Croatian Journal of Philosophy

Vol. XXI · No. 62 · 2021

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. XXI · No. 62 · 2021

Articles

Book Review

Croatian Journal of Philosophy
Vol. XXI, No. 62, 2021
https://doi.org/ 10.52685/cjp.21.62.1
Received: October 17, 2020
Accepted: December 15, 2020

Are Capabilities Compatible with Political Liberalism? A Third Way

THOM BROOKS
Durham Law School, Durham, United Kingdom

This article explores the relationship between capabilities and political liberalism. There are two views about how they might be compatible: Sen claims capabilities should be seen as a revision of primary goods while Nussbaum argues capabilities should form part of an overlapping consensus. It is argued they are both right—and incorrect. Whereas Sen identifies where compatibility might best be found, it is Nussbaum's conception of capabilities that is able to overcome Rawls's objections to Sen's proposal. This provides a new third way of conceiving how capabilities and political liberalism might address these concerns that is more compelling for how Sen and Nussbaum claim. The two rivals can come together, but not in the way that either of its most well known champions have argued.

Keywords: Capabilities; overlapping consensus; political liberalism; primary goods.

1. *Introduction*

It is striking that the leading proponents of the capabilities approach—Amartya Sen and Martha Nussbaum—each see their distinctive approach as compatible with John Rawls's political liberalism and not as an alternative to it.[1] Interestingly, both Sen and Nussbaum each see the compatibility of capabilities with political liberalism in very different ways. While Sen argues that his view of capabilities are best

[1] While Amartya Sen speaks of his capability approach and Nussbaum of her capabilities approach, I will use 'capabilities approach' to capture them both unless otherwise noted.

incorporated within Rawls's primary goods, Nussbaum claims that her different view of capabilities are better understood as forming part of an overlapping consensus.

This article considers Rawls's political liberalism and these different views on capabilities claim to locate themselves as a part of political liberalism in very different ways. It is argued that neither of these two models is successful and, instead, a new third option is more promising that lies between them. My claim is that Nussbaum's approach fits better with political liberalism, but only in the place that Sen locates for his own approach. Sen and Nussbaum are correct to highlight the compatibility of the capabilities approach with political liberalism, but not about how or where either find these two compatible. Instead, uniting capabilities and political liberalism is only possible through this new, third way which Rawls and they have overlooked.

The structure of my argument is as follows. I begin by describing Rawls's political liberalism before proceeding to discuss how Sen believed his capability approach is compatible with it. I then explain Nussbaum's arguments about how her approach is compatible with Rawls's political liberalism. The final section critically examines their claims and argues for a new third way bringing together elements from each showing why this different view is more compelling and overcomes objections that Rawls has about the relationship between political liberalism and capabilities.

2. *Political liberalism*

John Rawls (1996: xviii, see 3–4) came to believe that his *A Theory of Justice* suffered from a 'serious problem' concerning political stability. He recognized that citizens are deeply divided by reasonable and potentially incompatible religious, philosophical or moral comprehensive doctrines (Rawls 1996: 13, 168; Rawls 2001a: 193).[2] This 'fact of reasonable pluralism' about the 'reasonable comprehensive doctrines' that citizens have is neither exceptional nor rare, but an inescapable and ineliminable fact about democratic societies (Rawls 1996: 24–25n227, 63–64, 129, 140, 144, 147–48, 172; Rawls 1999: 131; Rawls 2001aa: 3–4, 33–34, 36, 40, 84). We cannot wish our reasonable differences away.

Reasonable pluralism presents a potential threat to political stability because if citizens are to be free and equal—as Rawls presupposes—their reasonable disagreement about preferences for different and opposing comprehensive doctrines can undermine political stability over time. This creates a problem for how political stability may be possible without denying the equality of citizens. Rawls formulates the challenge like this: 'How is it possible that deeply opposed though reasonable comprehensive doctrines may live together and all affirm the

[2] Rawls's political liberalism rejects our taking account of unreasonable comprehensive doctrines. I will use 'reasonable comprehensive doctrines' and 'comprehensive doctrines' interchangeably.

political conception of a constitutional regime? What is the structure and content of a political conception that can gain the support of such an overlapping consensus?' (Rawls 1996: xx)

For Rawls, disagreement about comprehensive doctrines matters. The public can be divided about whether their political community should support policies like capital punishment or criminalizing euthanasia without exception from the comprehensive doctrines they endorse.[3] Rawls's political liberalism respect the equality of citizens by forbidding our prioritizing any one comprehensive doctrine over others. Otherwise, the doctrines held by some citizens would have a privileged political status above the doctrines held by others—and their equality would be respected. This creates the need to find some way to strike this delicate and potentially complex balance.

Rawls argues that it is a fact that different citizens will endorse a range of different reasonable comprehensive doctrines. This fact of reasonable pluralism poses a threat to political stability over time requires a solution—otherwise, political stability over time will be very difficult, if not impossible, to maintain. Rawls argues that citizens must have a way to determine political judgements acceptable to all while treating all reasonable comprehensive doctrines on an equal footing to respect the equality of citizens. He says:

> Thus I believe that a democratic society is not and cannot be a community, where by a community I mean a body of persons united in affirming the same comprehensive, or partially comprehensive, doctrine the fact of reasonable pluralism which characterizes a society with free institutions makes this impossible. (Rawls 2001a: 3)

So disagreement over comprehensive doctrines is not to be addressed by privileging one over others or by aiming to have all citizens affirm doctrine. We must find some other way to navigate reasonable pluralism.

Rawls's solution is to defend the idea of an overlapping consensus (Rawls 2001a: 32–38; Rawls 1996: 133–72; see Freeman 2007: 366–71; Maffetone 2010: 261–74; Scanlon 2003: 159–61). This consensus is a common platform that all citizens can accept no matter which comprehensive doctrine they endorse. The consensus is conceived as a 'political conception of justice' (Rawls 2001a: 20). For Rawls, the threat to political stability posed by reasonable pluralism is 'a problem of political justice, not a problem about the highest good' and so requires a 'political' solution (Rawls 1996: xxvii, see 140).

An overlapping consensus is a shared political conception of justice that citizens can reasonable accept irrespective of which comprehensive doctrine they endorse. Rawls says that 'political liberalism looks for a political conception of justice that we hope can gain the support of an overlapping consensus of reasonable religious, philosophical, and moral doctrines' (Rawls 1996: 10). Citizens can affirm an overlapping

[3] For example, on capital punishment see Brooks 2004; Brooks 2011.

consensus as reasonable without rejecting their acceptance of any reasonable comprehensive doctrine.

They create a consensus through the use of public reasons. These reasons are claims we may reasonably offer to others for mutual acceptance available to every citizen (Rawls 2001b: 208). We can contrast public reasons with non-public reasons (Rawls 2001a: 92; Rawls 1996: 213, 220–22). The latter are reasons that might be acceptable to some, but not all, comprehensive doctrines. For example, providing a reason to endorse or reject a public policy because of the view of an organized religion's doctrine is a non-public reason because its acceptance requires our support for a particular religion's authority on that matter—and so incompatible with other reasonable comprehensive doctrines citizens may endorse (Rawls 1999: 169–70; Rawls 1996: lv–lvii). An overlapping consensus is only possible when all citizens can reasonably accept the public reasons offered to justify a political conception of justice all can support notwithstanding their reasonable differences. Public reasons may not settle all, or almost all, political questions we face. However, for Rawls, they are the only reasons we have to perform this important task (Rawls 2001a: 26–27, 91; Rawls 1996: 163).

An overlapping consensus built on public reasons creates a shared political conception of justice without setting up a new, rival comprehensive view (Rawls 1996: xxix). A consensus is freestanding because its acceptance does not entail any special commitment to any particular doctrine (Rawls 2001a: 33, 37; Rawls 1996: 13). In this way, Rawls claims that an overlapping consensus justifies political stability 'for the right reasons' as it respects the equality of citizens and their reasonable differences over the good (Rawls 1996: xxxix, xli–xliii).

3. *Sen on political liberalism*

Rawls's proposal of political liberalism as a solution to the problem of political stability in light of reasonable pluralism has attracted criticism. Some, such as Kurt Baier (1989), Brian Barry (1995) and George Klosko (1994), have argued that political liberalism's overlapping consensus is unnecessary for securing political stability. They argue that there are resources in Rawls's theory of justice—such as a commitment to two principles of justice—that can forge stability despite disagreements about the good. Others like Kent Greenawalt (1995), Michael Sandel (1994) and Leif Wenar (1995) claim an overlapping consensus is too fragile to secure political stability. They accept that the consensus is an important bridge connecting citizens across their reasonable differences over the good, but skeptical about how strong the ties that bind them together can be. In previous work, I have argued that we can accept both sides: Rawls does have resources in his theory to build solidarity beyond what others have identified, such as reciprocity, that can strengthen the ties that an overlapping consensus offers (Brooks 2012).

One especially interesting line of criticism comes from proponents

of the capabilities approach claiming that Rawls's political liberalism can be improved by incorporating capabilities into his account. How his political liberalism might be compatible with capabilities will be the subject for the rest of our discussion.

Amartya Sen argues Rawls's list of primary goods are best understood in terms of Sen's understanding of capability and so he should revise his list accordingly (Sen 1985: 199–201, see Nussbaum 2006: 141; Nussbaum 2000: 68; Scanlon 2009: 197–99). In contrast, Rawls claims that citizens must be guaranteed primary goods above a social minimum. These primary goods are defined as 'what persons need in their status as free and equal persons, and as normal and fully cooperating members of society over a complete life' (Rawls 1971: xiii). These goods include the following:

A. Basic rights and liberties, also given by a list;
B. Freedom of movement and free choice of occupation against a background of diverse opportunities;
C. Powers and prerogatives of offices and positions of responsibility in the political and economic institutions of the basic structure;
D. Income and wealth; and finally,
E. The social bases of self-respect. (Rawls 1996: 181)

If a state is unable to guarantee a social minimum of these primary goods, then it fails to satisfy the constitutional essentials of a just polity (Rawls 1996: 228–29). This is because our capacity to understand, to apply and to act from—and not merely be in accordance with—the principles of political justice require a social minimum is achieved and makes possible what Rawls calls 'a decent human life' (Rawls 2001a: 18–19, 129).

Sen argues that Rawls should revise this account of primary goods to remove a problematic ambiguity at its centre that would help clarify their role in securing individual freedom and well-being (Sen 1999: 56; Sen 1995: 33; Sen 2008: 24–25; Sen 2009: 238). Sen says:

> Some primary goods (such as "income and wealth") are no more than means to real ends … Other primary goods (such as "the social basis of self-respect" to which Rawls makes an explicit reference) can include aspects of the social climate, even though they are generalized *means* (in the case of "the social basis of self-respect" means to achieving self-respect). Still others (such as "liberties") can be interpreted in different ways: either as a means (liberties permit us to do things that we may value doing) or as the actual freedom to achieve certain results. (Sen 1999: 306–7)

Sen's argument is that Rawls understands primary goods too narrowly—as a means to satisfactory human living, but not its end (Sen 2009: 234). The problem with primary goods, for Sen, is that they fail to capture an important distinction between our '*doing* something' and our 'being *free* to do that thing' (Sen 2009: 234, 237).

Sen argues this problem can be solved by revising Rawls's account of primary goods to become capabilities (Sen 1995: 87; Sen 2009: 64).

Sen claims that this would 'not be a foundational departure from Rawls's own programme, but mainly an adjustment' (Sen 2009: 66). This is because, for Sen, 'basic capabilities can be seen as a natural extension of Rawls's concern with primary goods'—and a more robust account of them (Sen 1980: 218–19; see Sen 1999: 74, 78; Sen 2009: 262). Rawls should have recognized that institutions are not always required to secure primary goods in some cases: understanding them as capabilities would correct this mistake (Sen 2009: 90).

In summary, Sen claims that Rawls's political liberalism can be compatible with capabilities. Sen argues that capabilities provide a more robust—and more consistent—understanding of primary goods within Rawls's theory. If we revise Rawls's account of primary goods, we can fit capabilities into his political liberalism and close gaps and inconsistencies in Rawls's account.

4. *Nussbaum on political liberalism*

Martha Nussbaum also argues that Rawls's political liberalism is compatible with her own approach to capabilities—which is in a different way from Sen's (Nussbaum 2006: chpts 1–3; Nussbaum 2000: 5, 14, 59, 4–75, 105; Nussbaum 2011: 19, 79, 89–93, 182). Where Sen believes his capability approach can be best incorporated as an improved modification of Rawls's primary goods, Nussbaum claims her capabilities approach can be a part of an overlapping consensus—and so her different conception of capabilities is argued to fit best in a different part of political liberalism.

Nussbaum disagrees with Sen about understanding primary goods as capabilities because she believes it could jeopardize the 'desired simplicity' that Rawls aspired to with his theory of justice 'both in indexing relative social positions and in describing the point of social cooperation' (Nussbaum 2006: 142). While she acknowledges that perhaps Rawls's theory could be made more compelling, it would come at a cost—so Sen is incorrect to argue that Rawls should accept this revision even if there is good reason for him to accept it.[4]

Moreover, Nussbaum claims that if Rawls made this revision it 'would require a major overhaul of the theory [of justice], particularly as a theory of economic justice' (Nussbaum 2006: 146). For Nussbaum, individual decisions about conceptions of the good are left by Rawls to citizens whereas the capabilities approach endorses a shared, public conception of justice where the good of others is built into the good of each citizen (Nussbaum 2006: 158). Furthermore, capabilities are not merely instrumental to human dignity, but 'as ways of realizing a life

[4] I find this criticism unconvincing because it is unclear that any rendering of capabilities—such as a list like Nussbaum's—would jeopardize desired simplicity in a list like Rawls's. This criticism may be aimed at a particular understanding of capability, namely, Sen's, but does not clearly concern alternative understandings about capabilities.

with human dignity' (Nussbaum 2006: 161). The right and the good are inseparable and they 'seem thoroughly intertwined' (Nussbaum 2006: 162–63). Capabilities are 'fundamental entitlements of citizens' and all capabilities are 'necessary for a decent and dignified human life': 'If people are below the threshold on any of the capabilities, that is a failure of basic justice, no matter how high they are on all the others' (Nussbaum 2006: 166–67; see Nussbaum 2000: 73; Nussbaum 2011: 36). In short, while Nussbaum recognizes the potential overlap between primary goods and capabilities identified by Sen, she rejects it as it largely leaves Rawls's contractarianism intact whereas capabilities are 'fundamental entitlements' that must be secured.

She argues that capabilities 'can become the object of an overlapping consensus among people who otherwise have very different comprehensive conceptions of the good' (Nussbaum 2006: 70). Nussbaum says:

> The political principles of the capabilities approach are supported by independent arguments about human dignity. We do not try to generate principles out of compassion alone, but, instead, we seek to support them and render them stable through the development of a compassion that is attuned to the political principles for which we have argued. (Nussbaum 2006: 91)

For Nussbaum, capabilities can be part of an overlapping consensus because both enjoy a freestanding justification and she claims capabilities are compatible with any reasonable comprehensive doctrine (Nussbaum 2006: 79, 304–5; Nussbaum 2011: 89–92). While she provides a list of ten capabilities, Nussbaum is also very clear that the list is not 'final' nor set in stone: 'if it turns out to lack something that experience shows to be a crucial element of a life worthy of human dignity, it can always be contested and remade' (Nussbaum 2011: 15).

In summary, Sen claims that Rawls should revise his account of primary goods to conform to Sen's views of capability—Sen claims this would improve Rawls's account and show how capabilities and political liberalism can be made compatible. Nussbaum claims that Rawls should look to capabilities as a part of any overlapping consensus and that this is the best space to bring capabilities and political liberalism together. She argues that capabilities can fulfil the function of an overlapping consensus because any reasonable comprehensive doctrine can connect and support with capabilities. So Sen is correct to claim capabilities and political liberalism are compatible, while Nussbaum claims they are compatible in a different part of Rawls's theory (*e.g.*, an overlapping consensus) than argued for by Sen (*e.g.*, primary goods). They cannot both be correct.

5. *A third way*

I believe Sen and Nussbaum are both correct to argue Rawls's political liberalism is compatible with capabilities, but not in the way that either Sen or Nussbaum claims. Sen is correct that capabilities are best

placed as a more robust modification of primary goods, but Nussbaum's understanding of capabilities are more compatible with Rawls's theory of justice to serve as this modification. In short, Sen identifies where capabilities should be located and Nussbaum provides the better fitting view of capabilities to fulfil this role—but not *vice versa*.

To begin, Rawls considers and rejects Sen's proposed revision, but the reason for this rejection is illuminating. Rawls recognizes the significant overlap across his account of primary goods and Sen's account of capability. The issue is that the latter is broader, but in a problematic way. Rawls notes that 'I hope that now our views are in accord on the topics that concern us here, though his view has more broader aims than mine' (Rawls 1996: 179). Rawls says that 'I agree with Sen that basic capabilities are of first importance and that the use of primary goods is always to be assessed in the light of assumptions about those capabilities' (Rawls 1996: 183). However, Rawls argues:

> In reply, it should be stressed that the account of primary goods does take into account, and does not abstract from, basic capabilities: namely, the capabilities of citizens as free and equal persons in virtue of their two moral powers. It is these powers that enable them to be normal, and fully cooperating members of society over a complete life and to maintain their status as free and equal citizens … These remarks locate the role of primary goods within the framework of justice as fairness as a whole … we see that it does recognize the fundamental relation between primary goods and persons' basic capabilities. In fact, the index of those goods is drawn up by asking what things, given the basic capabilities included in the (normative) conception of citizens as free and equal. (Rawls 2001a: 169–70)

Rawls argues that 'Sen might accept the use of primary goods, at least in many instances': primary goods already incorporate some substantive connection with capabilities that does not require further revision (Rawls 2001a: 170). Primary goods have flexibility in application even if not explicitly open to future revision over time and changing circumstances.

However, the difference is that primary goods are more determinate and easier to apply than capabilities (Rawls 1996: 185). Rawls says: 'A scientific (as opposed to a normative) measure of the full range of these capabilities is impossible as a matter of practice, if not theoretically as well' (Rawls 2001a: 171). For Rawls, concepts, such as 'well-being', are 'not sufficiently determinate' (Rawls 1971: 283; see Cohen 2011: 40–43, 47–48. 50–51). Primary goods are more attractive because they offer an account that speaks to some measure of well-being in a way that is more relevant for application to practices. Primary goods, not capabilities, satisfy the publicity criterion whereby claims of injustice are easily accessible and verifiable by all.

This criticism of Sen's account is much less of a problem for Nussbaum's account of capabilities as it presents a list. So one criticism of the capabilities approach is that it is too imprecise and does not offer 'workable criteria for interpersonal comparisons that can be publicly

and, if possible, easily applied' (Rawls 1996: 186). This objection can be overcome by defending a more determinate account of capabilities—like Nussbaum's—that avoids this problem.

For Rawls, every citizen is guaranteed a social minimum of primary goods. These include (a) basic rights and liberties, (b) freedom of movement and choice of occupation, (c) political and economic freedoms, (d) income and wealth and (e) 'the social bases of self-respect' (Rawls 1996: 181). Thus primary goods represent a package of essential rights and freedoms, opportunities, basic needs and self-respect.

Nussbaum's capabilities approach captures this conception in a more robust alternative form. Her proposed list of ten capabilities includes Life; Bodily Health; Bodily Integrity; Senses, Imagination, Thought; Emotions; Practical Reason; Affiliation; Other Species; Play and Control Over One's Environment.[5] Neither Nussbaum's capabilities approach nor Rawls's social minimum are meant to offer a complete account of social justice—although both claim to provide us with an essential component for any such account of 'minimum core social entitlements' (Rawls 1971: 244–45, 279; Nussbaum 2006: 75).

Capabilities and the social minimum address the same primary goods, but capabilities provides more clarity—or what Nussbaum calls 'a rather ample social minimum' (Nussbaum 2011: 40). While both capture a minimum of basic rights and liberties, only the capabilities approach is explicit in its relationship to human rights and rights more generally (Nussbaum 2006: 78; Nussbaum 2011: 62). They each address freedom of movement and occupational choice, yet capabilities develops greater specification of their importance for human flourishing and related goods, such as affiliation, recreation and some measure of control over political and material environments. Thus capabilities—as understood by Nussbaum—do not merely map onto primary goods, but the former provide an extended view of the latter. Nussbaum's list of capabilities is a better fit with Rawls's list of primary goods than Sen's capability approach which eschews such lists and does not focus on the need to satisfy a threshold minimum like Nussbaum's capabilities approach and Rawls's social minimum of primary goods does.

The second and more crucial reason why Rawls rejects Sens's proposed revision of the primary goods as capabilities is because Rawls understood capabilities as a kind of comprehensive doctrine. Rawls argues that political liberalism 'presupposes no particular comprehensive view, and hence may be supported by an enduring overlapping consensus of reasonable doctrines' (Rawls 2001a: 37). Rawls believes that primary goods have a more limited nature than capabilities. These goods are understood within a political conception of justice that address the needs of citizens and 'not anyone's idea of the basic values of human life and must not be so understood' (Rawls 1996: 188). Rawls says: 'Justice as fairness rejects the idea of comparing and maximizing overall

[5] On Nussbaum's list, see Brooks 2020.

well-being in matters of political justice': primary goods should not be understood in terms of 'anyone's idea of the basic values of human life … however essential their possession' (Rawls 1996: 188). Political liberalism would then best respect the fact of reasonable pluralism and endorse a political conception 'that is mutually acceptable to citizens generally' (Rawls 1996: 188).

The problem with this objection is that it assumes without argument that the capabilities approach is an overly substantive view about the good that a person might reasonably reject. The primary goods are compatible with any reasonable comprehensive doctrine. So Rawls cannot be opposed to any endorsement of goods for fear that they might be incompatible with reasonable pluralism without denying primary goods altogether. Rawls is clear that primary goods overlap to some substantial degree with capabilities, such as the need to secure the guarantee of moral powers for each individual. Rawls appears to claim that primary goods are different from capabilities because they provide a sufficiently 'thin' conception of the good endorsable by all reasonable persons. Primary goods are a sufficiently thin conception because they are compatible with any reasonable comprehensive doctrine. But what Rawls needs to argue is not that primary goods are sufficiently thin, but rather that capabilities are too 'thick': the issue is then not whether capabilities are more robust than primary goods, but incompatible with any reasonable doctrine. If they are not, then his objection fails—and it does fail. The capabilities approach is not a fully comprehensive doctrine as even its leading critics accept (Pogge 2010: 19–20).

Rawls's concern is directed towards a specific understanding of capability, namely, Sen's approach. This is a more 'thick' conception than Nussbaum's insofar as only the latter has a particular focus on satisfying a threshold in a manner not dissimilar to how Rawls employs primary goods and their social minimum. This minimum is potentially compatible for all in a way that a view of capability without specified thresholds does not. In this way, Nussbaum's list makes a better fit with Rawls's list without importing a full comprehensive doctrine—a risk that Rawls thought possible with Sen's capability approach.

However, Nussbaum did not accept Sen's view that capabilities are best incorporated into political liberalism as a revised view of primary goods, but instead as part of an overlapping consensus, she acknowledges the close connection between her list of capabilities and Rawls's primary goods and she says her list 'could figure as an account of primary goods' (Nussbaum 2006: 116, see Nussbaum 2000: 5, 74–75). But she could also have noted a further substantive connection between capabilities and primary goods in that both are understood in terms of threshold satisfaction: what matters for Nussbaum is that opportunities to exercise capabilities above a threshold can obtain and what matters for Rawls is primary goods can be enjoyed above a social minimum. As Nussbaum notes, 'the notion of a threshold is more important in my account than the notion of full capability equality' (Nussbaum 2000:

12). We look to ensuring we are all above a threshold as a fundamental concern of justice for every individual without trade-offs between capabilities—and likewise between primary goods. For these reasons, Nussbaum's account of capabilities seems the better fit despite her reservations about revising Rawls's primary goods in terms of capabilities.

Nussbaum's argument for understanding capabilities as part of an overlapping consensus is problematic—and because of the specific content she builds into her capabilities. For example, the capability of Bodily Integrity includes a right to 'choice in matters of reproduction' (Nussbaum 1999: 41). If her capabilities approach is to be a part of an overlapping consensus, then it must be acceptable to any reasonable comprehensive doctrine—and these doctrines include all major world faiths (Rawls 1996: 59). The problem here is not that Bodily Integrity is a capability, but that Bodily Integrity is given with the specific content of providing for a right to reproductive choice. This is because not all major world religions—Roman Catholicism as only one of several examples—would accept this right upfront. It is possible through the use of public reasons to make the argument for reproductive choice in a way that Catholics could access—perhaps even on grounds of Bodily Integrity—even if most, it not all, did not find these reasons compelling. The possibility of Catholicism does not rule out the community's acceptance of reproductive rights, but any acceptance will need to be achieved through public reasons and not assumed—or given—through a foundation built on an overlapping consensus with parts *prima facie* objectionable to Catholics. Nussbaum's problem is easily avoided by not so narrowing construing this capability in a way that cannot be accepted by every reasonable comprehensive doctrine.

There are also more controversial issues regarding other capabilities. Another capability on Nussbaum's list is Play which provides individuals with a minimum right 'to enjoy recreational activities' (Nussbaum 1999: 41–42). All citizens are guaranteed at least a minimum to ensure each has a minimally decent life. However, not all major world religions accept leisure as a good—including the Puritans who helped found America.[6] They may be wrong about the importance of play, but this is to be weighed up—in Rawls's political liberalism—through the interplay of public reason and not a given upfront.

In summary, Nussbaum's capabilities are provided with content that may clash with central tenets of the major world religions included in Rawls's list of reasonable comprehensive doctrines. It is clear that each can have their objections to policies over reproductive choice or leisure activities through the interplay of public reason. However, the content of Nussbaum's capabilities seems to close off this conversation before it begins. Perhaps if left more 'thin' and specified more minimally there would be less concern. As stated, the content of Nussbaum's capabilities does not make obvious their being automatically a part

[6] I am grateful to Derek Matravers for raising this objection to me.

of any reasonable comprehensive doctrine. So while the list provides a more robust list of primary goods that we might accept in an original position, her list seems too thick to be accepted as an overlapping consensus by individuals from any comprehensive doctrine. Nussbaum is correct that capabilities can be compatible with Rawls's political liberalism and she offers an account that can achieve this result, but not where she thought (*e.g.*, overlapping consensus) but instead somewhere else (*e.g.*, primary goods).

6. *Conclusion*

This article has explored the relationship between capabilities and political liberalism. Both Amartya Sen and Martha Nussbaum argue that their different views on capabilities are compatible with political liberalism in different places. Sen claims capabilities should be seen as a revision of primary goods while Nussbaum argues capabilities should form part of an overlapping consensus.

I have argued that they are both right—and incorrect. Sen correctly identifies where capabilities and political liberalism are most compatible. While Rawls raises objections specifically about how Sen's capability approach is too 'thick' and closer to a comprehensive doctrine, these objections can be met by Nussbaum's capabilities approach which is more 'thin', uses a list, focuses on meeting a minimum threshold and so a more ready fit with Rawls's list of primary goods. So Sen is correct about where to find compatibility, but it is Nussbaum who has the capabilities approach that is the most compatible at that specific point.

Nussbaum argues compatibility between capabilities and political liberalism is best found in an overlapping consensus. However, I have argued that the content of her capabilities is problematic as it includes a core that clashes with the reasonable comprehensive doctrines that should be able to accept it. This is not to say that Nussbaum is mistaken to argue for a right to reproductive choice or for leisure. But it is to say that, on Rawls's account, such content is problematic for prioritizing some doctrines over others. While Nussbaum is correct that capabilities and political liberalism are compatible, I do not agree this is found in the location she identifies.

This article has not considered the merits of bringing capabilities and political liberalism, such as how political stability might be better secured over time as I have argued elsewhere (Brooks 2015). It is clear that capabilities and political liberalism need not be viewed as rivals, but can be brought together except only not how or where Sen and Nussbaum thought– using Nussbaum's list in where Sen identified compatibility in Rawls's primary goods. Thus we need to look for a third way fusing the two to render this compatibility possible.[7]

[7] Many thanks to Maria Dimova-Cookson, Peter Jones, Pauline Kleingeld, Martha Nussbaum, Avital Simhony and Martin van Rees for constructive comments on earlier drafts.

