudice created in closed informational bubbles and echo chambers. In such conditions, reasonable agreement does not have the potential to be truth-conductive.

## 3. *The division of epistemic labor between citizens and experts*

In his article "The Limits of Expertism," Nenad Miščević embraces—or, to be precise, does not even raise—the question of the epistemic justification of democracy, assuming that the legitimacy of democratic processes and practices must depend on the epistemic quality of their decisions. Miščević unambiguously accepts the stance that the epistemic quality of decisions and beliefs is closely tied to the epistemic value of truth. He also agrees that the epistemic virtues of some social practice, procedure, or institution, are intimately related to generating true beliefs, i.e., decisions based on true premises. Miščević correctly labels his and my position by calling us "truth-respecting theoreticians." From our agreement on this matter, I derive an understanding that Miščević does not even mention, and which is related to my previous discussion about the position of epistemic proceduralism. Since epistemic proceduralism rejects the procedure-independent and intrinsic standard of truth as the criterion of epistemic legitimacy, I assume that he would, like I have done, develop a critical stance towards epistemic proceduralism of any kind. Namely, his article makes it clear that the epistemic quality of democratic procedures is not ensured simply by making them fair, but that there is some external, procedure-independent, and objective criterion for assessing the epistemic quality of decisions.

Miščević focuses on the question of how to organize democratic procedures to yield the highest epistemic quality or to attain the value of truth. He seems to accept that this question demands we assess the role of experts in democratic deliberation and decision-making, as they are, as individual epistemic agents, the best guides to the truth—i.e., they are the most truth-conducive epistemic agents. My book devotes a lot of attention precisely to the role of experts. Contrary to epistemic proceduralists, who perceive any inclusion of experts in decision-making processes as a threat of unfair privileging or epistocracy, my model of reliability democracy attempts to show that the expert participation in the democratic decision-making process improves the reliability of procedures, and, in turn, the epistemic quality of the final decisions. However, the critical question is how to divide epistemic labor between citizens and experts without stripping procedures of their political justification (Christiano 2012). For epistemic proceduralists, this question is meaningless, as they use different methods to shun this possibility as both politically harmful and epistemically inefficient. While Miščević, on the other hand, agrees with my proposal to include experts to make the final decisions/beliefs/judgments more epistemically valuable, he suggests a different division of labor between citizens and experts. I will try to show that his model of the division of labor leaves less space to citizens, and is, thus, more expertist than my suggestion. I hold that, as such, he argues in favor of some kind of epistemic instrumentalism because he sacrifices political justification for epistemic values, while barely adding anything new to epistemic quality.

Miščević correctly interprets my stance that consensualism would be the ideal decision-making practice were we to live in ideal circumstances that satisfy the epistemic preconditions for participating in public deliberation. In other words, democratic deliberation and democratic procedures would have complete epistemic legitimacy if they, as processes, possessed the relevant epistemic features that made them reliable. In other words, participants in a debate should be (i) adequately informed about the topic they are discussing, which can be labeled *the condition of adequately informed participants*, and (ii) they must not be egoistically or emotionally tied to their stance in such a manner that they will immediately reject any opposing view, which we can call *the anti-dogmatic condition for participants* (Kitcher 2011, Lehrer and Wagner 1981). However, as it is unrealistic to assume that everyday democratic decision-making and voting procedures will meet these demands—which has been proven by ample empirical evidence—the fulfillment of these conditions can be considered an ideal scenario (Ahlstrom-Vij 2012, 2013, Sustain 2006). In ideal circumstances, democratic procedures and the resulting consensus would generate epistemically valuable decisions.

Nonetheless, in real or sub-ideal cases, people are usually inadequately informed about some of or all the topics they are deciding about, they have seldom been taught to absorb the detailed data needed for making decisions, are not motivated to form beliefs of high epistemic quality correctly, and do not have the time to do thorough research (Goldman 1991). What is more, in everyday decision-making processes, citizens are pliant to many biases, stereotypes, and prejudice, which they are (sometimes) unaware of and which they (voluntarily or automatically) do not control, which casts serious doubt on the condition of adequately informed participants, as well as on the anti-dogmatic condition for participants (Dunning and Kruger 1999, Fricker 2007).

There is also a myriad of structural social limitations in transmitting and filtering knowledge, and in communicating in a globalized world that relies on social networks, the Internet, and non-transparent algorithms for selecting and disseminating information, which leads most of us to live within echo chambers. Such a non-ideal conversational context for fulfilling the conditions of adequate knowledge and openness necessarily thwarts the epistemic quality of the decisions generated through fair democratic procedures. And finally, democratic decision-making itself—or the famous "wisdom of crowds"—has its internal deficits and limitations related to the flattening of beliefs to those which are understandable to everyone, and which are often not of the highest epistemic quality (Gigone and Hastie 1993, Prijić-Samaržija 2005, Prelec, Seung, and McCoy 2017). More succinct, in real-world situations that we describe as sub-ideal epistemic circumstances, it is difficult to expect that public deliberation will automatically generate an epistemically valuable or truth-conducive consensus. This is precisely why the distinction between ideal and sub-ideal epistemic conditions is crucial to understanding my position and the proposal of reliability democracy (Goldman 2010). Since we live in sub-optimal epistemic conditions, democratic deliberation will not automatically—merely by including all citizens in fair procedures—generate epistemic quality. What we need is to design democratic processes in such a way to make them as reliable as possible, i.e., to make them ensure the highest possible epistemic quality.

Miščević agrees that the difference between sub-ideal—or real-world—epistemic circumstances and ideal epistemic circumstances is vital for defining the division of epistemic labor. While I, within my real-world approach, focus on the question of epistemic relationships and the division of labor in sub-ideal circumstances to generate decisions of the highest epistemic quality, he suggests we should keep this parallelism in mind by assessing sub-ideal epistemic conditions as approximates to their ideal theoretic counterpart. While I suggest we explore how best to satisfy epistemic norms and which processes of dividing epistemic labor generate the highest epistemic quality while preserving the democratic rationale, Miščević recommends that we "project the notion of rationality downwards" from ideal circumstances to sub-ideal circumstances to ascertain how approximate they are to their idealized counterpart. Given that, contrary to epistemic proceduralism, I maintain the concept of procedure-independent epistemic quality (truth-conduciveness), which helps us ascertain whether our real and non-ideal circumstances meet specific epistemic standards. However, I hold that Miščević's example is excellent theoretical and methodological support. While I have attempted to show which aspects of the real world muddle epistemic quality, Miščević provides a useful methodological toolkit for establishing the epistemic quality we are after: we can imagine an ideal situation as a thought experiment and analyze whether our real circumstances are at all close to ideal. Both my empirical (or naturalistic) approach and his rationalist (or normative) approach could be methodologically beneficial for attaining more epistemic quality.

While we agree that epistemic justification is intrinsic, and while we share the same theoretical framework, we diverge on concrete proposals of procedures that would, in sub-ideal circumstances, ensure the highest epistemic quality. Simply put, we split on the question of the division of labor between citizens and experts. I begin from the attitude that, in sub-ideal circumstances, "the wisdom of crowds" will use no invisible hand to generate the epistemic quality of beliefs and decisions. Instead, we need to ensure it by including experts. My example of the division of labor between citizens and experts is as follows: through *consensus*, the procedures of public deliberation, and majority voting, citizens define the problem they need resolved and oversee the experts by confirming or rejecting their solutions to the problem. Experts, on the other hand, as agents expertly trained to solve problems within their area of expertise, seek answers to the suggested issues and present them to citizens. It is crucial to note that I think citizens should be the ones who will, through *consensus*, choose the experts they best trust to resolve their problems. Here I echo Elizabeth Anderson's empirical example that citizens with a minimum education and access to the Internet can select a trustworthy expert on the topic of, for example, global warming (Anderson 2011).

Miščević believes that the division of labor and the decision-making process should occur differently. First, he does not think that all citizens are capable of defining "goals and values" because the limitations—the fact they do not fulfill the epistemic conditions—that constrain them in resolving the detected problems will equally restrict them in defining the issues that the experts should address. Second, he beliefs that not all citizens, for the same reason, will be capable of detecting reliable experts. Instead, they will prefer those who share their stances and whose stances they can recognize. For these reasons, Miščević suggests that citizens, within their interest groups, choose experts who will represent them in later deliberations where experts will (i) define their goals and values, and (ii) best resolve their problems. In other words, citizens participate in debate within their groups—class, ethnic, gender, religious, or like—where they choose experts who will join experts from other—class, ethnic, gender, religious, or like—groups in resolving problems. According to Miščević, we can consider a situation where the chosen expert representatives from different groups deliberate about which issues should be fixed and then solve them some kind of optimum approximation of the ideal state, because experts are more capable of rational debate than citizens, which makes it more likely for them to satisfy our epistemic conditions.

Why do I think that Miščević's example is more expertist than my own? First, he reduces civic participation and deliberation between citizens to the choice of an expert who will (with other experts) assume the entire epistemic labor of defining the goals, resolving them, and overseeing how they are resolved. Miščević seems even more distrustful towards the wisdom of crowds than I am because he believes that the epistemic potential of collective intelligence and, in particular, the epistemic diversity of perspectives will not be able to generate epistemic quality in any aspect other than the choice of representative experts. Although all citizens are included, their role in decision-making is far more limited than in my suggestion, and the part of experts is increased. This reduction of civic participation in epistemic labor to the selection of representatives is unacceptable for three reasons.

First, the role of citizens in deliberation and decision-making is reduced in favor of experts, which upsets the balance between democratic and epistemic rationale or justification. In my book, I endorse a hybrid approach that simultaneously assesses the epistemic and the ethical/political justification of processes, practices, and institutions. As I have mentioned above, I believe that epistemic instrumentalism, which sacrifices political goals for epistemic values, is not an appropriate approach. Likewise, I hold that political instrumentalism, where epistemic values are sacrificed for the political, is equally unacceptable. This stance is why I characterize expertism (and its radical form, epistocracy)—the position where experts have the central role in decision-making—as epistemic instrumentalism. Thus, as my book asserts that, in sub-ideal circumstances, there is a structural conflict between political and epistemic values (because the political right to participate does not generate decisions of the highest epistemic quality), the hybrid model is a conscious and conscientious quest for a balance that maintains both political and epistemic values. This binds us to sacrifice the highest possible epistemic quality to preserve the democratic rationale but also to sacrifice political values by giving a unique role to experts. Miščević's proposal sacrifices political values for the epistemic to the extent that disbalances political and epistemic demands and, aiming to approximate the ideal of rationality, establishes a stronger expertism than I am willing to propose.

Second, I hold that Miščević has overlooked the epistemic potential of citizens, focusing on their epistemic deficits in sub-ideal circumstances. Just like the collective intelligence of crowds has its deficiencies, individual (expert) intelligence also entails its limitations, which urges us to find an appropriate balance or, more succinct, an antidote for both deficits. Since knowledge is dispersed throughout society (Hayek 1945), "many minds" know little about a lot, while experts know a lot about little (R. E. Goodin and K. Spiekermann 2018). There are many indicators that collective intelligence—due to cognitive diversity and the diversity of perspectives, heuristics, evidence, interpretations, and even biases—sometimes generates solutions better than those made by individual experts (Goodin 2006, Hong and Page 2004, Landemore 2013, 2014, Mercier and Sperber 2011, Page 2007, 2008, Zollman

2010). Moreover, randomly formed collectives are even more epistemically successful than structured collectives such as interest groups. Experts, on the other hand, belong to the homogenous world of the highly educated and the materially well off. Miščević attempts to secure the condition of diversity in his deliberative groups of experts by stressing that they come from different ideological groups, and preserves the desired level of rationality by only including experts.

However, advocates of collective intelligence claim that knowledge is dispersed through society, and experts cannot fully satisfy the condition of cognitive diversity. Randomly formed collections of citizens ensure a degree of diversity that makes them more reliable truth-trackers than groups of experts who advocate for different comprehensive doctrines. Keeping this in mind, we need to give citizens space where their epistemic advantage of diversity will yield the best epistemic results, which is in the areas where there is no highly sophisticated factual knowledge (Zubčić 2020, Janković 2020). This is precisely the space I recommend for citizens, who should have a crucial role in defining goals/problems, choosing the experts who will resolve those problems, and conducting a second-order assessment of the consensus of trustworthy scientific experts (Anderson 2011). In short, citizens' epistemic potential is underestimated and reduced to their choice of an ideological representative who will define their problems and then resolve them. Unlike Miščević, I can easily imagine that I, as a non-expert and a citizen, could choose a climatologist who does not belong to my ideological group if they could reasonably be tasked with resolving the previously defined problem of divesting from fossil fuels and transitioning to renewable resources. I can also imagine myself overseeing whether she is appropriately solving this problem. Likewise, I do not think anyone would struggle with choosing trustworthy macroeconomists, who might not belong to their worldview, if we have previously defined the issue of increased economic inequality as the problem he needs to resolve, nor would they struggle with a second-order assessment of whether the work is done. The role of experts lies in providing a technical solution to a problem based on factual knowledge—regardless of whether we are talking about science or morals and political questions. This is precisely why, in my proposal, *all* citizens choose experts who are not selected as the best representatives of their group interests but only as people who can best solve their problem. I believe this model preserves both the epistemic potential of the diversity present in collective intelligence and the epistemic potential of the factual knowledge embodied by individual intelligence.

Third, it is worth asking whether, in Miščević's division of epistemic labor, the chosen experts will be constrained in their representative role while making decisions and resolving problems. Namely, I am wondering about the condition of being non-dogmatic during deliberation. Since they are chosen to advance the group's interests, their potential for rational discussion is limited not only by their value judgments but by the fact they need to represent group values. The question is whether, in circumstances of disagreement, they are allowed to be open towards other experts. Their role is to represent their group's stances, which is why—even if they are, as experts, considerate of rational discourse and the strength of evidence—they cannot give up their initial stance. Their situation is an illustration of the Steadfast View in the debate about disagreement. According to the Steadfast View, the fact a peer disagrees with you is irrelevant. Because disagreement, even among peers, does not warrant a response, there may thus be cases where the uniquely rational response for both parties to a dispute is to stick to their initial beliefs (Kelly 2005). In this plot of deliberation between representative experts, we can even imagine the Extra Weight View, the stance that it is rational to give more weight to your own belief simply because it is yours (Wedgewood 2010). The potential of this approach to rationally resolve a disagreement is as small as possible and entails some worryingly skeptical conclusions—since both parties in deliberation stick to their original beliefs, both p and not-p can be considered equally epistemically valuable. In everyday real-world situations, this means there are no solutions to disagreements, and, thus, no solutions to problems. Should we, as Miščević writes, need a solution to the migrant crisis, representative experts from different groups will not be able to suspend their stance and the stance of the group. In other words, Miščević's representative experts do not satisfy the epistemic condition of a non-dogmatic approach to deliberation, which hampers the epistemic quality of their solutions, decisions, and beliefs. An expert chosen by everyone, on the other hand, does not have these constraints, and thus, despite all the restraints limiting them as an individual epistemic agent, he comes closer to the epistemic conditions for generating truth-conducive beliefs/decisions/solutions to problems.

## 4. *Conclusion*

In the wake of the culture of ignorance and the crisis of enlightenment, the epistemic justification of democracy as a system that makes its decisions through democratic procedures is of utmost importance (DeNicola 2017). However, it is equally important not to perceive epistemic justification only as a byproduct of fair democratic processes but as an intrinsic value related to objective and procedure-independent epistemic value, or, more succinct, to truth-conduciveness. In this sense, the decision to exclude experts from democratic decision-making procedures—a choice we find in epistemic proceduralism as a critique of both epistocracy and more moderate forms of expertism—is not only unjustified but entails unwanted consequences by increasing distrust towards experts and their expertise. We live in a time when citizens distrust experts or all the wrong reasons: citizens do not doubt experts because they have, as individuals or groups, shown they have not been able to solve the citizens' problems (which is a reason why they indeed should not be granted trust) but because their central virtues of expertise and objective epistemic value have been brought into question.

Today, people distrust experts because they generally do not believe in expertise, which places us firmly within a culture of ignorance. This predicament is particularly harmful because the expertise only real experts can exercise, and which we need for the epistemic quality of our decisions/beliefs, cannot be substituted by the fairest and the most democratic procedures. For this particular reason, we need to reconsider the role of experts in democratic processes and, to ensure epistemic quality, make room for real expertise and those experts who reliably practice it. However, the part of experts in the division of epistemic labor must be appropriately balanced with democratic procedures and civic participation. Experts should present themselves not only through truth-revealing situations that paint them as those who solve their problems but as responsible professionals. In simpler terms, experts must show they are aware of their value limitations, that they acknowledge the citizens' goals and concerns, and that they are non-dogmatic (meaning, that they are willing to resolve disputes by altering their position, rather than by sticking to their original stance). To craft the best division of epistemic labor, we must acknowledge that civic participation is not only the space of the political justification of democracy. Instead, it also contributes to epistemic justification, which is why we must give citizens a fitting role in improving the epistemic quality of democratic procedures.

## *Bibliography*

Ahlstrom-Vij, K. 2012. "Why Deliberative Democracy is (Still) Untenable?" *Public Affairs Quarterly* 26 (3): 199–220.

Ahlstrom-Vij, K. 2013. "Why We Cannot Rely on Ourselves for Epistemic Improvement." *Philosophical Issues* 23: 276–296.

Anderson, E. 2011. "Democracy, Public Policy, and Lay Assessments of Scientific Testimony." *Episteme* 8 (2), 144–164.

Christiano, T. 2012. "Rational Deliberation between Experts and Citizens." In J. Mansbridge and J. Parkinson (eds.). *The Deliberative Systems. Deliberative Democracy at the Large Scale.* Cambridge: Cambridge University Press.

DeNicola, D. R. 2017. *Understanding Ignorance.* Cambridge: MIT Press.

Estulnd, D. M. 2008a. "Epistemic Proceduralism and Democratic Authority." In R. Geenens and R. Tinnevelt (eds.). *Does Truth Matter? Democracy and Public Space.* Springer: Dordrecht.

Estulnd, D. M. 2008b. *Democratic Authority. A Philosophical Framework.* Princeton: Princeton University Press.

Fricker, M. 2007. *Epistemic Injustice.* Oxford: Oxford University Press.

Gigone, D. and Hastie, R. 1993. "The common knowledge effect: Information sharing and group judgment." *Journal of Personality and Social Psychology* 65: 959–974.

Goodin, R. E. 2006. “The Benefits of Multiple Biased Observers.” *Episteme* 3 (3): 166–174.

Goodin, R. E. and Spiekerman, K. 2018. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.

Goldman, A. I. 1991. “Epistemic Paternalism: Communication Control in Law and Society.” *Journal of Philosophy* 88: 113–131.

Goldman, A. I. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.

Goldman, A. I. 2010. “Why Social Epistemology is Real Epistemology?” In A.Haddock, A. Millar, and D. Pritchard (eds.). *Social Epistemology*. Oxford: Oxford University Press.

Hong, L. and Page, S. E. 2004. “Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (46): 16385–16389.

Jankovic, I. 2020. “Epistemic Feature of Democracy: the Role of Expert in Democratic Decision Making.” *Philosophy and Society* 31 (1): 37–42.

Kitcher, P. 2011. *Science in a Democratic Society*. Amherst, New York: Prometheus Books.

Kruger, J. and Dunning, D. 1999. “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Incompetence Lead to Inflated Self-Assessments.” *Journal of Personality and Social Psychology* 77 (6): 1121–1134.

Landemore, H. 2012. “Why the Many are Smarter than the Few and why It Matters.” *Journal of Public Deliberation* 8 (1): 7.

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

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

Mladenović, I. 2020. “Democracy, Truth, and Epistemic Proceduralism.” *Croatian Journal of Philosophy* 19 (2): 171–182.

Page, S. E. 2007. “Making the Difference: Applying a Logic of Diversity.” *The Academy of Management Perspectives* 21 (4): 6–20.

Page, S. E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton: Princeton University Press.

Peter, F. 2008. *Democratic Legitimacy*. New York: Routledge.

Peter, F. 2013. “The Procedural Epistemic Value of Deliberation.” *Synthese* 190 (7): 1253–1266.

Prelec, D., Seung, S. H. and McCoy, J. 2017. “A solution to the single-question crowd wisdom problem.” *Nature* 541 (7638): 532–535.

Prijić Samaržija, S. 2017. “The Role of Experts in a Democratic Decision-making Process.” *Etica & Politica / Ethics & Politics* 19(2): 229–246.

Prijić Samaržija, S. 2018. *Democracy and Truth: The Conflict Between Political and Epistemic Virtues*. Milano, Udine: Mimesis International.

Sunstein, C. 2006. *Infotopia: How many Minds Produce Knowledge*. New York: Oxford University Press.

Zollman, K. 2010. “The Epistemic Benefit of Transient Diversity.” *Erkenntnis* 72 (1): 17–35.

Zubčić, M. L. 2020. “Social Epistemic Inequalities, Redundancy and Epistemic Reliability in Governance.” *Philosophy and Society* 31 (1): 43–55.