References

Baier, K. 1989. "Justice and the Aims of Political Philosophy." *Ethics* 99: 771–90.

Barry, B. 1995. "John Rawls and the Search for Stability." *Ethics* 105: 874–915.

Brooks, T. 2004. "Retributivist Arguments against Capital Punishment." *Journal of Social Philosophy* 35: 188–97.

Brooks, T. 2011. "Retribution and Capital Punishment." In M. D. White (ed.). *Retributivism: Essays on Theory and Practice*. Oxford: Oxford University Press: 232– 45.

Brooks, T. 2012. "Reciprocity as Mutual Recognition." *The Good Society* 21: 21– 35.

Brooks, T. 2015. "The Capabilities Approach and Political Liberalism." In T. Brooks and M. C. Nussbaum (eds.). *Rawls's Political Liberalism*. New York: Columbia University Press: 139–73.

Brooks, T. 2020. "Capabilities, Freedom and Severe Poverty." In T. Brooks (ed.). *The Oxford Handbook of Global Justice*. Oxford: Oxford University Press:199–213.

Cohen, G. A. 2011. *On the Currency of Egalitarian Justice and Other Essays in Political Philosophy*. Princeton: Princeton University Press.

Freeman, S. 2007. *Rawls*. London: Routledge.

Greenawalt, K. 1995. "Some Problems with Public Reason in John Rawls's Political Liberalism." *Loyola of Los Angeles Law Review* 28: 1303–17.

Klosko, G. 1994. "Rawls's Argument from Political Stability." *Columbia Law Review* 94: 1882–97.

Maffetone, S. 2010. *Rawls: An Introduction*. Cambridge: Polity.

Nussbaum, M. C. 1999. *Sex and Social Justice*. Oxford: Oxford University Press.

Nussbaum, M. C. 2000. *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

Nussbaum, M. C. 2006. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge: Belknap/Harvard University Press.

Nussbaum, M. C. 2011. *Creating Capabilities: The Human Development Approach*. Cambridge: Belknap/Harvard University Press.

Pogge, T. 2010. "A Critique of the Capabilities Approach." In H. Brighouse and I. Robeyns (eds), *Measuring Justice: Primary Goods and Capabilities*. Cambridge: Cambridge University Press: 17–60.

Rawls, J. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.

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

Rawls, J. 1999. *The Law of Peoples*. Cambridge: Harvard University Press.

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

Rawls, J. 2001b. "Justice as Fairness." In *Collected Papers*. Cambridge: Harvard University Press: 47–72.

Sandel, M. 1994. "Political Liberalism." *Harvard Law Review* 107: 1765–94.

Scanlon, T. M. 2003. "Rawls on Justification." In S. Freeman (ed.). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press: 139–67.

Scanlon, T. M. 2009. "Value, Desire and Quality of Life." In M. C. Nussbaum and A. Sen (eds.). *The Quality of Life*. Oxford: Oxford University Press: 185–200.

Sen, A. 1980. *Equality of What?* Cambridge: Cambridge University Press.

Sen, A. 1985. "Well-being, Agency and Freedom: The Dewey Lectures 1984." *Journal of Philosophy* LXXXII: 169–221.

Sen, A. 1995. *Inequality Reexamined*. Oxford: Oxford University Press.

Sen, A. 1999. *Development as Freedom*. Oxford: Oxford University Press.

Sen, A. 2008. "The Economies of Happiness and Capability." In L. Bruni, F. Comim and M. Pugno (eds). *Capabilities and Happiness*. Oxford: Oxford University Press: 16–27.

Sen, A. 2009. *The Idea of Justice*. London: Allen Lane.

Wenar, L. 1995. "*Political Liberalism*: An Internal Critique." *Ethics* 106: 32–62

Croatian Journal of Philosophy
Vol. XXI, No. 62, 2021
https://doi.org/ 10.52685/cjp.21.62.2
Received: August 19, 2020
Accepted: December 15, 2020

J. S. Mill on Higher Pleasures and Modes of Existence

TIM BEAUMONT
School of Foreign Languages, Shenzhen University, Shenzhen, PR China

The passage of Mill's Utilitarianism that sets out the condition in which one pleasure has a superior quality than another stokes interpretive controversy. According to the Lexical Interpretation, Mill takes one pleasure, P_1, to be of a superior quality than another, P_2, if, and only if, the smallest quantity of P_1 is more valuable than any finite quantity of P_2. This paper argues that, while the Lexical Interpretation may be supported with supplementary evidence, the passage itself does not rule out qualitative superiority without lexical dominance, as it only requires P_1 to be more valuable than any quantity of P_2 that it is possible for someone to experience. Some will object that this concession to opponents of the Lexical Interpretation still renders Mill's condition for qualitative superiority too demanding to be plausible. However, if Mill's qualitative rankings apply to higher-order pleasures taken in modes of existence as such rather than to the pleasures of different activities chosen from within these modes, the objection loses much of its force. One upshot is that Mill may have more to contribute to debates in contemporary population axiology than is usually acknowledged.

Keywords: John Stuart Mill; pleasure; qualitative hedonism; repugnant conclusion; population axiology; utilitarianism.

1. *Introduction*

Mill's *Utilitarianism* (1861) is generally taken to defend a hedonistic axiology according to which all and only pleasurable and painful *experiences* are bearers of final (or non-instrumental) value and disvalue respectively (X: 209; Beaumont 2018a, 2019).[1] It is also usually interpreted as rejecting a form of *quantitative* hedonism that takes the final value of a pleasure to be proportional to its "quantity", where this quantitative value is constituted by the product of two sub-values: intensity

[1] All references to Mill are to volume and page number of his (1963–1991).

and duration (X: 213, 236; Crisp 1997: 32; Wilson 1990: 277–8). This is replaced with a form of *qualitative* hedonism that takes a pleasure's final value to be a function of its "quality" as well as its "quantity" (X: 211). Since Mill proceeds as if he accepts quantitative hedonism's conception of quantitative value, his qualitative hedonism is then taken to imply that the final value of pleasurable experiences is a function of their quality *as well as* their intensity and duration (Donner 1991: 41). On this view, the final (dis)value of a life is an aggregative function of the final value of its pleasures and the final disvalue of its pains, and the final (dis)value of a set (or population) of lives is an aggregative function of the final (dis)value of each individual life it contains. This paper treats the preceding interpretive framework as a working assumption and as its point of inception.

One interpretive controversy that arises therefrom concerns the veracity of the Lexical Interpretation of Mill's qualitative hedonism, which takes it to imply that one pleasure, P_1, is of a superior quality than another, P_2, if, and only if, the smallest quantity of P_1 bears more final value than any finite quantity of P_2 (proponents include: Brink 1997: 153; Crisp 1997: 40; and Riley 2003: 418). In contemporary population axiology, Derek Parfit also seems to endorse the Lexical Interpretation, and to intimate that it allows Mill to evaluate the aggregated pleasures of distinct populations in such a way as to avoid the:

> "*Repugnant Conclusion*": ceteris paribus, given two possible substantial populations, *A* and *B*, of a numerical size n_A and n_B respectively, such that the members of *A* enjoy the highest quality of life possible but those of *B* a quality of life that merely crosses the minimal threshold to make it worth living, the existence of *B* may still be of greater final value provided that n_B is sufficiently larger than n_A for the aggregate welfare of *B*'s population to outweigh that of *A*. (2004: 9–11, 17–20; 1984: 413–4)[2]

However, while many—but by no means all (Arrhenius, Ryberg, Tännsjö 2017: 2.8)—would see this as an advantage, it has not drawn many population axiologists to Mill's position (including Parfit himself, although see his 2016). Aside from an aversion to hedonism as such, the main reason for this may lie in the thought that the (supposedly) lexical character of Mill's qualitative hedonism could only generate this advantage at the cost of generating what we can refer to as the:

> *Extreme Conclusion*: ceteris paribus, given one possible *population*, *C*, of a numerical size, n_C, whose members enjoy a good quality of life, and one possible *person*, *d*, who enjoys a quality of life that is of a marginally superior quality, the existence of *d* would be of greater final value than the existence of *C* regardless of how much greater n_C is than 1 (see also Arrhenius, Ryberg, Tännsjö 2017: 2.1.1, 2.2).[3]

[2] Unlike Parfit's exposition, this one builds the final explanatory ('provided') clause into the conclusion itself (for the sake of convenience). It should also be noted that this reasoning assumes that the hedonist will cash out the notion of 'quality of life' in terms of the final (hedonic) value thereof.

[3] Some may also worry about the way in which a lexical qualitative hedonism will evaluate pains, but an examination of such concerns must be left for another study.

Nevertheless, this Extreme Conclusion objection at the level of *inter-personal* evaluation—pertaining to the comparison of the final value of the lives of *distinct* people or beings (including sets thereof)—has been underexplored by Mill scholars as such (although see Miller 2010: 58).[4]

The main reason for this seems to be that most Mill scholars think that if the Lexical Interpretation were true, his qualitative hedonism would be falsified by an analogously 'extreme' implication at the antecedent *intra-personal* level of evaluation—pertaining to the comparison of the final value of different ways in which *one* person (or being) could live his or life—namely, that it can never be optimal for an individual to sacrifice hedonic quality for hedonic quantity (Anderson 1991: 9; Crisp 1997: 41; Hauskeller 2011: 433–5; Schaupp 2013: 275–6).[5] For example, many will intuit that, ceteris paribus, while Life 1 below is hedonically preferable to Life 2, Life 3 is hedonically preferable to either:

Life 1: with an abundance of 'higher quality' pleasure in books but no 'lower quality' pleasure in dessert.

Life 2: with no 'higher quality' pleasure in books and an abundance of 'lower quality' pleasure in dessert.

Life 3: with slightly less of the 'higher quality' pleasure in books than in Life 1 and as much of the 'lower quality' pleasure in dessert as Life 2 (adapted from Anderson 1991: 9).

Supposing the Lexical Interpretation were true, then, one might take this to show that, while Mill's distinction between quality and quantity is intuitive in terms of helping to explain why Life 1 is preferable to the comparatively 'repugnant' Life 2,[6] the requirement that quality lexically dominate quantity is too strong—or 'extreme'—to be plausible.[7] Moreover, with this conclusion at hand, it has been claimed that the

[4] Those who doubt the commensurability of Mill's position with those of more recent philosophers may wish to consult Beaumont (2018b).

[5] 'Antecedent' because this is the level upon which Mill focuses when he offers the condition for qualitative superiority (X: 211, see below).

[6] For the distinction between 'repugnance' at the inter-personal and intra-personal levels, see Arrhenius & Rabinowicz (2015: 226–8) and Parfit (2016: 119).

[7] Anderson also concludes that Mill's appeal to 'quality' is an appeal to the value of something other than pleasure, and thus illicit given his official commitment to hedonism (1991: 10). Non-hedonists may be correct to think it is necessary to appeal to non-hedonic *value-bearers* to explain the comparative value of such lives (Schaupp 2013: 267; Skorupksi 1989: 299–303). However, this paper assumes that Mill is a consistent hedonist (see Beaumont 2019). Moreover, since it is designed to explore Mill's *qualitative* hedonism rather than question hedonism as such, it will assume that when two lives differ in final value, Mill is right to think it is in virtue of differences in the value of their pleasures. This much is compatible with the pleasures being valuable *because* they are pleasant (*full* hedonism) or because they are *made* valuable by non-hedonic *value-makers* (*partial* hedonism) (Crisp 1997: 26). The full hedonist view is also compatible with *internalist* or *externalist* conceptions of pleasure (Miller 2010: 35). This paper remains agnostic regarding these other debates concerning the character of Mill's hedonism (but see Beaumont (2019) for further discussion).

Lexical Interpretation should be rejected on grounds of interpretive charity (Saunders 2016: 504; West 2004: 63–4).

This paper will re-examine the key passage of *Utilitarianism* in which Mill explains when, and why, one pleasure has a superior quality than another, and provide a technical taxonomy of some of the interpretive moves that must be made to defend the Lexical Interpretation (Section 2). It also defends two of these moves by arguing that the Lexical Interpretation is correct to interpret Mill as (1) offering a necessary, as well as a sufficient, condition for qualitative superiority (Section 3); and (2) taking the qualitative superiority of P_1 to P_2 to imply that the smallest quantity of P_1 bears more final value than any finite quantity of P_2 that a competent judge's "nature is capable of" (X: 211) (Section 4). However, without denying the truth of the Lexical Interpretation, the paper notes that it cannot rest on the passage in question alone, as any finite quantity of which a competent judge's *nature is capable* is less than *any* finite quantity as such.[8] In consequence, the strongest conclusion that the passage supports in isolation is that the smallest quantity of P_1 must bear more final value than any (finite) quantity of P_2 that could be experienced in a *single life* (Lifetime Interpretation).

Of course, given the way that the intra-personal version of the Extreme Conclusion objection is framed above, in terms of the choice between Lives 1–3, Mill would remain vulnerable to it given the Lifetime Interpretation. However, Section 5 of the paper argues that Mill's illustration of qualitative superiority with reference to different capacity-based "modes of existence" (X: 213) indicates that the objection may be ill-framed. Since modes of existence contain sets of pleasures (and pains), such as pleasures taken in books and dessert, pleasures taken in these sets of pleasures can be viewed as higher-order pleasures (or enjoyment) taken in the mode of existence as such.[9] In consequence, if Mill's qualitative rankings are taken to apply at the higher-order level (Inter-Modal Interpretation), the supposed optimality of the trade-off between pleasure in books and dessert (as previously framed in terms of the choice of Life 3 over Life 1) would not challenge his position.

Given the Inter-Modal Interpretation, this optimality could be reframed in terms of an (intra-modal) preference for greater variety in the lower-*order* pleasures one experiences *within* a given mode of existence, rather than an occasional preference for lower over higher *quality*

[8] As highlighted by the possibility of inter-personal aggregation.

[9] West (2004: 64, 67, 69) also distinguishes between "first-order" and "second-order" pleasures, on the one hand, and links the latter to Mill's "modes" or "manner[s] of existence" (X: 211, 213), on the other. While his use of these terms is perfectly legitimate, it is slightly different to the higher- and lower-order distinction in this paper. For example, whereas he seems to use "second-order pleasures" to refer to *momentary* pleasures taken in one's "self-image" as one enjoys a first-order pleasure in conformity therewith, this paper takes higher-order pleasures to consist of the *extended* pleasure or enjoyment of sets of momentary lower-order pleasures over time.

pleasures as such. Moreover, this much is compatible with it remaining plausible that the value of the smallest quantity of the higher-order pleasure taken in a mode of existence incorporating reading outweighs that of the maximal quantity of pleasure that could be taken in a life devoid of the capacity for reading but replete with pleasure in dessert. In consequence, when combined with the Inter-Modal Interpretation, the Lifetime Interpretation (if not also the Lexical Interpretation) renders it plausible—as opposed to absurdly 'extreme'—to think that relations of qualitative superiority could emerge between the higher-order pleasures taken in those two modes of existence. Moreover, since the Lifetime Interpretation of Mill's qualitative hedonism does not automatically require Mill to embrace the Repugnant Conclusion, it raises the possibility that he found a way to navigate between 'repugnance' and 'extremity' at the inter-personal level that has yet to be fully explored.

2. *Mill on quality and quantity*

Mill sets out the condition for the qualitative superiority of one pleasure over another as follows:

> [A] If I am asked, what I mean by difference of quality in pleasures, or what makes one pleasure more valuable than another, merely as pleasure, except its being greater in amount, there is but one possible answer. [B] Of two pleasures if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure. [C] If one of the two is, by those who are competently acquainted with both, placed so far above the other that they prefer it, even though knowing it to be attended with a greater amount of discontent, and would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a superiority in quality, so far outweighing quantity as to render it, in comparison, of small account. (X: 211)

The relationship between Passage A, on the one hand, and Passages B and C, on the other, is confusing for several reasons. Firstly, while Passage A says that "there is but one possible answer" to its question, Passages B and C provide what look like different, and potentially conflicting, criteria. According to the interpretation that can be referred to as B-Sufficiency, Mill's one answer is given by Passage B, thus implying that it provides a sufficient condition for qualitative superiority. It then takes the more demanding condition constituted by the combination of Passage B and C to be sufficient for one pleasure to have *much* greater quality than another (Miller 2010: 58; Saunders 2011: 188–90; Schaupp 2013: 268; Schmidt-Petri 2003: 102–4). In contrast, according to the interpretation that can be referred to as C-Sufficiency, Passage B provides a necessary but insufficient condition for qualitative superiority, and thus Passages B *and* C provide the sufficient condition *jointly* (Anderson 1991: 9; Crisp 1997: 29; Hauskeller 2013: 433; Riley 2003: 410; Riley 2008: 63; Sturgeon 2010: 1711).[10]

[10] While this interpretation is labelled as 'C-Sufficiency' for short, it is important to emphasize that it takes Mill's condition to incorporate the material from Passage

A further interpretive puzzle pertaining to Passages B and C is whether Mill's sufficient condition for qualitative superiority is also a necessary condition thereof. The interpretations that can be referred to as B-Necessity and C-Necessity respectively both answer in the affirmative, with the former appending this claim to B-Sufficiency (Miller 2010: 58) and the latter to C-Sufficiency (Anderson 1991: 9; Crisp 1997: 29; Riley 2003: 410; Riley 2008: 63).[11] In contrast, the interpretation that can be referred to as Anti-Necessity rejects both B-Necessity and C-Necessity on the grounds that Passage B or C respectively would only have committed Mill to the corresponding positions if he had rephrased them in a biconditional form (Schmidt-Petri 2003: 102–4; 2006: 166).[12] The next two sections of the paper argue that Anti-Necessity can be set aside (Section 3), and that C-Sufficiency and Necessity should be endorsed instead (Section 4).

3. *An argument against anti-necessity*

One argument for C-Sufficiency is that, since Passage B only purports to provide a criterion for judging which of two pleasures, P_1 and P_2, is the "more desirable," and all agree that if P_1 and P_2 have the same qualitative value, P_1 could be more desirable than P_2 in virtue of superior quantitative value *alone*, the passage cannot provide a sufficient condition for qualitative superiority (Riley 2003: 412). Thus, the argument implies, to generate a sufficient condition Passage B must be supplemented with Passage C.

One reply on behalf of B-Sufficiency is that when Mill uses the clause "except its being greater in amount" in Passage A, following "what makes one pleasure [P_1] more valuable than another [P_2]", he is not simply asserting that the criteria for qualitative difference must be distinct from the criteria for quantitative difference, but also indicating that the criteria given in Passage B should be read as presupposing that the quantities of P_1 and P_2 are already held fixed as equal (Miller 2010: 57–8). In consequence, the response maintains, once this implicit presupposition is read into Passage B, the greater value of P_1 as pleasure can *only* be explained by its superior quality, and hence Passage B *can* provide a sufficient condition after all (cf. Riley 2008: 62–3).

In the discussion that follows it will be important to distinguish between Mill's *truth* and *justification* conditions, which is to say the conditions in which he takes it to be the case that P_1 *is* of a superior quality to P_2, on the one hand, and the conditions in which he deems it justifiable to *believe* that this is so, on the other. The preceding defence

B as well, as the latter includes a "moral obligation" clause that is not explicitly restated in Passage C (see footnote 13).

[11] Miller seems to endorse B-Necessity when he says of the condition in Passage B that "its satisfaction is all that is required" (2010: 58).

[12] Schmidt-Petri only directs this objection at C-Necessity. However, ceteris paribus, its validity would undermine B-Necessity as well.

of B-Sufficiency suggests that the best way to do this is to cash it out in terms of its attribution to Mill of the following claims:

> *Ontological B-Sufficiency*: if a quantity, *x*, of P_1 is more valuable as pleasure than the same quantity, *x*, of P_2, P_1 *is* of a superior quality to P_2.
> *Epistemic B-Sufficiency*: if all or almost all competent judges prefer a quantity, *x*, of P_1 as pleasure to the same quantity, *x*, of P_2, the belief that P_1 is of a superior quality to P_2 is justifiable.[13]

This also allows B-Necessity to be cashed out in terms of the attribution to Mill of the corresponding pair of claims, Ontological B-Necessity and Epistemic B-Necessity, consisting of the preceding claims but with the direction of the conditionals reversed (see below).

In contrast, in the case of C-Sufficiency the corresponding theses would be as follows:

> *Ontological C-Sufficiency*: if any quantity of P_1, however small, is more valuable as pleasure than any quantity of P_2 "which their nature is capable of" (Passage C), P_1 *is* of a superior quality as pleasure to P_2.
> *Epistemic C-Sufficiency*: if all or almost all competent judges prefer any quantity of P_1, however small, as pleasure to any quantity of P_2 "which their nature is capable of", the belief that P_1 is of a superior quality to P_2 as pleasure is justifiable.

As before, this allows C-Necessity to be cashed out in terms of the attribution to Mill of the corresponding pair of claims, Ontological C-Necessity and Epistemic C-Necessity, consisting of the preceding claims but with the direction of the conditionals reversed (cf. Riley 2003: 418; Schmidt-Petri 2006: 166).

When formalized thus, it is possible to reduce, say, the epistemic dimension of the debate between B-Sufficiency and C-Sufficiency to the question of whether Mill would endorse Epistemic B-Sufficiency.[14] Likewise, the debate over Anti-Necessity reduces to the question of whether it would be legitimate to: (1) infer Mill's commitment to Ontological and Epistemic B-Necessity from his commitment to Ontological and Epistemic B-Sufficiency respectively; and (2) infer Mill's commitment to Ontological and Epistemic C-Necessity from his commitment to Ontological and Epistemic C-Sufficiency respectively. In the remainder of this section, each of these inferences will be examined in turn.

Firstly, given the assumption that Mill is a *qualitative* hedonist, it should be uncontroversial that if one adopts Ontological B-Sufficiency, one must also adopt:

[13] These formulations oversimplify somewhat by abstracting from the "moral obligation" clause (Passage B) and the "discontent" clause (Passage C) (Anderson 1991: 9), the examination of which are left to another study. They also employ the "all or almost all" competent judges of (X: 211) rather than the mere "majority" thereof of (X: 213), but not *much* hangs on this for the argument that follows (see footnote 16).

[14] Note that the advocate of B-Sufficiency will take Mill to endorse *Epistemic* C-Sufficiency, whilst maintaining that C-Sufficiency *as such* is mistaken for taking Mill to reject *Epistemic* B-Sufficiency.

> *Ontological B-Necessity*: if P_1 *is* of a superior quality to P_2, a quantity, x, of P_1 is more valuable as pleasure than the same quantity, x, of P_2.

After all, given a commitment to Ontological B-Sufficiency, and its attendant assumption that it is possible to compare and commensurate the hedonic quantities of pleasures of different hedonic qualities, Mill could only reject Ontological B-Necessity by denying that the qualities of pleasures affect their final value. However, that would be to abandon *qualitative* hedonism in favour of *quantitative* hedonism by implying that if P_1 and P_2 are equal in quantity, they are of equal final value.

Secondly, the *epistemic* transitions from endorsing either Epistemic B-Sufficiency, on the one hand, or Epistemic C-Sufficiency, on the other, to endorsing the following theses respectively should also be straightforward:

> *Epistemic B-Necessity*: if the belief that P_1 is of a superior quality to P_2 is justifiable, all or almost all competent judges prefer a quantity, x, of P_1 as pleasure to the same quantity, x, of P_2.
> *Epistemic C-Necessity*: if the belief that P_1 is of a superior quality to P_2 as pleasure is justifiable, all or almost all competent judges prefer any quantity of P_1, however small, as pleasure to any quantity of P_2 "which their nature is capable of" (Passage C).

Of course, in each case the obvious objection will be that Mill could take an individual's hedonic beliefs to be *justifiable* when these beliefs defer to the preferences of all or almost all competent judges, without taking the beliefs to be *unjustifiable* insofar as they defy that verdict.[15] For instance, one might think that Mill posits another mode of justification besides epistemic deference to the verdict of the competent supermajority, that is compatible with epistemic defiance thereof. However, shortly after Passage A–C, Mill rules out this possibility—by ruling out a justification for such epistemic defiance—when he declares that "[f]rom this verdict of the only competent judges, I apprehend there can be no appeal" (X: 213).[16]

This leaves the final controversy to be considered, over whether a justification for taking Passage A–C to support the attribution of Ontological C-Sufficiency to Mill would justify attributing the following to him as well:

> *Ontological C-Necessity*: if P_1 *is* of a superior quality as pleasure to P_2, any quantity of P_1, however small, is more valuable as pleasure than any quantity of P_2 "which their nature is capable of" (Passage C),

[15] Such deference would not entail altering one's preference for P_1 over P_2 when almost all competent judges prefer P_2 to P_1, but rather refraining from taking one's preference to *justify* a belief that P_1 is qualitatively superior to P_2.

[16] A possible objection is that formulating the epistemic theses in terms of a "majority" of competent judges (see footnote 13) would have invalidated these transitions when: (1) the judges' preferences generate a tied qualitative verdict vis-à-vis two pleasures; and (2) it is nevertheless justifiable to believe that one is qualitatively superior. However, Mill gives no indication that (2) could be true given (1), even though it would not require an illicit counter-majoritarian appeal. Moreover, even if sound, the objection would not invalidate the transitions in standard cases in which there is a majority verdict.

Of course, it would be "a mistake in propositional logic" (Schmidt-Petri: 2006: 170) to take the former to entail the latter directly.[17] However, the validity of the inference does not depend upon Mill's introduction of an *explicit* biconditional in Passages B & C as it is provided *implicitly* by Passage A's claim that Passage B or Passages B & C provide(s) *the* "*one* possible answer" (emphasis added) to the question of when differences of quality obtain. After all, as Mill acknowledges while discussing causation in the *System of Logic*, if Q is the *only* condition that will suffice for R, Q is also necessary for R, and hence the satisfaction of the *only* sufficient condition for R can be inferred from R itself (VII: 438).[18] In other words, given Passage A, the sufficient condition for qualitative superiority provided by Passage B or Passage B & C must also be a necessary condition thereof, and the *Logic* shows that Mill recognized the validity of this kind of inference. In consequence, the grounds of the debate shift to which of B-Sufficiency or C-Sufficiency should be endorsed in the first place.

4. *An argument for c-necessity*

An advocate of B-Sufficiency (and hence, given the argument of Section 3, B-Necessity) might object to C-Sufficiency (and hence C-Necessity) on the grounds that if Mill had taken his "*one* possible answer" (emphasis added) to be given by Passages B & C *jointly*, he would have combined them into a *single* clause. However, this objection overlooks Mill's criticisms of Bentham's "intricate and involved style", which Mill claims to have rendered Bentham's "later writings books for the student only, not the general reader":

> He [Bentham] could not bear, for the sake of clearness and the reader's ease, to say, as ordinary men are content to do, a little more than the truth in one sentence, and correct it in the next. (X: 114–5)

This passage is also grist to the mill of C-Sufficiency as it suggests that Mill could also view Passage B as a "sentence" containing "a little more than the truth," and thus requiring correction "in the next."[19]

That Passage C should be read this way is also suggested by a diary entry from 1854, in which Mill claims that:

> [C*] Quality as well as quantity of happiness is to be considered; less of a higher kind is preferable to more of a lower. [B*] The test of quality is the preference given by those who are acquainted with both. (XXVII: 663)

Discussions of this passage tend to focus on whether it supports the Lexical Interpretation directly, by implying that it can never be opti-

[17] The following argument would also ground the three preceding inferences.

[18] That is, Mill notes that the *causal* case in which Q is the "only possible cause" of R is also a *logical* case of Q if, and only if, R (VII: 438). Similarly, for Mill, seeing α is not only sufficient for proof of α's visibility, but also necessary qua constituting the *only* form of proof possible (X: 234).

[19] Note that in *On Liberty* Mill also formulates his "one very simple principle" cumulatively across two sentences, with an essential 'civilizational' scope restrictor added in the second (XVIII: 223).

mal to sacrifice quality for quantity (Riley 2003: 415; Saunders 2011: 195; Schmidt-Petri 2006: 176). However, the passage could also support the Lexical Interpretation indirectly through what it reveals about the genealogy of Passages B and C. On the one hand, since Passage B* has a similar content to Passage B, and Passage C* has a similar content to Passage C, but the diary entry reverses their order, we cannot simply take Passage B to provide Mill's sufficient condition because it is prior to Passage C. On the other, since Passage B* elaborates and clarifies the condition introduced in Passage C*, the genealogical origin of Passage C also supports the claim that its purpose is to elaborate and clarify the condition introduced in Passage B. In consequence, an examination of Passage C* and B* weakens the case for B-Sufficiency, and strengthens that for C-Necessity, even before we consider how far the meaning of Passages C* & B* aligns with that of Passages B & C.

To see the significance of the point, we need to view the diary passage in a broader context. In *Utilitarianism* Mill makes clear that happiness consists of a positive balance of pleasure over pain, but that if this positive balance is not somehow predicated upon higher quality pleasures, it can only be considered to be what the *Logic* describes as happiness in the "humble" as opposed to the "higher" sense of the term (VIII: 952; see also X: 211, Beaumont 2018a: 454). In consequence, by maintaining that "less of a higher kind" of happiness "is preferable to more of a lower", without adding any further quantitative qualifications (Beaumont 2019: 559 n.26; Riley 2003: 415; cf. Saunders 2011: 195), the diary entry supports the Lexical Interpretation by implying that the value of the higher quality pleasures that help to constitute the higher form of happiness, lexically dominates the value of the lower quality pleasures that help to constitute the lower form of happiness.

While defending C-Necessity, Jonathan Riley argues that Passage C should be read in the same lexical manner as the diary entry, by cashing out the quantity "which their nature is capable of" clause as "*any* finite amount" (2003: 418, emphasis added). In response, some object that, since some such finite quantities would transcend the capacities of a single person to experience them—for example, even at the intra-personal level one can imagine a quantity achieved by extending its duration into an afterlife, or by ratcheting up its intensity to a level requiring a super-human constitution (see also Crisp 1997: 23–5)—the "which their nature is capable of" clause introduces an implicit quantitative qualification that Riley's move ignores (Miller 2010: 57–8; Saunders 2011: 193; Sturgeon 2010: 1711 n.29). However, since the genealogy of the diary entry suggests that Passage C should be read as elaborating on Passage B, on the one hand, and the diary entry's condition for qualitative superiority is much closer to that provided by C-Necessity, on the other, the most that the preceding objection could be claimed plausibly—as opposed to definitively—to show is that C-Necessity should not be extended into the Lexical Interpretation. For example, one might claim that, while the diary entry presents a lexical requirement for qualita-

tive superiority, this was simply the original seed from which the position that Mill articulates in Passages B & C grew, and it is the weaker condition for qualitative superiority found therein, which he opted to publish, that constitutes his considered position.

Without such an extension into the Lexical Interpretation, C-Necessity alone would imply that, for P_1 to be of a superior quality to P_2, the smallest quantity of P_1 must bear more final value than any quantity of P_2 of which the competent judges are capable of—which is to say, at least a lifetime's worth of P_2—without requiring it to outvalue *any* finite quantity of P_2 as such (the Lifetime Interpretation). To illustrate, consider Mill's famous comparison of the superior quality of "mental" pleasures, unique to the human mode of existence, and the "bodily" ones, to which a swine's mode of existence is entirely restricted (X: 211–3). In this case, C-Necessity implies that, ceteris paribus, a life with the smallest possible quantity of mental pleasure consistent with the human mode of existence bears more final value than a 'swinish' life that is replete with bodily pleasure but devoid of such mental pleasure. In contrast, the Lexical Interpretation generates the stronger implication that, ceteris paribus, the former life bears more final value than *any number* of lives of the second kind (Beaumont 2019: 571–5). In consequence, the answer to the question of whether C-Necessity should be construed in terms of the Lifetime Interpretation, or extended into the Lexical Interpretation, may be important for understanding the exact implications of Mill's axiology for his utilitarian animal ethics (Section 6).

5. *Hedonic quality and modes of existence*

Some deem the Lexical Interpretation to be uncharitable to Mill because the conditions for qualitative superiority it attributes to him are too demanding—or 'extreme' in their implications—to be plausible (Saunders 2016: 504; West 2004: 63–4). Since it could be claimed that this objection continues to apply in the case of C-Necessity alone—and thus the Lifetime Interpretation—its weaker requirement notwithstanding, it is worth outlining why its requirement may be much less demanding than it first appears.

Sometimes Mill writes about 'pleasures' of different qualities as if we can choose between them on a day-to-day basis. In these cases, he seems to use 'pleasure' to refer to short-term pleasurable experiences (X: 212–3), including the experience of activities (X: 235). However, his key examples of differences of hedonic quality involve comparisons of the pleasantness or enjoyableness of different experiential "modes of existence" (X: 213). For example, in addition to his comparison of the human and swinish modes mentioned above, he also compares the modes of the "intelligent human being" and the "fool", the "instructed person" and the "ignoramus", along with the mode of the person of "[moral] feeling and conscience" as contrasted with that of the person who is purely "selfish and base" (X: 211–2).

As thus construed, a mode of existence consists of the experience of the multifarious potentialities born of the possession and exercise of a set of capacities over time. Given the taxonomy of experiences in Mill's *Logic*, different modes will include different kinds and levels of "sensations, thoughts, emotions, and volitions" (VII: 64). Moreover, some modes will require specific activities to exercise and sustain the capacities upon which they are predicated (X: 213). However, in the case of most human beings at least, a mode of existence will be much broader than any flash of consciousness or activity it incorporates (X: 215), in virtue of embodying the higher-order experience of experiencing *many* sensations, thoughts, emotions, and volitions *over time*, along with the sense of self that emerges in the process (XXXI: 138).

This much suggests that Mill could deem some of these higher-order experiences, consisting of *sets* of (lower-order) experiences of pleasures, to be higher-order pleasures that transcend the sum of their parts. After all, in the *Logic* Mill introduces the notion of "mental chemistry" to explain how comparatively complex feelings or experiences can be *generated*, as opposed to simply *constituted*, by sets or combinations of comparatively simple feelings or experiences (VIII: 854). If this were correct, Mill's focus on modes of existence would be explicable in terms of the fact that it is ultimately the higher-order pleasures taken therein to which his hedonism assigns its qualitative rankings (the Inter-Modal Interpretation).

To illustrate, consider the following passage from *Utilitarianism*:

> A state of exalted pleasure lasts only moments, or in some cases, and with some intermissions, hours or days, and is the occasional brilliant flash of enjoyment, not its permanent and steady flame. Of this the philosophers who have taught that happiness is the end of life were as fully aware as those who taunt them. The happiness which they meant was not a life of rapture; but moments of such, in an existence made up of few and transitory pains, many and various pleasures, with a decided predominance of the active over the passive, and having as the foundation of the whole, not to expect more from life than it is capable of bestowing. (X: 215)

One way to read this passage is to take the "brilliant flash of enjoyment" to refer to a higher quality pleasure than that embodied in the "permanent and steady flame". However, the alternative proposed here is to read Mill as ascribing a high-level quality to the "enjoyment" of the "permanent and steady flame", and to take the "flash" to represent a brief increase in its intensity. On this view, the high-level quality is ascribed to an extended enjoyment of—or higher-order pleasure taken in—an "existence" with "moments" of "rapture", "many and various pleasures, with a decided predominance of the active over the passive" etc.[20] Since Mill also describes this existence as including only "few and

[20] Crisp (1997: 27–8) notes that Mill's hedonism can only be understood properly when the term 'pleasure' is taken to include 'enjoyment'. Moreover, he also observes correctly that we can enjoy extended experiences that include some pain or suffering, such as struggling to climb a mountain. In consequence, one might take

transitory pains", the quantitative preponderance of pleasure over pain it entails makes it a happy one (X: 210). Moreover, this is the "higher" form of happiness that Mill refers to in the *Logic* (VIII: 952) because the said preponderance incorporates higher-order pleasures of a relatively high quality (X: 211–2). In consequence, this account also aligns with the diary entry's reference to the "*quality* as well as quantity of *happiness*" (XXVII: 663; emphasis added) because it explains how the quality of the higher-order pleasures taken in a mode of existence can pass over into the quality of the happiness of that mode once its pains are also taken into account.[21]

Importantly, once C-Necessity is combined with the Inter-Modal Interpretation its condition for qualitative superiority becomes far less contentious at the intra-personal level. People may disagree over the size of the smallest possible *durational* quantity that P_1 would have to take for it to constitute a genuine experience of, say, the human mode of existence, as opposed to an experience of nothing more than one of the fleeting sensations, thoughts, or activities that are performed therein. However, regardless of whether one judges the correct answer to be a day, a week, or longer, it would not be possible to refute Mill by claiming that a lifetime of swinish pleasure is more valuable than the pleasure of a snapshot of the consciousness of a human being (supposing this is actually true),[22] as the quality of the pleasure of the human *mode of existence* would not be captured through such a comparison.[23] Moreover, nor would it be possible to refute Mill by insisting that the more *cerebral* nature of the pleasure of reading is insufficient to ensure that it is *always* more valuable than the comparatively *sensual* pleasure of eating dessert, irrespective of how reading-rich and dessert-poor one happens to be at the time (cf. Anderson 1991: 9; Crisp 1997: 40–1). After all, if the pleasures of activities are indexed to the pleasure of the mode of existence in which they are undertaken—compare the

this to show that higher-order pleasures should be construed as pleasures taken in sets of lower-order pleasures *and pains*. However, that interpretive option has been rejected here for several reasons. Firstly, not all pains are enjoyable. Secondly, those which are enjoyable can be reconceived as pleasures using Mill's notion of complex ideas (VII: 57). And thirdly, Mill uses 'happiness' to encompass sets of pleasures and pains in which the former pre-dominate over the latter. In consequence, to take the higher-order pleasures, of which the Inter-Modal Interpretation claims 'quality' rankings to be predicated, to include pains, would be to muddy the waters between happiness and higher quality pleasure. In contrast, the version of the Inter-Modal Interpretation presented here clarifies the distinction whilst explaining the tightness of the connection between higher quality pleasure and higher quality happiness (see below).

[21] Hoag (1987: 418) refers to Mill's conception of happiness as a "higher-order […] end of life".

[22] See also the example offered by Miller (2010: 58).

[23] In consequence, to employ the technical terms employed by Griffin (1986: 83–5), when the smallest possible unit of a pleasure-type is sufficiently large, the distinction between "trumping" and "discontinuity" may become less significant.

mindset of Socrates eating dessert to that of a pig doing so—the difference between the pleasure of reading and eating dessert within the same of mode existence may be one of mere quantity, with the relative values depending upon the context of the choice.[24]

At an interpretive level, the preceding would also explain why Mill is more concerned to deny the existence of *voluntary* descent from a higher to a lower mode of existence than to insist that all highly cultivated people with strength of will are motivated to devote every drop of their time and energy to the most demanding intellectual, aesthetic, or moral activities (X: 213). That said, it is no doubt significant that one route to such descent could lie through choosing the likes of dessert over reading *consistently*, thus failing to sustain the capacities upon which the advanced mode is predicated. As a result, the Inter-Modal Interpretation can also explain Mill's occasional practice of referring to (intra-modal) activities as higher and lower pleasures, in terms of their role as indicators of the quality of the modes that people occupy, or to which these activities may lead them to fall or rise.

6. *Conclusion—Looking ahead*

This paper has provided a defence of C-Necessity, and thereby argued that Mill's qualitative hedonism should be interpreted in terms of the Lifetime Interpretation or the Lexical Interpretation. The choice between the two must ultimately be settled with reference to Mill's broader corpus. All that has been claimed here is that you cannot get to the Lexical Interpretation via (X: 211) alone.

The paper has also argued that, given the combination of the Lifetime and Inter-Modal Interpretations, the objection that Mill's condition for qualitative superiority is too 'extreme' or demanding loses much of its force at the intra-personal level. However, in doing so, the intent was not to suggest that *this* gives the Lifetime Interpretation much of an advantage, as the Lexical Interpretation can also be combined with the Inter-Modal Interpretation. Moreover, turning back to population axiology, advocates of the Lexical Interpretation may argue that, once it is combined with the Inter-Modal Interpretation, its extra vestigial demandingness gives it an interpretive advantage over the Lifetime Interpretation. After all, they may claim, it is the Lexical Interpretation *alone* that can explain how Mill's qualitative hedonism can avoid certain variants of the Repugnant Conclusion that prompted him to embrace *qualitative* hedonism in the first place. For example, they may argue that it is less plausible to interpret Mill as believing that the smallest quantity of the pleasure of the *human* mode is more valuable than that of the fullest pleasure taken in *one swinish life only,* than to interpret him as taking the former pleasure to outvalue that taken in *any finite number of swinish lives* (see Beaumont 2019: 571–5).

[24] Recall that Mill refers to "happiness" as "an existence made up of […] many and various pleasures" (X: 215). See also West (2004: 62) and Saunders (2016: 515).

In consequence, it is important to note that it is not (actually) self-evident that the Lifetime Interpretation would preclude such a *judgment*. After all, the Lifetime Interpretation can deny that Mill takes relations of qualitative superiority to *entail* lexical dominance, without denying that he takes *any* relations of qualitative superiority to rise to that level. On this view, Mill could take some qualitative differences to be far more significant than others, but for this to be manifest only once we cross the threshold from intra-personal to inter-personal evaluation. One reason this is important is that it is also far from self-evident that the Lexical Interpretation has an interpretive advantage at the inter-personal level when it comes to Mill's other axiological comparisons of modes of existence.

Recall that Mill also posits relations of qualitative superiority in the pleasures taken in the following modes: that of the *intelligent* being over the *fool*, that of the *instructed* person over the *ignoramus*, and that of the person of *feeling and conscience* over that of the person who is *selfish and base*. In the case of each of these pairs, is it more plausible to take Mill to believe that his qualitative superiority claim *entails* only that the smallest quantity of the former bears more final value than an entire lifetime of the latter (Lifetime Interpretation), or that he also takes it to *entail* that the smallest quantity of the former bears more value than that taken in *any finite number of such lives* (Lexical Interpretation)? When the choice between the Lifetime and Lexical Interpretation is framed thus, it is apparent that the Lexical Interpretation will be more prone to generate 'extreme' results, of doubtful consistency with Mill's own judgments, the further up the modal ladder of rank the (allegedly) lower quality pleasure of a given pair is located. That said, exactly how prone will depend upon an answer to a question that may be important for future research, namely, that of how Mill's individual comparisons of pairs of modes of existence are supposed to relate to each other. For example, is the qualitative superiority between the pleasure in the mode of existence of the instructed person and that of the ignoramus supposed to hold when the latter is also a person of feeling and conscience (supposing Mill takes that to be possible)?[25] Ceteris paribus, an answer in the affirmative would make the Lexical Interpretation far less radical, and thus far less prone to generating 'extreme' results, than an answer in the negative.

Nevertheless, it is far from clear that the Lexical Interpretation can contain the danger when Mill's qualitative judgments are taken into consideration.[26] In an intriguing letter to Thomas Hare written in 1865, Mill reports how reading Plato in Avignone, while staying there with his wife, is not quickening his "zeal" in his "own cause, as a candidate" for parliament:

[25] Consider (XVIII: 31; XIX: 390, 402).

[26] For one attempt at containment, resting on further interpretive claims the plausibility of which it is not possible to examine here, see Riley (2009: 131–3).

> It is an *infinitely pleasanter mode* of spending May to read the Gorgias and Theatetus under the avenue of mulberries which you know of, surrounded by roses and nightingales, than it would be to listen to tiresome speaking for half the night in the House of Commons. The only disagreeable thing here is having to choose between pleasures. (XVI: 1061, emphasis added)

Perhaps Mill is simply exaggerating here but, given the shortage of textual evidence with any potential to *confirm* the truth of the Lexical Interpretation (Schmidt-Petri 2006), it would be difficult for its advocates to dismiss the passage in this way, as opposed to taking it as *the* confirmation that Mill views qualitative differences between pleasures in terms of an infinite—or unbridgeable—gap in the value of their pleasantness.[27] However, while advocates of B-Necessity may think that even the Lifetime Interpretation will generate an uncharitably 'extreme' implication in this case, the Lexical Interpretation is clearly in greater danger of doing so. After all, could Mill really be implying that, ceteris paribus, no number of lifetimes worth of the pleasures he could take in his parliamentary mode could bear more final value than the pleasures he could derive from *a month* in his philosophical mode with his wife at Avignone (cf. Parfit 2004: 18)?[28]

Of course, ultimately interpretations of Mill's qualitative hedonism also require evaluation in terms of whether, or how far, they can be made to cohere with his moral judgments. This can only be done with the aid of bridging principles predicated on specific interpretations of Mill's utility principle, on the one hand, and his theory of justice and rights, on the other. Since those interpretations will be highly contested (Cooper et al 1979; Lyons 1997), we should be careful to avoid jumping to simplistic conclusions about the *moral* implications that would flow from Mill's hedonism *as such* given the adoption of one interpretation thereof as opposed to another (Skorupski 2000: 259–60). At the same time, the examples above should make clear that there is at least the potential for the choice between the Lifetime and Lexical Interpretation to have a significant impact upon how far Mill's utilitarianism is taken to imply that differences in beings' modes of existence can generate differences in their moral status, and thus the attractiveness of his position for contemporary philosophers interested in population *ethics* as such.[29]

[27] For example, Riley's interpretation is that "any higher kind is infinitely more pleasant than any lower kind" (2009: 128).

[28] Here it may be worth noting a couple of other interpretive possibilities worth exploring: (1) that the size of the value gap is due to the fact that Mill is actually comparing a *mode* with a mere *activity* (this may be congenial to advocates of B-Necessity who take Mill to allow for *some* relations of qualitative superiority to rise to the level of lexical dominance (cf. Miller 2010: 56; see also Beaumont 2019: 571–5)); and (2) that the description of the parliamentary mode (or activity) as 'tiresome' implies that it is actually best conceived as a pain (this may be congenial to advocates of the Lexical Interpretation who want to limit its 'extremeness').

[29] I would like to thank Christoph Schmidt-Petri and Dale Miller for comments on and criticisms of an earlier version of this paper, as well as the two anonymous

References

Anderson, E. 1991. "John Stuart Mill's Experiments in Living." *Ethics* 102 (1): 4–26.

Arrhenius, G. Rabinowicz, W. 2015 "Value Superiority." In I. Hirose and J. Olson (eds.) *The Oxford Handbook of Value Theory*. Oxford University Press: 225–48.

Arrhenius, G. Ryberg, J. Tännsjö, T. 2017. "The Repugnant Conclusion." In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/repugnant-conclusion

Beaumont, T. 2018a. "J. S. Mill's Hedonism: activism, experientialism, and eudaimonism." *British Journal for the History of Philosophy* 26 (3): 452–474.

Beaumont, T. 2018b. "A Perennial Illusion? Wittgenstein, Quentin Skinner's Contextualism, and the Possibility of Refuting Past Philosophers." *Philosophical Investigations* 41 (3): 304–28.

Beaumont, T. 2019. "J. S. Mill on Calliclean Hedonism and the Value of Pleasure." *Dialogue: Canadian Philosophical Review* 58 (3): 553–578.

Brink, D. 1997. "Mill's Deliberative Utilitarianism." In D. Lyons (ed.) *Mill's Utilitarianism: Critical Essays*. Rowman & Littlefield Publishers: 149–84.

Cooper, W. Nielson, K. Patten, S. 1979. *New Essays on John Stuart Mill's Utilitarianism* (ed.). Canadian Journal of Philosophy—Supplementary Volume.

Crisp, R. 1997. *Mill on Utilitarianism*. London: Routledge.

Donner, W. 1991. *The Liberal Self: John Stuart Mill's Moral and Political Philosophy*. Ithaca: Cornell University Press.

Griffin, J. 1986. *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford: Oxford University Press.

Hauskeller, M. 2011. "No Philosophy for Swine: John Stuart Mill on the Quality of Pleasures." *Utilitas* 23(4): 428–46.

Hoag, R. 1987. "Mill's Conception of Happiness as an Inclusive End." *Journal of the History of Philosophy* 25: 417–31.

Lyons, D. 1997. *Mill's Utilitarianism: Critical Essays* (ed.). Rowman & Littlefield Publishers, Inc.

Mill, J. S. (1963–1991) *Collected Works of John Stuart Mill* (Vol. 1–33). Ed. John Robson. University of Toronto Press. https://oll.libertyfund.org/people/john-stuart-mill

Miller, D. 2010. *J. S. Mill: Moral, Social and Political Thought*. London: Polity Press.

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

Parfit, D. 2004. "Overpopulation and the Quality of Life." In J. Ryberg and T. Tännsjö (eds.). *The Repugnant Conclusion*. Dodrecht: Kluwer Academic Publishers: 7–22.

Parfit, D. 2016. "Can We Avoid the Repugnant Conclusion?" *Theoria* 82: 110–27.

Riley, J. 2003. "Interpreting Mill's Qualitative Hedonism." *The Philosophical Quarterly* 53(212): 410–8.

referees for the *Croatian Journal of Philosophy*. Responsibility for the views expressed is purely my own.

Riley, J. 2008. "What Are Millian Qualitative Superiorities?" *Prolegomena* 7 (1): 61–79.

Riley, J. 2009. "Millian Qualitative Superiorities and Utilitarianism, Part II." *Utilitas* 21(2): 127–43.

Saunders, B. 2011. "Reinterpreting the Qualitative Hedonism Advanced by J. S. Mill." *Journal of Value Inquiry* 45: 187–201.

Saunders, B. 2016. "Recent Critics of Mill's Qualitative Hedonism." *Philosophy* 91 (4): 503–521.

Schaupp, K. 2013. "Books Before Chocolate? The Insufficiency of Mill's Evidence for Higher Pleasures." *Utilitas* 25(2): 266–76.

Schmidt-Petri, C. 2003. "Mill on Quality and Quantity." *The Philosophical Quarterly* 53 (210): 102–4.

Schmidt-Petri, C. 2006. "On an Interpretation of Mill's Qualitative Hedonism." *Prolegomena* 5(2): 165–177.

Skorupski, J. 1989. *John Stuart Mill*. London: Routledge.

Skorupski, J. 2000. "Quality of Well-Being: Quality of Being." In R. Crisp and B. Hooker (ed.). *Well-Being and Morality: Essays in Honour of James Griffin*. Oxford: Clarendon Press.

Sturgeon, N. 2010. "Mill's Hedonism." *Boston Law Review* 90: 1705–1729.

West, H. 2004. *An Introduction to Mill's Utilitarian Ethics*. Cambridge: Cambridge University Press.

Wilson, F. 1990. *Psychological Analysis and the Philosophy of John Stuart Mill*. University of Toronto Press.

Croatian Journal of Philosophy
Vol. XXI, No. 62, 2021
https://doi.org/ 10.52685/cjp.21.62.3
Received: November 9, 2020
Accepted: July 27, 2021

Aesthetic Eating

ADAM ANDRZEJEWSKI
University of Warsaw, Warsaw, Poland

The aim of this paper is to sketch a framework for perceiving the act of consumption as an aesthetic phenomenon. I shall argue that, under some circumstances, it is possible to receive aesthetic satisfaction from the act of eating food, in which the object of one's appreciation is, for the most part, considered separately from what is actually eaten. I propose to call such a process "aesthetic eating" and argue that due to its aesthetic autonomy it might be a potential factor in enjoying certain kinds of food. This phenomenon is apparent in the case of the types of food that are acquired tastes. It is plausible that distinguishing the aesthetic pleasures of food from the ones associated with the act of eating can not only enrich our aesthetic life but also deepen the aesthetics of our overall gustatory experience.

Keywords: Gustatory aesthetics; food; aesthetic experience; aesthetic pleasure; Richard Shusterman.

"I am a bread of life"
John 6:48

1. *Introduction*

In the film *Gold Rush* (dir. Charlie Chaplin, 1925) viewers can admire a famous scene in which the main character portrayed by Chaplin is cooking and eating his own shoe. The boot is prepared in a pan with much attention and care. When it is ready and seasoned with watery gravy, it is served to the table, divided into two portions and consumed. At the very beginning, the protagonist seems to be slightly reluctant to the taste but after a few bites the situation changes drastically and a proper feast begins. Chaplin enjoys the shoe with true devotion, tastes all of its fragments and combines different elements to maximize the

flavour. His gentle moves resemble eating a roasted chicken, fried fish, or delicious spaghetti, and the scene ends as a picture of full gustatory immersion into the unusual meal.

Despite the undoubtedly humorous aspect of the described scene, there is philosophical content to be found in it as well. It seems that Chaplin takes pleasure not only from the object consumed (sic!) but also from the way in which it was consumed.[1] Eating a shoe is—naturally—a substantially exaggerated example but thanks to its extremity, it provokes one to ask the question: What are the aesthetic and gustatory qualities of the act of eating and of the dish itself? In particular, (how) is it possible to separate pleasures of the act of eating from what is eaten? So far, food aesthetics have concentrated on the artistic and aesthetic status of cuisines, exploring their potential to exist within the world of art and the gastronomy scene, as well as on food's sensory dimension (see Perullo 2016; Mancioppi and Perullo 2020; Jacques 2015; Jacques and Adrià 2015). Although a great level of thought precision has been reached when it comes to edible objects, e.g. whether they are artworks (Andina and Barbero 2018) or culinary recipes, e.g. how they exist (Borghini 2014, 2015), not much attention has been given exclusively to the aesthetic potential of the act of eating itself.

Recently, this gap has been filled to some extent by the contribution of Richard Shusterman. By developing this topic Shusterman attempts "[…] to introduce a bit more precision in gastronomical theory by focusing on the art of eating in a more restricted sense and by distinguishing that sense from other meanings of the term." (2016: 261). In particular, he wants to distinguish—as he labels it—the art of eating from the art of cuisine (culinary objects, *resp*. dishes) and the art of food appreciation and food criticism (food writings). In his proposal, Shusterman highlights the features of the art of eating as a process of food consumption that goes beyond the qualities exemplified in food; a number of reasons is given for distinguishing the art of eating from other ways of food ingestion.[2]

In this paper I would like to go one step further and make room for a more liberal take on food consumption than the one proposed by Shusterman. While I do not intend to criticize Shusterman's proposal *per se*, I will offer its substantial re-reading that goes beyond his initial claims. My observation is that Shusterman is still too focused on food itself in his proposal. Therefore, I shall argue that it is possible to think of such a way of eating in which what gets appreciated is, for the most part, considered *separately* from what is actually eaten. That is,

[1] It would like to distinguish between the pleasure taken from watching x and the pleasure received when actually doing x. Thus, the subject of my considerations are not the aesthetic experiences and pleasures received when watching someone else eating.

[2] I use the term "food consumption" as an umbrella concept covering *all* modes of eating and drinking (e.g. aesthetic, religious, social, nutritional etc.). For the sake of simplicity I qualify beverage consumption as part of food consumption.

in such a way of eating we value the qualities of the process and not the qualities of food itself (although such a situation is quite rare in everyday life). I suggest to label this process "aesthetic eating" and shall propose two understandings of it: a weak and a strong one. My thesis is supported theoretically by recent developments in process aesthetics (Nguyen 2020). The paper is structured as follows. In §2 Shusterman's definition of the art of eating is presented and analysed.In §3, I list the theoretical challenges of his proposal. In particular, I claim that his definition is still too tightly linked to food as the object consumed. In addition, a handful of reasons for not seeing the act of eating as art are provided. Next (§4), by shifting from "the art of eating" into "aesthetic eating", I show how my position overcomes the difficulties faced by Shusterman's proposal. Lastly, the paper ends with the summary of my argument (§5).

2. *Shusterman's definition of the art of eating*

The definition established by Shusterman consists of two steps. First, the simple act of eating is distinguished from the more conscious and socially-oriented act of food consumption. Humans, unlike other animals, eat not only to satisfy hunger or thirst. They do it for other reasons as well (political, religious, moral and the like) and human food consumption is marked and shaped by culture to a high degree (Telfer 1996). This means that for the perception of food as objects to be eaten, aspects that go beyond the act of consumption itself are important (Korsymeyer 1999). Moreover, culture provides numerous tools, traditions and habits to critically discuss food and evaluate our gustatory choices and tastes. According to Shusterman, an aware form of eating ("eating know-how") is food consumption equipped with a certain kind of knowledge and sensibility regarding the history of gastronomy, cultivation of food ingredients and food preparation, as well as general rules of food pairing. Shusterman then takes a step further and sets up another level of eating:

> I would propose a further distinction: between gastronomes who simply know how to select and enjoy good food (and who master the art of eating in this important but basic sense), and those gastronomes, who also know how to eat aesthetically in the fullest sense—beyond making good food choices and combinations. By this I mean those gastronomes whose knowledge of food and sensitive tasting is translated into an art of eating focused also on the aesthetic elements and qualities of the experience of ingesting food. (2016: 263)

Here we are provided with an important clarification, which might be treated as a sketch for the definition of the art of eating. The ability of selecting and enjoying food as well as knowledge in gastronomical history and practices is *not* enough for the art of food in a proper sense. What is needed is the transition from the consciously undertaken food choices and ways of gustatory enjoyment into discovering the aesthetic

potential of food eaten *and* aesthetic qualities driven from the act of food consumption, as well as the fact that food is incorporated into the body. According to Shusterman, the art of (fine)[3] eating is characterized by the following two features.