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *Which Theory of Public Reason? Epistemic Injustice and Public Reason*

ELVIO BACCARINI
*University of Rijeka, Rijeka, Croatia*

*Rawlsian public reason requires public decisions to be justified through reasons that each citizen can accept as reasonable, free and equal. It has been objected that this model of public justification puts unfair burdens on marginalized groups. A possible version of the criticism is that the alleged unfairness is constituted by what Miranda Fricker and other authors call epistemic injustice. This form of injustice obtains when some agents are unjustly treated as not reliable, or when they are deprived of epistemic resources to utter their claims or burdened when they need to express demands. I show that the Rawlsian model can stand the objection. Restricting justificatory reasons, at least when basic issues of human rights, liberties and opportunities are at stake, is needed in order to warrant a stable society as a fair system of cooperation among free and equal citizens.*

**Keywords:** Epistemic injustice, Miranda Fricker, public reason, John Rawls.

## *1.*

According to the Rawlsian view of political legitimacy, at least fundamental public decisions related to basic rights and liberties must be justified through reasons that all agents can accept as reasonable, free and equal. Namely, justifying a public decision through reasons that one could reject as a reasonable, free and equal person would risk enforcing decisions that infringe basic rights, liberties and opportunities, and endangering society as a stable system of cooperation among free and equal persons (Rawls 2005). Such a view is challenged in various ways. Among other objections, one states that a principle of legitimacy which insists on such justificatory reasons discriminates minorities, or disadvantaged groups that feel uneasy, or are not able, to make use of the mainstream political language and conceptual scheme of egalitarian liberal societies (Peñalver 2007, Dyer and Stuart 2013).

In the present paper, I examine whether such objection correctly attributes epistemic injustice to the Rawlsian proposal. The concept of epistemic injustice is first introduced and defined by Miranda Fricker (2007), and it indicates situations where agents are treated epistemically unfairly, or are in an unfair epistemic position due to their group belonging. After Fricker formulated the theory, there have been further developments (Anderson 2017, Dotson 2008, 2011, 2012, 2014, Medina 2013, 2018).

Fricker and other authors engaged in epistemic justice debates do not explicitly criticize the conception of public reason from the perspective of epistemic injustice, and, as far as I know, the two debates have never been put explicitly in relation. However, formulating the debate in terms of epistemic injustice seems as a possible interpretation of some of the criticism of public reason. Thus, I think that it is helpful to analyse the issue explicitly in terms of epistemic injustice in order to evaluate the public reason theory. Answering to the challenge of epistemic injustice is very important for the theory of public reason in virtue of its commitment to freedom and equality because epistemic injustice harms protection of such ideals. This is why it is crucially important for the theory to defeat this objection.

If the objection of epistemic injustice succeeds, public justification would have to be more open to a variety of reasons. Among them, there is the convergence theory of public reason, which requires that public decisions must be justified through reasons that each qualified agent can accept, but it is not necessary that the reasons are shared, because convergence of different reasons is sufficient (Gaus 2011); the accessible reasons view of public justification that says that the necessary and sufficient condition for being a valid reason is that each qualified agent can interact with it, by understanding, debating, commenting, analysing, criticising, etc. it (Laborde 2017); substitution of public reasoning to public reason, i.e. establishing rules that agents can follow in the process of public justification, instead of defining in advance the reasons that can be used (Chambers 2010); substituting Rawls's static view of public reason, with a more dynamic view of reasons that must be constantly tested through the principle of rational verification and probability (Ferretti 2019).

In the paper, I proceed as follows:

1. I describe Rawls's conception of public reason.
2. I describe Fricker's conception of epistemic injustice.
3. I put forward a form of criticism of Rawls's conception of public reason that could be interpreted as a challenge of epistemic injustice.
4. I investigate whether Rawls's public reason is a form of epistemic injustice. I show that there are no elements in the epistemic injustice debate which can represent a basis for criticism of Rawls's public reason. I explain that Rawls's requirements of public reason are needed in order to secure freedom and equality.

5. I describe the demands of epistemic virtue and distributive epistemic justice in fair social interaction, as well as in the context of Rawls's political philosophy.

## *2.*

The theory of public reason is a theory of public justification, i.e. of justification of public rules. Its core idea is explained by the liberal principle of legitimacy: "Our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason" (Rawls 2005: 137). The principle, thus, requires restraint to use, in the process of justification of public decisions (that for Rawls are limited to constitutional essentials) reasons that not all reasonable agents can share as free and equal. In the light of Rawls's explanation of public reason, the principle of restraint can be interpreted, in a sense, like a principle of translation as well. Namely, citizens can use non-public reasons in their public justification of decisions, provided they offer translation in public reason terms and conceptual scheme in due time (Rawls 1999: 584, 591–593). Here a specification is needed. In Rawls's original view, the restraints of public reason apply primarily to constitutional essentials and "in other cases insofar as they border on those essentials and become politically divisive" (Rawls 2001: 117). The limitation to such domain, however, is not so strict for all authors. Some of them explicitly extend public reason to all laws and public policies (Quong 2011). Gerald Gaus even extends public reason to cover all social morality (Gaus 2011). These, however, are not issues that I will adjudicate here. In this paper, I cover all these cases.

Among reasons that can be employed in public justification, according to the liberal principle of legitimacy are, primarily, ideals like that of society as a fair system of social cooperation, certain basic rights, liberties and opportunities, and concepts related to these basic organizing ideas. Further public reasons are represented by "presently accepted general beliefs and forms of reasoning found in common sense, and the methods and conclusions of science when these are not controversial" (Rawls 2005: 224). Note that these reasons are dynamic. Methods and conclusions of science can, obviously change through development of science. Likewise, this happens for general beliefs and forms of reasoning in common sense. In particular, this holds when we consider possible pressure to common sense beliefs that can come from better consideration of what is coherent with the organizing political ideas of reasonable, free and equal persons seen above.

Public reason is characterized by epistemic responsibility, as well. Persons who participate in it are in a condition of disagreement about general doctrines not because of epistemic irresponsibility, but in virtue of burdens of judgment, i.e. difficulties in reasoning about moral issues. An example of this is empirical underdetermination (Rawls 2005: 54–58).

Think about the debate on abortion. As Rawls famously said, it is legitimate to justify a law on abortion by appealing to reasonable public reasons like the value of human life, the equality of women, the right to choose in religious or moral issues. It is not legitimate to justify a law by appealing to comprehensive doctrines like religions or moral theories like utilitarianism (Rawls 2005: 243). In another example, it is legitimate to pass a law against racial segregation by appealing to values of freedom and equality of all people, but not to a religious doctrine that says that we are all equally God's children. Jonathan Quong offers an instructive example of application of the Rawlsian view of public reason.

The example regards passing a law that affirms the legitimacy of same-sex marriage (Quong 2013: 1). Imagine that some people say that marriage is one of the aspects of human flourishing. Imagine that the same people say that same-sex marriage fully realizes this. They pass the law on the basis on these beliefs. Quong, by instantiating the requirements of Rawlsian public reason, says that such a law is not legitimate, because it is supported by sectarian reasons. In other words, it is not admitted to appeal to a kind of controversial view of human flourishing in order to pass a law.

It is important to keep in mind what is the central rationale for public reason: to ground the project of building a stable cooperative society of free and equal persons, by avoiding threats that could result from leaving public decisions to the contingencies of various worldviews.[1] This rationale for the Rawlsian view of public reason is sustained by the occasions in the history of democracy where minorities have been deprived of their rights by majority votes. Think about the racial segregation in USA, an output of democratic representative system. The Rawlsian view of public reason answers to this demand by denying legitimacy to public decisions justified through reasons that not all reasonable agents as free and equal, as well as cognitively responsible, can accept.

## *3.*

Let's see, now, what is epistemic injustice, in its various manifestations. Fricker speaks about three distinct forms of epistemic injustice. One is testimonial injustice. It "occurs when prejudice causes a hearer to give a deflated level of credibility to a speaker's word" (Fricker 2007: 1). Fricker exemplifies this form of epistemic injustice through a conversation that includes characters in *The Talented Mr. Ripley*. The detective, Greenleaf, a character in the story, was investigating a case of murder. He disregarded the epistemic attitude, i.e. the suspicions,

[1] My proposal is particularly inspired by Jonathan Quong's theory (2011).

of the victim's fiancé toward a potential criminal with the sentence "Marge, there's female intuition, and then there are facts" (Fricker 2007: 14), where obviously the intention was to dismiss her as a reliable provider of testimony and epistemic contribution. This is clearly a case of epistemic injustice, because on the basis of identity prejudice the detective did not even take in consideration the cognitive contribution of the woman as a possible candidate for truth that deserves to be explored. Epistemic injustice consists in excluding the possibility that a woman can rationally consider evidence, and assuming that she just has unfounded intuitions. The exclusion was clearly gender oriented and the detective would not dismiss in a similar way a male colleague.

Hermeneutical injustice manifests itself because of an unfair relation, where the privileged have at their disposal communicative resources to interpret and express social experiences that are significant to them, while the less advantaged are deprived of these capabilities (Fricker 2007: 155). A clear example of hermeneutical epistemic injustice is represented by sexual harassment. Before the phenomenon was individuated, the members of the discriminated group had not had the resources to render intelligible something that was harmful to them. Victims were pressed to describe the harm in already publicly managed and shared descriptions of harms. In their absence, they were unable to express the harm that they suffered. Hermeneutical epistemic injustice was constituted by the absence of such resources that caused strong burdens in communicating meaning relevant for them.

Members of the privileged group were cognitively deprived of the understanding of the social phenomenon as well, but they had benefits from the cognitive lacuna in their social context. As a consequence of this condition, harms of sexual harassment were underestimated for a long time. This was due to unfair relations of power that caused unfair epistemic interpretative relations (Fricker 2007: 156–157).

There is a third form of epistemic injustice remarked by Fricker, i.e. distributive epistemic injustice, that consists in "the unfair distribution of epistemic goods such as education or formation" (Fricker 2013: 1318).

In more recent debates, some other forms of epistemic injustice have been discussed. Kristie Dotson has conceptualised contributory epistemic injustice. An important novelty of her contribution is remarking that there are different hermeneutical resources utilised by different groups. Here, there is an important difference from Fricker's theory. In Fricker's description, even the disadvantaged group does not have the hermeneutical resources needed for fully expressing their condition. Hermeneutical injustice is represented by this shared deficit in society. In Dotson's description, the worse off dispose of resources to describe their own condition. They are, however, forced to abstain from using it. Instead, they are forced to use a conceptual scheme that does not entirely express their condition. This is because of wilful ignorance and structurally prejudiced hermeneutical resources (which can be even unintentional) of the other side. An example can be represented by women of a minority who are inhibited in testifying family violence of which they are victims, in order to avoid to strengthen prejudices that concern their community as being particularly constituted by violent persons. Inequity is not represented by the general absence of hermeneutical resources that has unequal consequences for different groups. It is manifested by a reduction of the ability of some agents to participate in an epistemic community because the other side persists in wilful ignorance and structural prejudices (Dotson 2012, 2014).

Jose Medina has described further manifestations of epistemic injustice. He has interpreted testimonial and hermeneutical epistemic injustice like flaws in proper recognition. Agents, or actions, are recognized in the wrong way (Medina 2018). For example, civil disobedience inspired by achievement of equality is recognized like sedition.

A specific form of epistemic injustice has been denounced by Derek Anderson (2017). This is conceptual competence injustice, i.e. unwarranted denial of competence in managing a priori claims that cannot be assessed empirically. In such a form of epistemic injustice, a representative of a marginalized group is denied conceptual competence, as a result of systematic (economic, educational, etc.) oppression of the group to which she belongs. Conceptual competence injustice matters for the present paper, because, like Andersons says, it regards, among else, competence in managing concepts relevant for the present discussion, like assessments of justice, or injustice. However, I do not discuss it specifically, because it is a kind of testimonial injustice, as Anderson says.

## *4.*

I show now the possible connections between public reason and epistemic injustice. I focus here on the claim that public reason's restraint requirement deprives people of expressive resources that they need, because they are not able to use other resources, or, because these resources are not replaceable for them. Some critics of public reason say that the restraint and translation rule puts particular burdens exactly on the discriminated minorities. It leaves unrecognized important parts of the experience of those parts of society who are already disadvantaged. For example, in USA, "African-Americans and the poor, both of whom benefit enormously from churches' egalitarian inculcation of civic engagement and skills" (Peñalver 2007: 539).

Such a criticism of public reason is instantiated by Eduardo Peñalver who says that the principle of restraint "of public reason can work to silence the central, and perhaps most compelling, elements of religious speakers' political beliefs and motives" (Peñalver 2007: 533).

Peñalver indicates the example of Karl Barth's evangelical moral theology, according to which people's knowledge of the good is rooted in the perception of a command of God, apprehended in an immediate, direct and intimate personal account with God. God's will is known by a single person, but this is not visible to other persons, nor "available to his own reflection. A believer committed to Barth's conception of ethics would not be able to offer any publicly cognizable reason for her behaviour, even in the form of a mediating principle" (Peñalver 2007: 534). As a consequence, such believers are excluded from the process of public deliberation.

Even when some people or communities are able to translate their claims, something may be, and sometimes will be, lost in translation. When religious people translate their arguments into the language of public reason, the arguments are less compelling. "The assertion of the less persuasive public arguments—the price for admission into public discourse for an inclusive system of public reason—could undermine the credibility of the non-public arguments" (Peñalver 2007: 535). Peñalver's criticism shown above suggests that the restraint rule of public reason corresponds to hermeneutical epistemic injustice, or, perhaps more precisely, to contributory epistemic injustice. In his description, the principle of restraint exactly deprives agents of the resources, or the best resources, that they have, for the expression of burdens, harms, disadvantages, or unfairness, that they endure.

Further problems, claims Peñalver, are that those who must translate their original messages will in some sense be epistemologically stigmatized, or, more precisely, their non-public discourse will be stigmatized. In addition, it can be possible that the stigmatization extends to the use of public reasons of such persons. For example, Alf can discredit Betty's attempt to evaluate an action as unjust by the employment of some public reason concepts, in virtue of the non public reasons that she embraces.[2] Injustice is manifested in comments like "We can't take seriously a claim of justice, when it is expressed by a person who believes in such an unreliable religious doctrine". Here, it seems that we see the charge of testimonial epistemic injustice (in Betty's case, we have a conceptual competence testimonial injustice).

Critics of public reason could press with the objection despite a general disanalogy between the definition of epistemic injustice and the restraint rule of public reason. An important component of epistemic injustice, as Fricker indicates, is an identity prejudice, at the core of the discrimination of some groups. There is no such prejudice implied or present in the theory of public reason. Its inspiration is addressed against segregation, harassment, etc. The intention of the theory of public reason is to hamper the mechanisms of majoritarian aggregative democracy that risk to found a society ruled by principles different than those based on the ideal of society among free and equal. In Rawls's view, this is achieved since according to his theory laws are legitimate

[2] An analogous objection has been raised by one of the reviewers of the article, to whom I express my thanks.

only if each and every person can accept the sustaining reasons of the laws as reasonable, free and equal. But the final result, say the critics, is marginalization of some minorities and depriving them of resources to oppose damages and unfairness, and to protect themselves.

## *5.*

In what follows, I argue that the restraint rule of public reason does not cause epistemic injustice. Firstly, I reject the accusation of testimonial epistemic injustice that is related to the charge of stigmatization. In my view, the restraint and translation rule does not represent nor favour stigmatization of persons, or the doctrines that they endorse, because it does not exclude them as valid public reasons in public justification in virtue of their epistemic weakness. On the contrary, they can even be recognised as sophisticated intellectual constructions. But they are still under the constraint of showing that they are reasonable in Rawls's sense. In other words, they need to show that they are suitable as justificatory reasons in the project of a stable cooperative society among free and equals. In order to achieve this, they need to show that they can be put in positive relation with the organizing ideas of such a society, which implies their translation in public reason terms.

In reply to the other charge of stigmatization, I remind the intention of my paper. This is to discuss whether requirements of public reason as such represent epistemic injustice. But, far from corresponding to a requirement of public reason, a reaction like the one indicated above, "We can't take seriously a claim of justice, when it is expressed by a person who believes in such an unreliable religious doctrine", is simply unreasonable and cannot function as a justificatory reason, in the public reason model. Thus, although the present objection indicates a realistic problem in the process of public justification, it is not an objection to public reason as such, but to a deviation from public reason.

In reply to the stigmatization objection, I remind also about Rawls's proviso. As I have shown above, Rawls admits the public employment of sectarian doctrines. Such employment can, even, be helpful. They are thus not stigmatized, but respected, and, in some cases, even welcome when they can help public reasons. The requirement is not to ban them, but to translate them in due time to public reason terms. Restraint regards only the last stage of the decision-making process, when proposals must be verified as suitable for a stable society of free and equals.

Still, there could be a problem of testimonial injustice in the fact that maybe some non public reasons correspond to truth, and, despite this, it is not allowed to use them in public justification. One could, even, object that there is contributory injustice, because other persons persist in their ignorance, instead of acknowledging true doctrines. But I think that there is no injustice here, provided that other persons have listened carefully and bona fide bearers of such a doctrine, and, still, there is reasonable disagreement due to burdens of judgment. In such a case, it is needed to show the positive relation of the doctrine with the organizing public reasons in order to ground a stable cooperative society of free and equals.

In fact, Fricker does not see an instantiation of epistemic injustice, neither in its hermeneutical, as well as in its testimonial form, in the constraint which requires from agents that they translate demands or complaints into shareable terms. Correspondingly the strategies of discriminated groups to express the harms that they are subjected to that Fricker illustrates are not constituted by expressing their stance in specific terms that it is not possible to share with others. On the contrary, she shows attempts to shape their social experience in a way that can become part of shared meanings. In the harassment case, for example, the strategy appears to be that of creating public awareness of a peculiar way of how people can be victims of humiliation, of commodification, of denial of autonomy, etc.

An example of this is, in Fricker's view, the achievement of a group of women in determining that the term "harassment" is the proper concept for describing the harm they suffer. Here an extensive description (quoted by Fricker in her book) of the way how 'sexual harassment' was endorsed, is illustrative: "'Eight of us were sitting in an office of Human Affairs', […] 'brainstorming about what we were going to write on the posters for our speak-out. We were referring to it as 'sexual intimidation', 'sexual coercion', 'sexual exploitation on the job'. None of those names seemed quite right. We wanted something that embraced a whole range of subtle and unsubtle persistent behaviors. Somebody came up with 'harassment'. Sexual harassment! Instantly we agreed. That's what it was" (Fricker 2007: 150).

A critic of public reason could still protest by saying that it puts on the victim a further burden, that of shaping the debate in Rawlsian public reason terms, i.e. in terms around the organizing idea of society as a fair system of cooperation among free and equal citizens. Valid public reasons must be related to such organizing idea, as well as to basic rights, liberties and opportunities.

My reply is that once harassment and domestic violence have been explained in terms of humiliation, harm to autonomy, harm to integrity, etc., and once it has been explained that not leaving the husband is not a sign of complacence, the explanation in terms of society as a fair system of cooperation among free and equal citizens, as well as certain basic rights and liberties, becomes feasible. Further, and more fundamentally, the public reason restraint and translation requirement does not represent injustice, because it is not arbitrary. Some constraints are, always, needed for participation in public deliberative process. For example, the requirement to be familiar with a common language could be justified in some conditions. The request is not unjust, if it has justification. Similarly, engagement in public reason terms is justified by the commitment to a stable society among free and equals, and thus requiring it does not represent injustice.

## *6.*

I further defend the restraint and translation requirement of public reason, by addressing a challenge to its critics. The basic problem for those who criticize the public reason model of justification of public decisions for causing epistemic injustice is to provide a model of public justification that is better in protecting freedom and equality. For the strand of criticism that I show in this paper, the apparent proposal is to let each group use its own resources, which means that they are not required to relate their justification to the idea of society as a fair system of cooperation among free and equals, nor to the basic rights, liberties and opportunities, as organizing ideas of public justification. But, then, the question is how to distinguish legitimate from illegitimate claims. There is the serious problem of not being able to rebut discriminatory claims, for example.

The translation in public reason terms is important in order to distinguish real cases of discrimination or unfairness from improper requirements that are exactly addressed to justify injustice, and, even, oppression. Think about the dramatic problem of infibulation. A case in Seattle, described by Jacob Levy is instructive (2000). Communities that immigrated there, practice infibulation, which, as we all know, is an extremely cruel ritual. "Those who do not die of blood loss or infection face a life of great pain during sexual intercourse and great danger during childbirth" (Levy 2000: 54). As Levy indicates, there was a debate among the committee of the Medical Center and representatives of the community in order to look for a compromise that, although not able to affirm the equality of women in that community, at least will save the functionality of their bodies, and will avoid pain and dangers related to the absence of hygienic conditions. A base of compromise was found, because at least some representatives of the communities agreed that sunna circumcision (judged by medical experts as analogous to male circumcision) in appropriate hygienic conditions would be sufficient to meet the cultural and religious requirements met by infibulations. Here representatives of the minority communities expressed something that is important for them, i.e. acceptance of signs of the supremacy of males, and their messages were understood. Such matters important for communities and their members are not understood when the communities are hermeneutically marginalized and are interpreted only through the categories of the mainstream egalitarian liberal culture. But what are the consequences of this understanding, in cases like infibulation?