First, the art of eating is essentially temporal and this feature is manifested at two levels. The first level is concerned with the obvious fact that eating is a temporally structured phenomenon: there is a sense of sequencing in the meals appearing at the table (e.g. tea is served after dinner) or the special order of dishes within a specific meal (e.g. appetizer is served before the main course). The second level refers directly to the sense of timing understood in bodily terms. That is, our body by necessity needs time to absorb and digest food: "[…] one mouthful leading to the next, or, more precisely, one mouthful leading to a complex sequencing of smelling, biting, tasting, chewing, swallowing, and breathing" (Shusterman 2016: 264). For example, when drinking a glass of chilled Riesling or Pinot Grigio our overall bodily (including gustatory) experience might vary greatly depending on the time taken between the sips or our posture (standing at the bar, sitting at the table or lying on the grass).

Second, in connection to its temporal nature, the art of eating is an artistic performance "[…] whose enjoyment is in the performative process of eating" (Shusterman 2016: 264). What is valued here is the act of food consumption. It also means that the art of eating displays different aesthetic features that the eaten objects does. Shusterman says: "Appreciating food's crucial contribution to the art of eating, I nonetheless argue that the art of eating goes well beyond the aesthetic qualities of the objects eaten" (Shusterman 2016: 264). This is the crucial point for defining the art of eating and Shusterman—being aware of that—immediately refers to the possible objection that might be raised in order to challenge that feature. That is, one could claim that the enjoyment is rooted in the objects eaten (dishes) and not in the act of eating itself (I shall return to it later on). To address this potential issue he proposes to take a lesson from the philosophy of theatre, making the following analogy: the art of eating *deepens* the valued qualities of food and also refers to qualities that *exceed* the objects eaten in the same way the theatrical performance deepens the artistic value of the script. Thus, we can receive aesthetic enjoyment "[…] that go[es] beyond the tastes, smells, and visual forms of our food objects […]" (Shusterman 2016: 264).[4]

[3] Shusterman refers here to "fine" art, but he actually does not make any extensive clarification about this notion further. I shall leave this matter unresolved as it is not entirely relevant to this paper. I suppose, however, that the predicate "fine" is used by Shusterman in order to emphasize that he is interested in art in a proper (and not just honorific) sense and I follow this way of understanding when reconstructing his position.

[4] Shusterman enumerates here the bodily movements connected with the act of eating, such as chewing, sipping, swallowing and the like. The specific elements of

To sum up, while Shusterman postulates the existence of the art of eating, his goal is to provide the definition of this phenomena which will bring more precision to the philosophy of food and widely understood gustatory discourse. The art of eating is characterized by the two main features: (i) temporality and (ii) performativity. Moreover, the most philosophically important result of Shusterman's position is the claim that the aesthetic enjoyment of the art of eating goes beyond the objects eaten.

3. *Challenges*

Without doubt, Shusterman's view of the aesthetic process of consumption brings a considerable amount of novelty and precision into the contemporary philosophy of food and aesthetic theory. Shusterman is very much right in highlighting the temporal and performative aspects of the act of eating. I also concur with the analysis of the value of the art of consumption as a phenomenon contributing not only to the enrichment of our aesthetic life but also of human health.

In this section, however, I shall outline some tensions faced by Shusterman's definition. My main issue is with the mentioned proposal being still too focused on food appreciation and, thus, the process of eating not being (despite the explicit declaration) liberated from the objects eaten (dishes). These critical remarks will serve me as the ground for further inquiry where I would like to offer a more autonomous interpretation of the aesthetic act of eating (in §4).

Shusterman claims that the art of eating amounts to answering the question about the way in which an edible product should be consumed. In his words: "[…] how to eat and drink in terms of our modes and manners of ingestion can be constructed as the art of eating in its narrower and precise sense" (Shusterman 2016: 262). Since his proposal is to interpret at least some manifestations of the process of eating as an expression of *art*, I think one might ask for a certain sense of normativity of the aesthetic appreciation of that artform, e.g. if a given evaluation is correct or not (Zangwill 2003/2014). In other words, the aesthetic appreciation needs to be directed to the qualities of some entity *qua* the kind of the entity it is.[5] If this is correct, then "how to eat and drink" is determined by the food itself rather than the *way* in which the food is consumed. And this "how" allows for eliciting properties (both aesthetic and non-aesthetic) from the objects eaten. The primary function of food in the art of eating is explicitly noticed by Shusterman. I evoke

the art of eating are: (1) posture; (2) movements; (3) the accessories for eating (e.g. choice of eating instruments); (4) selection of foods and their sequencing (acts of non-selection as well; also spontaneous selection), and (5) perception (269–274).

[5] This idea is notably presented and defended by Kendal Walton. In short, he claims that in order to correctly perceive an artwork we need to identify its features and categories (1970: 356). For the same issue explicitly concerning food and cuisines see also Ravasio (2018).

his words: "Appreciating food's *crucial* contribution to the art of eating […]" (2016: 264, my emphasis). Thus, in the art of eating, the food goes first. I do not want to say that this claim is substantially wrong. On the contrary, it could be perfectly fine, *unless* we state that a certain version of food consumption is *art* in itself. If we are keen to treat the art of consumption as a legitimate kind of art, as Shusterman does, it seems we need to make room for its "liberation" from the objects eaten. Thus, we have to identify its distinct features that do not belong to other artforms (e.g. to the culinary art understood as edible objects), if we would like to discuss the art of eating in a "narrow" sense. Moreover, the art of eating must be perceived as aesthetic on an equal level with food (e.g. a dish belonging to culinary art) and not as something only giving us "access" to the objects eaten.

It should be noted that the philosophy of theatre has already dealt with a similar problem. In the traditional view on the ontology of theatre, a performance was just an interpretation of the literary work and as such was denied an artistic status. Its role was only to give the viewers "access" to the literary work and it was judged according to its truthfulness to the literary work (Hamilton 2009a). It was showing that a theatrical performance is not just an interpretation of the literary work that allowed for treating it as an independent form of art (Osipovich 2006; Hamilton 2009b). As has been mentioned earlier, Shusterman anticipates this kind of worry and proposes to see the relationship between the object eaten and the act of eating as analogical to the one between a work of literature and theatrical performance. Suppose that a person is consuming an apple. In the very act of eating they have access to the qualities connected with what an apple is (how it tastes, smells and the like) as well as the qualities of the process of consumption in itself (e.g. rhythm of chewing). Thus, the eaten object is somehow "present" in the process of eating and some of its aesthetic qualities are "inherited" by the process of eating (and, naturally, the process of eating has different qualities than the apple as well). This is quite true with theatre, where the play "inherits" some properties of the script as the script is just one of the ingredients necessary for creating a theatrical performance.[6] For example, Torvald Helmel's words "There can be no freedom or beauty about a home life that depends on borrowing and debt" are both a part of *A Doll's House* as Henrik Ibsen's work as well as of a theatrical performance under the same title (although, in the first case they are written, whereas in the second case they are spoken).

However, whereas in the contemporary philosophy of theatre a theatrical performance is free from any obligations regarding a literary work, in Shusterman's view one of the key aspects of the art of eating

[6] According to this view a performance is never a performance of some other work, nor is it ever a performance of a text. The text of the literary work could be used as one of the ingredients to create a performance. See Hamilton (2007: 31–33).

is to deepen the already existing properties of food. I do not think that such a role separates the act of eating from food in a sufficient degree. The potential aesthetic autonomy of the act of eating would be achieved only if we could value the act of eating without focusing so much on what we eat. Analogically, I can imagine a situation in which we are able to value a theatrical performance for its features no matter (mostly) what features are expressed by the work of literature.

In short, despite the undisputable value of Shusterman's proposal it suffers from being still too object-focused. Food plays an over-crucial role in the definition described above, which considers the art of eating something merely secondary and auxiliary when compared to culinary dishes. This does not seem enough for someone who, as Shusterman, truly wants to present the art of eating as an example art, not only in an honorific sense.

Unlike Shusterman, I suggest not to refer to the act of eating as a work of art. I do not claim that some processes of food consumption cannot constitute an art form. There is no doubt that the question of whether an act of eating can be art depends primarily on the definition of art that we agree to accept. In other words, if we accept a fairly liberal definition of art we can easily include the act of eating into the art world.

Having said that, I do not want to commit myself to any particular definition of art since, I hope, the more neutral my position is, the broader scope it has. Moreover, I am not entirely sure that seeing gastronomy (food as objects, eating as processes) through the lens of art ascribes any special value to it in itself.[7] I concur with Ferran Adrià, a Catalan chef and food theorist, who argues that gastronomy and art are marked by different sorts of creativity and performativity[8] and, as a result, "[g]astronomy is not an art and does not purport to be one" (Jaques 2018: 241). Not perceiving gastronomy within the art framework guarantees its autonomy and own language of communication. I do hope that my step is not a misinterpretation of Shusterman's intention. First, he does not give any definition of art at all (*Nb*. for someone this could be a reason for a potential critique) and seems to be more interested in art as a cultural practice emerging on different occasions. Moreover, Shusterman uses the term "art" as well as "fine art" to draw our attention to what he really thinks is truly treating the act of eating as a fully-fledged source of aesthetic experiences. I do share the opinion that the act of eating, under some circumstances, is a profound aesthetic phenomenon.

[7] In a similar spirit: John (2014).

[8] Various types of food and art histories are in length noted by Korsmeyer (1999: 141–145).

4. *Aesthetic Eating*

In this section, I would like to propose such a definition of the act of eating—I label it "aesthetic eating" instead of "the art of eating"—in which what is appreciated is considered somehow separately from what is actually eaten. That is, in aesthetic consumption we value the qualities of the process, and not the qualities of food itself. It is suggested that the new proposal enables us to overcome the challenges faced by Shusterman's position as well as sheds light on the aesthetic function of the process of eating. I shall offer two understandings of aesthetic eating: a weak and a strong one. The former is Shusterman's claim reconceptualised, whereas the latter goes toward the aesthetic autonomy of food consumption.

I propose *weak aesthetic eating* (henceforth WAE) to be a practice focused on eliciting aesthetic qualities of food and drinks by following certain ready-made categories. By "ready-made" I mean rules and patterns that are traditionally associated with various kinds of foods and drinks that (supposedly) maximize their gustatory (and, thus, aesthetic) experiences.[9] For example, we usually eat soup with a spoon or cut a beef steak with a knife, or have a glass of sparkling wine as an ice-cold drink, or a hot espresso as soon as it arrives at the bar counter. I shall then propose the following definition of WEA:

> *x* is WAE if: (i) *x* is an act of food consumption of a dish *p* by a person *O*; (ii) *x* focuses on eliciting *p*'s aesthetic qualities, and (iii) *O* aesthetically appreciates *p* through *x*.

WAE can be called "aesthetic" because it provides access to the aesthetic features of the eaten objects. (However, my definition does not presuppose that the process has or has not any aesthetic properties in itself.) WAE is an aware *way* of selecting and consuming food and drinks as well as a careful manner of detecting and enjoying the aesthetic potential of cuisines and food ingredients. In WAE the proper object of appreciation is a dish or a drink.[10] Even if we experiment with the way we eat by trying to apply rules other than the ready-made categories (e.g. drink a cold espresso or eat melted ice-cream), in the end we are interested in what has changed in our perception and/or gustatory experience *of* the food. In other words, the possible value ascribed to every violation of a well-established approach towards certain kinds of food is judged according to our renewed or reconceptualised culinary experience. It is easy to notice that WAE defined in such a way is linked to Shusterman's proposal. Similarly to it, food plays the main part here.

[9] Food's ability to elicit aesthetic experiences has been defined on many occasions. See e. g. John (2014), Jacques (2014), Adams (2018).

[10] This seems to be coherent with John Dewey's opinion on food and aesthetics: "[…] seeing, hearing tasting, become aesthetic when relation to a distinct manner of activity qualifies what is perceived" (2008: 55). Thus, the act of eating might become an aesthetic experience (and WAE remains neutral about that) only when we take into account what is actually consumed. See also Koczanowicz (2016)

However, Shusterman's proposal is stronger than WAE when treating the very act of eating as an object of aesthetic appreciation.

So far, I have referred to WAE, which has been motivated solely by my decision, but it is also worth mentioning that WAE can be somewhat "forced" by the unusual nature of the served dish. As an example I shall refer to Ferran Adria's *Margarita* (2005). It took the form of a snow-white cube with a dab of salty foam on top, which was eaten with a teaspoon. Its shape and consistency was meant to trick our intuitions about the margarita drink: an alcoholic beverage that must necessarily be in a liquid form. Imbibing this "drink" makes for an even more perplexing experience given the fact that it is served as an aperitif. Regardless if Adria's *Margarita* is a token of the so-called edible art or not (Andrzejewski 2018), we—as guests at the El Bulli restaurant—would have to decide in what way to approach this oddly shaped aperitif. Treating WAE in an experimental way might contribute to a discovery of new tastes, flavour combinations and modes of food preparation.

Now I would like to introduce the notion of *strong aesthetic eating* (henceforth SAE) understood as a practice focused on the aesthetic qualities of the act of eating. That is, SAE promotes paying attention and care to qualities belonging exclusively to the way of food consumption. SAE's definition is as follows:

> x is SAE if (i) x is an act of food consumption of a dish p by a person O; (ii) O aesthetically appreciates x, and (iii) x is appreciated mostly regardless of p.

In SAE we encounter an *alternative* object of appreciation that is the *way* of eating or drinking. Imagine that you are at quite a fancy restaurant enjoying contemporary Central European cuisine (say, a kind of "variation" on pierogi/pielmieni). In WAE you would enjoy the food as a wonderful dish, ideally re-interpreted according to the contemporary trends yet still faithful to its long and noble tradition of being a local specialty of the European periphery, whereas in the framework of SAE you would enjoy that dish *as well as* the manner in which you enjoy it.

I would like to support the argument for SAE's existence by referring to the recently emerged notion of the art of action (Nguyen 2020).[11] In short, the art of action is the art in which artefacts are intentionally meant to elicit first-hand aesthetic experience of mental and physical processes. That is, we "appreciate the aesthetic qualities which arise in their own action" (Nguyen 2020: 2). Contrary to the art of artefacts where we are focused on objects "[t]he art of actions, on the other hand,

[11] Nguyen uses the term "art" in a very broad sense and his definition is not very much attached to it, see (Nguyen 2020: 4). For him, the most important thing is to include some processes as belonging to the group that deserves aesthetic attention and consideration. Hence, I do not see how supporting SAE by the notion of the art of action would automatically contradicts my intuition that SAE is not a kind of art. What I am doing here is rather supporting my claim by the broad idea of the aesthetics of action.

are marked by distinctively self-reflective aesthetic appreciation. In these arts, the focus of appreciator's aesthetic attention is on aesthetic qualities of their own action." What should be stressed here is that in the art of action we gain access, evaluate and process first and foremost our own, personal experiences, which cannot be appreciated (and shared) by someone else. This is due to the definition, which states that only I have a full access to the qualities of the action I am currently performing.[12] My suggestion is that SAE exemplifies a phenomenon, which could match the above description. In other words, SAE is an act in which what we aesthetically appreciate is the qualities emerging from the process of eating, which are are to be separated from the qualities of the food itself.

One step in establishing my argument is to notice that in WAE th food will always play the main role, as it is the central object of our appreciation and focus, but in SAE we can at least imagine a situation in which the way of eating may become more "important" than the object eaten. I do not want to claim here that the food should be fully ignored; my point is only that *if* we want to aesthetically appreciate the *way* of eating then we need to give it privilege over the food eaten (otherwise what would be appreciated is the food solely). To explore such a possibility, I support my claim with an observation that Shusterman's proposal focuses *only* on cases when the art of eating *deepens* the aesthetic pleasures provided by food and drink and—by doing so—goes beyond these pleasures. Thus, it acknowledges only the "positive" relationship between the food and the act of eating. In such a case, it's quite intuitive to perceive the act of eating as still bound to the objects eaten: the act of eating contributes to something already enjoyable. However, we could easily think of a different scenario. Aesthetic eating may not increase my pleasure, but could save me during a totally disastrous dinner. In other words, if we could find a situation in which aesthetic eating would translate into the aesthetic satisfaction with the act of eating even if the object consumed (dish) it is not particularly tasty (or we just do not fancy it), then SAE would prove itself a valuable claim.

Again, imagine that you have been invited to dinner by a dear friend. She has chosen a small, cozy restaurant run by an acclaimed chef with a strange surname you cannot even pronounce. You have both ordered a tasting menu, and before you have finished your welcome drink, the first dish arrives. From that moment your horror is about to begin… There is an oyster on your plate! You are not a big devotee of seafood at all and you actually do not remember when exactly you have last eaten these monsters! Despite the tragic situation you do not want to hurt your friend's feelings and decide to eat the oyster. She shows you how to hold the shell, what gestures to adopt in order to eat, what the accurate timing is, and at which moment you need to put the shell back.

[12] Naturally, some actions are experienced and evaluated form the outside and the inside (Nguyen 2020: 3).

You follow her instructions accurately and the result is not shocking at all: you do not like the oysters. However, you actually *liked* the funny way of eating them and felt quite proud of yourself for stepping out of your comfort zone.[13]

My claim is that the specific nature of SAE reveals itself especially when it's accompanied with quite a trivial yet profoundly true observation about our live: it is evident that we do not always like certain objects or activities such as hiking, boat trips, silent cinema, karaoke, or a vast range of types of food and beverages. However, sometimes we find ourselves in a situation when we would like to change our attitude toward particular objects, tastes, flavours, or styles and undertake the effort to like these things even if we actually do not fancy them.[14] In such a case we deal with the so-called acquired taste. "Acquired taste" according to Kevin Melchionne "jump-starts new satisfactions where I do not initially find them" (2007: §1). It should not be confused with discovered taste. Unexpected flavour combinations, artworks or experiences one has never had before are subjects of taste preferences that are, metaphorically speaking, hidden in oneself. For example, when I was in primary school I never thought that a sandwich with salty crisps and olive oil could be my thing (I admit: it does not sound very appetizing). However, after the first bite I totally fell in love! For me this experience was an example of discovered taste: a certain situation just revealed some truth about my personal taste. It turned out that I have always liked that kind of sandwich but have just not realised that.

One could rise the objection that this is an argument in favour of liking the food and not the way of eating. This is partly true. However, I argue that if the aesthetic satisfaction driven from the act of eating is *able* to contribute to liking an initially disliked food then such way of eating is indeed a legitimate source of an aesthetic experience. What is more, in such a case we cannot simply say that aesthetic experiences are derived from food because we do not fancy it in the first place.[15]

[13] One could argue that in such a case we appreciate not the act of eating itself but rather a broader experience at the table that is distinct from food ingestion. This objection can be easily rebutted by noting the following two facts. First of all, a recent research has shown that eating experience is multisensorial in nature and brings many aspects that are not vividly (at first glance) connected with food consumption (see Spence, Youssef 2019). Secondly, Nguyen argues that in the aesthetic process we also include our experiences of the outside world that are related to the appreciated activity (2020: 9).

[14] An extreme example of such a case is analysed by Korsmeyer: "When disgust or revulsion is confronted and overcome, what was at first disgusting can become delicious" (2007: 149)

[15] One could rise another objection in a similar spirit: if the object of appreciation in SAE is not the food itself then it is a bit mysterious how SAE is supposed to help us acquire new taste. This potential issue might be undermined by noticing two points. First, in the process of acquiring new tastes (also in art) many factors are involved, e.g. decisions, intentions, social behaviours and project planning (Melchionne 2017, Korsmeyer 2002). Second, if we enjoy a certain way of eating then

For acquired tastes the element of decision is necessary. As Melchionne puts it: "To acquire taste, we must *decide* to change the facts of our mental life. Acquired tastes are taken up despite the fact that, at the outset, we did not like them. Involving *as if* activity, acquired tastes are, by definition, never immediate, direct, or simple" (Melchionne 2007: emphasis in the original). So, an individual has to know that they do not like *x* and decide to act as if they like *x*. It is worth noting that an acquired taste finally ends up in liking what we have previously only faked to like. Naturally, individuals cannot be forced (internally or externally) to like anything. Although, they might pretend to like something (a situation that results in a fake taste) or just stop making the effort to like something (acquired taste is not accomplished).

Having made a very brief outline of the issue of acquired taste, I suggest that aesthetic eating, if rightly adopted, is able to facilitate the process of developing acquired tastes. Firstly, WAE enables the subject to elicit the aesthetic properties from the object eaten. If we *decided* to really like something, then it would be valuable to maximize the aesthetic and gustatory impact of this object. For example, if someone has made a decision to like seafood, they should eat different kinds of seafood with attention and care in order to experience the seafood's taste fully. (Even though sometimes this might be really painful). It cannot be ruled out that new ways of eating seafood (with or without a spoon, with eyes closed, accompanied with some herb-based drinks and the like) might contribute to its overall taste and provide us with more reasons to become devoted admirers of seafood.

Secondly, in SAE we are focusing mainly on the *fact* that we're eating something and on the *way* in which we are doing it. For example, imagine that you are not a huge fun of whisky but many of your friends are and you often associate this alcoholic beverage with a nice vision of green Highlands or the rocky Islay island. What really might help you discover new experiences in the whisky world might be the way in which you consume that alcohol: your posture, other ingredients, the special way of pouring the mineral water, the manner of sipping, the timing in tasting and the mere fact that you're challenging yourself. In other words, it is not only the bodily movements that contribute to the aesthetic satisfaction but also the cognitive aspect of the whole process that plays an important role in the overall experience.[16] SAE invites one to value what could be called a meta-level of food consumption. It is true that we pay attention to how we eat but when it comes to the actual moment of consumption in many cases (however, surely,

we have to facilitate a certain kind of food in order to encounter this way of eating as frequently as possible. This might finally help us in acquiring a new taste.

[16] I do not claim that the bodily and cognitive aspects of our existence go apart (in particular, I do not believe in any sort of Cartesianism). However, in some cases we are able at least conceptually to distinguish between these two elements. For example, a decision to eat a certain dish belongs to the realm of cognition, yet, naturally, is very often facilitated by the bodily needs.

not always) the way of eating becomes quasi-transparent: it's treated as a sort of "means" enabling us to taste the dish (WAE). SAE focuses on the higher level in the sense that one constantly pays attention and takes pleasures in something that is, in most cases, dimmed by food's features.

Liberating substantially the way of food consumption from what is consumed in SAE may also have a moral dimension. There is no need to say much about the obvious fact that individuals consume different kinds of food of varying quality, as well as its various quantities depending on their geographical location, socio-cultural conditions and financial abilities. Sometimes we eat what we can and not necessarily what we really want to. During mid-80s in a number of European countries people had very limited access to some of nowadays common goods such as citrus fruit, coffee, tea and some kinds of meat, to name just a few. Preparing a dinner for Christmas Eve for many families was quite a difficult task to accomplish since the meal itself was rather modest. However, despite the vivid shortcomings, many people took a considerable pleasure (I may risk to call it aesthetic in some cases) in the way in which food was prepared, put great effort into finding the ingredients (not always through official distribution channels) and showed a sort of creativity when they needed to replace some unavailable ingredients with more popular substitutes. SAE would help in dealing with the situation of shortage and unfriendly circumstances by focusing on the way of eating, e.g. consuming with dignity, responding to the behaviour of our loves ones, taking satisfaction from the mere fact that we did everything we could to prepare something to eat etc., and not being too much linked to what we actually ate. It should be emphasized that the suggested role of SAE is rather modest here but it still adds something to our moral realm.[17]

SAE does not work without any limitations. It is hard to argue that we can *always* receive aesthetic pleasure from the way in which we consume *x*, while at the same time completely not paying attention to what *x* actually is. Although this is often the case, one can easily imagine a situation when we consume something so intensively flavoured or just profoundly not tasty that an individual simply cannot take even a tiny unit of pleasure from the act of eating in itself or they are not able to consume the food at all. Surely, a way of eating is ontologically dependent on what we eat: we can eat only when there is something to be eaten. Ontological priority does not mean, however, that judgments over the aesthetic qualities of food and the way of eating are not

[17] An analogy might be driven from the ritual of tea ceremony. In short, it expresses profound aesthetic and moral values and each element of the ceremony must be in the right place. However, perfection is not required. On the contrary, imperfection plays a crucial role for this ritual. Moreover, when done in the right way, the taste of the tea becomes of secondary importance compared to the way of preparing and consuming this warm beverage, which are the most important ones. See Saito (2007: 33–35; 2017).

separable. They often go hand in hand, but SAE shows that it is not a matter of necessity. Moreover, I want to emphasize that in SAE we are not "blind" to what we are eating. On the contrary, we do notice the consumed food but SAE invites us to experience something far beyond that.[18]

What Shusterman's position and SAE have in common is that the object of aesthetic appreciation is the process of eating. I fully concur with Shusterman in setting up the project going beyond the food and seeking to value the somatic aspects of food ingestion. On the other hand, as I pointed out in §3, his proposal still suffers from being too object-focused: the food plays too big a role in his definition of the art of eating and needs to be liberated. Unlike Shusterman's art of eating, SAE offers more independence to the act of eating, not only regarding its aesthetic autonomy but also by seeing the process of eating on its own terms. Such a way of its conceptualisation opens up the possibility of taking a genuine aesthetic pleasure from it and provides an opportunity for several improvements if needed. That is, when we are predominantly focused on the qualities of the act of eating we might notice its weak points.

5. *Conclusion*

The aim of this paper has been to sketch a framework allowing to see the act of eating as an aesthetic phenomenon. What I want to show is that, under some circumstances, it is possible to receive aesthetic satisfaction from the act of food consumption, in which what is appreciated is, for the most part, considered separately from what is actually eaten. I called such a process "strong aesthetic eating" and contrasted it with "weak aesthetic eating". The latter is a conscious way of eating that is primarily focused on eliciting food's aesthetic features. Giving a considerable aesthetic autonomy to strong aesthetic eating does not automatically mean that food should not be valued aesthetically at all or valued to a lesser degree than the process of eating. On the contrary, I have argued that thanks to its aesthetic autonomy strong aesthetic eating is able to contribute to liking certain kinds of food (especially when we think of acquired tastes). I do hope that distinguishing the aesthetic pleasures of food from these of the act of eating will be able to not only enrich our aesthetic life but also deepen the aesthetics of our overall gustatory experience.[19]

[18] This is exactly why we cannot claim that in SAE what is appreciated in the end is the food itself (as in the case of WAE). If SAE belongs to process aesthetics then the proper object of appreciation in SAE are qualities arising from the whole action of eating, and these cannot be reduced to the object eaten. Thus, the object eaten and the way of eating it are strictly connected to each other, yet separable in terms of the aesthetic appreciation.

[19] I would like to express my gratitude to Alessandro Bertinetto, David Collins, Max Ryynänen, Mateusz Salwa, Marta Maliszewska and Monika Favara-Kurkowski,

References

Adams, M. 2018. "The Aesthetic Value of Local Food." *The Monist* 101 (3): 324–339.

Andina, T., Barbero, C. 2018. "Can Food be Art?" *The Monist* 101 (3): 353–361.

Andrzejewski, A. 2018. "How to Frame Edible Art." *The Nordic Journal of Aesthetics* 27 (55-56): 82–97.

Bonsdorff, P. 2012. "Pending on Art." *Contemporary Aesthetics*. Special issue: Artification.

Borghini, A. 2014. "Metaphysis of Natural Food." In P. Thompson and D. Kaplan (eds.). *Encyclopedia of Food and Agricultural Ethics*. New York: Springer: 1378–1381.

Borghini, A. 2015. "What is a Recipe?" *Journal of Agricultural and Environmental Ethics* 28: 719–738.

Dewey, J. 2008. *Art as Experience*. In idem, *The Later Works, 1925–1953*. Vol. 10. Carbondale: Southern Illinois University Press.

Hamilton, J. 2007. *The Art of Theatre*. Oxford: Blackwell.

Hamilton, J. 2009a. "The Text–Performance Relation in Theater." *Philosophy Compass* 4: 614–629.

Hamilton, J. 2009b. "Theatrical Performance and Interpretation." *Journal of Aesthetics and Art Criticism* 49: 307–312.

Jacques, J. 2014. "Food." In M. Kelly (ed.). *Oxford Encyclopedia of Aesthetics*, 2nd edition. Oxford: Oxford University Press: 63–67.

Jacques, J. 2015. "Main Issues on Gustatory Aesthetics." *CosMo: Contemporary Studies in Modernism* 6: 173–185.

Jaques, J. 2018. "Stovetop Philosophy: An Interview with Ferran Adria." *The Monist* 101 (3): 237–246.

Jacques, J., Adrià, A. 2015. "For an Applied Philosophy of Gastronomy." *CosMo: Contemporary Studies in Modernism* 6: 163–172.

Koczanowicz, D. 2016. "John Dewey: A Culinary Perspective." *Pragmatism Today* 7 (2): 88–94.

Korsmeyer, C. 1999. *The Making Sense of Taste*. Ithaca: Cornell University Press.

Korsmeyer, C. 2002. "Delightful, Delicious, Disgusting." *Journal of Aesthetics and Art Criticism* 60 (3): 217–225.

Mancioppi, E., Perullo P., 2020. "Estetica aromatica. Odore, politiche dell'atmosfera e impegno percetivo." *Rivista di estetica*, 73 (1): 118–135.

Michel, C.,Velasco, C., Gatti E., Spence Ch., 2014. "A taste of Kandinsky: assessing the influence of the artistic visual presentation of food on the dining experience." *Flavour* 3 (7): 1–10.

Melchionne, K. 2007. "Acquired Taste." *Contemporary Aesthetics*, vol. 5.

Melchionne, K. 2017. "Aesthetic Chioce." *British Journal of Aesthetics* 57 (3): 283–298.

as well as two anonymous reviewers for their helpful comments and constructive criticism on this paper in draft. Special thanks go to Iris Vidmar Jovanović for her constant support and valuable advice. I am also indebted to the Faculty of Philosophy at the University of Warsaw for their generous financial and administrative support whithout which this research could not be conducted.

Nguyen, C. T. 2020. "The Arts of Action." *Philosophers' Imprint* 20 (14): 1–27.

Osipovich, D. 2006. "What is a Theatrical Performance?" *Journal of Aesthetics and Art Criticism* 54: 461–470.

Perullo, N. 2016. *Taste as Experience. The Philosophy and Aesthetics of Food*. Foreword by M. Montanari. New York: Columbia University Press.

Ravasio, M. 2018. "Food Landscapes: An Object-Centered Model of Food Appreciation." *The Monist*, 101 (3): 309–323.

Saito, Y. 2007. *Everyday Aesthetics*. Oxford: Oxford University Press.

Saito Y. 2017. "The Role of Imperfection in Everyday Aesthetics." *Contemporary Aesthetics* 15.

Shusterman, R. 2016. "Somaesthetics and the Fine Art of Eating." In S. Irvin (ed.). *Body Aesthetics*, Oxford: Oxford University Press: 261–280.

Spence, Ch., Youseef J. 2019. "Synaesthesia: The multisensory dining experience." *International Journal of Gastronomy and Food Studies* 18.

Telfer, T. 1996. *Food for Thought*. London: Routledge.

Walton, K. 1970. "Categories of Art." *The Philosophical Review* 79 (3): 334–367.

Zangwill, N. 2003/2014. "Aesthetic Judgment." *Stanford Encyclopedia of Philosophy*. E. Zalta, (ed.), accessed June 29, 2020.

Croatian Journal of Philosophy
Vol. XXI, No. 62, 2021
https://doi.org/ 10.52685/cjp.21.62.4
Received: August 19, 2020
Accepted: March 17, 2021

Chomsky's London

ZOLTÁN VECSEY
MTA-DE-SZTE Research Group for Theoretical Linguistics (ELKH), Hungary

Semantic externalism is the view according to which proper names and other nominals have the capacity to refer to language-independent objects. On this view, the proper name 'London' is related semantically to a worldly object, London. Chomsky's long held position is that this relational conception of reference is untenable. According to his internalist framework, semantics should be restricted to the examination of the informational features of I-language items. Externalists reject this restriction by saying that without employing the relational notion of reference, it would remain entirely mysterious how we can talk about our perceptible environment. This paper offers a novel argument for externalism. The basic idea is that external reference proves to be indispensable even for Chomskyans who regard our talk about the environment as irrelevant for the purposes of semantics.

Keywords: Nominal reference; externalism; invariantism; internalist semantics: indispensability.

1. *The semantic value of 'London' is not London*

Chomsky emphasizes in several of his later writings (e.g. 2000, 2013, 2016) that the standard externalist semantic theory of reference is untenable. Chomsky's main contention is not that semantics as a serious scientific enterprise is impossible *tout court* or that the technical term 'reference' is entirely meaningless. The problem lies rather in a lethal combination of two general linguistic points of view: externalism about semantics and invariantism about reference. Semantic externalism is based on a world-oriented perspective that goes beyond the mental states and processes of language users. Those who adopt this perspective assign worldly objects and events to expressions as their semantic

values. Proper names and other nominals are supposed to be related causally, historically or functionally to their referents which, in turn, are thought of as existing at least partly independently of the mental. The other point of view, invariantism about reference, is a natural companion of semantic externalism. If reference is conceived as being a certain kind of relation between expressions and language-external objects, then a second plausible hypothesis is that reference-apt nominals are related either to a unique individual or to a specific type of object. Accordingly, invariantism amounts to the claim that, vagueness and ambiguity aside, referential relations remain constant in all possible contexts of use. It does not really matter what is asserted about the unique bearer of a proper name, the name refers to *that* bearer in all of its contextual applications. And it does not matter in which sentential structure a particular nominal appears, it will refer always to the *same* type of object.

The combination of these ideas fits into a venerable tradition from Russell to Kripke, not to speak about some more recent developments in truth-conditional semantic theories.[1] Moreover, many of the research results of this tradition proved to be easily compatible with the hypotheses of other data-driven sciences that are devoted to investigate the cognitive and neurological basis of language.[2] If this is indeed the real situation of research, and there is a growing body of semantic-independent evidence that many items in the nominal domain have a stable relational structure, then it is hard to evade the question: what are Chomsky's theoretical reasons for rejecting this tradition? Are there persuasive arguments that force us to give up combining externalist and invariantist theories of reference entirely?

In an often-cited passage of his book *New Horizons in the Study of Mind and Language* Chomsky seems to offer such an argument:

> Referring to London, we can be talking about a location or area, people who sometimes live there, the air above it (but not too high), buildings, institutions, etc., in various combinations (as in *London is so unhappy, ugly, and polluted that it should be destroyed and rebuilt 100 miles away*, still being the same city). Such terms as *London* are used to talk about the actual world, but there neither are nor are believed to be things-in-the-world with the properties of the intricate modes of reference that a city name encapsulates. (Chomsky 2000: 37)

The first sentence of this passage reminds us that such ordinary proper names as 'London' can be used colloquially to refer to more than one object. Instead of being related to a unique referent (i.e. a city), 'London'

[1] Reference fixing is today one of the most intesively investigated issue in this field. See, for example, the contributions of Dickie (2015) and Gómez-Torrente (2019) to this theme.

[2] See, among others, the individual volumes of the Springer Series on *Studies in Brain and Mind*. To mention one of these studies, Calzavarini (2019) points out that the question whether humans have referential capacities can, in principle, be decided on the basis of neuroanatomical data.

appears to be a systematically polysemous expression. It is thus best conceived as a means which offers multiple assertoric possibilities for ordinary usage: we may talk about a location, about a certain set of citizens, about various institutions and buildings that belong to a local government etc. by applying one and the same expression. If this is so, as it appears to be, then the question 'What is *the* referent of 'London'?' is utterly misleading. We should allow for more than one potential referent for this name, and, as Chomsky's example nicely illustrates, there is no need to invoke distinct conversational contexts for demonstrating this claim. Different referents may occur even in a single sentential/predicative structure. Although it is far from being a knockdown argument against external referential invariance, the polysemy argument clearly shows one thing, namely, that without a careful analysis even the most simple and most plausible examples of this semantic phenomenon remain unpersuasive.[3]

The second sentence of the passage acknowledges that 'London' may be used to refer to language-external objects, but it rejects the assumption according to which these language-external objects have a real existence. Note that the qualification 'language-external' ought to be understood here as a mark of ontological status: if an object is external to language according to its ontological status, it must exist outside the mind. So what is rejected is the assumption that there is a determinate object (or there are determinate objects) outside the mind to which 'London' may refer in our referential talk. While this may sound for some as a sheer contradiction, this is not at all the case. The crucial element in the argument is the (age-old) distinction between *uses* of names for referring to objects and *names* referring to objects. One can see the referential use of names for our daily purposes as wholly unproblematic and at the same time deny that names are devices of reference according to their very nature. Language users are skilled enough to refer to the external world by using the proper name 'London' but from this it does not follow that there really is something in the external world that answers to their referential practices. For this being the case, Chomsky suggests, London as an external object of reference would have to possess apparently incompatible properties. For example, *is a location* and *functions as the centre of government* denote categorially different predicative properties, the first being a property of spatiotemporal objects, the second being a property of abstract artefacts. Yet we can rather easily construct a sentence in which these two properties are predicated simultaneously about the (putative) referent of 'London'—*London lies on the river Thames and functions as the centre of government*—without evoking any sense of conflict or anomaly. This shows, again, that in everyday usage names may re-

[3] Defenders of referential invariance may argue that 'London' uniquely and invariantly refers to a city, but this name might occasionally refer to other objects through the intervention of certain metonymical processes. For an argument of this type, see Vicente (2019).

fer to objects that have incompatible properties but these usage data do not entail that there are objects that exhibit incompatible properties.

The central idea behind this example is that the externalist viewpoint in semantics must be accompained with an extravagant ontology: if one takes nominal reference to be a kind of word-world relation, then one must populate the world with such weird objects that are, say, concrete and abstract at once. One immediate consequence is that the semantic value of 'London' is not and cannot be London. As mentioned above, Chomsky rejects externalist *cum* invariantist approaches to reference without claiming that semantic theory as such rests on a fundamental mistake. What is, then, the alternative picture about semantics?

Chomsky's opinion is quite precise in this regard. In order to make progress in understanding human language, linguistics in general and semantics in particular should define itself as a branch of naturalistic inquiry. This means, roughly, that linguistic research ought to apply the standard methods of natural sciences, and the explanatory models of linguistics should introduce only well-defined terms into their theoretical vocabularies. For such terms that are infected by human interests and unreflective thought, there is no place in a serious language science. According to Chomsky, all terms that are somehow associated with the hypothesis of *English as a public language* belong to this latter category. Thus, 'assertion', 'referring name', 'language-external object', and many more similar terms that are dependent on the legitimacy of the term 'public language' loose their explanatory power in a genuine naturalistic enterprise. What remains is the replacement of these "obscure" externally anchored terms with their internal equivalents. Chomsky argues forcefully in many places that the only viable alternative to the externalist tradition is to see language as an internal property of an individual. Seen from an externalist point of view, languages can be identified with classes of observable objects, phonetic and graphemic strings. If an approach to a natural language focuses on such public external objects, the target is called an E-language. In contrast, an internal-individual language—an I-language—should not be thought of as an object in the sortal sense of the word. It is not something that is represented in the mental sphere of an individual in the format of Fodorian Mentalese or of a shared, internalized language code. Rather, it is a provisional state of the mind/brain of a particular individual. And as such, it is part of the general biological makeup of the individual. It is worth stressing, however, that internalism does not lead to a solipsistic conception of language. Although I-languages are mind/brain states that are realized in distinct individuals, there are significant similarities between these states. Among other things, I-languages have a common biological base which guarantees that they share several biological/functional properties.

What is more important with respect to our initial problem—the natural-seeming semantic relationship between 'London' and London—is that I-languages are held to be abstractly characterizable sys-

tems that perform computational operations on syntactic items. Let us suppose that in order to use 'London' competently, one must possess an I-language item *london'*. If *london'* is a lexeme, then, according to the current Chomskyan background syntax, it is an atomic mapping of sound and meaning. If it is a more minimal syntactic object, then, perhaps, it is a category-neutral root. Either way, *london'* must provide some legible information to the articulatory and the conceptual interface systems of the mind/brain of individuals. We can say that, in the Chomskyan framework, the proper task of semantics is to investigate how the information at these interface systems enables the individual to *use* 'London' for talking about various aspects of the language-external world (Chomsky 2003b: 294–295).

In sum, within the internalist framework of linguistics, semantics should be conceived of as a form of syntax. This reinterpretation entails a narrowed domain of investigation in the sense that semantics must be restricted to the examination of the elementary informational properties of I-language items. The analysis of the nature of word-world relations—that is, the analysis of the mechanisms of nominal reference—falls outside of this domain. Which means, in turn, that semantics as a genuine naturalistic inquiry does not, and should not, incorporate substantive theses about the ontology of the mind-external world (Collins 2009: 56–57).

2. *Ineffective externalist responses*

Proponents of externalism might think that the "deontologization" of semantics is too a high price to pay to solve the problem generated by the incalculable referential behaviour of 'London' and other nominals. It could be argued that if one denies that there exists a relation between nominal expressions and objects that have a language-independent ontological status, then it remains peculiar or mysterious how we can talk about anything at all that is outside of our mind/brain. This kind of objection is not unfamiliar in the literature on reference. An eminent example is Burge (2003), who defends the traditional externalist standpoint in the following way: "I see no reason to think that there is anything scientifically wrong or fruitless in studying language–world relations, or with taking them to be part of the formal structures elaborated in semantical theory. Like aspects of the theory of perception, this aspect of semantics is not internalist, even in Chomsky's broad sense of internalist" (Burge 2003: 466). Silverberg's (1998) earlier counter-argument to Chomsky's internalist framework seems to be of the same type as that of Burge as he has said that "our perceptual abilities provide a basis … to our ability to refer to particular things and to indicate classes of things, even if our conceptions of these things and classes contain significant error" (Silverberg 1998: 231–232).

To posit a direct justificatory link between the theory of perception and semantic theory may appear to be a good argumentative strat-

egy in this context for two reasons. First, it is commonly assumed in cognitive science and epistemology that, under normal circumstances, processes of perceiving carry sensory information about objects that are external to our mind/brain. If perceiving is essentially a *relational* process, and our talk about the external world is closely and multiply intervowen with the informative content of these processes, then we might surely assume that out talk about the world has also a relational structure. Relationality in language may be taken, in the end, as a property that is inherited from direct sensory contacts with the world. Second, during perceiving how things are, individuals maintain a rather complex relationship to their environment. Presumably, the theory of perception has enough resources to specify which of the elements of this relationship are causally determined. On this basis, one can give a precise explanation for why a given individual in a given perceptual circumstance is in a particular perceptual state. And given that perceptual states are internal states of individuals, one can further argue that environmental causal factors are constitutive with respect to a certain range of internal states. This can be taken to show that some specific internal mind/brain states of individuals, including I-language states, would not exist without persistent environmental inputs.

It is questionable, however, whether this kind of argumentative strategy can be applied successfully against the central theses of Chomsky's internalism. For it is surely not part of the internalist framework that semantics should be entirely decoupled from the perceivable world of language-external objects. Undoubtedly, there are some passages in Chomsky's recent work that can be read in an opposite way. Such an impression may arise, for example, when he talks about science in general, and says that the scientific enterprise can be regarded as making use of our *innate* "cognoscitive powers", which are only marginally influenced by the environment.[4] This sounds as if the the proper aim of science were to disentangle our internal and innate capacities. But the crucial point lies elsewhere. That the internal mind/brain states of individuals may be determined externally is so obviously correct that it would be irrational to deny it. Chomsky mentions three factors that are relevant in this regard: (i) the human genetic heritage, which functions as a biological basis for language acquisition, (ii) environmental information that give an individual shape to I-languages, and (iii) certain cognitive and physical edge conditions without which I-languages were impossible (Chomsky 2007: 3). But none of these three external factors shows that I-language items *must* be endowed necessarily with world-related information. In fact, something like the opposite is the case. The objection levelled by Burge, Silverberg and others assumes that if there is such a thing in language as (inherited) external relationality, then it has an immediate explanatory relevance for the semantic properties of internally generated I-language items. In fact, beyond a

[4] See, for example, Chomsky (2003a: 41).

gesture towards perception, it is hard to find clear evidence that this is indeed a plausible assumption. According to the internalist view, semantic "theory" sees only the syntactically computable features of lexical items: *london'* is seen as a package of formal features that can be transferred to the interface systems in successive phases of derivation. The issue whether the surfaced item, 'London', is related in its interpretation to a worldly object, falls outside the scope of semantics proper.

As mentioned above, Burge and Silverman are not the only representatives of this questionable critical interpretation. Kennedy and Stanley (2009), in their otherwise excellent paper, raise the following objection against Chomsky's treatment of 'London':

> One class of example that Chomsky gives concerns sentences like 'London is a city in England.' According to Chomsky, native speakers will tell us that this sentence is actually true. But Chomsky thinks it is quite clear to all that the city of London, the standard semantic value of the noun phrase 'London', does not exist (Chomsky 2000: 37). We certainly do not accept his reasons for thinking so. Nevertheless, even if we did, this would not give us a reason to reject semantic theories that assign to the sentence 'London is a city in England' truth conditions that require there to be a genuine entity in the world that is actually called 'London'. It would just give us a reason to conclude that none of the non-negated sentences containing the word 'London' are true. (Kennedy and Stanley 2009: 586–587)

As Stoljar (2015) rightly observes, Kennedy and Stanley here misinterprets what Chomsky is saying. They suggest that, according to Chomsky's opinion, London, the city, does not exist. This is not quite correct. Chomsky never contends that there is no such thing in the world as London. In fact, he emphasizes just in the quoted passage that *London is not a fiction*. What does not and could not exist in reality is, according to him, a determinate naturalistic object, which can be taken to be the contextually invariant referent of 'London'.

There is a second misinterpretation in Kennedy and Stanley's line of thought. Chomsky says that native speakers judge the sentence 'London is a city in England' as true. Note, however, that this contention is based on an empirical conjecture about how speakers *use* that sentence. Disturbingly, Kennedy and Stanley take this example as if it concerned the semantic status of the name 'London'. It is not so. At least, it is not so according to an I-language based model of semantics. In that model, 'London' has no semantic status at all.

King (2018) seems also to misunderstand Chomsky's London example. King claims that it is not lethal for an externalist semantics that speakers can use a particular proper name for referring to distinct types of objects. The adequate externalist response to this problem is to relax the strong invariantist condition on reference (King 2018: 781–782). The essence of King's proposal is that one should attach multiple polysemous meanings to names. If 'London' can be used for referring to a location, then one ought to say that it has a meaning *London1*. If

'London' can also be used for referring to a certain set of citizens, then one ought to say that it has another meaning, a meaning, which may be called *London2*, and so forth. This multiplicity does not generate puzzling beliefs or other opaque attitudes in speakers, argues King, since they are fully aware that *London1*, *London2*, etc. are related to different objects. Moreover, one might assume that *london'* possesses exactly so many internal semantic features that are required by these diverse external objects. If one thinks that the observable multiplicities of meaning must somehow be mirrored at the I-language level, this is a reasonable assumption. Yet it is not completely clear, how this proposal would solve the original problem. Recall that internalism rejects public usage as a reliable data source for semantics because ordinary "usage facts" are highly opaque phenomena, and thus the term 'public language' cannot be precisely and coherently defined for scientific purposes.[5] In this regard, King's proposal seems to rest on a problematic argumentative step: we are invited to draw a conclusion through a backward inference from public usage to I-language semantics. After all, given the scientifically intractable character of public languages, it might have been better for King to avoid premising his argument on facts of ordinary usage.

3. *Departing from the inside-out direction*

I do not wish to claim that all of the extant externalist responses to the London example are ineffective; there might exist a powerful counter-argument to Chomsky's internalism that has simply escaped my attention.[6] I want, however, to call attention to a hitherto (so far as I can find out) undescribed way of defending the traditional doctrine of semantic externalism. The kind of defense I have in mind need not be motivated by arguments from perception or from other epistemically significant relations to the environment. And it also does not need to be based on arguments that attribute significant explanatory value to the observable behaviour of speakers. As we have seen above, Chomskyan internalists can hope to gain easy victories in debates, where they are confronted with critical challenges that they reject as scientifically irrelevant or inappropriate from the start.

The character of the debate about 'London' (and London) can be changed, however, when one takes into closer consideration how in-

[5] Though it is a side issue here, let us note that it would be quite important to understand in what sense linguistic theories are dependent on the reliability of their data sources. For a systematic treatment of this issue, see Kertész and Rákosi (2012).

[6] Since this paper focuses primarily on the problem of reference, I do not want to discuss the recent turns in the debate about copredication. Of course, one cannot examine the phenomenon of copredication without touching the possibilities of nominal reference. On this, see Collins (2017), Gotham (2017), Liebesman and Magidor (2017, 2019), and Ortega-Andrés and Vicente (2019).

ternalists make use of their *own* arguments. One interesting aspect of Chomsky's argumentative strategy is that in order to deny that London is the semantic value of 'London', he seems to be compelled to accept that under certain circumstances 'London' has to stand for London, the city, and for *nothing* else. Such an invariantist figure of speech must be legitimate even according to his strict naturalistic standards of scientific reasoning. For if it were ambiguous or otherwise misleading to say that 'London is a city in England', then it would also be pointless to claim that London cannot function as an invariant semantic value for 'London'.

There are two inter-related reasons for this. First, names are specific nominal items in the sense that they can become part of a natural language only under relatively transparent conditions. When a newly introduced name is intended to stand for a settlement, there must be an appropriate sortal object, a highly populated area, that somehow can serve as the unique bearer of the name. And similarly, when a newly introduced name is intended to stand for a location, there must be an appropriate sortal object, typically a geographical space, that somehow can serve as the unique bearer of the name. It does not really matter in which way the intended bearer of a new name is accessible in the context of name-giving. In cases where it can be identified unproblematically with perceptual means, there is no need for linguistic intervention. In other cases where direct perceptual identification is impossible, the intended bearer should be identified by applying descriptive sentences. What is important is the (relative) epistemic clarity of the name-giving situation: in order to attach a name to a particular object, one should be aware both of the uniqueness and of the sort of the object in question. Probably all ordinary proper names acquire such a privileged sortal value due to the originating act by which they are introduced into the lexicon of a particular speech community. It would certainly be surprising, if it turned out that, instead of focusing on a single object, 'London' has been introduced by a group of speakears conjunctively for standing for a settlement *and* a location *and* a set of people *and* many other sort of things.[7]

Second, in denying that 'London' has an invariant semantic value, one cannot but presuppose the sortal referential capacity of the name which figures in the act of denial. That is, in order to make the nega-

[7] As a referee points out, it is not mandatory to see name-giving as providing a deciding evidential support for externalism, because introductory acts can also be construed internalistically. Hinzen (2007) could be cited as an example in this respect. Hinzen's main contention is that external name-giving presupposes the pre-existence of individual concepts, and individual concepts are accomplished by narrow syntactic processes. From this it follows that name-giving is an internal process regarding its informational sources. Of course, one can follow this argumentative path but then the burden is to explain the possessing conditions of the individual concept LONDON without involving mind-external factors. It is not clear to me whether Hinzen's internalist approach can overcome this burden.

tive verdict—'The semantic value of *London* is not London'– a non-ambiguous scientific claim, the object language term 'London' should be taken as having a pre-established external significance. Without making this presupposition, without implicitly granting that there is a privileged value attached to the name, one cannot even consistently formulate such a claim. The most plausible candidate for this value is, of course, the object that has been selected for and identified in the original name-giving situation. So, in spite of all protests, the internalist approach to nominal reference must take at face value the 'London'/London uniqueness relation *prior* to making any verdict about the I-language features of *london'*.[8]

Note, however, that what is at stake here is *not* primarily the questionable status of public language. One can find a rather close analogy to this situation in the literature on basic logical inferences.[9] There the question is: how can one explain the phenomenon of inferential validity. If it is known, for example, that from *A* and *If A then B* one may safely infer *B*, then there must be some kind of explanation of why one can confidently rely on this rule of inference. The difficulty is that the most straightforward explanation of this fact deploys the same rule of inference in an indispensable way. The derivation of the conclusion from the premises starts from the assumption of the two premises, then it follows the step of modus ponens, and the process of derivation ends with a twofold deployment of conditional proof.[10] Since this series of steps includes the target of the explanation—*A* and *If A then B*, then *B* –, the derivation of the validity of the rule proceeds in a circular manner. One possible view concerning the source of this particular difficulty says that basic inferential validity is presuppositional in the sense that the logical enterprise is committed to certain transitions between sentences in the form of rational trust.

Issues of rational trust aside, it is not hard to state an analogy between contested semantical and logical phenomena: external referential invariance appears to be indispensable for basic semantic explanations in the same way as inferential validity appears to be indispensable for

[8] If this or something similar is the correct way of understanding nominal reference, then the following question arises: how can 'London' be used to refer to so many sort of things? It is worth stressing that identifying privileged semantic values for names is compatible with many proposals about our referential practices. It can be argued, for example, that 'London' might change its value due to pragmatic processes like meaning modification or reference-shift. Such pragmatic processes are rather complex and supposed to involve many contextual factors.

[9] For more on this, see the debate between Boghossian (2003) and Wright (2004, 2018).

[10] Wright (2004: 173) presents the derivation in the following manner:

1	(i) P	Assumption
2	(ii) If P, then Q	Assumption
1,2	(iii) Q	(i), (ii) Modus Ponens
1	(iv) If (if P, then Q), then Q	(ii), (iii) Conditional Proof
	(v) If P, then if (if P, then Q), then Q	(i), (iv) Conditional Proof

basic logical explanations. To repeat, external referential invariance, as it is taken here, is not a marginal semantic assumption that can be suspended as soon as one leaves behind the suspicious dimension of public language. Rather, it is an inherent, *sui generis* capacity of nominals that cannot be considered as a secondary reflex of a more specific or more fundamental semantic capacity. It is hard to imagine how one can cast doubt on the existence of this capacity without undermining the intelligibility of those parts of semantic theory, which are dependent on the exercises of that very same capacity. In light of this, one can pose a touchstone question that all parties of the debate must answer:

> *The Question of Explanatory Indispensability*. Is there any coherent way to deny the possibility of external referential invariance without applying nominals that have the capacity to refer invariantly to a particular external object?

Traditional externalists are in a position to answer this question with a resolute 'no'. They hold that reference as a basic dyadic relation between words and the world must, on pain of intelligibility, be presupposed in semantic theorizing. They also hold that there are no available strategies to suspend this presupposition. This is so because all opposing arguments must at some point involve constitutively the relata of certain dyadic external relations. Externalists are not absolutists in the sense that they do not claim that every nominals must exhibit this dyadic property in every possible context.[11] That would obviously be an unreasonable demand. Names of failed posits of scientific inquiry such as 'phlogiston' or 'ether', character names of literary fiction such as 'Sherlock Holmes' or 'Hamlet', metaphors and other forms of figurative language illustrate quite well how wide the range of possible exceptions is. Thus the claim of the externalist is a modest one, namely that in basic semantic explanations one has to presuppose that *some* nominals have a relational semantic structure even in cases where one makes an attempt to deny the mind-external existence of the objects of this very relation.

Of course, one might try to say, following Chomsky, that there is no such object in the world, which corresponds to the intricate modes of the reference of the name 'London'; and therefore the presupposition in question is a mere illusion. So the Question of Explanatory Indispensability should be answered with 'yes'. But this would not work. If it is *not* somehow presupposed that 'object', 'world' etc. have the capacity to refer to something language external, then the argument from the intricate modes of reference cannot be interpreted as a coherent reply.

Are there other strategies to answer the Question of Explanatory Indispensability with 'yes'? There appear to remain only two options.

[11] This idea is clearly stated in Borg (2009: 41), where she says that externalism "is the theory that *for at least some expressions*, what they mean is determined by features of the agent's environment, but this clearly leaves room for other types of expression in natural language whose meaning is determined in other ways."