There may be practical consequences in conflict management (Ceva 2016). The goal, in such a context, is to render possible to communities to speak with each other in a fair interaction. Further, there could be good practical consequences. An agreement can be reached that can save the lives and the quality of life of many girls. But nothing really substantial is obtained from the point of view of justification of public rules or policies as far as justice is concerned. As opponents of the compromise say, "[t]he cut might have only been symbolic, but it was symbolic of a tradition that insisted on controlling girls, symbolic of a particularly brutal kind of repression. The cultural need that the hospital was seeking to fill was seen as an illegitimate one, the need to have at least the symbols of control over the sexuality of girls and women" (Levy 2000: 54).

The humiliation, or sense of oppression, suffered by people who are forbidden to practice a substitute of infibulation, as required by their tradition, is properly judged as not giving the entitlement to a claim of justice. The public reason strategy indicates as the test for the legitimacy of a claim the possibility to be described in the language of reasonable public values, as related to the organizing idea of society as a fair system of cooperation among free and equal citizens, and certain basic rights, liberties and opportunities. The substitute of practice of infibulation, even after proper engagement in interpretative efforts, is still properly interpreted as a claim for a privilege to oppress. Thus, the practice is not rejected as a claim of justice because of hermeneutical marginalization of a minority community. The reason of exclusion is the impossibility to translate the claim for the practice in terms respectful of freedom and equality.[3] This is a reasonable and justified test.

The same happens, for example, in the cases of social movements and communities that require laws that discriminate against homosexuals, denying to them, for example, public offices and positions of teachers.

Obviously, not all religious requirements are so horrible like infibulation, or clearly discriminatory, like the one shown above.[4] Some appear as less harmful and potentially legitimate even to some liberals, like denying same-sex marriage, or exposition of religious symbols in public institutions (Laborde 2017). In some cases, religious appeals are clearly good. Dyer and Stuart's reference to Martin Luther King indicates such a prominent case (Dyer and Stuart 2013). But still, the important message of the clearly horrible or harmful examples is that we cannot take religions (or other non-public doctrines) like self-authenticating sources of public norms. Their claims must be assessed through public justification that warrants protection of a stable order protective of freedom and equality. Otherwise, we lose a criterion to determine when appeals to religion are discriminatory (like forbidding public employment for sexual minorities), when they are not (like, maybe, in the requirement to expose religious symbols in public institutions), or when

[3] Thanks to a reviewer for the suggestion of this formulation.

[4] Thanks to a reviewer for pressing this point.

they even contribute to proper human rights, liberties and opportunities (like in King's case). The needed warrant is achieved through the public reason constraint.[5] This can put additional or specific burdens on some communities, until they develop capabilities to express their claims in public reason terms, but this is a price that needs to be paid, for the sake of stable protection of the order of freedom and equality.

To be sure, I do not think that this unfortunate consequence regards claims of racial equality. Even without entering in analyses of real-world cases, like claims of the civil rights movement, or court decisions, we can confirm this thesis by looking at theoretical disputes. Think about an actual debate on Rawls's theory of justice and racial justice. I will present it here in the shortest term, just to give a brief illustration. Tommie Shelby says that everything needed for a proper approach to racial justice is present in Rawls's theory of justice, in particular a non-discriminatory principle of liberty, as well as a principle of fair equality of opportunity (Shelby 2013). On the contrary, Charles Mills says that the excessive abstraction of Rawls's theory of justice does not permit to deal in appropriate way with questions of racial justice (Mills 2017). It is necessary to take in consideration real life facts, like the history of injustice and needs for correction.[6]

What matters, for the present discussion, is not whether Rawls's theory of justice meets satisfactorily questions of racial justice, because I do not discuss Rawls's theory of justice. I am concerned, here, with Rawls's theory of legitimacy of public decisions, concretely, in the actual example, with the question whether claims of racial justice can be properly framed in public reason terms. The Mills vs Shelby debate shows that it can. In fact, both authors use proper public reasons, i.e. the kind or reasons that can be addressed to persons as reasonable, free and equal. These are various concepts and theses that can be coherent with the fundamental idea of society as a stable system of cooperation among free and equals, as well as with some basic ideals like rights, liberties and opportunities, or can even be such ideas and ideals themselves. Such concepts are for example appeals to past injustice and requirements of corrective justice.

## *7.*

There is still the problem that some people may be (and in fact, are) unfairly situated in matching the requirements of public reason. Here some important lessons can be drawn from Fricker's discussion of epistemic virtue and distributive epistemic injustice (Fricker 2013). The requirement of epistemic virtue is that representatives of the mainstream group or community be engaged in giving due respect and attention to the expressive resources of minorities in order to understand the sig-

[5] My claim here is inspired by Quong (2011, 2013).

[6] Thanks to a reviewer for the indication of relevant authors.

nificance of some social phenomena, a social atmosphere, a system of regulation, or a policy for them. As Fricker says: "The form the virtue of hermeneutical justice must take, then, is an alertness or sensitivity to the possibility that the difficulty one's interlocutor is having as she tries to render something communicatively intelligible is due not to its being a nonsense or her being a fool, but rather to some sort of gap in collective hermeneutical resources" (Fricker 2007: 169). Importantly, nothing at all here is said against public reason's requirement of translating claims into the conceptual scheme of public reason. The reaction against the specific burdens of some people in front of the requirements of public reason could be not to give up public reason, but to do the outmost to alleviate such burdens, and prospectively to eliminate them.

Here it is important to remember that some of the critics of public reason say that it is unfair, and certainly not correspondent to a reasonable interpretation of the duty of civility, to put all the burdens of understanding in communication on the minorities that speak in the language of their comprehensive doctrines, typically, religious doctrines (Waldron 2012). One of Fricker's solutions is to attribute the duty of epistemic justice to the advantaged. In the present discussion, this means that citizens familiar with the language and conceptual scheme of public reason have the duty to contribute to the attempts of translation of minority claims in the language of reasonable public values, or to help minorities in this translation. They must participate in explaining why and how these claims are related to the basic organizing idea of society as a fair system of cooperation among free and equals, as well as to basic rights, liberties and opportunities.

Epistemic virtues include also a contribution to the proper recognition of persons and facts. This is because, as Medina indicates, epistemic injustice derives from, among else, wrongful recognition. Such is, for example, interpretation of agents as violent, instead of as being engaged in protests that aim to achieve equality. A similar misrecognition involves their actions (Medina 2018). Epistemic virtue requires that privileged members of society be engaged in the needed public shift of perspective and interpretation. This shift is then a precondition for the proper framing of debates in public reason terms.

However, the primary requirement of distributive epistemic justice is to provide discriminated groups with the capabilities to participate actively in public justification. I relate the requirement to Rawls's idea of fair value of political liberties (Rawls 2005: 5). This idea requires the background of a basic structure of society that ensures resources to the discriminated group in order to help them to develop the ability to articulate their meanings in the language of reasonable public values. Formal equality is not sufficient. Substantial equality is needed, as well, i.e. equality from the standpoint of social and material resources in order not to be hermeneutically marginalized, as well as the possibility to be educated in order to have the possibility to learn the language of human rights and reasonable public values as a necessary condition to express claims for recognition of rights. Fair access to public media is needed, as well. Although, in the absence of the ability of marginalized groups to express their claims in public reason terms, there is a duty for mainstream groups to make the translation, the most complete accomplishment of epistemic justice is constituted by enabling minorities to express their claims in the public reason conceptual scheme. Despite compatibility with epistemic justice, public reason implies the exclusion of some groups from the process of public justification. Such groups are those that are definitely unable to translate their moral claims in the language of public reason, nor others can do this for them, because of the nature of the relevant experience of these people. Their moral experience is particularized and personalized and they cannot share it with the others. This is the case of Karl Barth's evangelical moral theology, as described by Peñalver (2007).

Their position, although unfortunate, is not unjust however. The reason is the same as the one I have indicated above in reply to the objection that public reason puts additional or specific burdens on some people. Participation in the process of public justification requires some justified preconditions, and one of them is the capacity to offer reasons to others as reasonable, free and equal.

Note that the alleged condition of epistemic injustice of Karl Barth's evangelical moral theology is different from Marge's, in Fricker's example. By assumption, the moral theologist in the example relies only on unshareable, not public and personal insights. Marge, on the other hand, could support her intuition in some way, or offer it only as the motivation to steer research in a specific direction, not as the only or ultimate cognitive resource. Epistemic injustice, in her case, is present because she is simply immediately excluded from cognitive contribution, irrespectively of her merits. If Marge insisted that her intuition is the only, or ultimate, source of justification, her opinion could be legitimately neglected, and it would be irrational to consider such exclusion unjust.

## *8.*

In conclusion, I do not claim that all public communication related to policy making must be always in terms of public reason. Practical needs could require, and frequently do require, strategies of interaction that neglect public reason, for the sake of communication with citizens who obstinately do not endorse its fundamental ideals. For example, some strategies of conflict management could be required (Ceva 2016). This is a valid reaction, sometimes the best available. It is not valid however, when we want to establish what is just, because it endangers the recognition of all citizens as free and equal, as we have seen in the infibulation example. Threatening the values of freedom and equality is not admissible in the justification of a conception of justice.

Finally, we see that Rawlsian public reason does not represent a case of epistemic injustice. On the contrary, it places reasonable constraints on agents to treat each other as free and equal. Further, I have shown reasons in its favour. In this way, I have contributed to its acceptance as the suitable form of public justification and public reason.[7]

## References

Anderson, D. E. "Conceptual Competence Injustice." *Social Epistemology* 31 (2): 210–223.

Ceva, E. 2016. *Interactive Justice. A Proceduralist Approach to Value Conflict in Politics*. London: Routledge.

Chambers, S. 2010. "Secularism Minus Exclusion: Developing a Religious-Friendly Idea of Public Reason." *The Good Society* 19 (2): 16–21.

Dotson, K. 2008. "In Search of Tanzania: Are Effective Epistemic Practices Sufficient for Just Epistemic Practices?" *The Southern Journal of Philosophy* 46 (S1): 52–64.

Dotson, K. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236–257.

Dotson, K. 2012. "A Cautionary Tale. On Limiting Epistemic Oppression." *Frontiers* 33 (1): 24–47.

Dotson, K. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28 (2): 115–138.

Dyer, J. B. and Stuart, K. E. 2013. "Rawlsian Public Reason and the Theological Framework of Martin Luther King's 'Letter from Birmingham City Jail.'" *Politics and Religion* 6 (1): 145–163.

Ferretti, M. P. 2019. *The Public Perspective. Public Justification and the Ethics of Belief*. London: Rowman and Littlefied.

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

Fricker, M. 2013. "Epistemic Justice as A Condition of Political Freedom?" *Synthese* 190 (7): 1317–1332.

Gaus, J. 2011. *The Order of Public Reason. A Theory of Freedom and Morality in a Diverse and Bounded World.* Cambridge: Cambridge University Press.

Laborde, C. 2017. *Liberalism's Religion*. Cambridge: Harvard University Press.

[7] This article is an outcome of research funded by the Croatian Science Foundation (HRZZ) (Project RAD, Grant IP-2018-01-3518). My gratitude goes to all colleagues and friends who have discussed with me previous versions of this paper, and in particular to Aleksandar Šušnjar who read the final version. Many thanks for precious comments to Miranda Fricker, as well as to other participants to the "Miranda Fricker symposium", held in Rijeka, in November 18th, 2013. Likewise, my thanks goes to all colleagues and friends who commented my talk held at LUISS Rome in January, 17th, 2014. I address my thanks to all colleagues and friends who participated in the discussion of my talk in May 2019, and in particular to Massimo Reichlin and Roberta Sala, during my Erasmus visit to the University San Raffaele in Milan. Many thanks to two reviewers of this journal.

Levy, J. T. 2000. *The Multiculturalism of Fear*. Oxford: Oxford University Press.

Medina, J. 2013. *The Epistemology of Resistance. Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford: Oxford University Press.

Medina, J. 2018. "Misrecognition and Epistemic Injustice." *Feminist Philosophical Quarterly* 4 (4): article 1.

Mills, C. W. 2017. *Black Rights/White Wrongs: The Critique of Racial Liberalism*. Oxford: Oxford University Press.

Peñalver, E. 2007. "Is Public Reason Counterproductive?" *West Virginia Law Review* 110 (515): 515–544.

Quong, J. 2011. *Liberalism without Perfection*. Oxford: Oxford University Press.

Quong, J. 2013. "Liberalism without Perfection. A Précis." *Philosophy and Public Issues* 2 (1): 1–6.

Rawls, J. 1999. "The Idea of Public Reason Revisited." In S. Freeman (ed.). *Collected Papers*. Cambridge: Harvard University Press.

Rawls, J. 2001. *Justice as Fairness. A Restatement*. Cambridge: Harvard University Press.

Rawls, J. 2005. *Political Liberalism*. New York: Columbia University Press.

Shelby, T. 2013. "Racial Realities and Corrective Justice." *Critical Philosophy of Race* 1 (2): 145–162.

Waldron, J. 2012. "Two-Way Translation: The Ethics of Engaging with Religious Contributions in Public Deliberation." *Mercer Law Review* 63 (3): 845–868.

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *Unscrutable Morality: Could Anyone Know Every Moral Truth?*

MARCUS WILLIAM HUNT*
*Tulane University, New Orleans, USA*

*To begin to answer the question of whether every moral truth could be known by any one individual, this paper examines David Chalmers' views on the scrutability of moral truths in* Constructing the World. *Chalmers deals with the question of the scrutability of moral truths ecumenically, claiming that moral truths are scrutable on all plausible metaethical views. I raise two objections to Chalmers' approach. The first objection is that he conflates the claim that moral truths are scrutable from PQTI with the claim that moral truths are scrutable from non-moral truths. The upshot of this objection is that Chalmers has not in fact shown the scrutability of moral truths from the scrutability base from which he proposed to do so, PQTI. The second objection concerns his handling of moral sensibility theory, which fails to take into account certain features of the emotions—features which generate what I term synchronic and diachronic emotional co-instantiation problems. The upshot of this objection is that we have good reason to deny that any one individual could ascertain all moral truths, if moral sensibility theory is true, no matter how idealized the emoter.*

**Keywords:** David Chalmers, moral sensibility theory, moral epistemology, moral psychology, philosophy of emotion.

## 1. *Introduction*

It seems many philosophers would agree that it is important to know moral truths. Some might think that knowing moral truths is important for the purposes of correct moral behavior and moral evaluation of oneself and others, but others might allow that knowing moral truths is important as an end-in-itself; moral truths are the sorts of things that it is just good to know. This raises the question of whether there

* For his comments on an earlier version of this paper thanks to Kevin Morris, and thanks to the reviewers.

are any constraints on the moral truths any one individual could know. As the limiting-case, could any one individual could know every moral truth? In this paper I begin an investigation of this question by examining one influential epistemological proposal that has the upshot that all moral truths could be known; David Chalmers' discussion of "scrutability" in *Constructing the World*. I first outline Chalmers' view that, on all plausible metaethical theories, moral truths are scrutable from PQTI. I then offer two objections. First, Chalmers conflates the claim that moral truths are scrutable from PQTI with the claim that moral truths are scrutable from non-moral truths. Second, Chalmers' discussion of moral sensibility theory fails to take into account certain features of the emotions which prevent any given person from ascertaining all moral truths—emotions generate what I term synchronic and diachronic emotional co-instantiation problems. I conclude that, on at least one influential metaethical theory, it is not possible for someone to know every moral truth.

## 2. *Chalmers on Scrutability and Moral Truths*

According to Chalmers, given certain truths a hypothetical ideal reasoner would be able to know certain other truths. The former truths would be a "scrutability base," and the latter truths would be "scrutable" from the former. Chalmers has in mind an epistemological analogue to the idea of supervenience; just as facts of type x determine facts of type y, so too given truths of type x a hypothetical ideal reasoner could know truths of type y. Chalmers suggests that all truths are scrutable from four classes of truths; physical truths (P), phenomenal truths (Q—qualia), indexical truths (I), and a "that's all" sentence (T) (Chalmers 2012: 22).

To help illustrate the idea of scrutability, Chalmers uses two imaginative devices. One is the familiar idea of the Laplacean demon who, given a scrutability base consisting of all the truths about the present state of the universe, is able to scrute all the truths about the past and future states of the universe (Chalmers 2012: xiii–xv). The second is the idea of a Cosmoscope, a device that contains and displays to its user all the truths contained in the scrutability base; "information about the distribution of matter…[and]… a virtual reality device to produce direct knowledge of any phenomenal states" (Chalmers 2012: 114). Compared with the Laplacean demon, the Cosmoscope "simply offloads some of the work [of calculation, of imagination] from ourselves onto the world" (Chalmers 2012: 116). Such a device "will deliver a sort of supermovie of the world" (Chalmers 2012: 118). So, according to Chalmers, someone using a Cosmoscope that displayed to them all the PQTI truths, would in principle—given sufficient time—be able to ascertain all truths.

Not wishing to have to take on the burden of arguing for a specific ontological view about moral truths but wishing to argue that moral truths are scrutable from PQTI, Chalmers notes the most influential types of views of the ontological status of moral truths. He then briefly examines whether, on each of these types of views, moral truths are plausibly scrutable from PQTI. He notes five types of view;

1. Types of anti-realism.
2. Types of moral relativism, in which "moral sentences are adjudged true insofar as they are true according to an appropriate standard (that of a speaker, or an assessor)" (Chalmers 2012: 265).
3. Types of moral realism based on *a posteriori* identities between non-moral and moral expressions.
4. Types of moral sensibility theories on which one "must have a certain sensibility (certain emotional responses, say) in order to appreciate moral truths" (Chalmers 2012: 265).
5. Types of moral realism in which "moral truths that are not knowable even on full knowledge of nonmoral truths and ideal reflection" (Chalmers 2012: 266).

On (1), there are no moral truths to be scruted. On (2), moral truths are scrutable from social truths about the standards of the speaker. On (3), moral truths will be scrutable insofar as they are identifiable with certain non-moral truths, and insofar as we have access to the non-moral truth that these non-moral truths regulate our positive moral responses (Chalmers 2012: 265). Chalmers suggests in relation to (4) that, if it is the right metaethical view, then the scruting process "may have less of a rationalist upshot than one might have supposed… ideal reasoning will require the right sensibility, involving components that one might take to be emotional as well as traditionally rational" (Chalmers 2012: 266). He claims that (5) is independently implausible because he supposes it to involve a problem of the knowability of moral truths, when "the best reason for being a moral realist stems precisely from our apparent knowledge of moral truths" (Chalmers 2012: 266). With this, Chalmers takes himself to have shown how moral truths, on the gamut of metaethical views, are plausibly scrutable from PQTI or, as with (5), why such metaethical views are implausible.

## 3. *Objection 1—The Conflation of PQTI with Non-moral Truths*

Chalmers conflates the hypothesis that moral truths are scrutable from PQTI with the hypothesis that moral truths are scrutable from non-moral truths. He begins his discussion by noting that "One could ask the question: are moral truths scrutable from *PQI*? But it is easier to ask the more general question: are moral truths scrutable from non-moral truths?" (Chalmers 2012: 264). If non-moral truth were synonymous with PQTI this would be unproblematic. But non-moral truth is a broader scrutability base than PQTI and includes the truths of the numerous other "hard cases" that Chalmers discusses; mathematical truths, other normative and evaluative truths (epistemic and aesthetic), ontological truths, modal truths, intentional truths, social truths, deferential terms, names, metalinguistic truths, indexicals and demonstratives, vague truths, truths about secondary qualities, and counterfactual truths.

This conflation is especially problematic because Chalmers seems to make similar conflations with regards to some of his other hard cases. For instance, he suggests that "deferential truths are scrutable from nondeferential truths" (Chalmers 2012: 281) and that "metalinguistic truths will be scrutable from nonmetalinguistic truths" (Chalmers 2012: 285) rather than "x truths will be scrutable from PQTI." In each case, it is questionable whether Chalmers is dialectically entitled to make use of all the "non-x truths" in scruting the "x truths," and it might be that the scrutability of each hard case from PQTI rests on an assumption of the success of the scrutability of the other hard cases from PQTI, a disturbing circularity.

With regards to some of these other hard cases—such as names—whether they can or cannot be scruted from PQTI—rather than "non-name truths"—is unlikely to have any bearing on the question of the scrutability of moral truths from PQTI rather than from non-moral truths. However, with regards to some of the other hard cases, proving their scrutability from PQTI rather than from "non-x truths" does seem necessary to Chalmers' discussions of the metaethical views on offer. To wit; it seems that the inscrutability of intentional truths from PQTI alone would impact upon the scrutability of moral truths on (2), on most forms of (3), and on (4), since these metaethical views typically suppose that one must have intentional states of various sorts in order to access moral truths. The inscrutability of social truths from PQTI would impact upon the scrutability of moral truths on (2) and on some forms of (3), e.g. social truths about one's role that ground associative or contractive moral duties. The inscrutability of normative epistemic and modal truths from PQTI would have sundry impacts. Therefore, proceeding immediately to the scruting of moral truths from PQTI alone, only on (1) do they remain obviously scrutable, an unremarkable conclusion.