First, Chomsky can argue that 'reference' is an *informal* term, which has no equivalent in a serious science of language (Chomsky 2000: 31–32 and Stainton 2006: 921–922). If this were right, then the statement that 'London' refers to London' would be relegated from serious scientific discourse to informal usage. What consequences would that have for the investigation of reference? Informal usage is appropriate when one wants to talk about what language users do with nominals, or about what purposes they have if they express their thoughts by using nominals. Reflecting on these topics can surely contribute to the common sense understanding of reference. But that is all. Seen from the perspective of serious science, informal 'reference' is a vague term, unsuitable for making sufficiently clear statements about the rules that govern the core computational processes on I-language informational structures. And because it belongs to the class of vague, common sense terms, informal 'reference' is dispensable. Externalists can reject this line of thought in the same way as before. If Chomsky, or one of his followers, contends that 'reference' is an informal term, having no scientific equivalent, then this contention can have any weight only if it has already been presupposed that (i) 'reference' establishes a dyadic word-world relation, and (ii) the relata of this relation have or can have invariant external significance.

The second option for Chomsky would be to claim, that the idea of nominal reference is an *artefact* of the theory (Collins 2004). Let us imagine that internalists state their overall view in a normatively promiscuous language: while the scientifically strict part of their theoretical language is devoted to the analysis of I-language computations, the other, non-scientific part is devoted to the enlightening of the common sense aspects of language. And let us suppose that the second part has been generated as an unavoidable effect of the first part. Then it may be claimed that reference is merely a byproduct of theoretization, not a real component of I-languages. One might wonder, however, how to mitigate the following tension: if 'reference' is an artefact phenomenon, which belongs to the non-scientific vocabulary of internalism, then this term (or expression) has no identifiable semantic features (has no meaning); and if 'reference' has no identifiable semantic features (has no meaning), then it is hard to understand what kind of thing is it that must ultimately be discredited from the scientific discourse of internalism.

Now, it seems that all possible efforts to deny the indispensability of external reference for basic semantic explanations must consist of the following series of argumentative manoeuvres. First, accept the fact that nominals, *in a certain sense*, have the capacity to refer invariantly to language-independent objects. After that deny that external reference, *in that sense*, is the proper object for a serious language science. And, lastly, from the above two contentions as premises infer immediately that external reference is dispensable in a serious language science. The conclusion resulting from these manoeuvres is fallacious. Chomskyan internalists take it for granted that the first two conten-

tions are strongly motivated by the cumulative insights of generative syntax, which are themselves thought of as wholly independent from the issues of reference.[12] But, as we have seen, there is sufficient reason to be skeptical about the existence of such an explanatory independence.

4. *Conclusion*

It is important to stress what it means to answer The Question of Explanatory Indispensability with 'no'. It certainly does not mean that the internalists' approach to language as a whole ought to be seen as indefensible. That would be an overreaction. The negative answer has more moderate consequences. What should be realized, in any case, is that the internalist conception of semantics *in its present form* lacks the appropriate means for arguing against the existence of a unique 'London'/London relation. Externalist views are committed to the existence of word/world relations; consequently they are in a better position to give an adequate account of the nature of reference.

But why should one exclude the possibility that there are alternative ways of developing an internalist semantics? Why not think, for example, that I-language items can be associated with some minimal packages of relational information? Perhaps it can be shown that certain subsystems of the mind/brain interact with the conceptual interface system so that, in the case of nominals, a referential capacity is mandated for producing coherently interpretable computational outputs. This capacity could be postulated as a formal feature which must be valued at a given phase of derivation largely in the same manner as it happens with the traditional nominative feature in the course of case assignment. The only difference would be in the kind of information that the conceptual interface system transfers to generated surface structures. This adjusment, if possible, would indicate an interesting change in theoretical perspective. In this imagined situation internalists could respond with a cautious 'no' to The Question of Explanatory Indispensability without giving up their deeply entrenched view about the scientific primacy of I-languages.[13] The result would be a more powerful internalist semantics. In the end, it might turn out that Chomsky's London poses much fewer theoretical problems than it has been supposed in the past.[14]

[12] We should keep in mind that current generative syntax is a research program, not a theory. However, this otherwise important distinction plays a minor role in the present context. For more on this, see, for example, Chomsky (2007: 4).

[13] Pietroski (2017) makes a somewhat similar conjecture regarding an externally oriented I-language semantics, but he remains unconvinced.

[14] I would like to thank to an anonymous referee for his/her very useful comments on an earlier version of this paper.

References

Boghossian, P. 2003. "Blind Reasoning." *Proceedings of the Aristotelian Society Supplementary Volume* 77: 225–248.

Borg, E. 2009. "Must a Semantic Minimalist be a Semantic internalist?" *Proceedings of the Aristotelian Society Supplementary Volume* 83: 31–51.

Burge, T. 2003. "Psychology and the Environment: Reply to Chomsky." In M. Hahn and B. Ramberg (eds.), *Reflections and Replies. Essays on the Philosophy of Tyler Burge*. Cambridge: MIT Press, 541–470.

Calzavarini, F. 2019. *Brain and the Lexicon*. Cham: Springer.

Chomsky, N. 2000. *New Horizons in the Study of Mind and Language*. Cambridge: Cambridge University Press.

Chomsky, N. 2003a. "Reply to Lycan." In L. M. Antony and N. Hornstein (eds.). *Chomsky and his Critics*. Oxford: Blackwell: 255–263.

Chomsky, N. 2003b. "Reply to Ludlow." In L. M. Antony and N. Hornstein (eds.). *Chomsky and his Critics*. Oxford: Blackwell: 287–295.

Chomsky, N. 2007. "Approaching UG from below." In H-M. Gärtner and U. Sauerland (eds.). *Interfaces + Recursion = Language*? Berlin: De Gruyter: 1–30.

Chomsky, N. 2013. "Notes on Denotation and Denoting." In I. Caponigro and C. Conchetto (eds.). *From Grammar to Meaning the Spontaneous Logicality of Language*. Cambridge: Cambridge University Press: 38–45.

Chomsky, N. 2016. *What Kind of Creatures are We*? New York: Columbia University Press.

Collins, J. 2004. "Faculty Disputes." *Mind & Language* 19: 503–533.

Collins, J. 2009. "Methodology, not Metaphysics: Against Semantic Externalism." *Aristotelian Society Supplementary Volume* 83: 53–69.

Collins, J. 2017. "The Copredication Argument." *Inquiry* 60: 675–702.

Dickie, I. 2015. *Fixing Reference*. Oxford: Oxford University Press.

Gómez-Torrente, M. 2019. *Roads to Reference*. Oxford: Oxford University Press.

Gotham, M. 2017. "Composing Criteria of Individuation in Copredication." *Journal of Semantics* 34: 333–371.

Hinzen, W. 2007. *An Essay on Names and Truth*. Oxford: Oxford University Press.

Kennedy, Ch. and Stanley, J. 2009. "On 'Average'." *Mind* 471: 583–646.

Kertész, A. and Rákosi, Cs. 2012. *Data and Evidence in Linguistics: A Plausible Argumentation Model*. Cambridge: Cambridge University Press.

King, J.C. 2018. "W(h)ither Semantics!(?)." *Nous* 52: 772–795.

Liebesman, D. and Magidor, O. 2017. "Copredication and Property Inheritance." *Philosophical Issues* 27: 131–166.

Liebesman, D. and Magidor, O. 2019. "Copredication, Counting, and Criteria of Individuation: A Response to Gotham." *Journal of Semantics* 36: 549–561.

Ortega-Andrés, M. and Vicente, A. 2019. "Polysemy and Co-Predication." *Glossa: A Journal of General Linguistics,* 4(1).

Pietroski, P. 2017. "Semantic Internalism." In J. McGilvray (ed.). *The Cambridge companion to Chomsky*. Cambridge: Cambridge University Press: 196–216.

Silverberg, A. 1998. "Semantic Externalism: A Response to Chomsky." *Protosociology* 11: 216–244.

Stainton, R. 2006. "Meaning and Reference: Some Chomskian Themes." In E. Lepore and B.C. Smith (eds.). *The Oxford Handbook of Philosophy of Language*. (Oxford: Oxford University Press, 913–940).

Stoljar, D. 2015. "Chomsky, London and Lewis." *Analysis* 75: 16–22.

Vicente, A. 2019. "Chomskyan Arguments against Truth-Conditional Semantics Based on Variability and Co-Predication." *Erkenntnis*, https://doi.org/10.1007/s10670-019-00138-x.

Wright, C. 2004. "Intuition, Entitlement and the Epistemology of Logical Laws." *Dialectica* 58: 155–175.

Wright, C. 2018. "Logical Non-Cognitivism." *Philosophical Issues* 28: 425–450.

Croatian Journal of Philosophy
Vol. XXI, No. 62, 2021
https://doi.org/ 10.52685/cjp.21.62.5
Received: September 30, 2020
Accepted: December 14, 2020

Two Accounts of the Problem of Enhanced Control

DAMIR ČIČIĆ*
Belgrade, Serbia

According to event-causal libertarianism, an action is free in the sense relevant to moral responsibility when it is caused indeterministically by an agent's beliefs, desires, intentions, or by their occurrences. This paper attempts to clarify one of the major objections to this theory: the objection that the theory cannot explain the relevance of indeterminism to this kind of freedom (known as free will). Christopher Evan Franklin (2011, 2018) has argued that the problem of explaining the relevance of indeterminism to free will (which he calls "the problem of enhanced control") arises because it is difficult to see how indeterminism could enhance our abilities, and disappears when we realize that beside the relevant abilities free will requires opportunities. In this paper, I argue that the problem occurs not because of the focus on abilities, but because of the difficulty to explain how indeterminism could contribute to the satisfaction of the sourcehood condition of free will in the framework of event-causal theory of action.

Keywords: free will, indeterminism, control, event-causal libertarianism, ability to do otherwise, alternative possibilities, sourcehood.

1. *Introduction*

According to the 'standard version' of event-causal libertarianism, a person acts with free will when the person's action is caused indeter-

* I would like to express my deep gratitude to Professor Gary Watson for his encouragement and conversations about the ideas I present here. I would also like to thank Professor Taylor Cyr, Dr. Maria Sekatskaya, and Dr. Münir Cem Kayaligil for their comments on earlief r versions of this paper.

ministically (and non-deviantly[1]) by his or her mental states of the relevant kind (his or her desires, beliefs, intentions or their occurrences).[2] This theory is appealing to many philosophers because it combines incompatibilism (the thesis that free will requires the falsity of determinism) with a metaphysically modest theory of human agency—the event-causal theory.[3] However, many philosophers find it problematic, among other reasons, because it seems unable to explain adequately the relevance of indeterminism to free will (control necessary for moral responsibility) (Pereboom 2001: 49; Clark 2003). According to event-causal libertarian theory, agents in *deterministic* worlds don't have free will because of determinism, *even if* their actions are caused by their beliefs and desires. However, if the causal relations between their mental states and actions are indeterministic—the theory says—their actions *are* free actions. Many critics find this claim difficult to accept. They find it incredible that mere chance in the causal relation between one's mental states and actions could account for the difference between an unfree and a free action. They ask the following question: how could the mere *possibility* of an agent's mental states causing an alternative decision entail that the *agent* exercised *more* control (than he or she would have exercised otherwise) in making the decision he or she actually made?

Christopher Evan Franklin (2011, 2018) has offered an interesting reply to this challenge which he named "the problem of enhanced control". According to Franklin, critics of event-causal libertarianism are right that indeterminism *by itself* cannot enhance control: no one has more control simply because it is *possible* that he or she will do otherwise. However, in his view, indeterminism is not the only difference between free and unfree agency on event-causal libertarian theory. For in virtue of indeterminism, he argues, event-causal libertarian agents

[1] A piece of behavior could be caused by an agent's mental states in a way that undermines its being a free and intentional action. Thus, an agent may desire very much to do something and believe that she can do it by moving her body in a certain way, and that desire and belief could cause her body to move in that way. However, the relevant bodily movement would not be an intentional and free action if it were caused by the desire and belief via a causal chain which involves a reflex reaction of the agent's motor system triggered by the agent's excitement at the prospect of performing the action in question (see Davidson 1980: 78–79).

[2] Another version of this theory locates indeterminism in the processes that give rise to beliefs and desires (see Clarke 2003: Ch. 4).

[3] This is so because it explains free action in terms of the causal relations between events, which is the dominant way of explaining processes in nature, and because causation is widely regarded as a relation between events. In contrast, the agent-causal libertarian theory, according to which free actions are caused directly by agents as substances, is regarded by many philosophers as obscure and empirically implausible, or even as incoherent (Clark 2003: Ch. 10; Pereboom 2001: Ch. 3). For arguments in favor of the agent-causal theory see Taylor (1966) and O'Connor (2000). For arguments in favor of event-causal libertarianism see (Kane 1996; Balaguer 2004; Mele 2006: Ch. 5; Van Inwagen 1983: 137–50;).

have the *opportunity* to do otherwise which compatibilist agents lack.[4] This is significant, he claims, because the opportunity in question allows the former agents to exercise their ability to do otherwise and thus control how their lives unfold.[5] In his view, critics have failed to notice this due to their focus on *abilities:* not being able to see how indeterminism could provide new abilities, they concluded that *indeterminism* constitutes the only difference between event-causal libertarianism and compatibilism, and that indeterminism *per se* enhances control according to the former theory.

However, is the focus on abilities really what has led the critics to this conclusion? And, can the problem of enhanced control be solved by invoking the opportunity to do otherwise? I argue in this paper that the answers to these questions are negative. I argue that the problem of enhanced control does not arise because of the focus on abilities in general, or the ability to do otherwise in particular, but because it is not clear how indeterminism by itself, or *via* indeterministic alternative possibilities (ability and opportunity to do otherwise), contributes to the satisfaction of the sourcehood condition of free will—the condition which says that in order to have free will one must be an appropriate source of action.[6]

I begin my discussion of the problem of enhanced control by presenting Harry Frankfurt's famous argument against the relevance of alternative possibilities to moral responsibility. I do that because his argument rests on certain assumptions about free will—assumptions involving sourcehood—that, in my view, give rise to this problem. By considering event-causal libertarianism in the light of those assumptions, I provide an account of the problem that puts emphasis on the notion of sourcehood.[7] Next, I compare my account with Franklin's

[4] Franklin provides an argument for the incompatibility of determinism and the opportunity to do otherwise which he calls "The No Opportunity Argument" (Franklin 2011: 699–705). Van Miltenburg and Ometto (2016: 9–11) present an objection to his argument.

[5] Franklin suggests that the word "free" in the expression "free will" refers to the relevant opportunity, whereas "will" refers to the relevant ability (or abilities) (Franklin 2018: 60).

[6] The problem in the focus of this paper is an instance of a "general problem of enhanced control" that threatens any theory of free will in which mere indeterminism makes a difference between free and unfree agency. Richard Taylor's agent-causal libertarian theory, for example, is a theory of this type. Consequently, it faces the problem of enhanced control too (see Gary Watson 2004b: 193–194). My analysis of the problem as applied to event-causal libertarianism was inspired by Gary Watson's thesis that any theory which considers alternative possibilities and sourcehood (self-determination in Watson's terminology) as logically independent conditions, is bound to fail (Watson 2004b: 193).

[7] What exactly is the sourcehood condition? As I mentioned above, it is a condition according to which free will requires that one is, or that one has the power to be, an appropriate source of action (in the sense relevant to freedom and responsibility). But what does that mean? Perhaps the best way to answer this question involves giving examples of failures to be such a source. It is clear, for instance, that when one moves as a result of being pushed by another person or does something as a result of

(in key respects) and argue that my account explains better than his the worry that indeterminism constitutes the only difference between event-causal libertarianism and compatibilism. Finally, I consider potential objections to my account and reveal an important connection between the problem of enhanced control and the debate about the possibility of control over "passive omissions".[8]

2. *Indeterminism as a Frankfurt-style intervener*

Central to Frankfurt's argument against the relevance of alternative possibilities to moral responsibility is the following example:

> Suppose someone—Black, let us say—wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way… Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. (Frankfurt 2003/1969: 21–22)

This scenario supposedly supports the thesis that someone or something could deprive an agent of alternative possibilities while being totally irrelevant to the explanation of the agent's behavior. Because of Black, Jones could not have done otherwise, but that has no relevance

an irresistible desire, one does not originate the movement and the action in question, in the way necessary for being morally responsible. The same is the case when one acts under some kind of hypnosis or as a result of certain kinds of manipulation. In those cases one does not have control necessary for moral responsibility because one does not determine for oneself what one does, and one is not a true author of one's action. Thus, to have control necessary for moral responsibility one must satisfy the sourcehood condition, and to satisfy that condition certain factors *must not play* a role in the origination of one's action. What exactly those factors are, do they include determinism, what factors *must play* a role in the origination of one's action, and what role must they play for the sourcehood condition to be satisfied are difficult questions. Agent-causal libertarians, for instance, argue that factors which play a role in origination necessarily include the agent-causal power, while event-causal libertarians argue that those factors must involve only certain psychological events. For an in-depth analysis of the notion of sourcehood see Timpe (2012).

[8] Ishtiyaque Haji has offered yet another explanation of the worry that event-causal libertarianism cannot provide an account of enhanced control (Haji 2013: 227–246). According to Haji, this theory cannot provide an adequate account of enhanced control because indeterminism *diminishes* control. However, I will not discuss this explanation here for two reasons. First, if the account of the problem of enhanced control I will present is correct, there is no need to appeal to the alleged control-diminishing effects of indeterminism to explain that problem. Second, Haji's explanation does not concern the worry that indeterminism is the only difference between event-causal libertarianism and compatibilism, and, explaining that worry is one of my main goals in this paper.

to the explanation of Jones' action. For Jones performed the action on his own, for his own reasons.

Why is this important to the question about the relevance of alternative possibilities to moral responsibility? It is important, Frankfurt points out, because to be relevant to moral responsibility, something must bear on, or explain what actually happens, that is, explain how an action occurs (2006/1969: 23). This principle, together with the example above, leads to the conclusion that having alternative possibilities is irrelevant to moral responsibility. That is so because, if the example is coherent, that is, if Black's depriving Jones of alternative possibilities really has no relevance to the explanation of Jones' behavior, given this principle, Black's activity, and consequently, Jones' lack of alternative possibilities, are irrelevant to Jones' moral responsibility.

The fact that the principle in question seems intuitively very plausible, explains why the debate about Frankfurt's argument has been focused on the coherence of his example.[9] However, the same principle, or more precisely, a version of the principle in which the phrase "moral responsibility" is replaced by the phrase "free will", explains why the problem of enhanced control is so difficult, and why the reference to alternative possibilities (the ability/opportunity to do otherwise) cannot solve it.[10] In the light of this principle, which I will call "E" (as a shorthand for the phrase "Explanatory Principle"), an alternative possibility (AP) is relevant to *free will* only if it plays some role in the production of an action (and thus explains the occurrence of that action). Thus, according to this principle, an AP that indeterminism provides according to even-causal libertarianism is relevant to free will only if it is relevantly different from Frankfurt's Black. But, is that the case? In my view, the answer is "no". For just like Black in Frankfurt's example, an indeterministic AP plays no role in the production of an event-causal libertarian action since it has no bearing on the reasons for which the action was performed, or on the powers exercised in its performance.[11]

[9] For an objection to this principle see Widerker (2003: 61–62), and Ginet (2003/1996: 82–83). I present and respond to the objection in question in section 4.

[10] The following argumentation does not depend on the soundness of Frankfurt's argument (which I actually don't find convincing). It relies only on the plausibility of the principle in question (which I find very convincing). I don't find Frankfurt's argument convincing because I don't believe that a Frankfurt-style scenario—a scenario in which something deprives an agent of alternative possibilities without having an effect on what actually happens—is in fact possible. For this reason, I don't consider all the objections to Frankfurt's argument in this paper, but only those that question the relevant principle. For an analysis of the debate concerning Frankfurt's argument see Widerker and McKenna (2003: 1–13).

[11] Alternative possibilities that agents *take into account* when deliberating do bear on the reasons for which actions are performed. Thus, one might say that he or she chose to perform an action because he or she found the action better than an alternative action. However, the fact that an alternative is *indeterministic* adds nothing to the reasons one would otherwise have. The reasons we have for actions (the reasons we consider in deliberation, not the objective reasons) depend on what

That is so because an indeterministic AP, on event-causal libertarian theory, has no (positive) effect on what actually happens (i.e. it doesn't provide new causal powers or new reasons for performing actions).[12] Therefore, according to E, there is no relevant difference between an event-causal libertarian AP and Frankfurt's Black. In other words, E suggests that event-causal libertarian AP-s (the ability and opportunity to do otherwise) are irrelevant to free will.[13]

One could object, however, that, perhaps unlike Frankfurt's Black, an event-causal libertarian AP plays a role in the production of an action, although only a negative one: it implies that the causes of an action did not have to be its causes, that is, did not have to cause it. However, this does not make the AP in question relevantly different from Black (in the light of E). For what makes E plausible is the notion that having free will means being a true or appropriate *source* of action. This notion, which is one of the central notions in the contemporary free will debate, implies that in order for something to be relevant to free will, that something must contribute to one's being an appropriate source of actions. And, to do that, it seems that the thing in question must contribute to one's being (more) involved in the *production* of one's actions. However, the fact that one's actions merely lack deterministic causes implies nothing about one's involvement in the production of those actions.[14] Consequently, in the light of E, the fact

we believe the alternatives are, not on their metaphysical properties.

[12] Robert Kane has argued that indeterminism contributes to sourcehood by making agents "the ultimate creators (or originators) and sustainers of their own ends and purposes" (Kane 1996: 23). However, as Randolph Clarke has noticed, this 'ultimacy' "opens alternatives but does not secure for the agent the exercise of any further positive powers to causally influence which of the alternatives left open by previous events will be made actual" (Clarke 2003: 108). I will discuss Kane's view in more depth in section 4.

[13] An anonymous reviewer has objected that "a libertarian (event-causal or otherwise) would reject" the principle E, and that "the principle just seems to be a rejection of libertarianism—and as such, seems to beg the question against Franklin." However, not all versions of libertarianism are incompatible with the principle E. The traditional (or standard) versions of agent-causal libertarianism, for example, *are* compatible with this principle. According to the theories of this type, freedom relevant to moral responsibility requires the power of agents to cause their actions *qua* agents—the agent-causal power. And having this power, or being able to exercise it, requires the absence of determination of agents' actions by prior events, that is, it requires the existence of indeterministic alternative possibilities. Now, since the agent-causal power explains how (free) actions occur in the agent-causal libertarian framework (they are exercises of, or results of the exercises of, the agent-causal power), and since the alternative possibilities contribute to the existence of that power, alternative possibilities also explain (free) actions in that framework (indirectly). Therefore, traditional agent-causal libertarian indeterministic alternative possibilities satisfy the principle E. For a useful discussion of agent-causal libertarianism see (Nelkin 2011: Ch. 4).

[14] Derk Pereboom has made a similar point: "We have already seen that by incompatibilist intuitions, an agent is not responsible for decisions determined by factors beyond his control. However, if these factors, rather than determining a

that indeterminism bears on what happens negatively, makes indeterminism no different from Black, that is, it has no relevance to whether an action was performed freely.

Therefore, to understand why indeterminism seems irrelevant to free will (the control necessary for moral responsibility) in the event-causal libertarian setting, we do not need to consider the relation between indeterminism and abilities in general, as Franklin suggests. It is sufficient to pay attention to the idea of sourcehood. When we do that, indeterminism appears irrelevant to event-causal libertarian free will just as Frankfurt's "counterfactual intervener" Black is irrelevant to Jones' moral responsibility. For just like Black in Frankfurt's example, indeterminism in the event-causal framework has no impact on the satisfaction of the sourcehood condition of freedom of the will.

In what follows, I argue that this account of the problem of enhanced control is superior to Franklin's because it explains better than his account the worry that indeterminism constitutes the *only* difference between event-causal libertarianism and compatibilism.

3. *No new abilities vs no difference in sources*

As I mentioned in the introduction, Franklin suggests that critics of event-causal libertarianism perceive indeterminism as the only difference between event-causal libertarianism and compatibilism because they cannot see how indeterminism could provide new abilities. But is it obvious that indeterminism could not provide new abilities in the framework of the event-causal theory of action? In my view, it is not so because it does not seem inconceivable that indeterminism could do that, and because it is plausible to think that changes in the laws of nature could affect our abilities.

Franklin believes that indeterminism cannot contribute to abilities because he believes that abilities supervene only on intrinsic properties of agents, i.e. the ways they are constituted, and because indeterminism (or determinism for that matter) does not bear on intrinsic properties (Franklin 2011: 694–5; 2018: 59–72). However, this view is problematic for two reasons. First, as I said, it is plausible to think that abilities depend on the laws of nature, and the laws of nature are not intrinsic to agents. Franklin could respond that abilities supervene on intrinsic properties *holding the laws of nature fixed*. However, it is not clear if this reply is legitimate in the present context because we clearly do not hold fixed the laws of nature in discussions of the effects of indeterminism on abilities (i.e. the effects of *replacing* deterministic laws with indeterministic laws or vice versa). Second, the existence of some of our abilities could depend on our *environment* in the way in which some of our dispositions depend on our environment. For ex-

single decision, simply leave open more than one possibility, and the agent plays no further role in determining which possibility is realized, then we have no more reason to hold him responsible than we do in deterministic case (Pereboom 2001: 48).

ample, the existence of my key's disposition to open my lock seems to depend on the existence of my lock. That seems to be the case because the content of the disposition in question refers to my lock (Shoemaker 1980). Similarly, our abilities could depend on external things to which their contents are referring.[15]

The standard response to this argument is that my key does not have the disposition to open my lock, but rather the disposition to open locks of a kind to which my lock belongs (Molnar 2003). Franklin adopts a similar strategy when it comes to abilities. However, he admits that it is not obvious that this reply settles the dispute (Franklin 2018: 63).

Thus, it is not clear whether indeterminism could provide new abilities or not. Certainly, the appeal to intrinsic nature of abilities cannot settle this issue (at least, not in the current state of the debate). Consequently, when we focus on the relation between indeterminism and abilities in general, we don't see clearly that indeterminism is the only difference between event-causal libertarianism and compatibilism. For it is not obvious that indeterminism could not provide an agent with new abilities. However, when we compare the sources of event-causal libertarian and compatibilist actions we see immediately that indeterminism is the only difference between the types of agency in question. For the only difference between the sources of the event-causal libertarian and compatibilist actions is the absence of deterministic causes in the causal histories of the former.

That said, it *is* perhaps the case that some critics worry *both* about the alternative possibilities *and* sourcehood when it comes to the event-causal libertarianism. They worry that adding indeterminism to other conditions postulated by this theory cannot produce *any* new *powers* relevant to freedom and moral responsibility. Nevertheless, defining the problem of enhanced control simply in terms of powers would be wrong. For even if one could show that indeterminism provides the ability to do otherwise in the event-causal framework, or the opportunity to exercise it, it would remain unclear how that ability or opportunity could enhance an agents control over his or her actions. For it is not clear how the combination of the ability and the opportunity in question could make an agent a more appropriate source of action.[16]

Let us now look at some potential objections to this account.

[15] Ferenc Huoranszki argues that abilities are extrinsic (2011: Ch. 3). However, Huoranszki also argues that our abilities do not depend on whether the laws of nature are indeterministic or deterministic (2011: Ch. 2).

[16] Although I have argued that it is not obvious that indeterminism could not provide the ability to do otherwise in the event-causal framework, I am inclined to think that it could *not* in fact do that. I believe that the ability to do otherwise *is* significant to free will. If indeterminism provided the ability to do otherwise it would surely enhance control. But, indeterminism in the event causal framework could not enhance control, for the reasons I present in this paper. Thus, by *reductio ad absurdum*, indeterminism cannot provide the ability to do otherwise in the relevant framework.