## 4. *Objection 2—Sensibility and Scrutability*

Views falling under (4) refuse "the concession that values are not genuine aspects of reality" (McDowell 1998: 143), and assert that moral claims are truth apt, but that they are not knowable by the more usual epistemological means. Rather, they are learned from emotional, affective, and sentimental reactions. For simplicity of expression I will use the term "emotions" in subsequent discussion—this term should be understood loosely, as including a variety of affective states, sentiments, moods, feelings, and so forth. The person with refined moral emotions has "a reliable sensitivity to a certain sort of requirement which situations impose on behaviour" (McDowell 1979: 331–332), and through their emotional reactions to various life-situations comes to learn moral truths. Views falling under (4) are by no means rare or niche; flowering in the hands of British Enlightenment figures such as Adam Smith, David Hume, and the Earl of Shaftesbury, they remain important contemporary metaethical theories.

On (4), imperfect reasoners and imperfect emoters like us can often learn some moral truths by having the right sorts of emotional states in the various situations that we find ourselves in. The question for Chalmers' project is whether an ideal emoter—with "a big heart as well as a big brain" (Chalmers 2012: 266)—would be able to scrute all moral truths in this manner if they were to use the Cosmoscope to insert themselves into enough of the circumstances and perspectives in which the various the salient emotional states arise. I suggest that the answer to this question is "no." I specify and give criticisms of the scrutability of moral truths on two differing understandings of the Cosmoscope, before proceeding to some more general criticisms.

I explore two alternatives about what using a Cosmoscope would be like for the ideal emoter, since Chalmers' account is ambiguous. Let's call the first possibility the *engrossed* option. On this option, the Cosmoscope "will enable us to know just what it is like to be that subject" (Chalmers 2012: 275). That is, when using the Cosmoscope the user experiences a phenomenologically perfect reconstruction of someone else's experience. The user is immersed in all and only the physical truths, phenomenal truths, and indexical truths that the original experiencer had access to. For instance, if the user explored "what it was like to be Agamemnon in battle" they would have access to all and only the phenomena, and so on, that Agamemnon had. For this to happen it seems necessary that in the engrossed option the ordinary aspects of one's own self such as one's own memories, desires, beliefs, proprioception, etc., would be completely unavailable, on pain of one's experience being phenomenologically unfaithful to Agamemnon's.

In the alternative *unengrossed* option, the user retains their ordinary sense of self even as they experience all the phenomena, and so on, of Agamemnon in battle. The unengrossed option is much easier to imagine. In the unengrossed case the user of the Cosmoscope has an experience of this event that is not quite faithful to Agamemnon's experience, because they remain aware that they are in fact not in ancient Greece on a battlefield, but sat in a basement somewhere; there remains a doubleness of perspective that is absent in the engrossed option. In the unengrossed option the user is able to react to the experiences given by the Cosmoscope from their own ordinary perspective. For instance, such a user would be able to think "How curious to see ancient Greece, and through the eyes of a Greek hero, whilst sat in this basement!," which would not have been a thought available to one whose only perspective on the world was one patterned on Agamemnon's, as in the engrossed option.

## 5. *Criticism of the Engrossed Cosmoscope*

It seems that there is a tension between the faculties of the ideal emoter and the engrossed-Cosmoscopic experience. Suppose that the ideal emoter uses the engrossed-Cosmoscope to experience "what it was like to be Vlad the Impaler." Presumably an ideal emoter would look upon the impaled with emotions of horror, revulsion, disgust, and sadness. But to experience these emotions whilst using the Cosmoscope would mean that the ideal emoter's experience of "what it was like to be Vlad the Impaler" would be very unfaithful to Vlad's own experiences. This would mean that certain moral truths would be unavailable to them. For instance, let's suppose that the real Vlad the Impaler grew less bloodthirsty for a time as he came under the sway of a pacifist preacher—he became troubled, then rueful, then repentant, of his old ways. However, when he campaigns, he finds that he lapses back into violence with an especially wild abandon, and when he returns home becomes repentant again. Presumably, the ideal emoter's responses to the Cosmoscopic experiences would differ very drastically from the warps and wefts of Vlad's very unideal emotional life, but in doing so would necessarily miss out on much—in having nothing to repent of, in not knowing what it is like to take joy in moral violation, what it is like to repent of taking joy in moral violation, what it is like to be morally unstable, etc. On the other hand, if the emotional responses of the ideal emoter are absent, and only the emotional responses of various historical and future figures to their circumstances are present, it is unclear that all moral truths (the relevant standard for Chalmers' project) would ever be revealed by use of the engrossed-Cosmoscope, since the user would only have access to a huge array of the unideal emotional responses of actual historical and future persons, lacking access to whatever the ideal emotional responses to their various situations were.

## 6. *Criticism of the Unengrossed Cosmoscope*

In discussing the Cosmoscope, Chalmers says that there are certain emotional states, for example of "anger or of stupor," the entering of which "may undermine the capacity for reasoning" such that it is best to "think of the Cosmoscope as inducing imaginative states… without actually having those experiences" (Chalmers 2012: 116). On such a supposition, it seems that the user of the Cosmoscope would retain their own perspective whilst entertaining these imaginative states. It seems, *prima facie*, then, that the unengrossed option avoids my objections to the engrossed option.

However, on the unengrossed option it seems that the user of the Comoscope, even an ideal emoter, would not ever experience all the emotional states requisite for scruting all moral truths. Due to the presence of their own perspective assessing the deliverances of the Comoscope, the emotions the user experienced would not simply be copies of the emotions experienced by the historical (and future) figures over whose shoulders they were peering. For example, take a case of righteous indignation which reveals certain moral truths. It seems that for the user of the unengrossed-Cosmoscope vivid imaginings of the emotional states of the righteously-indignant would involve very different emotional states than those occurent in the righteously-indignant themselves. The unengrossed user might admire the righteously-indignant person or be inspired with a feeling of elevation by their example. These emotions may indeed reveal moral truths, but not necessarily the very same moral truths as the righteous-indignation. The righteously indignant person was presumably not self-admiring, nor inspired by their own example. Imagining the emotional states of others and experiencing one's own emotional reaction in response to these is clearly different than having the very emotional states to which one is reacting. There is therefore no guarantee that an ideal emoter would eventually be able to scrute all moral truths through the use of the unengrossed-Cosmoscope, gaining instead an enormous repository of their own emotional reactions to the emotional states of others.

A similar problem is that plausibly the unengrossed-Cosmoscope user would lack access to a host of moral emotions relating to the felt exercise of agency. For instance, for Agamemnon the faculty of acting and emoting were likely richly interwoven; the swinging of the blade is colored and aided by his righteous anger, which in turn is modified and strengthened or satisfied by carrying out this action—the doing is essential to the distinctive sort of emoting. Whilst an engrossed-Cosmoscope user could experience the illusion of this exercise of agency and the sentiments that arise in relation to it, an unengrossed-Cosmoscope user could not; at best being able to imagine what these states are like, as with any cinema-goer.

## 7. *Moral Emotions and Co-instantiation Problems*

I turn now to outlining some features of the emotions that, as well as constituting problems for Chalmers' account, seem more broadly to pose impediments to any articulation of the claim that, on (4), any one individual could know every moral truth.

The nature of the emotions makes the idea of one person being able to know all moral truths impossible. To see this, compare the emotions with more squarely cognitive states such as belief. From our own experience we know that it is possible to entertain numerous beliefs at once. From our own experience we know that it is possible to have certain different emotional states at once; it seems we can feel both angry and sad at once. But it seems that certain emotional states are incompatible and cannot be experienced simultaneously. For instance, it seems that one cannot feel both jovial and reverent, or feel both malicious and compassionate, simultaneously, even with regards to different objects. Moreover, even of emotional states that seem compatible when only two or three of them are experienced simultaneously, it seems like there is an upper limit on the number of emotions we can experience simultaneously. For instance, though I can feel sad as well as feeling angry, or hopeless, or bitter, or detached, or self-pitying, or resigned, or bored, or afraid, it seems hard to imagine that someone could feel all of these simultaneously, even if nothing about any one of these emotions seems to exclude experiencing any one of the others. From this, it seems impossible that one could experience every emotional state simultaneously, or, more weakly, that one could experience every emotional state salient to scruting moral truths simultaneously. Let's call this the *synchronic emotional co-instantiation problem*.

One question is whether the synchronic emotional co-instantiation problem is something contingent to human or non-ideal emoters, or something in the nature of emotional states themselves. It is hard to adduce a case in which something about "belief x" seems of its nature to exclude the simultaneous entertainment of "belief y." It seems in the case of beliefs that our inability to have millions of occurrent beliefs all at once is merely a contingent fact about us. The impossibility of my having occurrent beliefs about "the ontological argument," "the factors affecting the price of fish," and "every capital of Africa" seems to be a limitation about me, rather than something to do with the nature of these beliefs or the nature of beliefs or concepts in general; "there are no concepts whose possession is mutually incompatible" (Chalmers 2012: 114). So, there doesn't seem to be a *prima facie* bar here to the idea of an ideal reasoner able to apprehend all beliefs, and so truths, at once. In the case of the emotions it seems that there is something about the phenomenological experience of certain emotional states that necessarily excludes the simultaneous experience of other emotional states, and about the nature of emotions which puts a cap on the number that can be experienced simultaneously. It is difficult to prove that this is not merely a contingent limitation of ours, but I offer three speculative indications for thinking that it is not.

A first indication is given by the way in which, in ordinary experience, most emotions seem to modify one another. For example, rather than saying that one has an emotion of joy and also an emotion of amazement that stand separately in the same phenomenological field, it seems more accurate to say that one experiences a "joyful amazement" or an "amazed joy"—a single sort of compound or alloy-emotion. Our emotions themselves tend to blend with one another or contaminate one another. Were the emotions of an ideal emoter not to do this, we might plausibly say that the ideal emoter simply did not have the same sorts of emotions as us, and so perhaps was missing out on whatever class of moral truths our own alloy-emotions illuminate.

A second indication is that there is something confused about the idea of experiencing some emotions simultaneously on various theories of the purported constituent features of the emotions. According to various theories, emotions necessarily or paradigmatically involve certain patterns of attention or interpretations of experiences (Roberts 2003), certain judgements (Nussbaum 2004), or action tendencies (Frijda 2008), or bodily feelings (James 1884). Take fear and "feeling safe." It seems that one cannot both attend to and not attend to potential dangers, or endorse the constitutive judgements of fear and the feeling of safety at once ("The fearful is (not) present"). Likewise, the action tendency of fear and the feeling of safety seem opposed; being inclined to run away and to stay put. Likewise, the bodily feelings of the two are very different. For the constitutive features of one to be present means for the constitutive features of the other to be absent.

A third indication is given by the way in which we talk about emotions as against more squarely cognitive states such as belief. John believes that Copernicus was Polish. Most of the time, this is a dispositional belief for John, very rarely becoming an occurrent belief. We nevertheless say of John at any given time that he believes that Copernicus was Polish; it is a belief that we may say John has at any time of day. John also has the capacity to experience many different emotions. We talk about the frequency with which John experiences these various emotions in terms of his dispositions, his traits, or his character; if he often gets angry he is an angry man. Nevertheless, it seems that we do not speak of these dispositions as being the emotional states themselves. The disposition explains or summarizes John's frequent bouts of anger but is itself is not the anger. Whereas numerous items of dispositional belief can be said to co-exist in the same mind at once even when they are not called to attention, there seem to be no analogous "dispositional emotions" of which the same can be said. It would be bizarre to say, for instance, that John has dispositional envy even though he hasn't had an episode of envy for two years, whereas it makes sense to say that John has a dispositional belief that Copernicus was Polish even though this belief hasn't become occurrent for two years.

To use a metaphor, beliefs are like tabs in a minimized computer window—they are there, even if you haven't checked them in a while. Emotions are like the wavy patterns displayed by a 1990s screensaver; one is now displayed, the patterns that were displayed no longer exist, and the patterns that will be displayed do not exist yet, even though we can say which are likely to be displayed next, or which patterns are often displayed. If this observation is correct then its upshot is to reinforce the first two indications by showing that there is no way to sneak in "dispositional emotions" that are all somehow "had" by an emoter yet lie dormant and do not modify one another or prompt contradictory action tendencies, bodily feelings, etc.

One response to the synchronic emotional co-instantiation problem is to note that even if an ideal emoter could not experience every emotion at once, they are at any rate surely able to experience every emotion sequentially. Once a sufficiently long series of the appropriate emotions have been experienced, these emotional reactions will have scruted every moral truth. Against this, I suggest instead three reasons for thinking that there is also a *diachronic emotional co-instantiation problem*—meaning that it seems impossible that one could experience every emotional state sequentially, or, more weakly, that one could experience every emotional state salient to scruting moral truths sequentially.

First, some emotional episodes alter our characters, tending to give rise to recurring emotional episodes, or staining a broad spectrum of future emotional episodes in very specific ways. That some emotional episodes have this effect is an important aspect of the moral truths they seem to reveal (or hide). As well as making the idea of learning moral truths by sequentially undergoing differing emotional experiences overly simplistic, this fact plausibly excludes the learning of every moral truth by any given individual. The person whose character develops in one particular direction may not be able to access the moral truths learned by someone whose character develops in a very different direction.

Second, it would be a mistake to think that every given moral truth can be revealed by a single episode of an emotion. Rather, certain moral truths may only be learned through repeated emotional episodes and through complex patterns of emotional episodes.

Third, the memory of an emotional episode is not the same as experiencing the emotional episode itself. Whilst, plausibly, most moral truths are such that they can be revealed by an emotional reaction and then recorded by the memory, plausibly not all moral truths are like this. Some moral truths may be revealed in an emotional reaction but not admit of being added to a permanent stock of belief or knowledge, being instead temporary and situational. For instance, it is plausibly only during the-moment-at-which-you-think-you-will-now-die or only during the religious experience or only during the DMT-trip that you could be said to fully grasp the moral truths that the emotional aspects of these experiences conveyed.

To illustrate these points three I give two literary examples. First, the character des Esseintes from Joris-Karl Huysmans' *Against Nature* explaining his motivations for bringing a young boy to a brothel:

> … I'm simply trying to make a murderer of the boy. See if you can follow my line of argument. The lad's a virgin and he's reached the age where the blood starts coming to the boil. He could, of course, just run after the little girls of his neighbourhood, stay decent and still have his bit of fun, enjoy his little share of the tedious happiness open to the poor. But by bringing him here, by plunging him into luxury such as he's never known and will never forget, and by giving him the same treat every fortnight, I hope to get him into the habit of these pleasures which he can't afford. Assuming that it will take three months for them to become absolutely indispensable to him—and by spacing them out as I do, I avoid the risk of jading his appetite—well, at the end of those three months, I stop the little allowance. I'm going to pay

> you in advance for being nice to the boy. And to get the money to pay for his visits here, he'll turn burglar, he'll do anything if it helps him on to one of your divans in one of your gaslit rooms... I shall have contributed, to the best of my ability, to the making of a scoundrel, one enemy the more for the hideous society which is bleeding us white. (Huysmans 2003: 66)

And, from Friedrich Hölderlin's *Hyperion*:

> I know as well as you do that I could still trump up some kind of existence for myself, could, now that life's meal is eaten, still sit playing with the crumbs; but that is not for me, nor for you. Need I say more? (Hölderlin 1990: 116)

It seems that the emotional reactions of des Esseintes' boy-victim would cascade throughout his life in complex ways. Plausibly, he would learn moral truths unavailable to someone who enjoyed a life of voluntary celibacy, and likewise the moral truths learned by the celibate would be unavailable to this boy. Similarly, it seems that the sentiments of Hölderlin's character are provoked by an entire life of varied experiences, will color the remainder of their life, and could not be viscerally experienced by, say, a young person who faces a terminal illness and is eager to eat as much of "life's meal" as they can. Importantly, note that simply reading about these characters, or receiving testimony of the experiences of real people, although no doubt crucial for learning many moral truths, are not completely adequate substitutes for having these experiences; although readers may be able to imagine something of what it would be like to be des Esseintes' boy-victim, there is surely much we cannot begin to grasp.

Lastly, some hold that different sorts of emotions are appropriate for those of different ages, professions, genders, without any given constellation of emotional responses being better or worse or more complete (Grimshaw 1993, Smith 2002). The concept of an ideal emoter is therefore indeterminate, admitting of multiple differing instantiations, in a way that the concept of an ideal reasoner is presumably not. For instance, Seneca writes:

> It is a matter of great prudence, for the benefactor to suit the benefit to the condition of the receiver: who must be either his superior, his inferior, or his equal; and that which would be the highest obligation imaginable to the one, would perhaps be as great a mockery and affront to the other (Seneca 1882: 44).

One and the same emoter cannot be the superior, the inferior, and the equal, of the gift giver, and so cannot experience the emotions appropriate to each station in life.

## 8. *Conclusion*

It seems that not even an ideal emoter with a Cosmoscope would be in a position to scrute all moral truths if (4) is true, due to the synchronic and diachronic emotional co-instantiation problems. Whereas there seems nothing in-principle impossible about the idea of a single person having access to all the other sorts of truths, moral truths are a special case. If an ideal emoter, with Cosmoscopic access to all the PQTI truths—or, to widen the scrutability-base beyond Chalmers' claim, all non-moral truths—could not scrute all moral truths, then *a fortiori* non-ideal emoters without Cosmoscopic access to these truths could not scrute all moral truths.

For his own purposes, Chalmers might not be too bothered by this conclusion, since despite the co-instantiation problems, it might still be the case that although no individual could scrute every moral truth on (4), the human community might be able to scrute every moral truth, these truths being known in a disaggregated way by millions of individuals, each having a sliver of these truths—just as presumably no individual will ever know everything there is to know about chemistry and anthropology, but the human community or an ideal reasoner might. For our purposes, however, it is a significant result that on (4) not every moral truth can be known by a given individual, even an ideal emoter. It is presumably not at all troubling that each of us is not able to know every truth about chemistry. A dispersion of knowledge and a division of intellectual labour is both inevitable and useful. There are lots of crucial chemical and anthropological truths that it is probably important only that a few people know, and there are lots of trivial chemical and anthropological truths that it is probably not important that anyone know. However, it is potentially troubling that each of us is not able to know every moral truth because we tend to think that knowing about moral truths is each individual's business, whatever else they wish to know or do, and that no moral truth is too trivial that it is not worth knowing since it is very important that we always behave permissibly and that we properly evaluate the behavior of ourselves and others.

The finding that, on (4), any one individual cannot know every moral truth, should give us pause to consider what the proper ends of moral inquiry and moral learning may be; perhaps we should come to think that only knowing certain moral truths is our business, or that some moral truths are too trivial to be worth knowing. We should also pause to consider how this conclusion bears on our moral evaluations of others; if ignorance of a moral truth can be a moral excuse, and if we are all necessary ignorant of some moral truths, then we all enjoy slightly differing sets of moral excuses, excuses which in turn it may be very difficult for one another to know about.

I note that my conclusions here are limited in their extent in that further work is necessary to find out whether any individual can know every moral truth on other metaethical theories.

## *References*

Chalmers, D. 2012. *Constructing the World*. Oxford: Oxford University Press.

Frijda, N. 2008. “The Psychologists’ Point of View.” In L. Feldman Barrett, J. Haviland-Jones, and M. Lewis (eds.). *Handbook of Emotions*. New York: Guilford Press.

Grimshaw, J. 1993. “The Idea of a Female Ethic.” *Philosophy East and West* 42 (2): 221–238.

Hölderlin, F. 1990. *Hyperion and Selected Poems*. New York: Continuum.

Huysmans, J.-K. 2003. *Against Nature*. London: Penguin Books.

James, W. 1884. “What is an Emotion?” *Mind* 9: 188–205.

McDowell, J. 1979. “Virtue and Reason.” *The Monist* 62 (3): 331–350.

McDowell, J., 1998. *Mind, Value, and Reality*. Cambridge: Harvard University Press.

Nussbaum, M. 2004. “Emotions as Judgements of Value and Importance.” In R. C. Solomon (ed.). *Thinking About Feeling*. Oxford: Oxford University Press.

Roberts, R.C. 2003. *Emotions: An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.

Seneca, 1882. “Of Benefits.” In R. L’Estrange, tran. *Seneca’s Morals of A Happy Life, Benefits, Anger, and Clemency*. Chicago: Belford, Clarke and Co.

Smith, A., 2002. *The Theory of Moral Sentiments*. Cambridge: Cambridge University Press.

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *The Non-Identity Problem and the Admissibility of Outlandish Thought Experiments in Applied Philosophy*

ADRIAN WALSH
*University of New England, Armidale, Australia*
*University of Gothenburg, Gothenburg, Sweden*

*The non-identity problem, which is much discussed in bioethics, metaphysics and environmental ethics, is usually examined by philosophers because of the difficulties it raises for our understanding of possible harms done to present human agents. In this article, instead of attempting to solve the non-identical problem, I explore an entirely different feature of the problem, namely the implications it has for the admissibility of outlandish or bizarre thought experiments. I argue that in order to sustain the claim that later born selves cannot be harmed (since they are in fact different persons), one must rule inadmissible certain kinds of modally bizarre imaginary cases. In this paper I explore how one might justify such a constraint on outlandish cases and, in so doing, develop the outline of a model for distinguishing between admissible and inadmissible imaginary cases in philosophical debate.*

**Keywords:** Thought experiments, non-identity, philosophical methodology, moral luck.

## 1. *Introduction*

The non-identity problem directs our attention to the obligations we might have towards people: (i) whose existence we cause in some relevant sense; (ii) whose circumstances, while tolerable, are less than ideal but (iii) who would not exist if those less than ideal circumstances were improved. In such cases determining whether or not a harm has been done to a particular person must be constrained by considerations of whether or not less harmful alternative courses of action would have led to that person existing. Parfit's description of the problem in *Reasons and Persons* begins with the assertion that the issue of which future people will exist is dependent (at least in part) on when exactly the procreation takes place (Parfit 1984: 358). Parfit claims that in assessing the *relative harm* of the circumstances of any birth, one cannot compare them with other scenarios involving alternative policies and actions in which one would not have been born at all.

Much of the subsequent discussion has focussed on either the question of what the *proper object* of moral concern might be or the implications it might have for our obligations to future generations. Should it be states of affairs or individual persons? However, there is another significant feature of this line of reasoning that such discussions ignore; namely regarding the set of admissible scenarios for moral assessment. The debate tacitly assumes that *modal harms*—that is, counterfactual harms that are either nomically impossible or practically infeasible in the circumstances—are not relevant to the assessment of individual welfare. The refusal to countenance the outlandish or morally bizarre, I shall refer to as the 'nomic constraint'.[1] While Parfit directs our attention away from worries about individual welfare, in this paper the critical focus will be on one of the methodological assumptions that drive arguments for shifting away from individual welfare.

The distinctive feature of the discussion herein is that it treats the non-identity problem as requiring, amongst other things, constraints upon the relevant thought experiments and imaginary scenarios one might employ. The outlandish would appear to be ruled out. I suggest that reconsidering the non-identity problem in terms of the limits it places on what counterfactuals are relevant, sheds fresh light on our understanding of the proper role of thought experiments in our moral reasoning and, in particular, the role of outlandish or bizarre thought experiments play in our reasoning more generally.[2] It is not, as some would have it, that there is a level or degree of outlandishness beyond which we should venture, but rather that the admissibility of the outlandish depends on the argumentative context—or so I will argue.

## 2. *The standard analysis of non-identity and its relevance to applied ethics*

The non-identity problem consists in the claim that determining whether or not a harm has been done to a particular person must be constrained by considerations of whether less harmful alternative courses of action would have led to that person existing. Parfit's description of the problem in *Reasons and Persons* begins with the claim

[1] Jakob Elster explores the admissibility of the outlandish in his 2011 article. See also Wilkes 1988 and Pogge 1990.

[2] Herein I argue for thought experiments having a variety of roles in philosophical reasoning. However, writers in this area typically pick out only one such role; so, for instance, Elster (2011) treats them as primarily means of generating intuitions for testing our moral principles.

that whether a particular person exists is dependent (at least in part) on when exactly the procreation takes place.

> *The 14-Year-Old Girl*: This girl chooses to have a child. Because she is so young, she gives this child a bad start in life. Though this will have bad effects throughout the child's life, his life will, predictably, be worth living. If this girl had waited for several years, she would have had a different child, to whom she would have given a better start in life. (Parfit 1984: 358).

However, Parfit asserts that in assessing the relative harm of the circumstances of any birth, one cannot *compare them with other scenarios* involving alternative policies and actions in which one would not have been born at all. Thus the child cannot be said to be harmed by his mother's action of conceiving him at fourteen, for the very reason that he would not have been born at all if she had conceived later in life.

Notice that what we are concerned with here is one type of *counterfactual or modal harm*. It is not what the girl did to her child directly that is under scrutiny but what she didn't (namely not having him when she was eighteen). The focus here is on what an agent could have done which was less harmful than what he or she did in fact do. We might call this "could-have-done-better" counterfactual harm. We can contrast this with cases where we might hold someone responsible for some event that they were lucky did not occur, even though it might well have. For instance, if I run through a series of red lights without stopping, my action would be held to be morally blameworthy even when no harm comes to me or anyone else. This we might call 'there-but-for-the-grace-of-god' harm. In this paper it is the former type of harm and consequent blame with which we are concerned.[3]

Parfit raises the non-identity problem to argue for what he calls *impersonal* as opposed to *person-affecting* moral frameworks. He claims that we should reject the view that an outcome can only be worse if it is worse *for someone* and, further, that we act wrongly only by making a *particular* existing person worse off. Instead we have an obligation to do what would maximise overall happiness rather than the happiness of particular people.

In recent years the problem has taken centre stage in debates in applied ethics concerning the use of new reproductive technologies, whether in some cases one would be better off not being born and environmental debates about our obligations to future generations.[4] In the area of bioethics, it is commonly invoked in discussions concerning the harms, benefits and duties associated with the manipulation and alteration of the genetic make-up of future individuals. If we transform the genetic make-up of a future being in such a way that it is no longer the same individual, do we harm the child who would have been born had

[3] Whilst this might seem a little odd, it is quite common to define harm in term of counterfactuals. See for instance, Feinberg 1986.

[4] See, for instance (as just a glimpse into this vast literature) Roberts and Wasserman 2009 and Archard and Benatar 2009.

we not interfered? And who do we benefit by such interference? There is also a debate in the bioethical literature about wrongful life in which the non-identity problem has had a significant role. (Archard 2004) How can an individual be said to be harmed by being brought into existence when the contrast is with a state of affairs in which they do not exist and hence one cannot compare their relative state of well being?

In the area of environmental ethics, writers explore the implications of choosing conservation or degradation of the environment (Carter 2001). Suppose we have a choice between these two outcomes. Whichever choice we make has consequences, not only for the quality of life of those who inhabit the future, but the very identity of those future citizens. Suppose we choose degradation. Clearly the lives of those who inhabit a degraded environment are worse than those who inhabit a non-degraded one. But if they would not have existed in the non-degraded environment, and their lives are still sufficiently good that they are better off being born than not being born at all, then can we say that they are some how harmed by our choosing the degraded option?

## 3. *Non-identity and thought experiments*

The bulk of the specifically philosophical discussion of the non-identity problem has focused on whether one should indeed become an 'impersonalist' or whether there might be a rights-based solution to the problem. However, these competing responses to the problem all appear to accept implicitly a thesis upon which I wish to place some pressure: namely that nomic claims about identity should constrain the kinds of imaginary scenarios which we might employ to judge the relative welfare of a person.[5]

To see this let us reconsider our initial scenario regarding the fourteen-year-old mother and compare it with the following case.

> *Kangaroo-like Gestation* Imagine that human beings had similar reproductive capacities to those of kangaroos. Kangaroos have the ability to delay gestation of fertilised embryos, during lean periods, until such time as conditions are suitable for the bearing of healthy offspring. If human beings were like kangaroos then it would be possible for the fourteen-year-old mother to delay the birth of her child until she was eighteen, thus ensuring a better life for her child.

In this case it would be the identical child that is born four years later.

[5] In the entry on the Non-identity problem in the *Stanford Encylopedia of Philosophy*, M.A. Roberts notes in passing that in these cases the range of courses of action must be consistent with existing medical and genetic technologies. Later Roberts notes that since Parfit first described the case reproductive technologies have advanced in a way that makes it, at least, theoretically possible for the 14 year-old girl to tbe same child as the later born child. Nonetheless Roberts claims that probabilities need to be taken into account here and, even with new technology, it is *very probable* that the two children would be non-identical.

Accordingly, in this scenario if she chose to have the child at 14 when she could have had the identical child at say 18, we would much more willing to say that she has harmed that particular child.

Far less fancifully, we can readily imagine technology that would allow us to take fertilised embryos from the wombs of recently impregnated women and freeze them in a storage facility where they would be kept until such time as the woman was ready to gestate the future child. So here again our fourteen-year old girl would be able to bear the identical child in circumstances more conducive to his over-all wellbeing.

There are also cases we could consider that are much closer to the kinds of worries about identity we find in the bioethics literature, where the harm involves some genetic disadvantage. To see this consider another scenario:

> *The Bridesmaid's Dress*: a woman knows that if she conceives in the present month, the child conceived will suffer some serious genetically-based ailments, but not so severe that the child would be better off not existing. However, if she delays conception she will not fit into her bridesmaid's dress at wedding scheduled for nine and half months time. The woman decides to conceive and gives birth to a child.

Following the reasoning of the non-identity problem we cannot say that the child is harmed by comparing his welfare with that of a hypothetical child born a month later, since our knowledge of biology leads us to the belief that it would not be the same child. The standard and quite sensible view is that different eggs and different sperm make for a different genetic identity. Moreover, genetic identity is assumed to be a necessary, if not a sufficient, condition of personal identity: that is to say, one cannot be the same person without being genetically identical.[6]

But it is, of course, possible to imagine cases whereby the child who is born a month later would in fact be genetically identical. The woman need only be capable of delaying gestation for the later child to be genetically identical. What these kinds of cases demonstrate is that the non-identity of, for instance, later born children is not a *necessary* truth about the world, for there are possible cases where the child born a month later or four years later, or even a hundred years later would be the same person or at least genetically identical. The child born of the woman at the age of fourteen and the child born four years later are not *necessarily non-identical.*

In order to rule out such counter-examples, advocates of the non-

[6] In saying this we need to acknowledge explicitly the further problem concerning the role of genetic identity in personal identity: merely being genetically identical does not make an agent the identical person. However, for our purposes we can put that question to one side since the thought in the non-identity problem is simply that we do not have genetic identity in later born children and hence no claim of harm can be made in relation to such children.

identity problem must deem inadmissible, when determining relative welfare, any counterfactual states of welfare found in conceivable worlds that defy our current views of what is *possible in the circumstances* and which will usually be covered by the nomically possible. It is not simply *non-identity* that is determining whether or not harm has occurred in our world, but non-identity in circumstances closely resembling ours. Let us call this the nomic constraint on the use of thought experiments in the assessment of harm (NC). We might formulate it as follows:

> *The Nomic Constraint*: In assessing the relative moral status of any action we should not compare it with counterfactuals that are nomically impossible or are impossible in the present circumstances.

Another way of putting this point is that the non-identity problem rules out certain 'modal harms' in the assessment of relative welfare. When we decide whether a person is harmed by our choosing (or not choosing) a particular course of action, their well-being in possible worlds in which, given our current levels of understanding and technology or given our understanding of the laws of nature, they could not exist, are not relevant.

This is not uncontroversial. Philosophers, especially in the area of metaphysics, make regular use of examples that defy the laws of physics, as we know them, whilst others—who are in a minority—reject their employment. It is useful here to distinguish, roughly following Elster (2011: 242), between *conceivabilists* and *realists*.[7] According to the conceivabilist, so long as a case is conceivable then it is legitimate to use it as a means of testing our theories. According to the realist, on the other hand, only cases that could plausibly occur given the world as it is should be used to test our theory. In the area of moral philosophy this view is sometimes defended on the grounds that since moral principles are meant for guiding action in the world, cases drawn from other worlds are irrelevant. Kathleen Wilkes (1988) makes a similar point in relationship to questions of personal identity. Other realists argue that we lack the cognitive capacity to apply our intuitive faculties to outlandish cases (Elster 2011: 242).

Clearly the nomic constraint fits broadly within realism. Curiously, given the context of our discussion, Parfit's general view on the matter would appear to be essentially conceivabilist. In *Reasons and Persons*, in the midst of a discussion of whether Nozick's imagined Utility Monster is deeply impossible (and hence irrelevant to moral debates on public policy), Parfit notes that 'even an impossibility may provide a test for our moral principles'. He claims that "[W]e cannot simply ignore imagined cases." (Parfit 1984: 389). Yet, the non-identity problem

[7] In the main Elster treats imaginary cases as being used to elicit intuitions to test our moral principles, a claim which I will challenge later in the paper. However, at various points he does acknowledge different roles that thought experiments might play.

only generates the problems it does if we rule out the kinds of imaginary cases outlined above.

Whatever one's stance in the debate between realism and conceivabilism, it is somewhat anomalous given the regular use of modally outlandish cases by philosophers, that those exploring the non-identity problem have accepted so readily the idea that genetic identity determines personal identity and hence constrains what might count as a possible form of harm.

Let us summarise the line of reasoning thus far. According to Parfit non-identity means that the child born of the 14 year-old girl cannot complain of some harm being done to him or her. It is not the case that Parfit claims there is no harm; instead it is harm in terms of the overall sum of welfare, rather than harm to any particular single individual. However, the non-identity that underpins this argumentative move is—as the case of the kangaroos demonstrates—*contingent* not *necessary*. In focusing solely on *contingent non-identity* we must rule as inadmissible evidence any counter-examples from far-off possible worlds. This mode of ruling inadmissible I shall refer to as the 'nomic constraint'.

## 4. *The Non-identity problem and modal moral luck*

It would appear then that those debating the implications of the non-identity theory are agreed on one point at least, namely that nomic claims about identity should constrain the kinds of imaginary scenarios that we can legitimately employ to judge the relative welfare of a person. Let us focus a little more closely on this feature of the debate. What we have here is a *nomic constraint* on the imaginary scenarios and thought experiments that will count as relevant cases in the assessment of individual harm. The suggestion is that we cannot employ nomically impossible examples to determine the moral status of action or future policy and hence our use of thought experiments in moral philosophy will be constrained here by the extent to which those thought experiments introduce cases which are nomically possible.

One further interesting consequence of this is that it appears to entail a commitment to the idea of what we might call "modal moral luck", according to which the moral status of an action is contingent upon which particular possible world one inhabits. The term is related to the idea of 'moral luck' that Bernard William famously proposed to cover cases where the moral status of an action or even a whole life depends upon how things turn out. Moral luck describes cases where "an agent can be *correctly* treated as an object of moral judgement", despite the fact that a significant aspect of what that agent is assessed for "depends on factors beyond his control" (Nelkin 2004). In the Kantian tradition matters are entirely different for there it is assumed that we are only judged in terms of the motives with which we act. Our intuitions might well be thought to support the Kantian view since the very idea of moral luck sounds oxymoronic. Further it seems unfair to be judged by circumstances that are outside of our control. Yet at the same time our ordinary moral thinking seems to suggest that there is such a moral phenomenon. As Thomas Nagel notes:

> Where a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgement, it can be called moral luck. (Nagel 1979: 59)

In order to illustrate his point, Williams tells the story of the painter Paul Gauguin. Gauguin left his wife and family in France to pursue his career as a painter. Williams' point is that if it had turned out that Gauguin was a mediocre painter rather than a gifted artist, then we would judge his life rather differently. The general idea then is that we are more hostages to moral fortune than the standard Kantian analysis would have us believe.

Modal moral luck, by way of contrast, involves cases where the moral assessment will differ according to what world, of all those possible, it is in which the action occurs. Thomas Nagel raises a related (although far less extensive) idea when he discusses the category of 'circumstantial luck' (Nagel 1979). Circumstantial luck is luck about the circumstances in which one finds oneself. Nagel's example concerns Nazi collaborators in Germany during the period of the Third Reich. We condemn them for the morally appalling acts they performed, but if they had been shifted to South America in, say, 1929, perhaps they might have led morally exemplary lives. If we provide a different moral evaluation of the lives of the Nazi collaborators with their hypothetical counterparts in South America, then we have a case of what Nagel calls circumstantial moral luck. Modal moral luck extends the idea of relevant hypothetical counterparts to a much wider range of possible worlds, including those that we might think of as being nomically impossible.

According to the idea of modal luck then the moral status of one's action is contingent upon what possible world we inhabit. To illustrate the idea let us return to Parfit's original example of the fourteen-year old girl. Does the fourteen-year old girl harm her son by having him when she is fourteen? Is he harmed by not being born four years later? On this line of reasoning the moral status of her action depends upon what possible world it is that she inhabits. In the world where human beings cannot delay gestation of fertilised embryos then she does not harm him because he would not exist. In the world where human beings can perform this procreative trick, then she does harm him. The moral status of her action is thus contingent upon which possible world she inhabits. It is, we might say, a matter of modal moral luck whether she can be said to have harmed her son.

Modal luck would cover what we might call 'epistemic cases' where our lack of knowledge makes an outcome impossibly remote for us. There will be cases where our ignorance makes it impossible for us to take advantage of some harm-lessening process. Imagine that if we drink the right admixture of iodine and calcium that we can delay gestation for up to ten years. In this case again the fourteen year old could take this concoction and give birth to the same child four years later when she is better placed to raise him. In a world where this technique is widely known we might well judge the action of the woman of not delaying the pregnancy quite differently from we would in a world in which this information was not available. Again the fact that the same action is evaluated quite differently is a matter of modal moral luck.

Modal luck has interesting implications, not merely for debates regarding how we might determine whether harm has occurred and what kinds of counterfactual considerations are relevant, but also for our use of thought experiments in ethics more generally. Think about this in relationship to one of the more notorious thought experiments in the ethical literature, Michael Tooley's case of the superkittens (Tooley 1983). Tooley's target was the potentiality principle that is often used to oppose abortion. In attacking the idea that it is wrong to kill foetuses because they are potential persons, Tooley employs a thought-experiment involving highly rational cats to generate a putative *reductio ad absurdum*. He writes:

> Suppose that at some time in the future a chemical is discovered that, when injected into the brain of a kitten, causes it to develop into a cat possessing a brain of the sort possessed by normal adult human beings. Such cats will be able to think, to use language, to make decisions, to envisage a future for themselves, and so on—since they will have all of the psychological capacities possessed by adult humans. If one maintains that it is seriously wrong to kill adult members of the species *Homo sapiens*, one must also.....hold that it would be seriously wrong to kill cats that have undergone such a process of development. (Tooley 1983: 191)

He then claims that it follows that it is *prima facie* no more seriously wrong to kill a human organism that is a potential person, but not a person, than it is intentionally to refrain from injecting a kitten with the special chemical, and to kill it instead. This he suggests shows why the potentiality principle is wrong for in the example provided above it would be intrinsically wrong to refrain from injecting a kitten and killing it instead. The claim that it is *intrinsically* wrong to kill a foetus is therefore putatively reduced to absurdity.

However, we should be extremely wary of accepting this thought-experiment as a *refuter*. To do so would be to overlook the fact that we are making decisions in this world and the contingent facts about how this world is actually constructed are relevant to how we frame our decisions. In the possible world in which we possessed such a chemical, refraining from injecting a kitten and subsequently killing it *would be morally equivalent* to killing a foetus (and that says nothing about whether or not it would be intrinsically wrong). We might think of this as a form of *modal moral luck*. For the kitten-killer it will be a matter of 'modal moral luck' whether or not he inhabits a possible world in which killing kittens is equivalent to killing foetuses. But in our world they are not morally equivalent along the lines of comparison that Tooley discusses.

We can restate the point in the following way. Imagine two different people in two different worlds, World A and World B, who both accidentally run-over a young kitten. In world A we have the drug to turn cats into supercats whereas in B we do not. In the world of potential supercats the action has more bad-making features than in world B. The actions may well be equivalent in terms of their responsibility, causality and intention. But we might say that killing A is morally worse and, that it is so for the driver concerned, is a matter of modal moral luck.

## 5. *Modal Constraints and the Assessment of Harm*

However, nomic constraints on counterfactuals are decidedly odd in philosophy. How far would we want to generalise this 'nomism'? Would we want to adhere to it as a general principle, for bizarre counterfactuals are commonplace in philosophical debate. Think, for instance, in metaphysics of discussions of the swampman who emerges out of the swamp with all of the same properties as an ordinary human.[8] Equally, in the philosophy of mind there has been a great deal of discussion of zombies who behave as if they are conscious but in fact are not.[9]

If we turn our attention to ethics and applied philosophy again we find considerable use of what one might think of as 'bizarre examples'. Think, for example, of Judith Jarvis Thomson's case of the people-seeds that she raises when discussing the morality of abortion (Thomson 1971). These people-seeds float around in the atmosphere and if one does not place the appropriate guards on one's windows then they will float inside one's house, attach to the carpet and begin growing into people.

The use of bizarre examples, then, is standard-practice in many areas of philosophy in general and many would regard it as intellectually productive. Although there are realist critics of the modally bizarre—most notably Kathleen Wilkes and Thomas Pogge—the demand for constraints in terms of nomic possibility is somewhat anomalous. Realism of this variety would be thought by many as stymieing philosophical analysis and thus would seem to require justification.[10] The default position within philosophy is surely that the modally bizarre are admissible.

A further possible problem with *the nomic constraint* concerns its implications for commonly employed notions in moral philosophy such as *universalizability*. Does it restrict the circumstances in which it is

[8] See, for instance, Davidson 1987 and Millikan 1996.

[9] For useful surveys of this literature see Kirk 2005, Marcus 2004, Cottrell 1999 and Dennett 1995, 1991.