4. *Objections*

I have argued that the focus on *sources* of event-causal libertarian and compatibilist actions allows us to see clearly that indeterminism cannot provide enhanced control. However, some philosophers find this claim problematic. They argue that the absence of deterministic causes in the causal histories of actions is significant because, in virtue of their absence, event-causal libertarian actions are not mere outcomes of the past and the laws of nature, and that such agents are not "pushed" or "manipulated" by the laws and the past.[17]

This response, however, does not help event-causal libertarianism. For to say that event-causal libertarian actions are not mere outcomes of the past and the laws of nature suggests that compatibilist actions *are* mere outcomes of the past and the laws. In other words, it suggests that compatibilist agents are not really agents but mere transmitters of external influences. However, this suggestion is misleading since, according to the event-causal theory of action, compatibilist agents are agents for the same reason for which event-causal libertarian agents are agents: their behavior is caused non-deviantly by the relevant mental states.[18] On the other hand, to say that, due to indeterminism, event-causal libertarian agents are not pushed by the past and the laws of nature is useless to event-causal libertarians. For, as Randolph Clarke has pointed out, the fact that agents are not pushed by anything external, does not entail that they are more pushed by themselves (Clarke 2003: 106). Similarly, it is clear that the fact that they are not manipulated by the laws and the past (if "manipulation" just means causal determination) does not entail that they are more "manipulated by" or "controlled" by themselves.

[17] For the claim that, given determinism, "the agent is pushed by previous events into preferring and acting exactly as she does at each moment", see Ekstrom (2000: 190). For the claim that, given indeterminism, actions are not mere outcomes of the past and the laws of nature, see Haji (2013).

[18] An anonymous referee has observed that this "is, in effect, a one-line rejection of the so-called Consequence Argument (Van Inwagen), the point of which is precisely to show that under determinism (but not under indeterminism), actions are mere outcomes of the past and the laws." My response to this observation is that, if the point of the Consequence Argument were really what the referee claims it is, my argument would indeed constitute its refutation. For even under the assumption that determinism is incompatible with free will, it is plausible to make a distinction between an intentional action and something that merely happens to someone. For example, it is plausible to say that it is not the same thing if I move to another side of the street as a result of my intention to cross the street, and if that happens as a result of me being pushed by a strong wind (even if determinism precludes free will). For a similar point see (Huoranszki 2011: 13). However, there are other, more plausible, interpretations of the Consequence Argument that are immune to my "one-liner". One of those interpretations is an interpretation by Van Inwagen, according to which, the argument shows that under determinism, *no one can do otherwise* (Van Inwagen 1986: 68).

An event-causal libertarian could respond to this that the absence of manipulation or pushing provides the enhanced control, not by itself, but in conjunction with the alternative possibilities that indeterminism provides. For instance, he or she could argue that the absence of manipulation and pushing, together with genuinely open alternatives, enhances agents' control by making agents the *ultimate* origins and *true* authors of their actions (Haji 2013: 239–40). However, it is not clear that being an ultimate origin or a true author in this sense implies enhanced control. For, again, it is not clear how the alternative possibilities open to event-causal libertarian agents by themselves, or in conjunction with the absence of manipulation and pushing, contribute to the satisfaction of the sourcehood condition of free will.[19]

To check this claim let us consider Robert Kane's view of free will, which is, arguably, the most sophisticated event-causal libertarian view. According to his view, the "ultimate responsibility" for actions derives from the existence of a special type of undetermined decisions which he calls the "self-forming actions" or "SFA-s" (he calls them so because in making such decisions we are supposedly setting our goals and shaping our characters). We make these decisions (assuming that indeterminism is true), when we are torn between conflicting motives (e. g. moral reasons and egoistic inclinations, or "present desires and long-term goals") and we have to make an effort to end the state of indecision. Decisions we make in such circumstances are rational (made for reasons) and voluntary, whether we end up pursuing moral reasons and long term goals, or egoistic reasons and our present desires (Kane 2007: 26). Because of that, Kane argues, indeterminism involved in SFA's, does not diminish our control over our actions, as some philosophers think. However, does it enhance our control? In my view, the answer is "no". The key thing to observe here is that the state of being torn between conflicting options, where one has to make an effort to overcome indecision, does not require indeterminism. An agent in a deterministic world could be in a psychological state identical to that of an agent performing an SFA. So, the old question arises again: how could mere indeterminism give one more control over his or her decisions?[20]

[19] This is so because it is not clear how the combination of these factors could make an agent more involved in the origination of his or her actions. The factors in question do not give agents new powers to produce actions nor enable them to exercise the existing powers of that kind. In addition, they do not create new causes of actions or new reasons for actions. For the event-causal libertarian theory, unlike the traditional agent-causal theory, does not postulate the existence of any special power to originate actions (nor any reason or causal feature) that could not be possessed by an agent in a deterministic world.

[20] Daniel Dennett has made a similar point in his discussion of Kane's theory: "An indeterministic spark occurring at the moment we make our most important decisions couldn't make us more flexible, give us more opportunities, make us more self-made or autonomous in any way that could be discerned from *inside or outside,* so why should it matter to us? How could it be a difference that makes a difference?

One could argue that to understand Kane's answer to this question we have to consider his view in more detail. One could point out that, according to Kane, determinism is not just a mere addition to other conditions of free will (in particular the effort to make a decision), because it is *fused* with the effort to make a decision (or choice): "[t]he choice one way or the other is undetermined because the process preceding and potentially terminating in it (i.e. the effort of will to overcome temptation) is indeterminate" (Kane 1996: 128). Moreover, according to Kane, when making a free decision, the agent is making a dual (or double) effort; at the same time the agent is trying to make two incompatible decisions for which he or she has conflicting motivations. Thus, one could argue that Kane's theory explains how indeterminism produces a qualitative change in the origins of actions, thus contributing to the satisfaction of the sourcehood condition of free will.

However, the appeal to indeterminacy and doubling of effort does not explain how event-causal libertarian agents could be more involved in the production of their actions in comparison to their compatibilist counterparts. For, as Randolph Clarke has pointed out, indeterminate efforts explain free decisions only if they themselves are free. However, Kane's theory does not explain how indeterminism makes the later free (Clarke 2003: 89–90). Besides, it is questionable whether trying to make a specific decision (as opposed to trying to make *some* decision), and simultaneously trying to make an incompatible decision is psychologically possible, and whether it would aid or threaten freedom (Clarke 2003: 88).[21]

Now, one might wonder why indeterminism has to contribute to sourcehood in the event-causal framework in order to enhance control?

(Dennett 2003: 136) Consider also the following passage by Randolph Clarke: "But now, when we imaginatively compare this deterministic agent with another who differs only in that in the production of her decision there is the indeterminacy and indeterminism required by Kane, and we suppose that the former does not act with the freedom that would (given an ordinary moral capacity) suffice for responsibility, then we must say that the latter does not either. The required indeterminacy and indeterminism give us an ultimacy in the latter case that is not there in the former. But this ultimacy is wholly negative: it is merely the absence of any deterministic cause of the decision. It opens alternatives but does not secure for the agent the exercise of any further positive powers to causally influence which of the alternatives left open by previous events will be made actual" (Clarke 2003: 107–108). For a similar view see (Watson 2004a: 206).

[21] Mark Balaguer has developed a version of event-causal libertarianism that is also worth considering (see Balaguer 2010). However, in aspects key to the debate about the problem of enhanced control his account does not differ significantly from Kane's account. Like Kane's theory, his theory explains at best why indeterminism does not have to diminish control we have over our actions, but does not show how indeterminism could allow us to be more involved in the production of our actions. Surely, both of these theories deserve more attention, but I do not discuss them further here because I find another threat to my account more serious (the omissions-involving counterexamples to the principle E), and because I find the objections to Kane's view presented here very convincing.

Couldn't a "compatibilist sourcehood" in conjunction with alternative possibilities provided by indeterminism be sufficient for enhanced control? The answer is "no" because event-causal libertarian alternative possibilities are irrelevant to free will according to the principle E. But, why should an event-causal libertarian accept that principle? There are two reasons for that. First, the principle E is intuitively very plausible. To see this, one just has to consider one's own intuitions about Frankfurt's Black. Second, there are no good arguments against E. One way in which some philosophers have tried to challenge it is by pointing out that things which do not help to explain why an action was *omitted,* may be relevant to whether that action was *omitted* freely. This challenge is based on the assumption that control over actions and control over omissions are essentially the same thing. In other words, some philosophers have tried to challenge E by making a parallel between actions and omissions. Consider in that context the following example:

> *Broken Phone*: Smith witnesses a man being mugged outside his building. He knows he could easily dial 911, but, not wanting to be inconvenienced, decides to let sleeping dogs lie. Unbeknownst to Smith, however, and through no fault of his own, his telephone was not working. So he could not have called the police even if he had tried.[22] (Capes and Swenson 2017: 974)

The fact that the phone was broken made it impossible for Smith to call the police, but it played no role in Smith's decision to omit calling the police. Nevertheless, it seems clear that due to this fact, Smith is not responsible for his omission to call the police and did not freely omit doing it. For, in the given circumstances, it would clearly be wrong to demand of Smith to call the police and later blame him for not doing so, even though it would not be wrong to demand of him to *try* to call the police and blame him for failing to try. Therefore, the principle E appears to be false when it comes to omissions: it seems that something *can* be relevant to whether an action was omitted freely even though it did not actually play a role in the omitting. And that begs the following question: if freedom and responsibility are the same phenomenon when it comes to actions and omissions, is it really the case that something must play an explanatory role when it comes to actions to be relevant for freedom and responsibility?

To answer this question, it is important to notice that Smith's omission to call the police is not a *basic* omission. It is an omission which is

[22] This is a version of an example due to Van Inwagen (1983: 165–166). Neither Van Inwagen nor Swenson and Capes, however, use this example to challenge Frankfurt's version of the principle E, at least, not explicitly. Van Inwagen presents the example to show that Frankfurt-style examples do not generate the intuition that responsibility does not require alternative possibilities when it comes to *omissions*. Capes and Swenson use it to highlight the existence of luck in cases of this type, which in turn serves their defense of the 'flicker of freedom' response to Frankfurt's argument. On the other hand, Ginet (2003/1996: 82–84) and Widerker (2003: 61–62) explicitly consider similar examples as challenges to E, or more precisely, Frankfurt's version of E. However, I chose this example because of its simplicity. I will say a bit more about Ginet's and Widerker's examples in footnote 24.

a result of an action—a decision to omit.[23] This is important because in discussions about free will and moral responsibility we are not interested primarily in non-basic actions and omissions, but in basic actions and omissions. For those discussions are focused on whether we have the control that constitutes free will, rather than on the scope of that control. And, having such control is possible only if we can have control over our basic actions or omissions. This raises another question: is a basic omission version of the above scenario, a scenario in which freedom and responsibility depend on some factor which does not actually play a role in 'bringing about' a *basic* omission, possible? To the best of my knowledge, nobody has yet presented such a scenario.[24] And, it is hard to imagine a coherent scenario of that sort for two reasons. First, such an example would have to be a case of someone omitting

[23] With this in mind, it shouldn't be surprising that Smith's responsibility for the omission to call the police depends on factors that did not lead to his omission. For what we can control by our decisions depends on the cooperation of our environment, which may or may not influence our reasons for doing what we do. Moreover, what we do and perhaps what we omit doing by making decisions is a result of that cooperation. John Martin Fischer makes a similar point in the following passage: "So what the agent is morally responsible for depends on things which are quite extraneous to the agent—for instance the state of the telephone line. This might seem to introduce an unacceptable kind of 'moral luck' into our system of evaluating agents. But, whereas our way of specifying what Smith is responsible for depends on such factors, a certain moral evaluation of Smith does not. Smith would be equally morally blameworthy in either case, and it would seem appropriate to apply the same kind of punishment (or blame) in both cases. So, whereas a certain kind of moral luck applies to the specification of the content of moral responsibility, it does not apply to the extent or degree of blameworthiness, and it does not apply to the evaluation of agents" (Fischer 1986: 256).

[24] Ginet has presented an example which, at first sight, seems like such a scenario (2003/1996: 83–84). In his example a person omits an action without deciding to omit, and there is a device in her brain which monitors her neural processes ready to manipulate those processes if the person were to decide to perform the omitted action or begin to try to perform that action. This example is not relevant because it does not even attempt to show that the person has no alternative to basic omitting (omitting without deciding to omit). For, in Ginet's example, the device does not eliminate the possibility of the agent deciding to perform the omitted action. However, even if the device did do that, such an example would not be effective. For, even if something eliminated the possibility of an agent deciding not to omit the relevant action, if it played no role in the agent's omission, it would seem irrelevant to the agent's free will (control necessary for moral responsibility). Widerker, on the other hand, has showed that some factors which play no role in the explanation of our action or omission may be relevant to our moral assessment (2003: 61–62). Thus, he has showed that one's awareness that an alternative is morally better, or the existence of objective reason for performing or omitting an action (of which one may not be aware) may affect the person's responsibility (for non-basic omissions and actions) and blameworthiness without playing a role in the explanation of the relevant action or omission. This is an excellent observation. However, it is irrelevant to the principle E. For the principle E concerns free will (control necessary for moral responsibility), and factors in question do not affect a person's control over his or her action or omission. Moreover, these factors also don't seem to affect one's responsibility but only the content of one's responsibility.

to do something without deciding to omit (deciding to omit in such a case would also count as doing otherwise), and without being able to decide to omit due to the presence of a counterfactual intervener or mechanism (intervener or mechanism that does not actually intervene but would intervene under certain "counterfactual" circumstances). To show that the principle E is false when it comes to basic omissions, example of this sort would have to generate the intuition that the person in question is *not* responsible for the omission (and did not omit freely). However, if an omission was not caused by an external factor, it is hard to see why the person (which is otherwise free) did not omit freely (i.e. exercised control in omitting). Second, the idea that it is possible to have control over a basic omission (the idea of free basic omission) is problematic. For, as Michael Zimmerman (2017: 101) has observed, *having* control over something seems to entail having the ability to *exercise* control over that something. And, since an exercise of control is an activity, and basic omissions do not involve any activity[25], it is not clear how one could be able to exercise (direct) control over a basic omission. Consequently, it is not clear how one could *have* control over such an omission.

A natural response to this argument is that having control over a basic omission does not actually require having the ability to *exercise* control over a basic omission, but only the ability (and opportunity) to perform the omitted action (or make a decision to perform that action).[26] To evaluate this response, consider the following scenario. You are on a freeway riding in a self-driving car programmed to go to a certain destination. You notice that your car is approaching an exit, and you have the ability to take the steering wheel and make a turn. However, you don't do that and do not make a decision to turn, and as a result you continue cruising down the freeway. If this happened, one could argue that your control over your omission in this situation consisted in your ability to make a turn (or to decide to make a turn). In addition, one could argue that the principle E is false because the ability to make a turn in this case was relevant to your control over the omission even though it played no role in the "origination" of the omission.

The problem with this response, however, is that it is actually hard to see how you could have had control over your omission in this scenario. To see why this is the case, notice that there seem to be only three possible explanations of your omission. First is that while deliberating about what to do, you missed the chance to turn, and thus lost the opportunity to decide. Second is that you *couldn't* really decide (like Buridan's ass) because you were torn by conflicting motives of equal strength. Finally, it could be that you were able to decide and were

[25] If an omission involved an activity such as a decision to omit or a trying to omit, it could be argued that it is not a basic omission, but an omission that derives from such an activity.

[26] Dana Key Nelkin and Samuel Rickless hold this view (see Nelkin and Rickless 2017: 120–132).

fully aware of the timing, but instead of deliberating about what to do, you simply waited for an urge to turn to occur in you, and since that urge did not occur, you continued cruising.[27] Given any of these explanations, you lacked (direct) control over your omission. Given the first explanation, that is so because, on that explanation, the above story does not even involve a basic omission. For, on that explanation, the omission was a result of an overextended deliberation about what to do (i.e. it was a result of an action). Given the second explanation, you also lacked control over your omission, for given that explanation, your omission resulted from an inability to decide, and, clearly, an omission cannot be under one's control if it results from one's inability to perform the omitted action. Finally, given the third explanation, you lacked direct control over your omission because it was not up to you whether you would perform or omit the relevant action, and that was the case because we do not have *direct* control over urges that may or may not occur in us.[28]

Thus, it is hard to see how one could have direct control over a basic omission in virtue of having the ability to perform the omitted action. Consequently, examples that supposedly involve such control are problematic and do not support the claim that the ability to do otherwise may be relevant to one's control over an omission (and on the assumption of equivalence between control over omissions and actions to actions) even though it plays no role in the explanation of the relevant omission. In addition, as we saw earlier, Frankfurt-style omission cases represent potential counterexamples to the principle E only when it comes to non-basic omissions. Therefore, instead of giving us reason to abandon the principle E, considerations of omissions only give us reason to modify (or clarify) it by adding that it applies only to basic actions and omissions.

[27] Perhaps there is some other possibility that I cannot see currently. I invite you to consider this issue.

[28] Randolph Clarke says the following in support of the thesis that there are free basic omissions for which we can be responsible: "...when I refrain from touching the freshly painted object, I need not have decided not to touch it, and there need be no action at all that is my not touching it. Nevertheless, I might freely refrain from touching it. And it certainly seems that I can be responsible—perhaps praiseworthy—for not touching the object and messing up the paint job. In this case, its seems, we might have something—an intentional omission or refraining—that isn't an action of any kind and for which one can be directly responsible (2014: 108–109). I find Clarke's view problematic for the reasons pertaining to the explanation of omission which I mentioned in the discussion of the car example in the text above. In particular, my worry is that Clarke's omission in his example is not free because his intention to omit either has its source in some earlier decision of his, or is something that just occurs in him—something over which he has no control. When we make decisions, in contrast, we are not just automatically following the urges that occur in us, we are considering reasons for and against certain action and consciously accepting or not accepting to act on certain motives. For a different objection to Clarke's view see (Zimmerman 2015: 366).

5. *Conclusion*

When we consider carefully the claim that indeterminism constitutes the only difference between event-causal libertarianism and compatibilism, we see that the problem of enhanced control occurs not because of the worries concerning the abilities in general or the ability to do otherwise in particular, but because of the worries related to sourcehood. More precisely, we see that what generates the problem is the worry that, in the event-causal framework, indeterminism plays no positive role in the processes that lead to actions. In addition, we see that this fact about indeterminism represents a problem because of the principle E, that is, because it seems necessary that something plays a role in the processes that lead to actions in order to be relevant to control. Some scenarios involving omissions suggest, at first sight, that E is in fact false. However, those scenarios either don't concern control necessary for moral responsibility, because they don't involve basic omissions, or seem to lack coherence because it is not clear that having control over basic omissions is possible.

References

Balaguer, M. 2004. "A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will." *Noûs* 38: 379–406.

_____2010. *Free Will as an Open Scientific Problem*. Cambridge: MIT Press.

Capes, A. J. and Swenson, P. 2016. "Frankfurt cases: the fine-grained response revisited." *Philosophical Studies* 174: 967–981.

Clark, R. 2003. *Libertarian Accounts of Free Will*. New York: Oxford University Press.

_____2014. *Omissions: Agency, Metaphysics, and Responsibility*. New York: Oxford University Press.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

Dennett C. D. 2003. *Freedom Evolves*. New York: Viking.

Ekstrom, W. L. 2000. *Free Will: A Philosophical Study*. Boulder: Westview Press.

Fischer, M. J. and Ravizza, M. 2000. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Fischer, M. J. 1986. "Responsibility and Failure." *Proceedings of the Aristotelian Society* 86: 251–270.

Frankfurt, H. 2003. "Alternate Possibilities and Moral Responsibility." In M. McKenna and D. Widerker (eds.). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate Press: 17–25, originally published in *Journal of Philosophy* 66 (1969): 829–39.

Franklin, C. E. 2011. "The Problem of Enhanced Control." *Australasian Journal of Philosophy*, 89/4: 687–706.

_____2018. *Minimal Libertarianism: Free Will and the Promise of Reduction*. New York: Oxford University Press.

Ginet, Carl. 2003. "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing." In M. McKenna and D. Widerker (eds.). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate Press: 75–90, originally published in *Philosophical Perspectives* 10 (1996): 403–17.

Haji, I. 2013. "Event-Causal Libertarianism's Control Conundrums." *Grazer Philosophische Studien* 88: 227–246.

Huoranszki, F. 2011. *Freedom of the Will: A Conditional Analysis*, New York: Routledge.

Kane, R. 1996. *The Significance of Free Will*. New York: Oxford University Press.

_____2007. "Libertarianism." In *Four Views on Free Will*, by J. M. Fischer, R. Kane, D. Pereboom, and M. Vargas. Oxford: Blackwell: 5–43.

Mele, A. 2006. *Free Will and Luck*. New York: Oxford University Press.

Molnar, G. *Powers: A Study in Metaphysics*. Oxford: Oxford University Press.

Nelkin, D. 2011. *Making Sense of Freedom and Responsibility*. New York: Oxford University Press.

O'Connor, T. 2000. *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

Pereboom, D. 2001. *Living Without Free Will*. New York. Cambridge University Press.

Shoemaker, S. 1980. "Causality and Properties." In P. van Inwagen (ed.). *Time and Cause: Essays Presented to Richard Taylor*. Dordrecht: Reidel: 109–135.

Taylor, R. 1966. *Action and Purpose*. Englewood Cliffs: Prentice-Hall.

Timpe, K. 2012. *Free Will: Sourcehood and its Alternatives*. London: Continuum.

Van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

Van Miltenburg, N. and Ometto, D. 2016. "The Libertarian Predicament: A Plea for Action Theory." *Synthese* https://doi.org/10.1007/s11229-016-1148-3. (online first). (196: 161–178).

Watson, G. 2004a. "Soft-Libertarianism and Hard-Compatibilism." In *Agency and Answerability*: *Selected Essays*. Oxford: Oxford University Press: 197–215. (First published in *Journal of Ethics* 3 (1999): 351–65.)

_____2004b. "Free Action and Free Will." In *Agency and Answerability: Selected Essays*. Oxford:

Oxford University Press: 161–196. (First published in *Mind* 96 (1987): 145–72.)

Widerker, D. 2003. "Blameworthiness and Frankfurt's Argument against the Principle of Alternative Possibilities." In M. McKenna and D. Widerker (eds.). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate: 53–73.

Widerker, D. and McKenna, M. 2003. "Introduction." In M. McKenna and D. Widerker (eds.). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate: 53–73.

Zimmerman, M. J. 2015. "Clarke on Omissions." *Criminal Justice Ethics* 34(3): 361–371.

_____2017. "Omissions, Agency, and Control." In D. K. Nelkin and S. C. Rickless (eds.). *The Ethics and Law of Omissions*. New York: Oxford University Press.

Croatian Journal of Philosophy
*V*ol. XXI, No. 62, 2021
https://doi.org/ 10.52685/cjp.21.62.6
Received: October 7, 2020
Accepted: May 13, 2021

The Priority of Common Sense in Philosophy

MARTIN NUHLÍČEK
Comenius University in Bratislava, Slovak Republic

The aim of this paper is to explore the issue of priority of common sense in philosophy. It is divided into four parts. The first part discusses examples of common-sense beliefs and indicates their specific nature, especially compared to mere common beliefs. The second part explores in more detail the supposed positive epistemic status of common-sense beliefs and the role they play in delimiting plausible philosophical theories. The third part overviews a few attempts to formulate a legitimate argument, or justification, in favor of the positive epistemic status of common-sense beliefs, none of which, however, appears to be clearly successful. Finally, the fourth part addresses the central issue of priority of common sense. Two different types of priority are introduced, epistemic and methodological, and it is argued that only the latter applies to common-sense beliefs. If so, then common-sense beliefs are not to be conceived as cases of knowledge but as the clearest cases of what we believe is knowledge.

Keywords: Common Sense; metaphilosophy; philosophical method; philosophy of common-sense.

Introduction

There are thinkers who postulate the priority of common sense in (and over) philosophy. Typically, they claim that certain propositions or beliefs which are embedded in ordinary common sense also enjoy a specific status in philosophy. G. E. Moore called such beliefs "truisms" and included among them beliefs as "There are other people", "We have hands and bodies", "The Earth has existed for many years", etc. Such beliefs are held widely, almost universally. Indeed, we would hardly find a (normal) person to claim the contrary. But this is not all. According to some, *common-sense beliefs* also represent cases of knowl-

edge. They are said to be the clearest and most elementary examples of what we know: we *know* that there are other people, we *know* that we have hands and bodies, we *know* that the Earth has existed for many years. This provides such beliefs with special epistemic status. The proponents of this view claim that common-sense beliefs form the basic data or background against which philosophical theories can be assessed. If any theory contradicts those beliefs, i.e. if a theory results in the exclusion of common-sense beliefs from knowledge, then it is a strong reason to reject the theory as implausible. In this sense, common sense regulates acceptable philosophical positions; in this sense, common sense takes priority over philosophy. This approach provides an easy solution to various traditional philosophical problems. An obvious example would be Cartesian skepticism, which, in this view, is reduced to a purely rhetorical practice, as it cannot really threaten our knowledge of the outside world.

However, as might be expected, many disagree. In their view, it is not permissible to solve complex philosophical problems simply by pointing to the widespread beliefs of ordinary people. Our common intuitions about knowledge are not skeptical, but therefore we cannot simply dismiss skepticism. We generally accept inductive reasonings, but that does not mean that such reasonings are therefore justified. From the traditional point of view, solutions to such problems require subtle philosophical arguments. Instead, basing solutions solely on the opinion of ordinary people seems unphilosophical and irrational. It would be rational only if we could present a convincing argument or evidence in favor of the positive epistemic status of common-sense beliefs. But the proponents of common sense themselves do not require such an argument or evidence; what is more, they doubt its possibility. Moore states: "We are all, I think, in this strange position that we do know many things, with regard to which, we know further that we must have had evidence for them, and yet we do not know how we know them, i.e. we do not know what the evidence was" (Moore 1959: 45). Despite that, the proponents of common sense insist that common-sense beliefs represent true knowledge and that philosophical theories must conform to them. Is there something wrong with them? Have they abandoned rationality? Or, on the contrary, have they discovered something important that traditional philosophy has long overlooked? Put in different words, *should philosophy respect the priority of common sense?*

The aim of this paper is to explore the issue of priority of common sense in philosophy. It is divided into four parts. The *first* part discusses examples of common-sense beliefs and indicates their specific nature, especially compared to mere common beliefs. The *second* part explores in more detail the supposed positive epistemic status of common-sense beliefs and the role they play in delimiting plausible philosophical theories. The *third* part overviews a few attempts to formulate a legitimate argument, or justification, in favor of the positive epistemic status of common-sense beliefs, none of which, however, appears

to be clearly successful. Finally, the *fourth* part addresses the central issue of priority of common sense. Two different types of priority are introduced, epistemic and methodological, and it is argued that only the latter applies to common-sense beliefs. If so, then common-sense beliefs are not to be conceived as cases of knowledge but as the clearest cases of what we *believe* is knowledge.

1. *The Nature of Common-Sense Beliefs*

The proponents of common sense cite a number of different propositions that represent the beliefs of common sense. They include commonly held beliefs about the outside world, the past, the mental states of oneself and other people, as well as epistemic propositions stating knowledge of these things.

G. E. Moore gives the following examples of common-sense beliefs (see Moore 1953):

1. There are in the Universe an enormous numbers of material objects (e.g. our bodies, other people, animals, plants, stones, mountains, rivers, seas, planets, tables, chairs, etc.).
2. Human beings have minds inasmuch as we have a variety of mental states, including acts of consciousness. We see, hear, feel, remember, imagine, think, believe, desire, dislike, will, love and so on.
3. All material objects are located in space inasmuch as they are located at a distance from each other.
4. Mental acts are attached to—contained within—certain kinds of bodies (human bodies and perhaps those of the higher animals).
5. Mental acts are ontologically dependent upon bodies.
6. Most material objects have no acts of consciousness attached to them.
7. Material objects can and do exist when we are not conscious of them.
8. There was a time when no act of consciousness was attached to any material body.
9. All objects and acts of consciousness are in time.
10. We know 1.– 9. to be true.

Noah Lemos describes what he takes to be the main features of the common-sense tradition in philosophy (Lemos 2001: 204–206):

1. In answering certain philosophical questions, commonsensism holds that it is appropriate or reasonable to take as data certain ordinary, yet widely and deeply held, beliefs.
2. The commonsense philosopher takes these beliefs as data without having any proof for them.
3. The commonsense tradition is not committed to the view that our commonsense beliefs are indefeasible or immune to revision, but it does assign to our commonsense beliefs a great deal of "weight" or importance.

There is no clear way to explicate the amount of "weight" placed on common-sense beliefs in philosophical discourse. But at least it is plausible to say that if a philosophical theory conflicts with one or another common-sense belief, the common-sense philosopher seeks to resolve the conflict in favor of common-sense beliefs. Therefore, "the commonsense philosopher seeks to be conservative in his revisions of his commonsense beliefs" (Lemos 2001: 206).