[10] See Wilkes 1988 and Pogge 1990.

possible for me to imagine myself? Universalizability involves (roughly speaking) the idea that in determining the moral status of a proposed action we need to ask ourselves firstly what the world would be like if everyone did it and secondly how would we feel if it were to be performed on us. Any action to which we would not assent in these hypothetical circumstances is morally impermissible. The nomic constraint is unlikely to create difficulties for the first element since universalising some action will rarely require invoking impossible states of affairs. However, problems do arise when we consider the second condition regarding thinking oneself into the shoes of another. If I am supposed to imagine what it would be like to be a member of a very different culture and imagine how I would feel about certain forms of discrimination, then this seems to require that I imagine something impossible, namely that I am a very different person than I am. If non-identity rules out children born four years later, then *a fortiori* it might seem to rule out such radical transformations. But perhaps this is a red herring since the non-identity problem concerns what are possible counterfactual conditions for a particular person, not how I might imagine myself. Be that as it may, the significant point for our purposes is that, in determining relative welfare, the non-identity problem deems inadmissible any harm found in conceivable worlds that defy our current views of the nomically possible. However, if this were to be taken as a more general principle in philosophy, would it 'poison the well' of moral theory, since it would undermine our use of thought experiments?

For realists, however, who wish to eliminate the use of the outlandish these are not costs but desirable outcomes. Yet, the view is controversial. Must anyone who regards the non-identity as raising genuine philosophical issues about the nature of harm thereby commit themselves to realism about thought experiments?

Fortunately, this would appear not to be the case. It need not have such extreme ramifications for philosophical method if we limit the scope of the nomic constraint that underpins it. Reconsider the topic out of which the non-identity problem arose. It concerns harm and, more specifically, *actionable* harm. In this case we are concerned with assigning responsibility—and perhaps blame—for harms done to future people by dint of the timing of their birth or their genetic make-up. In pursuing such questions we are constrained by what could *reasonably be expected* in the current circumstances, not by what is logically or metaphysically possible. Harm, in this context, is a practical notion that does not need to be evaluated in relation to every possible world.

The upshot is that we need to revise our view about what assumptions underpin the non-identity problem. Our original nomic constraint advised that when assessing the relative moral status of any action, we should not compare it with counterfactuals that are nomically impossible or are impossible in the present circumstances. However, in order to generate the non-identity problem this constraint need not be generalised. Relying on the nomic constraint in the context of debates about harm need not commit one to it as a more general principle for moral theory, or indeed for philosophical theory more generally. Cast in a more modest way, then, the constraint can be rewritten as follows:

> *A modal constraint with restricted scope*: In assessing the relative moral status of any action *in terms of actionable harm* we should not compare it with counterfactuals that are nomically impossible or impossible in the present circumstances.

This weaker version of the constraint on imaginary cases, unlike stronger forms of realism, does not have such serious implications for philosophical method since it allows for the continuation of bizarre thought experiments in general, but rules out specific cases where there is actionable harm. It rules out 'kangaroo-style' cases when discussing harm in the current context but does not rule out counterfactual forms of harm, such as in the case of the inebriated driver who fortunately harms nobody, from being relevant when assessing actionable harm. Crudely speaking, drunk drivers are in and kangaroos are out.

There remain, we must admit, vexed questions of how to determine what might be impossible in the current circumstances. There are two closely related parts to this difficulty. The first concerns how we operationalize the constraint. If one is assessing harm—and this is particularly true if that harm were to have any legal ramifications—how can one be sure what really is impossible? There is always the possibility that some states of affairs are in fact possible despite our ignorance of that fact. Secondly, rapid technological changes can mean that what had been difficult even to imagine at one point in history can change overnight. Nonetheless, as troubling as these considerations of the assessment of harm might be, they do not undermine the general point about the legitimacy of ruling some thought experiments inadmissible on such grounds.

## 6. *Modal constraints and the argumentative context*

Thus far I have argued that the modal constraints associated with the non-identity problem are justifiable because we are concerned with actionable harm. At this point I wish to draw a more general lesson about the admissibility of outlandish thought experiments. Some critics, such as Elster, focus primarily, when assessing admissibility, on the modality of particular thought experiments. However, I beg to differ. Outlandishness or bizarre modality is not the central issue in relation to the admissibility of thought experiments, rather it is the argumentative context. It is not the violation of what is *possible alone* that justifies rejecting the claim that the child is harmed by not being born four years later. Instead it is the violation of what is possible *in an argumentative context in which such violations matter* that will justify constraint. 'Modal violation' is but one possible relevant factor when assessing such admissibility but there will be many others. The general idea is that constraints can be justified in relation to the argumentative context.

The approach I sketch briefly below involves what we might call 'thought experimental pragmatics', in which the admissibility of a thought experiments will be determined not by the specific features or inherent modal nature of the thought experiment itself, but by the role that it plays in the argumentative context. We can illustrate this by drawing an analogy with pragmatics in linguistics and the philosophy of language (Lycan 1995). There the term 'pragmatics' refers to approaches that explore the contextual dimensions of our use of language and the context-dependence of various features of linguistic interpretation (Korta and Perry 2011). The analogous idea proposed here is that the need for modal constraints should be determined with reference to the argumentative context.

In order to explicate this approach I begin by considering the diverse roles that thought experiments play in argumentation. There are, I suggest, a number of distinct functions that thought experiments play in arguments, which importantly cannot be reduced to a single function. Below I identify three such functions; although the list is not intended to be exhaustive. One consequence of this approach is that there is unlikely to be a single characterisation of what makes for a successful thought experimentation. (Here I have in mind the kind of regimentation that Soren Haggqvist gives in his 1996 book *Thought Experiments in Philosophy*). Note also the difference with Roy Sorensen's account in *Thought Experiments* (1992) in which Sorensen argues that the single role of thought experiments is to test modal consequences. Sorensen leavens this conception of function with the comment that the 'apparent narrowness of its function eases once we realize that there are many kinds of necessity' (Sorensen 1992: 6). Jacob Elster notes that imaginary cases are used in ethics to elicit intuitions against which moral principles might be tested and presumably elsewhere in philosophy they are used to challenge other principles (Elster 2011: 241), be they metaphysical or epistemological and so on.[11] Again he only mentions one function. However, on the account suggested herein, there is no such single function, no matter how broadly construed.

There are, then, at least three roles that thought experiments play in philosophical arguments. First, some thought experiments function as counter-examples in philosophical disputations. In responding to a theory or a definition that is intended to be either necessarily or universally true one might attack the position either by providing a counter-example or by demonstrating that the theory has absurd consequences.[12] The use of such refuters is a commonplace in moral philosophy. For example, in responding to the claim that it is *always* wrong to lie, an opponent of this view might argue that at least some lies are per-

[11] As the article proceeds, Elster does mention various uses of imaginary cases that do not fit quite so neatly into this understanding of their function.

[12] Roy Sorenson refers to these thought experiments as 'refuters', although, as he notes, not all refuters involve thought experiments (see Sorenson 1992: 153).

missible and do so by providing a case where most people would admit lying was acceptable (such as when telling a so-called ‘white lie’). The second category of thought experiments involves what we might refer to as ‘intuition pumps’. This is a term that is used in a variety of ways by philosophers—sometimes as a synonym for thought experiments and sometimes to refer to what the author believes is a pernicious mode of reasoning—but one common usage is where it refers to a class of thought experiments that aim to lead us, via our reactions to a single thought experiment, towards some general kind of conclusion. Trolley problems might be a case in point. Here we are meant to infer from the fact that we would choose to save the five rather than the one person on the track that numbers do count morally.[13] A third category of use to which thought experiments are often put by philosophers is as ‘clarificatory devices’ where the aim is to clarify our views on a difficult topic. Perhaps the most widespread of these are the ‘commitment cleavers’, that is, cases where thought experiments are used to enhance our understanding by teasing apart distinct, but easily conflated, principles. Presumably this is at least part of the significance of the story in the *Republic* of the Ring of Gyges that Plato raise in the midst of a debate about the nature of justice (Plato 1974: 36). The tale of this ring enables us to distinguish between those who endorse justice on the grounds of prudence and those who do so because they regard acting fairly to be a fundamental moral obligation that holds regardless of any benefits that might or might not accrue from being seen to act justly. Through Plato’s use of this dramatic device, the interlocutors in the *Republic* are forced to be more specific about their ethical and political commitments. In the end is it mere prudence or our intrinsic duty to act justly that underpins their publicly avowed commitment to justice?[14] Sometimes these argumentative devices are designed so as to test or clarify what a theory—as opposed to a person—might be committed. One might, for instance, devise a thought experiment that illuminates the difference between a Kantian and a Utilitarian approach to moral issues.

With this taxonomy in mind—and, more importantly, being apprised of the thesis that thought experiments play a variety of roles in argumentation—let us now return to the issue of method in ethics and how the idea of modal constraints relating to possibility might best be understood. (My assumption continues to be that Parfit’s problem raises important questions for ethical methodology). The claim herein is that the thought experiments philosophers often use in ethics can—

[13] One difficulty with attempting to use thought experiments in this manner is that people’s responses to thought experiments are often surprisingly varied. Hence they do not always pump our intuitions in the direction that the interlocutor hopes or expects.

[14] As C. L. Ten notes such thought experiments help us to determine whether a particular principle or commitment is “fundamental or subordinate” (Ten 1987: 21).

and should—be constrained, if the argumentative context renders more modally extravagant cases to be irrelevant. To illustrate how this would work, consider the following cases. If, for example, we were debating the rightness or wrongness of incarcerating human beings without trial it would be inadmissible in this context to introduce, as a relevant case, the example of possible people who enjoy being incarcerated. Given the practical context, such beings would not provide counter-examples to claims that it is wrong to imprison without trial. What matters here is not so much the *content* of the example in and of itself, but rather the *context* of the topic under discussion. With respect to the non-identity problem the argumentative context involves actionable harm). Yet, there will be other cases in which we are considering general principles of morality where it might well be appropriate to raise the moral consequences of persons who enjoy being incarcerated.

Consider a further example. Suppose that two philosophers are debating the issue of whether or not a pregnant women should have more say in any decision about whether or not to continue with a pregnancy than the man who fathered the foetus. Now suppose that in the context of this debate one of the disputants (let's call him 'Jim' for the sake of the example) raises the following imaginary scenario.

> *Ectogenetic birth*: Imagine a world in which human beings can be gestated entirely outside of the womb, in perhaps an incubator of some kind. In such a world there is no reason to think that the female parent has more rights than the male parent in determining whether the gestation should continue.

Jim argues that this example undermines the claim that women should have greater say than men in the determination of whether gestation should continue. The idea would be that the moral claim of women having greater rights here does not hold in all possible worlds for the very reason that it is not true all possible worlds that women will be responsible for the gestation. So far so good, but in line with the preceding dictum about context (i.e. our modal constraint), our friend Jim cannot use the case to make claims about whether or not women should have more say *now* in circumstances where women are in fact responsible for gestation. In this case, contingent features of the problem—namely the fact that we do not currently have ectogenetic birth—are relevant to our moral deliberations and should not be over-ruled by merely logically possible cases. And again it is a matter of modal moral luck whether or not a man might be thought to have less rights in determining whether gestation should continue.

The more general lesson here is not that we should always constrain our thought experimentation by contingent empirical realities, nor that bizarre examples should be expunged from moral thinking but, simply, that there will be cases—particularly in applied ethics and political philosophy—where the argumentative context is such that assessment of contingent factors about *what is actually possible* matters.

It is the argumentative context not the outlandishness of any thought experiment that should be our primary concern in determining admissibility. We must not rule out outlandish experiments merely because they are outlandish.

## 7. *Concluding remarks*

The non-identity problem raises difficult questions about whether it is possible to harm another person who, if we had not acted as we did, would not have existed. A significant assumption here is that if the harmful action had not been undertaken then the person would not exist. Yet, in each of the cases raised, it is entirely possible to imagine future scenarios—which are admittedly outlandish—in which the person existed without the harm. In order to sustain the claim that we cannot harm such individuals, one must rule out such imaginary scenarios. But why rule out these imaginary cases? My aim was to discover what might justify such constraints that are curious given the kinds of scenarios regularly explored in philosophical debate.

After considering the 'nomic constraint' (which would rule out all modally bizarre scenarios) I proposed, by way of justification, a more moderate constraint that focuses on the fact that in this instance we are concerned with actionable harm and hence the kind of scenarios that are relevant should be restricted by that concern. Scenarios are restricted in the non-identity problem because of the argumentative context. This is the core idea of the approach of 'thought experiment pragmatics'.

The more general claim I made is that it is the argumentative context, not the modal content of a particular imaginary case, which determines whether it is admissible. The approach defended here is midway between, on the one hand, those who would adopt an 'anything-goes' policy and, on the other, those like Kathleen Wilkes who regard outlandish thought experiments as intellectually pernicious (Wilkes 1988). There are, I would suggest, genuine grounds for limiting in certain debates the range of thought experiments that are to be regarded as admissible. This is an important lesson for areas of applied philosophy. While matters are somewhat different when considering, for instance, general moral principles, when the issue is very much a practical question of applied philosophy, such as is the case with the assessment of responsibility and actionable harm, it is appropriate to limit the range of relevant cases. The default position should be that outlandish examples are admissible until such time as an interlocutor can demonstrate that, given the context, they are irrelevant to the debate at hand.

## *References*

Archard, D. 2004. "Wrongful Life." *Philosophy* 79: 403–420.

Archard, D. and Benatar, D. (eds.). 2009. *Procreation and Parenthood: The Ethics of Bearing and Rearing Children*. Oxford: Oxford University Press.

Brock, D. 1995. "The Non-Identity Problem and Genetic Harms: the case of Wrongful Handicaps." *Bioethics* 9: 269–275.

Carter, A. 2001. "Can We Harm Future People?" *Environmental Values* 10: 429–454.

Cottrell, A. 1999. "Sniffing the Camembert: on the Conceivability of Zombies." *Journal of Consciousness Studies* 6: 4–12.

Davidson, D. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Society* 60: 441–458.

Dennett, D. 1995. "The Unimagined Preposterousness of Zombies." *Journal of Consciousness Studies* 2: 322–26.

Elster, J. 2011. "How Outlandish Can Imaginary Cases Be?" *Journal of Applied Philosophy* 28 (3): 241–258.

Feinberg, J. 1986. "Wrongful Life and the Counterfactual Element in harming." *Social Philosophy and Policy* 4: 145–177.

Gendler, T. S.. 2000. *Thought Experiments: On the Powers and Limits of Imaginary Cases*. New York: Garland Publishing.

Hare, C. 2007. "Voices from Another World: Must we respect the Interests of People who Do Not, or will Never Exist." *Ethics* 117: 498–523.

Haggqvist, S. 1996. *Thought experiments in philosophy*. Stockholm: Almqvist and Wiksel International.

Harris, J. 2000. "The Welfare of the Child." *Health Care Analysis* 8 (1): 27–34.

Holtung, N. 2001. "On the Value of Coming into Existence." *The Journal of Ethics* 5: 361–384.

Kirk, R. 2005. *Zombies and Consciousness*. Oxford: Clarendon Press.

Lycan, W. 1995. "Philosophy of Language." In R. Audi (ed.). *The Cambridge Dictionary of Philosophy*. Cambridge: Cambridge University Press: 586–589.

Korta, K. and Perry, J. 2011. *Critical Pragmatics: An Inquiry into Reference and Communication.* Cambridge: Cambridge University Press.

Marcus, E. 2004. "Why Zombies are Inconceivable." *Australasian Journal of Philosophy* 82: 477–490.

Millikan, R. 1996. "On Swampkinds." *Mind and Language* 11 (1): 70–130.

Nagel, T. 1979. *Mortal Questions.* New York: Cambridge University Press.

Nelkin, D. K. 2013. "Moral Luck." In *The Stanford Encyclopedia of Moral Philosophy* (retrieved August 2013) http://stanford.library.usyd.edu.au/entries/moral-luck.

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

Plato. 1974. *The Republic.* Translated by D. Lee. Harmondsworth: Penguin.

Pogge, T. 1990. "The effects of prevalent moral conceptions." *Social Research* 57 (3): 649–663.

Sorenson, R. 1992. *Thought Experiments*. New York: Oxford University Press.

Roberts, M. A. 1998. *Child versus Childmaker: Future Persons and Present Duties in Ethics and the Law*. Lanham: Rowman and Litlefield.

Roberts, M. A. 2013. "The Nonidentity Problem." In *The Stanford Encyclopedia of Moral Philosophy* (retrieved October 2013). http://stanford.library.usyd.edu.entries/nonidentity-problem

Roberts, M., and Wasserman, D. (eds.). 2009. *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*. Dordrecht: Springer.

Ten, C.L. 1987. *Crime Guilt and Punishment.* Oxford: Clarendon Press.

Thomson, J. J. 1971. "A Defense of Abortion." *Philosophy and Public Affairs* 1 (1): 47–66.

Tooley, M. 1983. *Abortion and Infanticide.* Oxford: Clarendon Press.

Wilkes, K. 1988. *Real People: Personal Identity without Thought Experiments*. Clarendon Press: Oxford.

Woodward, J. 1986. "The Non-identity Problem." *Ethics* 96: 804–831.

Wrigley, A. 2006. "Genetic Selection and Modal Harms." *The Monist* 89 (4): 505–525.

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *Book Discusson*

# *Bending and Stretching the Definition of Lying*

MARTINA BLEČIĆ
*University of Rijeka, Rijeka, Croatia*

*One of the recent trends in dealing with the concept of lying has been to argue that the idea that one needs to deceive someone in order to lie has been accepted too hastily. In* Lying and Insincerity *Stokke shares this opinion and proposes a definition of lying based on the notion of common ground that includes bald-faced lies. Additionally, he rejects the idea that lying can be accomplished with pragmatic means such as conversational implicatures and proposes a formal distinction between lying and misleading. In this review, I present the content of Stokke's book and critically discuss the two points mentioned above.*

**Keywords:** Lying, misleading, implicature, common ground, pretence.

Andreas Stokke's book *Lying and Insincerity* (2018) is a valuable addition to the debate about lying and deception that proliferated in the last decade or so.[1] In what follows I will briefly present the content of the book and then I will lay out my thoughts on some topics about which I disagree with the author. This disagreement should be read as a praise of the engaging content and presentation of the book. First of all, a few preliminaries about the general discussion about lying are in order.

An analysis of lying can be focused on the moral or on the conceptual dimension of the phenomenon. The first approach deals with the moral (un)acceptability of lying and considers questions like the following: are all lies bad, are some lies worse than others, is misleading

[1] Another important one is the book *Lying: Language, Knowledge, Ethics, and Politics*, a collection of essays edited by Eliot Michaelson and Andreas Stokke, published also in 2018 by Oxford University Press.

better than lying? The second approach tries to provide a theoretical definition of lying and focuses on the features that differentiate it from similar phenomena, such as deception, misleading or bullshit. It could seem that the moral approach is normative in its nature and that the conceptual one is purely descriptive, but the strength of this kind of distinction should not be overestimated. In his book, Stokke focused on the nature of insincere speech, including the attitudes that lie behind them, and does not venture into the moral dimension of the discussion.

Having said that, we can briefly present the structure of *Lying and Insincerity*. The book is divided in two main parts: Language and Attitudes. Chapters 1–5 are devoted to questions of language, and chapters 6–10 to matters of attitudes that may lie behind insincere speech. As the author points out, insincere speech resides at the intersection between language and attitudes, which results in the two parts of the book being interrelated. More specifically, but still very generally, we can say that in the first part of the book Stokke defends the idea that the intention to deceive the addressee is not crucial for lying and proposes a distinction between lying and deception based on formal aspects of language, while in the second part he defends the idea that in order to lie one must only be insincere on a shallow, that is, conscious, level and analyses the connection between insincere attitudes and the phenomenon of bullshitting.

In Chapter 1, Stokke defends the idea that the intention to deceive the addressee should not be included in the definition of lying because it excludes cases that we would, according to the author, intuitively consider as instances of lying. That leads him to the conclusion that lies cannot be generally understood as a species of deception, even though lying is often aimed at deceiving its victims. Stokke endorses the position that lies are insincere assertions, in the sense that a lie is a statement that, although does not need to be false, has to be disbelieved by the speaker. As the author notes in the Introduction, "(t)he main challenge for a view of lying as insincere assertion is to spell out what it is to assert something in a way that is broad enough to capture the nature of lying and narrow enough so as not to obscure the distinction between lying and other kinds of insincere speech" (Stokke 2018: 6). As all definitions, a definition of lying should not be to narrow nor to broad, but as we shall see, the nature of lying it should capture is still a matter of debate. In the chapter, Stokke presents the following definition of lying: "A lies to B if and only if there is a proposition such that

(L1) A says that *p* to B, and
(L2) A proposes to make it common ground that *p*, and
(L3) A believes that *p* s false." (31)

According to this definition, a lie is an insincere assertion. In the following chapters he defends this definition and explains the key concepts it is built on.

One important task is to define what an assertion is. In Chapter 2 Stokke presents a Gricean view of assertions, which he then rejects as inadequate. According to Stokke, the Gricean proposal is "bound to fall foul of particular facts concerning bald-faced lies, and concerning the way such a definition must locate lying in relation to the saying-meaning distinction" (38). In Chapter 3, he proposes his preferred notion of assertion, one based on Stalnaker's notion of common ground. According to the author, common ground information is information that is accepted for the purpose of the conversation. Using this notion, Stokke presents lies as saying something and thereby proposing that it becomes part of the common ground between speaker and hearer. Chapter 4 is devoted to the notion of what is said, and Chapter 5 to the difference between lying and misleading. The author argues for a notion of what is said that is sensitive to questions under discussion, i.e. the topic of the conversation, while being constrained by linguistic meaning. His main reason for characterizing lying in terms of assertion is to differentiate lying from non-linguistic forms of deception and insincerity. The classic contrast between lying and merely misleading is the contrast between asserting disbelieved information and conversationally implicating such information by asserting something believed to be true.

In the second part of the book the author explores the relationship between what is communicated and the speaker's attitudes. He shows this relation using the notion of bullshit, which is used to illustrate the point that insincerity is a more complex phenomenon than communicating what one believes to be false. Sometimes people communicate certain contents while being indifferent toward their relation with the truth. He argues for a shallow view of insincerity according to which insincerity is a matter of speaking without a conscious intention to communicate something one assents to, that is, an utterance is insincere when it is not consciously intended to communicate something the speaker assents to.

Chapter 6 opens the discussion about the ways in which people sometimes speak while being indifferent towards what they say that extends to Chapter 7. In the next two chapters, Stokke defends a shallow conception of insincerity. According to this view, whether or not one speaks insincerely depends on his or her conscious state of mind, not on unconscious beliefs, hopes or desires. The final chapter of the book, Chapter 10, explores the way in which we use various linguistic forms other than simple utterances of declarative sentences to communicate our attitudes in language. According to Stokke, even though non-declarative utterances can be insincere, they cannot be used to tell lies.

After having briefly presented the content of the ten chapters, I will proceed to comment specific topics Stokke deals with, concentrating mainly on the first part of the book. The first one is related to his endorsement of a non-deceptionist account of lying. The second one is his formal distinction between lying and misleading, The notion of lying provided by Stokke rejects some features that lies are traditionally supposed to have, while arguing for the necessity of others. There is no philosophically accepted definition of lying, but following Mahon (2016) we can identify four necessary conditions for an expression to be considered a lie in the traditional sense. The first condition is the statement condition, according to which lying requires a person to make a statement. The second is the untruthfulness condition, that states that lying requires that the person believes the statement to be false. The third is the addressee condition, that is the idea that lying requires that the untruthful statement be made to another person. According to the last condition, lying requires that the person intends the other person to believe the untruthful statement to be true. This condition is labeled as the intention to deceive the addressee condition. If all the conditions are satisfied, we are faced with a lie in the traditional sense: a statement that is believed to be false made to another person with the intention that they believe that statement to be true.

Recently, various authors have challenged the fourth condition (see Carson 2006, Sorensen 2007 and Fallis 2009), claiming that lying does not necessarily involve an intention to deceive the addressee. Stokke adhered to this current in the debate in his previous work (see Stokke 2013) and explicates his position even further in this book.

Following Carson (2006), he presents The Cheating Student example, which should provide to the reader a clear example of a lie made without an intention to deceive. The example goes as follows.

> "A student accused of cheating on an exam is called to the Dean's office. The student knows that the Dean knows that she did in fact cheat. But as it is also well known that the Dean will not punish someone unless they explicitly admit their guilt, the student says,
>
> (1) I didn't cheat
>
> Although the student says something she believes to be false, she does not intend to deceive the Dean. Even so, the student is lying." (Stokke 2018: 17, 18)

This is a classic example of what Sorensen (2007) has labeled bald-faced lies. Stokke seems to presuppose that this idea will be accepted at face value by the reader. In the pages that follow after the example, Stokke defends his position on the basis of intuitions and on what he considers to be a standard sense of the word "lie". Here are some examples of the constructions he uses: "It is highly counterintuitive to insist that the student in Carson's example did not lie to the Dean" (19); "(…) the insistence that the student did not lie that relies on a non-standard sense of the word."(21); (…) in such cases, this statement is still intuitively a lie" (28); (…) the student's utterance is still clearly a lie" (29).

I would like to suggest that in order to include this kind of cases in the definition of lying, i.e. exclude from it the intention to deceive the addressee condition, empirical data regarding people's attitudes that would support this position should be provided, or the position should be backed up by arguments. Without any of these elements, readers who share the author's intuitions will be convinced that bald faced lies are in fact lies, but those who do not share them could remain unconvinced by the examples alone.

Stokke returns to this particular example later on and explains it using his preferred notion of "common ground". Following Stalnaker (2002), he does not view common ground in terms of a believed proposition, but he proposes to view it as a proposition accepted for the purpose of the conversation (or believed to be accepted as such, or even just believed to be available). Applying this notion to lying we should say that "to lie is to say something one believes to be false and thereby propose that it be accepted by the participants and commonly believed to be accepted" (Stokke 2018: 52). Again, this view points to the idea that the intention to install false beliefs in the hearer is not necessary for lying. What is important for Stokke is that this notion allows something to be part of the common ground even when it is believed to be false. This is needed to allow bald faced lies in the definition of lying.

It could be objected that this notion of acceptance is too weak. The hearer would accept the speaker's proposition that *p* every time he understands it and is aware of the fact that the speaker wants to make it common ground that *p*. But knowing what the expressed proposition means and recognizing the intention of the speaker does not mean allowing it into the common ground. It yet has to become information that will be jointly used in the conversation.

Returning to the cheating student case, the student wants it to be common ground that she did not cheat. That is, she does not intend for anyone to really believe it, but she wants it to be accepted for the purpose of the conversation. Still, the Dean does not have to agree to this. We can imagine different ends of the story: the Dean can refuse to accept what the student said and explain to her the repercussion of a false statement and schedule another meeting with her. In this case, the student's assertion has not become part of the common ground, on the contrary, it brought the conversation to an end. Otherwise, if the Dean accepts the assertion and goes along with it, the conversational exchange that follows could be interpreted as "pretending". That would make it similar to a play, in which none of the parties involved sincerely believe what they say, but they are pretending to do so in order to achieve some performative goal. Stoke acknowledges a similar objection and discusses Mahon's (2016) idea of "pretence". According to him, this kind of objection should be understood as maintaining that the kind of pretence involved is unserious. He rejects the objection so understood by presenting the fictional case of a trial held under a totalitarian regime (see Stokke 2018: 58). He invites us to imagine that someone is called to the stand to testify about something that is commonly known to be false, that is, to go on the stand and tell a bald-faced lie. The fact that people in real life situations had chosen to be executed rather than to do so should prove to us, Stokke claims, that such "pretence" is anything but frivolous. But this example could point in another direction. Being on the stand and telling a bald-faced lie would carry with it the additional information of accepting the authority of those in power. This is the massage they do not want to commit to. The most salient message differs from what is said and it is exactly what the person on the stand wants to convey and make part of the common ground.

The same goes for the cheating student case. In saying that she did not cheat, the student is conveying the additional information that she knows the rules and she is going to take advantage of them. I believe this to be more relevant and informative that claiming that she did not cheat while everybody knows she did. What would be the point of that if no additional information is intended? I believe that this information is calculable in Grice's sense, making it a conversational implicature. The information conveyed using conversational implicatures is exactly what the speaker wants to be part of the common ground, and she relies on the rational capacity of the hearer to reach this conclusion to bring the message across. Still, Stokke, rejects the idea that intended false implicatures should be considered lies.

At this point it is worth presenting what Stokke labels as "bald-faced false implicatures". He uses the following example to illustrate what he has in mind (55). Thelma knows that Louise knows that Thelma has been drinking. Louise asks: Are you OK to drive? And Thelma replies: I haven't been drinking. Thelma implicates that she is OK to drive. As in the case of bald-faced lies, Thelma is not trying to get Louise to believe that she is OK to drive. In this case the false implicature is derived from a bald-faced lie. It is interesting to notice that the utterance of "I haven't been drinking" is considered by Stokke a bald-faced lie, but the implicature "I am Ok to drive", even though it is the most salient piece of information, is not considered to be a lie. This reflects his acceptance and strong defence of the statement condition for the definition of lying.

I believe that this position could be challenged by the introduction of the notion of default meaning in the discussion about lying and misleading. Default meanings are those arising automatically in a given situation of discourse (Jaszczolt 2005, 2010). They are the most salient or relevant meaning in a particular context. The primary content of an utterance is its most salient meaning. According to Jaszczolt, this is so even when this meaning does not bear any resemblance to the logical form derived from the syntactic structure of the uttered sentence (Jaszczolt 2016). I believe that applying the notion of default meaning to the discussion about lying would shed new light to some problematic cases.

Stokke rejects cases of falsely implicating as instances of lying, providing the following quote from Fallis (2009) to support his view: "you are not lying if you make a statement that you believe to be true. In fact, you are not lying even if you intend to deceive someone by making this statement" (44). According to Stokke, including false implicatures in the definition of lying “rejects one of the most fundamental distinctions we make about verbal insincerity (44.)”. Still, the goal of the authors who have tried to include false implicatures in the definition of lying (Stokke mentions Meibauer 2005 and Dynel 2011) is exactly to question the assumed distinction between lies and other forms of verbal deception. It seems that Stokke rejects the suggestion that intended false implicatures could be considered lies because they are excluded from the traditional definition, but someone who is so eager to reject the idea that an intention to deceive is necessary for lying, even though this idea has been widely defended and accepted, should not reject other approaches on the basis of their unorthodoxy. Accepting that the cheating student is in fact implicating something, and that this implicature is in fact the most salient meaning of his utterance, we could have a good explanation for the idea that the student did not lie at all: the default meaning of her utterance can be paraphrased with a true proposition.

In Chapter 4 Stokke gives his definition of what is said, which is defined as the weakest answer to a question under discussion that either entails or is entailed by a minimal proposition expressed by the utterance in question, given the context. Stokke’s notion of what is said is compositional. What is said is expressed by the minimal proposition, that is, a proposition that is determined solely by the composition of the constituents of the relevant sentence. He believes that the account of what is said presented in Chapter 4 draws the line correctly between lying and other forms of misleading speech. Still, this presupposes that the distinction between what is said and what is conveyed less explicitly matches the distinction between lying and misleading. As I have noted above, if we change this formal notion of what is said with the notion of default meaning, which I believe to be more suited for assessing various communicational layers our results could be more encompassing. Using Stokke’s terminology, a default meaning could be defined as the content of an utterance that optimally responds to the question under discussion. According to the author, communication is a cooperative activity of information exchange aimed to discover how things are (see p. 81). It remains somehow unclear, at least to me, why confine the idea of “question under discussion” to a formal notion of what is said. During a regular communicational exchange speakers and hearers communicate explicitly and implicitly, creating meanings and trying to “discover how things are” jointly, mostly unaware of formal distinctions between semantics and pragmatics. What is more important, they communicate successfully, which means that the discovery of how things are can be achieved by implicit and indirect communicational means. Again, it could be objected that what is merely implicated somehow always remains uncertain. I believe that this could be put aside since, following Mercier and Sperber (2017), this “uncertainty” is a distinguishable feature of all every-day reasoning.

In short, my main point so far was questioning the adequacy of a pure formal distinction between lying and misleading since communicators are often unaware of it, and their reliance on other people's words is not exhausted by "what is said". The same goes for lying. If I form a false belief o the basis of another person's words because that person had an intention to affect negatively my epistemic condition is it really relevant if it was done with assertions or implicatures? Would I really care?

As it has been mention at the beginning, Stokke is not concerned with the moral aspects of lying. Still, it is important to note that the distinction between lying and misleading has important moral consequences. According to the traditional view, misleading is always better than lying. But this position should be critically assessed taking into account the fact that it was a response to a general religious condemnation of lying. What better way to evade this strong moral position than to have a narrow definition of lying? Still, the idea has been perpetuated by contemporary authors like Fricker, who claims that where what is conveyed is not explicitly asserted there is a diminution in the responsibility for the truth of what is got across incurred by the utterer (Fricker 2006). The idea that conversational implicatures can be easily denied, regardless of the plausibility of such denial certainly also helped view implicatures as a weak communicational strategy (see Pinker and Lee 2010).

I believe that a rigid distinction between semantics and pragmatics is certainly useful for a formal analysis of language and speech. Discovering the intricate interaction between implicit and explicit content we use in communication is fascinating and makes us eager to create new fine-grained distinctions, but most language users are not aware of these intricacies. They want to communicate, they want to say that someone told a lie if he or she communicated something believed to be false, regardless of the degree of expansion of the proposition expressed. They will tell the truth and lie using implicatures without a conscious effort to communicate implicitly and indirectly.

Finally, I would like to illustrate the points I tried to make using a literary example. Recall Shakespeare's Iago, a character called by many in the play "honest" but who is in fact plotting to convince Othello that his wife is having an affair. During the many dialogues between these two characters, Iago never utters an explicit lie about Desdemona's affair, he suggests and insinuates, corroding in this way Othello's belief in his wife's fatefulness. Near the end of the play, Emilia, Iago's wife, confronts her husband and asks him to explain why Othello sad to her that he made him believe "his wife was false". Iago replies: "I told him what I thought, and told no more than what he found himself was apt and true" (Shakespeare 2006: 384). Whit his utterance Iago tries to distance himself from the belief Othello formed on the basis of his words, shifting the responsibility to Othello himself. Unconvinced by his response, Emilia asks him directly if he ever told Othello that

Desdemona "was false". "I did", replied Iago. Emilia concludes that he has lied: "You told a lie, an odious, damned lie", she says (Shakespeare 2006: 384). Why would Iago admit having lied if he never said explicitly that Desdemona was not being faithful unless his insinuations do count as lies? Moreover, I believe that the audience, acquainted with Iago's malicious plans trough his monologues, would not say that Iago is lying only on the rare occasions in which he is using an explicit lie and that he is not lying when he is "merely misleading". The subtleties of his deception are what makes his character interesting, but he is universally considered to be a liar, which would be contradictory if one would claim that in fact he was not lying.

To conclude, Stokke presents his ideas clearly, backing them up with a multitude of examples useful for testing the reader's intuition about the matter at hand. Sometimes, my intuitions differ from the author's, but this is exactly what made the book engaging. It is a book dense with concepts and theoretical questions, still, Stokke manages to make them accessible and easy to follow even for readers that are not acquainted with the ongoing debate about the discussed topics. My main concern is that the current debate about lying relies too much on a format notion of what is said that does not reflect the way people use language in their everyday lives. In my view, this makes the definition of lying too narrow. Still, many of the same authors argue for a definition that includes bald-faced lies, which makes it simultaneously too broad.[2*]

## *References*

Carson, T. 2006. "The definition of Lying." *Nous* 40: 284–306.

Dynel, M. 2011. "A web of deceit: A neo-Gricean view on types of verbal deception." *International Review of Pragmatics* 3: 139–167.

Fallis, D. 2009. "What is Lying? "*Journal of Philosophy* 106: 29–56.

Fricker, E. 2006. "Testimony and Epistemic Autonomy." In J. Lackey and E. Sosa (eds.). *The Epistemology of Testimony*. Oxford: Clarendon Press: 225–250.

Jaszczolt, K. 2005. *Default Semantics: Foundations of a Compositional Theory of Acts of Communication*. Oxford: Oxford University Press.

Jaszczolt, K. 2010. "Default Semantics." In B. Heine and H. Narrog (eds.). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press: 215–246

Jaszczolt, K. 2016. *Meaning in Linguistic Interaction: Semantics, Metasemantics, Philosophy of Language*. Oxford: Oxford University Press.

Lee, J. and Pinker, S. 2010. "Rationales for indirect speech: The theory of the strategic speaker." *Psychological Review* 117 (3): 785–807.

Mahon, J. E. 2016. "The Definition of Lying and Deception", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/lying-definition/.

[2*] This work has been fully supported by the University of Rijeka under the project number 17.05.2.2.05.

Meibauer, J. 2005. “Lying and falsely implicating.” *Journal of pragmatics* 37: 1373–99

Mercier, H. and Sperber, D. 2017. *The Enigma of Reason*. Cambridge: Harvard University Press.

Shakespeare, W. 2006. *Othello, the Moor of Venice*. Michael Neill (ed.). Oxford: Calderon Press.

Sorensen, R. 2007. “Bald-faced Lies! Lying without the intention to deceive.” *Pacific Philosophical Quarterly* 88: 251–264.

Stalnaker, R. 2002. “Common Ground.” *Linguistics and Philosophy* 25: 701–21.

Stokke, A. 2013 “Lying and asserting.” *Journal of Philosophy* 110 (1): 33–60.

Stokke, A. 2018. *Lying and Insincerity*. Oxford: Oxford University Press.

*Croatian Journal of Philosophy*
Vol. XX, No. 59, 2020

# *Book Reviews*

## *David Hitchcock,* On Reasoning and Argument. Essays in Informal Logic and on Critical Thinking, *New York: Springer, 2017, xxv + 553 pp.*

David Hitchcock's rather comprehensive book *On Reasoning and Argument. Essays in Informal Logic and on Critical Thinking* is a (revised and supplemented) collection of the author's essays on these broad topics, which were published independently during his long-lasting career of almost fifty years. As such, it fulfils the role of Hitchcock's long-awaited monograph on central issues in informal logic and critical thinking, both of which he had been teaching as a university professor at McMaster University in Canada. Though he had (co)-authored two other books earlier (*Critical thinking: A guide to evaluating information*, 1983. and *Evidence-based practice: Logic and critical thinking in medicine*, 2005), former of them a textbook, only *On Reasoning and Argument* provides a systematic overview of his views on informal logic and critical thinking, emphasizing *inter alia* in the concluding part of the book that these two notions should be carefully distinguished; first one pertaining to "sub-discipline of philosophy that seeks to develop criteria, standards and procedures for the construction, identification, analysis, interpretation, evaluation and criticism of arguments" (511), different from the one employed in formal logic, and second one pertaining to "a process of reflectively thinking about an issue with a view to reaching a reasoned judgment on what is to be believed or done" (511), an educational ideal which is to be fostered at all levels. The book is divided into several parts (I–VII), each of which comprises several chronologically ordered chapters—essays, accompanied by *References* and ending in a *Postscript*, written in retrospect for the purpose of this publication. In the *Postscripts* Hitchcock summarizes his theses from the chapters-essays, providing additional information on their genesis and adding critical remarks to his earlier views where needed. The book is also equipped with *Index* of names and concepts and a helpful *Foreword* by J. Anthony Blair, who encouraged Hitchcock to publish it in the first place, and a *Preface* by Hitchcock himself.

In Part I (*Deduction, Induction and Conduction*), composed of two essays written almost four decades apart, Hitchcock dwells on two well-established distinctions in argumentation theory (or philosophy of argumentation, as he puts it); deduction vs. induction—when it comes to different types of argument validity—and linked vs. convergent—when it comes to two or more reasons supporting a claim. Concerning the first topic, Hitchcock in his original essay rejects objections to deduction vs. induction distinction on grounds (i) ‘that some traditionally inductive and some traditionally deductive arguments provide conclusive grounds for their conclusions and some do not’ (objection by Perry Waddle) and that (ii) reconstructing arguer’s intention is necessary to classify arguments as of one type or another. Ad (i), Hitchcock points out that reasons for a claim being conclusive is not equivalent to argument’s being deductive and that filling inductive arguments with premisses which would turn them into deductive ones is not possible due to premisses not being justified independently of the conclusion. Ad (ii), although conceding that appraisal is concerned with arguments, not arguers, he maintains that “using a version of the principle of charity in settling on the standards by which to assess an argument. That is, we should assess it by those standards which give it the best chance of being a cogent argument” (19). As far as the linked vs. convergent distinction is concerned, Hitchcock considers it useful but only when applied non-derivatively to types of support of premisses to a conclusion and only derivatively to argument structures. Main revision in the *Postscript* to the original text consists in defining inductive strength as a type of support or a standard of appraisal, not as type of validity (33).

Part II (*Material Consequence*), containing six chapters, is probably central to the book since it addresses issues which have, according to Hitchcock himself, occupied him throughout his career. Starting with a paper on enthymematic arguments, Hitchcock emphasizes two problems regarding them which haven’t been satisfactorily solved: (i) the demarcation problem, i.e. distinguishing enthymemes from deductively valid arguments on the one hand and mere *non sequiturs* on the other hand and (ii) the evaluation problem, i. e. how to evaluate the inference in an enthymematic argument (40). He rejects defining enthymems as arguments whose authors have omitted one or more premises for two reasons: (i) we are often not in a position to question the arguer about whether the arguer had another premise in mind and (ii) authors of acknowledged enthymemes often have no additional premise in mind (43). He accepts the alternative approach according to which the implicit assumption of an enthymematic argument is a rule of inference (non-formal rule since its statement includes at least one content expression) in virtue of which the conclusion follows from the premisses (53). In the following chapter he continues the same idea, discussing various conceptions of logical consequence (64–68) and opting for introduction of enthymematic consequence, he revises definition of logical consequence in the following way: conclusion is a consequence of given premisses in the revised generic sense if the argument has a general feature which is incompatible with the argument’s having true premisses and a false conclusion, even though it is both compatible with its having true premisses and compatible with its having a false conclusion (77), thus he is able to define enythmematic consequence as one where the general feature includes a reference to at least one extra-logical constant. Similarly, he defends Stephen Toulmin’s notion of warrants as general rules of inference, not implicit premisses, answering objections to his distinction between data or grounds and warrants. Hitchcock further advances an ontic, not epistemic conception of inferential support according to which the conclusion of an argument might have inferential support though an addressee of the argument is not aware of its having it. He requires that inference-licensing covering generalizations be not only true (or otherwise acceptable) but also capable of supporting counterfactual instances.

In Part III (*Paterns of Reasoning*), composed of seven essays, Hitchcock deals with various issues, starting with validity of non-deductive arguments. He rejects his earlier position of methodological deductivism, i. e. proposal that "non-deductive arguments could be treated as if they were deductive, as long as one recognized that the proposition one added to make the argument deductively valid was not entirely the responsibility of the arguer, that it could in certain respects be presumed to be true unless shown otherwise" (199) and proposes methodological conclusivism instead, i. e. treating non-conclusive arguments as conclusive if a proposition to which the author is committed by the argument is added, presuming the added proposition to be true until shown otherwise. Examining reasoning by analogy and acknowledging its various kinds, Hitchcock aims to give criteria for good analogical inference according to his general theory of good inference, but also to discard the thesis of epistemological subject specificity of analogical reasoning. In essay on Pollock's model of practical reasoning, Hitchcock applauds his point that "practical reasoning requires not only the beliefs and desires which theorists of practical reasoning have required for millennia, and not just the additional distinct category of intentions for which Michael Bratman has argued, but also likings" (223). He objects incompleteness of the model due to lack of interconnected features of communication between rational agents, social cooperation and the recognition of moral constraints, Hitchcock also considers Pollock's requirement of a cardinal measure of situation-likings applicable only to computational simulation of a rational agent, but no to human beings (222). In the following essay on argument schemes he argues for a combined top bottom and bottom up approach (e. g. Woods and Walton) due to theoretical arbitrariness of the former and empirical inadequateness of the latter, acknowledging that the system of schemes needn't be complete but comprehensive. In essays concerning instrumental rationality and practical reasoning, where "Instrumental rationality", i.e. the rational selection of means for achieving a given goal is analyzed as more complex than finding an effective means of getting to a chosen goal (ensurement of achievability of the goal and permissibility of means, determination that no alternative means is preferable, weighing side effects and benefits of achieving the goal etc.). Discussing what Trudy Govier labels "conductive arguments", Hithcock argues that they're better described as appeals to considerations or to criteria where the conclusion may follow either conclusively from its premisses or non-conclusively or not at all. He emphasizes that weighing the pros and cons is only one, and probably the last way to judge whether the conclusion follows.

The rather short Part IV (*Interpersonal Discussion*) is Hitchcock's enterprise in exploring dialectical aspects of argumentation, however emphasizing that although study of argument must take these into account, both descriptively and prescriptively, it often exaggerates in viewing all arguments as in a dialectical setting (336). He also stresses important features of arguments which are in common with monological reasoning, such as their inferential structure and their components' epistemic status. He is chiefly concerned not with rules which are supposed to guide a rational mutual inquiry (he acknowledges that the title is a bit misleading), but sets of principles to which such rules should comply (315). Hitchcock concludes that the study of formal dialectical systems can have both theoretical and practical benefits (clarification of dialectical concepts like proponent and opponent, exploring various commonly recognized fallacies, especially those like begging the question), many questions etc. which only occur in interpersonal discussion). Additionally, he proposes amendations to Ralph Johnson's *Manifest Rationality*, primarily concerning Johnson's use of term argumentation as a sociocultural activity of producing, interpreting and evaluating arguments; Hitchcock believes term argumentative discussion is more appropriate and in line with other authors' use (cites e. g. early Pragmadialectics). He also objects tp Johnson's positioning argumentative discussion prior to the concept of argument in the order of intelligibility, binding him to circularity in defining both concepts.

Part V (*Relevance*) discusses ontological status of relevance (relation, not a property), its relation to irrelevance (contradictory pair), types of relevance etc. Defining relevance as a triadic relation between an item, an outcome or goal, and a situation, Hitchcock distinguishes epistemic from causal and practical relevance, focusing on the first of the three. He describes epistemic relevance as irreflexive, symmetric and vacuously transitive in a strict sense, and in the loose sense it is either reflexive or irreflexive (depending on the epistemic goal), non-symmetric and transitive (357). Concerning relevance within argumentative setting, an argument is said to have an irrelevant conclusion "if its conclusion cannot be ineliminably combined with other potentially accurate information to achieve the epistemic goal to which the argument is addressed. It has an irrelevant premiss if the premiss cannot be ineliminably combined with other potentially accurate information to achieve the epistemic goal to which the argument is addressed" (367). Hitchcock discusses Locke's *ad* fallacies of relevance and acknowledges them as fallacies of relevance with respect to the epistemic goal of instruction (such appeals don't bring knowledge), but claims there are not necessarily irrelevant with respect to other epistemic goals, e. g. rational acceptance of a conclusion from authorities with expertise in a cognitive domain to which the conclusion belongs. In essay 'Good Reasoning on the Toulmin Model' Hitchcock examines individually necessary and jointly sufficient conditions for good reasoning (justified grounds, adequate information, justified warrant, justification in assuming no exceptions apply) in the Toulmin model, comparing it to parallel approaches of argumentation schemes and their critical questions.

In Part VI (*Fallacies*) Hitchcock discusses a more general issue of usefulness of teaching fallacies in teaching critical thinking and a more specific issue of *ad hominem* arguments. Inspired among other things by his own experience in teaching fallacies to students, Hitchcock advances several arguments against fallacies having central role in teaching critical thinking: (i) the correct identification of an argumentative move as a fallacy requires a complex apparatus of analysis, hence it makes more sense to teach the analytical apparatus for correct reasoning than to begin with the fallacies;

(ii) fallacy labels are not necessary to the exercise of critical thinking; everything that can be said with the use of these labels can be said without them (in general said more clearly); (iii) fallacies approach is unduly negative and fosters an attitude of looking for the mistake and labelling it, instead of dealing with the substance of what one is discussing; (iv) learning the fallacies is of no help in constructing good arguments of one's own and appreciating the merits of good arguments, which are components of critical thinking. Adopting Ennis' definition of critical thinking as reasonable and reflective thinking that is focused on deciding what to believe or do, Hitchcock emphasizes its constructive and reactive aspects which particularly goes against fallacies approach (425). He is also concerned with discrepancy between empirical data on types and frequency of mistakes in reasoning and traditional catalogue of fallacies (which John Woods named Gang of Eighteen). Discussing argument *ad hominem*, Hitchcock argues that it is not a fallacy in neither of its variants (abusive, circumstantial, *tu quoque*) due to the conception of a fallacy as a common mistake in reasoning that is commonly deceptive, but a legitimate dialectical strategy (similar to Woods' approach). However, he believes *ad hominem* attacks are necessary in teaching critical thinking since they concern finding good sources of information (students should learn under which conditions allegations of bias, incompetence or bad character are relevant to judging the quality of a source of information).

In the concluding Part VII (*Informal Logic and Critical Thinking*) Hitchcock discusses place of informal logic in philosophy, different concepts of argument which can be found within it, its relation to critical thinking and effectiveness of teaching critical thinking. As far as the first topic is concerned, Hitchcock is akin to classify informal logic as philosophy of argument, a sub-discipline of philosophy in its own right, particularly addressing often stated remarks on informal logic as applied or social epistemology (Battersby, Goldman), which he believes start from mistaken point of equating informal logic to critical thinking (which is a topic in philosophy of education). As mentioned above, Hitchcock accepts Ennis' definition of critical thinking and further differentiates it from the logical appraisal of arguments in extending beyond a single argument, thus having a creative component, and involving critical assessment of evidence. Critical thinking requires both skills, attitudes and dispositions which enable the critical thinker to think critically when required and do it well. Examining effectiveness of instruction in critical thinking, Hitchcock observes that its success is rather moderate, with more significant improvement in courses involving computer-assisted tutoring (argument mapping) or which are combined with writing instruction and practice (student discussion).

Although lacking a textbook structure, *On Reasoning and Argument* offers an informative overview of main topics in informal logic and critical thinking. It is probably more suitable for readers already introduced in these topics, although may appeal to novices. Hitchcock's careful approach is a fine example to younger scholars working in argumentation theory.

GABRIELA BAŠIĆ HANŽEK
*University of Split, Split, Croatia*

## *Michael E. Bratman,* Planning, Time and Self-governance: Essays in Practical Rationality*, New York: Oxford University Press, 2018, 272 pp.*

In his new book *Planning, Time and Self-governance: Essays in Practical Rationality* Michael E. Bratman tries to answer the following questions: *Why be a planning agent and what is the value of a planning theory of intention?* In order to answer these questions, he develops a diachronic account of rationality with the notion of self-governance. In order to fully understand what this means we need to look briefly at Bratman's earlier work. Over the course of the last three decades, Bratman has developed his theory of action and practical reasoning—a planning theory of intention. The planning theory of intention states the following. Human beings are planning agents. We have the ability to formulate and execute plans. Plans are types of intentions that are focused on future action i.e. future oriented intentions. Our capacity to form and execute plans stems from two general needs that we have as human beings: the need for deliberation or practical reasoning and the need for coordination. Our ability to deliberate would be of minimal use to us if we were doing it only moments before the time of action. In order to use our deliberate capacities to its fullest we deliberate in advance i.e. we plan. The second need that we, as human beings have is a need for coordination. We can distinguish between two types of coordination: personal coordination that we have with ourselves at different times (intrapersonal coordination) and coordination that we have with others (interpersonal coordination). Because we are limited creatures, both in cognitive and material recourses, we need, in order to achieve complex and temporally distant goals both types of coordination—intrapersonal and interpersonal. Plans are an essential part of human agency and practical reasoning. Our ability to make plans is something that separates us from animals (although not the only thing; others being our language capacity, reflection, higher-order cognition). Plans, as forms of intention, have distinctive normative properties like commitment and nonreconsideration. When we formulate plans we usually, if no new evidence, information or reasons arise, stick to them and do not reconsider them. This is because it would be almost impossible for us to manage our own lives if we would deliberate about every moment of every day on every decision we make. For example, if I want to go to the theatre I will do it in the following manner. I deliberate weather I want to go, decide on it, form an intention and then execute the action of going to the theatre. When all this is done, I will not, usually, reconsider every step of the way between my house and the theatre weather I should go or not. I consider that matter settled (although subject to change if I receive new evidence, information or reasons). From this Bratman builds the normative side of his planning theory of intention. The normative side consists of rational pressures that are put on the agent who identifies herself as a planning agent. Those pressures are intention consistency and intention stability. They state that an agent, if she is a planning agent, has to have intentions that are not contradictory. She cannot, simultaneously, have intentions that are not co-possible. For instance, an agent cannot intent and not intend to go to the theatre tonight. Also, her intentions have to have some level of stability i.e. she cannot suddenly drop her future-directed intention without any reason whatsoever. Lastly, this pressure gives rise to the norms of practical rationality. The most important norm for Bratman's theory of intention is means-end coherence. The norm, roughly states that an agent, if she intends some end E and is aware of the necessary means to E which is M she is rationally required to intent to M. This simply means that we need to intend the necessary means (if we know them) for our desired ends (goals). This is, very briefly, Bratman's theory of intention which he has developed in the last three decades and which had profound influence in the fields of philosophy of instrumental rationality and philosophy of action.

Now we can return to the book at hand—*Planning, Time and Self-governance: Essays in Practical Rationality*. In this book, Bratman tries to answer the questions *why to be a planning agent and what is the value of a planning theory of intention?* With his planning theory of intention Bratman has presented an in-depth and influential way of thinking about practical reasoning, instrumental rationality and everyday decision making. Bratman's model explains and offers normative structures for everything from simple everyday decisions like what to eat for lunch tomorrow, to choosing between different option regarding your carrier or planning your retirement years. But, according to some philosophers he has not answered the real question regarding rationality and action, that is why should we care about being a planning agent and what value does a planning theory of intention brings to our lives. The short answer, located in the title of the book, is *self-governance*. We all want, at least a certain amount of, coherence and stability in our own lives. Means-end coherence and stability of intention can certainly provide instrumental reasons for stability and coherence in our lives but it seems (according to Bratman's critics and Bratman's argumentation in this book) that we value governing our own lives noninstrumentally—and that value is self-governance. For the long answer to the questions *why to be a planning agent and what is the value of a planning theory of intention* we need to examine the book more closely.

Firstly, we shall take a look at the structure of the book i.e. how chapters align with one another, secondly we shall examine the content of all the chapters and lastly see how it all ties up together.

The book is comprised of a set of essays that can stand independently of each other. Each essay has a clear and precise line of argumentation that can stand on its own and serves as a point in the overall argumentation of the entire book. All of the essays, excluding the first essay (introduction) and the last essay, were published as independent papers elsewhere. These essays serve as chapters in this book and their order is chronological (with some exceptions). Bratman's argumentation in this book can be analyzed as having two parts with two small excursions.

In the first part, Chapters 1–4 (roughly), Bratman establishes the problem at hand. The first problem is, as I have mentioned at the beginning, why should someone be a planning agent or what is the value of a planning theory of intention. The second problem is a problem of instrumental rationality in general i.e. whether there is such a thing or can it be reduced to theoretical rationality. Bratman acknowledges that these are genuine problems for his planning theory of intention and that something needs to be done. In these chapters he is also laying grounds for the latter argumentation involving self-governance.

In the second part, Chapters 5–11 (roughly), Bratman proposes a brand new way of looking at his planning theory of intention and that is *Self-governance-Planning Agency*. The idea is to put some value into the planning theory of intention and that value is self-governance. We, presumably, find some value in governing our own lives in contrast to aimlessly going from one personal project to another not finishing any of them. Bratman is arguing, roughly, that in order to achieve what we value, and that is self-governance, we need to commit ourselves to the normative aspects of his planning theory of intention; means-end coherence and intention stability. Now I will briefly discuss the content of each of the chapters in the book.

Bratman's "Introduction" servers two main purposes. The first one is to offer a summary of all the other chapters in the book and the second is to present the *challenge*. The challenge is presented by Joseph Raz and Niko Kolodney and states that the idea that planning norms are norms of rationality is a myth. The rest of the book is Brtaman's response to that challenge.

In the second chapter "Intention, Belief, Practical, Theoretical" and the third chapter "Intention, Belief, and Instrumental Rationality" Bratman expands and explains the challenge presented to him. Bratman is claiming that all of his critics, or at least most of them, have one thing in common. They are reducing the requirements of practical rationality, like demands for consistency and coherence, to the requirements of theoretical rationality. He calls these authors *cognitivists*. Cognitivists are authors who claim that "practical rationality of one's system of intentions is, at bottom, theoretical rationality of one's associated beliefs" (19). There are at least three authors who can be classified as cognitivists: Gilbert Harman, J. David Velleman and R. Jay Wallace. In these chapters Bratman engages with the criticism of Gilbert Harman and J. David Velleman more thoroughly. Harman's basic idea, as Bratman calls it, is that when an agent intents some end E she is necessarily believing E. Bratman responds by arguing that sometimes we intent some end E and do not believe it—as in the case of forgetfulness. Velleman's critique of Bratman's work can be roughly summarized by the following question: "Why… should an agent be rationally obliged to arrange means of carrying out an intention, if he is agnostic about whether he will in fact carry it out?" (Velleman 2007: 205). This is an attack on Bratman's core normative requirement of practical rationality—means-end coherence. Bratman's response is that we are rationally obliged to the norm of means-end coherence because this norm stems from practical values like cross-temporal integrity, cross-temporal self-governance and sociality.

Chapter 4 "Intention, Practical rationality and Self-governance" is the core chapter of Bratman's book. In it he defends his planning theory of intention as an account of rationality and sets foundations for a diachronic account of rationality by introducing the notion of self-governance. Firstly, Bratman restates his norms of practical rationality, *means-end coherence,* and *intention consistency*, respectively. Then, he argues that we have a distinctive noninstrumental practical reason to oblige to these norms. That reason is cross-temporal self-governance. The concept of self-governance is something that Bratman derives from the works of Harry Frankfurt. Frankfurt's idea is that we need an account of what is it for an agent to identify with a certain thought or an attitude. In other words, what is it for a thought or an attitude to speak to an agent; which thought has agential authority for an agent. Frankfurt, and subsequently Bratman, states that the relevant question for our practical thought and action is for an agent to ask herself *Where do I stand?* with respect to my intentions, attitudes and desired ends. She does this via deliberation and reflection. When she has found "the place to stand" on some practical issue she can govern in a particular domain, "for self-governance where you stand guides your relevant thought and action" (97). Because, we as human beings, are planning agents, we have the reason to oblige to practical norms of rationality and that reason is self-governance.

In Chapter 5 "Agency, Time and Sociality" Bratman is introducing and reintroducing two ideas that will be relevant for his diachronic account of rationality. Those ideas are *shared intentional activity* or the ability to have *we-intentions* and self-governance at the time (synchronic) and self-governance over time (diachronic). He does not explore these ideas in substantial details in this chapter.

In Chapter 6 "Time, Rationality and Self-governance" Bratman expands on his notions of synchronic and diachronic self-governance. Synchronic self-governance is simply an agent's practical standpoint at the time i.e. "synchronic structures of attitudes that is sufficiently unified so as to constitute where the agent stands at that time" (144). Synchronic self-governance is a necessary but not sufficient condition for diachronic self-governance. In order to achieve diachronic self-governance several conditions need to be met. Those conditions are the diachronic notion of personal identity i.e. an agent needs to be the same person over certain period of time, psychological continuity of the agent's mental states over time, semantic interconnectedness of the agent's intentions and default stability of intention. Bratman argues that the agent's intentions need to be meaningfully connected in the context of her practical standpoint. He is doing that because he wants to avoid examples like the agent having a coherent and consistent set of weird and physically impossible fantasies. The agent's intentions need to be stable (in absence of supposed conclusive reasons for change) in order to persist over time. When these conditions are met we have a diachronic notion of self-governance. We can use this notion as a normative (noninstrumental) reason to conform to practical norms of rationality like means-end coherence and stability of intention.

As I mentioned before, there are two excursions in the second part of the book; Chapter 7 "Temptation and the Agent's Standpoint" and Chapter 8 "The Interplay of Intention and Reason". In Chapter 7, Bratman revisits one of the key issues of his planning theory of intention—the problem of temptation, which is, in short, a diachronic form of the "weakness of will" problem. In Chapter 8, Bratman engages in a discussion with David Gauthier's theory of deliberation and practical reasoning. Both of these chapters are excursions at least in two senses. Firstly, they tackle specific issues; the problem of temptation and David Gauthier's theory of deliberation and practical reasoning. Secondly, these chapters make the chronology of the book out of sync. That being said, the problems in these chapters are solved by the account of diachronic rationality using the notion of self-governance, so in some way they do fit with the rest of the second part of the book.

Chapter 9, "Consistency and Coherence in Plan" is written somewhat in a form of a dialogue between Bratman and a fictional planning agent named Kate. In this "conversation" Kate is asking two questions: is there any reason for her to be a planning agent and can she sometimes be a planning agent and sometimes not be a planning agent depending on her current preferences and whether it is advantageous for her at that particular time? Bratman's answer to the first question is that we should be planning agents because we should value governing our own lives. In other words, the reason to be planning agent is self-governance. The answer to the second question is that an agent, in this case Kate, cannot actually choose to *be planning agent*. Not all agents are planning agents but those who are cannot simply cease to be planning agents at will because planning agency is embedded in their psychic economy.

In Chapter 10 "Rational Planning Agency" Bratman develops his full-fledged diachronic account of rationality. Building on his notion of diachronic self-governance Bratman argues for diachronic plan rationality which consists of several normative constraints. *Practical rationality/Self-governance-Planning Agency* states, roughly, that it is pro tanto, defeasibly irrational to fail to have a coherent practical plan-infused standpoint or to choose contrary to that standpoint. *Diachronic Plan Rationality* states, roughly, that when the conditions for synchronic and diachronic self-governance are met it is defeasibly, pro tanto irrational to make choices that bock your continued diachronic self-governance. *Rational end of diachronic self-governance* states, roughly, that it is pro tanto, defeasibly irrational for a planning agent, capable of self-governance to fail to have an end of diachronic self-governance. These normative constraints (not exhaustively) constitute Bratman's diachronic account of rationality.

In the last chapter of the book, Chapter 11 "A Planning Agent's Self-governance Over Time" Bratman explores the merger of his two ideas: the diachronic account of self-governance and intentional shared agency. The result is acting "together" with oneself at different times: a shared agency model of diachronic self-governance. In other words, the idea is that an agent "cooperates" with himself from different times in a way that different agents cooperate with each another. The idea is not new *per se* because it goes back to the days of decision theorists and game theorists like McClennen, who argued that an agent is bargaining with himself from different times, but Bratman is revising the idea in a new light using the notion of diachronic self-governance.

Overall the book is very well structured and the argumentation is clear and precise. The chapters follow from one another nicely (with the possible exceptions of Chapters 7 and 8 which I have discussed). The book has two "flaws". The first is that the book is not kind to the readers that are not familiar with contemporary issues in philosophy of action and instrumental rationality. The second is that some chapter focus on specific issues, like temptation, and do little to contribute to the general argumentation presented in the book. The project in the book is quite ambitious. Bratman is presenting a new and fresh way of looking at practical rationality, normative reasons and philosophy of action. Whether his account of *Diachronic Plan Rationality* works or not is for the reader to decide.*

DAVID GRČKI
*University of Rijeka, Rijeka, Croatia*

* The writing of this paper was supported by the Croatian Science Foundation under grant number IP-06-2016-2408.