However, not every universally held belief is a common-sense belief. Or, to put it differently, common-sense beliefs do not gain their special philosophical status simply from the fact that they are universally held. Common-sense philosophers do not reject philosophical theories on the sole ground that they contradict some widely shared beliefs. It must be a *common-sense belief* that many know; if a theory suggests the opposite of such a proposition, it is a reason to reject that theory. But how are we to distinguish common-sense beliefs from universal beliefs that are not common-sense beliefs? Nicolas Rescher suggests making a distinction between *common beliefs* and *common-sense beliefs*. Whereas common beliefs are simply widely held beliefs, i.e. what everyone "knows", common-sense beliefs differ in their normative claim. They encompass items of information that everyone *should* know regarding the basic realities of human situation (Rescher 2005: 23). Thus, common beliefs are those which everyone accepts, while common-sense beliefs are those which every normal person should accept, were a question raised about them. In this sense, there is a connection between common sense and rationality. Mere common beliefs may also contain various mistakes and prejudices, as they did many times in history. Meanwhile, common-sense beliefs express what is ultimately *reasonable* to believe.

Now, what if someone denied common-sense beliefs? As Rescher points out, we would probably conclude that one did not first understand them. It is because there is "nothing sophisticated, complicated, or technical about common sense, and no special training or insight is needed for its realization. It relates to the sort of thing that anyone must realize who functions in the circumstances at issue" (Rescher 2005: 24). Another option is to just assume that the person in question is not mentally fit to grasp such propositions. Indeed, Rescher identifies "foolishness or idiocy" as one contrastive opposition to common sense (Rescher 2005: 26). Thomas Reid would also add to this group some philosophers who, with the help of philosophical arguments, allow themselves to arrive to opinions contrary to common sense. As he puts it, if someone were "to be reasoned out of the principles of common sense by metaphysical argument, we may call this *metaphysical lunacy*" (Reid 1823: 260). Of course, it does not mean that common-sense beliefs are utterly invariable or that they cannot be corrected or changed by philosophy. A critical examination may sometimes reveal discrepancies among common-sense beliefs and thus necessitate a revi-

sion. However, in that case, we would be leaving the domain of common sense in a different direction. Intellectual performance which requires an amount of training and specific skills constitutes the other limit of common sense. It is known as *expertise*, which is the second contrasting opposition to common sense according to Rescher (Rescher 2005: 26).

Common sense itself does not need to be defined as a distinctive human faculty. Reid sometimes does refer to a faculty of common sense; however, it is not clear whether he considers it to be a *sui generis* faculty. In contrast, at one point he suggests that common sense may be "only another name for one branch or degree of reason" (Reid 1969: 567). Thus, common sense can be understood as the correct exercise of our general reasoning abilities; or simply as correct *judgment*. This assumption can be supported by Reid's very definition of common sense: "Common sense is that degree of judgment which is common to men with whom we can converse and transact business" (Reid 1969: 557). As it appears, common sense focuses on obvious truths that are neither complex knowledge nor require special expertise. They are just a matter of ordinary "correct judgment". Speaking more theoretically, common sense represents a kind of *cognitive minimalism*, i.e. awareness of things that are so evident that people with normal cognitive abilities cannot fail to realize their truth. Probably that is why the propositions of common sense are called *truisms*.

Particular truisms can be generalized into global epistemic claims, each of which includes a whole class of possible individual common-sense propositions. For instance, common-sense propositions such as "I have hands", "It is day", "There are other people", and the like, fall under a general epistemic claim "Those things do really exist which we distinctly perceive by our senses, and are what we perceive them to be." We may recognize Thomas Reid as the author of this claim (Reid 1969: 625). It was him who famously named a total of twelve such general claims, calling them *principles of common sense*. Those principles may be viewed as attempts to summarize the variety of common-sense judgments into a number of basic "axioms" which are thought to represent the most fundamental features of human cognition. It should be noted, however, that the discovery of the principles of common sense is the result of a specific philosophical endeavor, and their possession is therefore not itself part of common sense. One requires special philosophical expertise to be able to articulate such principles; hence, those principles constitute a *philosophical* reflection of common-sense beliefs. Clearly, that goes beyond the scope of ordinary common sense.

2. *The epistemic status of common-sense beliefs*

Common sense is often connected with the practical affairs of human life. For example, Rescher says: "Common sense is a realistic guide in matters of what to think and what to do" (Rescher 2005: 55). He states that common-sense beliefs are *certain*; however, not in the sense of ab-

solute, anti-skeptical certainty but of practical, human-life certainty, which defines the standard of "being beyond reasonable doubt" (Rescher 2005: 29). Thus, common-sense propositions are not a matter of logical necessity. Nevertheless, they are strong presumptions that can hardly be denied: "To be sure, the claims of common sense do not have the backing of some sort of necessitation that guarantees their irrefragable certainty. But what they do have in their favor is a powerful presumption" (Rescher 2005: 57). Where does this presumption come from and what does it make of the epistemic status of such propositions?

To find an answer, we have to start with Reid's well-known characteristics of common-sense beliefs (as put in Rescher 2005: 35):

1. *universality* by way of being generally held;
2. *commonality* in reflecting the usage of all languages;
3. *undeniability* by way of being such that their contradictions are not merely false but absurd;
4. *irresistibility* so that even those who question them at the level of theory are compelled to accept them in conducting the practical affairs of life.

For the most part, it is the last characteristic, *irresistibility*, which plays the main role in determining the certainty of common-sense beliefs. In fact, we are not able to give up such beliefs as "There are other people". That makes them the fundamentals in our everyday reasoning, deciding, and acting. Any attempt to withhold common-sense beliefs, as is sometimes required by philosophical arguments, is therefore futile, because it is humanly impossible. As Lemos sums it up, it is impossible, first, psychologically, since those beliefs are irresistible. Second, it is impossible practically, as neglecting such beliefs would disrupt the course of everyday life. And third, it is supposed to be impossible also in a philosophical way, as those beliefs are referred to as inseparable parts of human nature (Lemos 2004: 73–74). Because common-sense beliefs are irresistible and impossible to give up, they are conceived to be *self-evident*. This means that their evidential status is not derived from an argument; that is, they neither require an argument nor do they admit one. No philosophical argument can make them more evident, just as no counter-argument can make them less evident. In general, the evidence of common-sense beliefs cannot possibly be altered on the basis of reasoning.

These features, I believe, lie at the heart of the presumed positive epistemic status of common-sense beliefs. Such beliefs are not subject to philosophical arguments, and yet, they basically constitute prime examples of what "evident" means. Because of that, they are supposed to enjoy a fairly high level of certainty. For what else could be labeled as certain, if not the most apparent self-evident beliefs? Therefore, common-sense beliefs are not taken as certain just because they are universally held. They are held to be certain because they are self-evident, and for the same reason they are also universally spread. What is more, common-sense

philosophers typically understand the propositions of common sense as something we *know*. Not only do those propositions exhibit certainty in a psychological way, but it is also assumed that they form the *epistemic basis* of our knowledge. That basis itself does not have to be proved; on the contrary, it may serve as the source of proof for other propositions. Thus, the propositions of common sense are thought to represent the deepest foundations of knowledge, which can only be identified but not challenged. It may be analogous to other situations: we are certain that something is immoral, beautiful, etc., even if we cannot explain why.

The described epistemic status of common-sense beliefs allows its proponents to easily answer the problem of *skepticism*. They agree with the skeptic that certainty is a condition of knowledge. But they insist that a distinction must be made between reasonable certainty and the absolute certainty which the skeptic requires. The certainty of knowledge is the realistic certainty of life. To say that something is certain is to say that it is as certain as anything of that kind can be. Common-sense beliefs are conceived as prototypes of such certainty. So, from the common-sense point of view, the skeptic demands certainty that is realistically unattainable, and, on the other hand, she tries to challenge beliefs that are in fact unquestionable. Therefore, the skeptic can be answered as follows: when we have taken all the steps to achieve reasonable certainty, then we have knowledge. If the skeptic is not satisfied with this, she places unreasonably high criteria on knowledge and demands evidence that is practically impossible to provide. In the eyes of common-sense philosophers, since skepticism proves itself as unreasonable, we can safely ignore it. Some have gone even further, and, from the position of common sense, they have tried to respond directly to the skeptical challenge. A shining example was G. E. Moore, who presented his "proof of an external world" by drawing attention to the belief that he had two hands (see Moore 1993). Nevertheless, he only provoked confusion, probably because he mixed two different views. From one point of view, his answer sounds quite convincing because he presents a simple common-sense belief; but from the other point of view, his common-sense answer does not provide the expected level of expert elaboration, which gives the impression of insufficiency. Hence, for the common-sense philosopher, it seems more appropriate to point out a conflict between skeptical requirements and common sense than to attempt to respond directly to the skeptic.

Common sense sets limits not only to skepticism but also to any philosophical theory whose consequences would contradict common-sense beliefs. The proponents of common sense typically argue as follows:

1. Proposition P is a common-sense proposition which everyone knows.
2. Theory T implies that no one knows P.
3. Therefore, theory T is false.

In this argument, we can see that the propositions of common sense are indeed attributed a high epistemic status. Contradiction with such

propositions alone is considered sufficient to reject the whole theory. It is precisely in this sense that common-sense beliefs serve as "hard data" to which philosophical theories have to conform.

The approach in which common-sense beliefs take epistemological precedence over philosophical theories is expressed in the works of several authors sympathizing with the common-sense view. For instance, we find Roderick Chisholm saying (Chisholm 1982: 113):

> It is characteristic of "commonsensism," as an alternative philosophical tradition, to assume that we do know, pretty much, those things we think we know, and then having identified this knowledge, to trace it back to its sources and formulate criteria that will set it off from those things we do not know.

Thomas Reid suggests that common sense creates the ultimate criterion for the acceptability of different philosophical positions (Reid 1823: 17):

> A traveller of good judgment may mistake his way, and be unawares led into a wrong track; and while the road is fair before him, he may go on without suspicion and be followed by others but, when it ends in a coal pit, it requires no great judgments to know he hath gone wrong, nor perhaps to find out what misled him.

Elsewhere, Moore writes (Moore 1922: 163):

> There is no reason why we should not, in this respect, make our philosophical opinions agree with what we necessarily believe at other times. There is no reason why I should not confidently assert that I do really *know* some external facts, although I cannot prove the assertion except by simply assuming that I do. I am, in fact, as certain of this as of anything; and as reasonably certain of it.

The epistemological function of common-sense beliefs is consistent with what Chisholm describes as *particularism*. This view argues that the study of knowledge begins with cases that are considered to be clear knowledge, and only according to them do we try to formulate criteria for knowledge (Chisholm 1973: 14–15). Of course, one may ask, how can we recognize that our initial "knowledge" is correct and unaffected by error? We cannot, the particularists reply. But they point out that there is no other reasonable way to conduct the study of human knowledge. We can only start with things we know and then try to examine and improve them.

Let us quote Chisholm again (Chisholm 1977: 16):

> We presuppose, first, that there is something that we know and we adopt the working hypothesis that *what* we know is pretty much that which on reflection we think we know. This may seem like the wrong place to start. But where else *could* we start?

3. *Justification of common-sense beliefs*

Common-sense beliefs are irresistible. They represent the highest level of practical certainty a belief can possibly acquire. But the question of their psychological irresistibility must be separated from the question of their

epistemic legitimacy. In a broader sense, we have to determine whether the psychological specificity of some beliefs justifies the special epistemic status they allegedly possess. Or, as Christopher Hookway puts it: "We must decide whether appeal to such 'common-sense' certainties embodies a response to fundamental epistemological issues or is simply an attempt to ignore them" (Hookway 1990: 397). Speaking about knowledge claims, they are generally associated with the possibility of being exposed to critical challenges. If someone answered the question of how she knows something, that she only insists she knows it, it would seem irrational. This is why some recent epistemologists have complained that G. E. Moore's anti-skeptical common-sense insistence that he knows there is a hand in front of him is more a stubborn refusal to take skepticism seriously than a philosophically sensitive response to it (Hookway 1990: 401). Therefore, according to Hookway, the philosophy of common sense does not end with simply naming certain common-sense beliefs. In addition, "a common-sense philosophy must explain why it is legitimate to trust these certainties. This is the fundamental difficulty facing a philosophical appeal to common-sense" (Hookway 1990: 399).

A philosophical appeal to common sense without justifying its contributions would be, no doubt, unsatisfactory. Immanuel Kant develops his well-known critique of common-sense philosophy exactly along the same lines. He writes (Kant 1950: 7):

> It is indeed a great gift of God to possess right or (as they now call it) plain common sense. But this common sense must be shown in action by well-considered and reasonable thoughts and words, not by appealing to it as an oracle when no rational justification for one's position can be advanced.

As we can see, Kant takes appealing to common sense as opposed to rational justification. In his view, reliance on common sense cannot be philosophically legitimate unless it is substantiated by justifying reasons. Although Kant's critical view of common sense may not be widely shared today, his assumption has been preserved; namely, that if common sense is to serve philosophical purposes, it must be appropriately justified. This assumption is also present in current accounts of common-sense beliefs. For example, Noah Lemos lists two essential points regarding the possibility of common-sense knowledge (Lemos 2004: 14–15):

> First, in order to have common-sense knowledge, one needs a general criterion that tells us that beliefs of such common-sense sort represent knowledge.
>
> Second, to fulfil the criterion, an argument—one free from epistemic circularity—is crucial for having the sort of knowledge that the common-sense philosopher claims.

Thus, the attempt to rationally justify common-sense beliefs amounts to searching for a rational argument that would support their positive epistemic status. It is crucial that the argument be "free from epistemic circularity", i.e. it must not presuppose the epistemic status of the beliefs whose epistemic status is to be proved. To presuppose something

which is just to be proved would be, of course, a fallacy of reasoning. So, can such an argument be effectively constructed? Let us review a few notable attempts.

William Alston, for example, admits that justificatory arguments in support of our cognitive abilities *are* epistemically circular, but at the same time he points out that no such non-circular arguments are available. Hence, we have to accept epistemic circularity, which, however, might not be irrational. Alston favors an approach which appeals to the "practical rationality" of our ways of forming beliefs. He argues that many of our belief-forming ways are firmly established. It does not seem to be in our power to avoid forming beliefs in the established ways and substitute them with entirely different ways. At least, it would probably be very difficult to do so. Moreover, the same problem of epistemic circularity that beset our practices would also confront these new alternatives. Given these facts, Alston concludes that it is "practically rational" for us to continue engaging in our established ways of forming beliefs (Alston 1996: 271).

Alston's argument may be viewed as pragmatic, provided that the only alternative is to admit our beliefs as widely unjustified. But does it offer a plausible solution to the problem of justifying common-sense beliefs? One might object that the method described could also allow justification for such things as crystal ball gazing. Of course, such objection would be unfair, as crystal ball gazing is certainly not an established way of how people form beliefs. Nevertheless, Alston's view invites perhaps a little more circularity than he intended to, since his very argument appears to be epistemically circular. His conclusion about which ways of forming beliefs we should continue to engage in is derived from a premise describing what ways of forming beliefs we actually engage in. In the words of Ernest Sosa: "If we push reflection far enough with regard to why we should accept the premises of *this* argument, don't we find ourselves appealing precisely to *its* conclusion?" (Sosa 1996: 315). If it is true that Alston's argument is circular, in a way it only emphasizes the need to find an argument that avoids circularity.

Let us proceed to Sosa's own view. Like Alston, he thinks that epistemic circularity in justifying our fundamental beliefs is inescapable, but he too suggests that it does not preclude those beliefs from being justified and, hence, rational. However, Sosa develops his argument from a more general position. He considers the totality of our belief-forming ways, which he refers to as W. When we take W, he says, then by using W we can know that W is reliable. The fact is that this conclusion is formed by the use of our best intellectual procedures. Although we have not avoided epistemic circularity, it is not at all easy to pinpoint what has been omitted or done wrong in this argument. Sosa therefore concludes that it is permissible to justify our belief-forming processes in a circular way, as we have no other overall way to do so (Sosa 1996: 318). This resembles Alston's argument on a larger scale:

we have no other totality of belief-forming ways than W, so it is permissible to prove with W that W is rational. But if there is a genuine similarity with Alston, then Sosa's argument is also circular itself.

Alston and Sosa both argue that from a practical point of view, we have no other option than to rely on the belief-forming processes that we naturally possess. As already pointed out, it is to be conceived as a pragmatic response that primarily seeks to avoid the skeptical alternative that our beliefs generally lack justification. Sometimes it is argued that any reasons that have not yet been challenged can be safely used for belief-justification even if it is circular. This view is based on the principle of "innocence" of reasons, unless proven otherwise. For instance, Michael Bergmann makes a difference between "malignant" and "benign" epistemic circularity. The former kind of epistemic circularity arises in a questioned source context, as he calls it, while the latter arises in an unquestioned source context. And while epistemic circularity in a questioned source context is obviously "bad" (as he calls it again), epistemic circularity in an unquestioned source context is not necessarily mistaken. In a questioned source context, we cannot use the source to verify its own reliability—precisely because it is questioned. We would need an independent argument for that. But if we are just reconstructing the reasons why we believe in the reliability of a source without questioning it, we do not need an independent argument (Bergmann 2004: 717–721). This allows the justification of beliefs, including common-sense beliefs, by any other beliefs and sources which are presumed to be justified, provided that they have not been challenged so far. Thus, Moore can be justified in the belief that there is a hand in front of him just on the basis of his plain sight, provided that his sight has not been questioned. And similarly, the reliability of his sight can be justified by the fact that it correctly informs him about external things such as hands—again, provided that this capability has not been questioned.

But the crucial question is not whether we are relying on circularity in justification, but whether circular justification is epistemically legitimate. Those are two independent questions. Even if the answer to the first one is affirmative, it does not necessarily imply that the answer to the second one would be affirmative as well. As for the second question, it is certainly true that to grant epistemic circularity is to grant a method which is generally conceived to be fallacious. However, there appear some options how to handle the problem of epistemically circular justification. One could suggest, for instance, that circular justification should be allowed where no non-circular justification is available. When there is no non-circular justification to the belief that "there is a hand in front of me", a circular argument can be used, such as: "There is a hand in front of me, which is justified by the reliability of my visual perception, which, in turn, is justified because it correctly informs me about external things, such as hands, being in front of me."

Even if we overlook the fact that such arguments look unsatisfactory at first glance, there is a deeper problem. If we allow circular justification in *some* cases, how do we prevent circular reasoning in *all* cases? Based on what criterion should we distinguish that we admit the circular justification of perception or memory but not, for example, of telepathy? It is not at all easy to answer such objections.

The fact is that most authors deny the possibility of epistemic circularity as a rational method of justification. Richard Fumerton is one of the most radical ones, who, to prove the point, compares circular justification of belief-forming processes to the justification of astrology (Fumerton 1995: 177):

> If a philosopher starts wondering about the reliability of astrological inference, the philosopher will not allow the astrologer to read in the stars the reliability of astrology. Even if astrological inferences happen to be reliable, the astrologer is missing the point of a philosophical inquiry into the justifiability of astrological inference if the inquiry is answered using the techniques of astrology.

In his particularly dramatic way, he concludes that epistemic circularity is completely inadmissible (Fumerton 1995: 177):

> You cannot *use* perception to justify the reliability of perception! You cannot *use* memory to justify the reliability of memory! You cannot *use* induction to justify the reliability of induction! Such attempts to respond to the skeptic's concerns involve blatant, indeed pathetic, circularity.

Fumerton apparently expects some independent, i.e. non-circular, argument in favor of the credibility of our cognitive sources. His expectation is probably derived from the fact that we usually require independent justifications of local beliefs. For example, a particular memory might be quite independently justified with photographs, written records, memories of others, etc. But when it comes to the general justification of our belief-forming methods as such (e.g. memory as such), it appears to be considerably difficult, if not impossible.

Thus, moving back to common-sense beliefs, it seems that there is no unproblematic way of justifying them with an argument free from epistemic circularity. This puts us in a position where we must look for some way to epistemically support common-sense beliefs other than on the basis of an independent argument. In the words of Noah Lemos: "If a philosophical curiosity about the reliability of our faculties could only be satisfied by an argument free of epistemic circularity, then it would seem to be a mark of philosophical wisdom to accept the fact that that cannot be done" (Lemos 2004: 51).

4. *The priority of common-sense beliefs*

What attitude should one take to the question of the legitimacy of common-sense beliefs? As we have seen, attempts to support it with an argument free from epistemic circularity seem problematic. One has only a few options how to deal with it. First, one can admit epistemic

circularity and insist that our beliefs are legitimately justified in a circular way. Some authors adhere to this position, yet it is not at all clear if the controversy which it contains can be resolved. Second, one can argue that there is no satisfactory way of supporting our beliefs and thus conclude that we do not know the things we believe to know. This is the position of skepticism that has notoriously little philosophical appeal, mainly because it refuses human knowledge instead of seeking for its explanation. There is also a third option. One can claim that some of our beliefs have a special epistemic status that allows them being justified even without a supporting argument. This approach is known as *foundationalist*, since it treats some beliefs as foundational, i.e. as epistemological "axioms" whose legitimacy is somehow based in themselves and is not to be proved by an argument. Such a view makes it possible to circumvent the issue of searching for a non-circular argument, and, at the same time, it is conceived as not inviting skepticism.

Common-sense philosophers typically consider common-sense beliefs to be foundational. Thomas Reid postulates them as ultimate sources which may provide justification for other beliefs but do not need such support themselves. He insists that justification "must stop only when we come to propositions which support all that are built upon them, but are supported by none themselves—that is, to self-evident propositions" (Reid 1969: 596).

Such propositions neither require nor admit of proof; their epistemic authority is independent and, in fact, *superior* to any argument that could be formulated in favor of them (Reid 1969: 116):

> [It] is not by any train of reasoning or argumentation that we come to be convinced of the existence of what we perceive; we ask no argument for the existence of the object, but that we perceive it; perception commands our belief upon its own authority, and disdains to rest its authority upon any reasoning whatsoever.

According to Reid, common-sense beliefs are self-evident and thus give us immediate knowledge. They are subject to no proof or reasoning, and yet they *are* justified. As he puts it, they somehow bear "the light of truth" in themselves: "[t]here is no searching for evidence; no weighing of arguments; the proposition is not deduced or inferred from another; it has the light of truth in itself, and has no occasion to borrow it from another" (Reid 1969: 593).

But Reid's view can only hold if we presuppose that our cognitive faculties generally work properly. Otherwise—if our faculties were thought of as possibly prone to error—their contributions could not be safely regarded as knowledge. And indeed, Reid states as a general principle that "the natural faculties, by which we distinguish truth from error, are not fallacious" (Reid 1969: 630). Now, the crucial question is: how do we know this principle to be true?

Reid holds that it is a *first principle,* simply known to be true. By holding that proposition to be a first principle, he suggests that the general statement about the reliability of our natural faculties is im-

mediately evident. It is neither inferred from any other proposition nor dependent on any argument. For Reid, such conviction would probably be the testimony of some natural faculty itself—or, as Lemos somewhat uncharitably puts it, "some faculty 'vouching' for itself" (Lemos 2004: 71). In any case, it is regarded as a foundational principle which, immediately known, provides epistemic justification for other, non-foundational propositions. It is in this way that Reid rejects the view that the only epistemically satisfactory way to know that our faculties are reliable is via a non-circular argument.

Moreover, he offers a variety of philosophical as well as practical reasons which imply that without such a principle, a non-skeptical epistemology could not be constructed. From one traditional point of view, epistemology struggles to combine three different claims:

(1) We know things.
(2) We can know things only if our cognitive faculties are reliable.
(3) We do not know if our cognitive faculties are reliable.

The skeptic accepts (2) and (3) and derives from them that (1) does not hold. Anti-skeptical authors, on the other hand, take (1) as a premise and rather attempt to modify (2) or (3) accordingly. What about common-sense philosophers? They claim that to give up (1) would be philosophically and practically unreasonable (or even absurd) and to give up (2) would be irrational. Therefore, they modify (3) to its exact opposite: We *know* that our cognitive faculties are reliable. Combined with (2), this allows them to logically support the common-sense conclusion (1).

In the previous section, we could see Alston and Sosa trying to prove the opposite of (3) with arguments which were, nevertheless, circular. A wholly different strategy was proposed in Sosa's later work, where he introduced a distinction between *animal knowledge* and *reflective knowledge*. In its simplest form, animal knowledge involves just knowing a thing, whereas reflective knowledge requires also knowing how and why we know that thing (see Sosa 2007: 24). Since for most ordinary purposes it is quite sufficient to exhibit animal knowledge, we can assert (1) without paying attention to (2) and (3). Thus, Sosa's approach may be understood as an attempt to "dissolve" the problem of the three claims for ordinary knowledge. But this is not to be confused with the position of common-sense philosophy. Its proponents typically acknowledge the importance of explaining our knowledge, i.e. of explaining the relations between the three claims. As already mentioned, in following the main aim of defending (1), they turn (3) to its opposite. They believe such move is legitimate, for as they see it, the general reliability of our faculties is a foundational truth.

From a methodological point of view, it is possible to place Reid (and other proponents of common-sense philosophy) in the before-mentioned position of particularism (see section 2). According to it, any relevant study of knowledge can be effectively initiated in no other way than by assuming that our cognitive faculties work (roughly) properly. There is literally no alternative to initially trusting at least some of our facul-

ties. But from a logical point of view, to hold that some propositions are foundational in character does not amount to a guarantee of them being true. For instance, a clear deliverance of the senses would certainly be referred to as foundational knowledge. However, since it is contingent, it is still logically compatible with the possibility of error, however slight. It is this feature that prevents many authors from accepting foundational beliefs as unproblematic knowledge.

As it appears, the debate on common-sense beliefs is twofold; first, common-sense beliefs are treated as starting points of epistemic scrutiny, and second, they are attributed with a peculiar epistemic status of foundational knowledge. These differences are well captured in John Greco's distinction between two kinds of priority held by common-sense beliefs:

> In brief, common sense beliefs enjoy an epistemological priority in that they constitute a foundation for knowledge: such beliefs enjoy the kind of evidential status required for knowledge, even without being grounded in further evidence themselves. Common sense beliefs enjoy a methodological priority in that they constrain philosophical theory: such beliefs serve as pre-theoretical commitments that philosophical theories ought to respect, at least in the absence of good reasons for rejecting them. (Greco 2014: 142)

Let us have a look at both kinds of priority in more detail.

Methodological priority, as already indicated, concerns common-sense beliefs as the starting points of philosophical scrutiny of human knowledge. It suggests that any philosophical theory has to respect common-sense beliefs in the sense that it should not arbitrarily contradict them or lead to a consequence that contradicts them. However, there is no absolute bar to violating common sense: a theorist may speak against common sense, but if she does, she has to provide very good reasons to do so. Thus, the methodological priority may be thought of as *prima facie*: one's theory might end by violating common sense, but it should not start off that way. The contrary approach, i.e. disregarding the principles of common sense and postulating philosophical constructions instead, would probably be not only ineffective in explaining real human knowledge but, as Reid famously points out, also potentially destructive for the philosophical effort itself: "Such principles are older, and of more authority, than Philosophy: she rests upon them as her basis, not they upon her. If she could overturn them, she must be buried in their ruins" (Reid 1823: 14).

Epistemic priority, on the other hand, is the thesis that common-sense beliefs are examples of foundational knowledge and are known directly even in the absence of supporting arguments. One interpretation suggests that this view is based on the principle that our beliefs should be treated as "innocent until proven guilty". But according to Greco, such principle explains methodological priority rather than epistemic priority (Greco 2014: 147–148). So how could we explain the alleged epistemic priority of common-sense beliefs? Three possible options come into consideration. First, a special epistemic status of

common-sense beliefs may be supported by an independent argument. As demonstrated before, such attempts end in epistemic circularity. Second, we can point to the irresistibility of common-sense beliefs and our incapability of giving them up without feelings of discomfort or absurdity. Especially Reid often resorts to this type of defense, but, as already stated, it is more a defense of methodological than of epistemic priority. To claim that we cannot easily abandon certain beliefs is to claim their strategic position in our cognitive systems, but not yet that they accurately represent the corresponding states of affairs. And third, common-sense beliefs may be declared as epistemically prior by definition: if a belief is identified as common-sense belief, it is assumed to have a positive epistemic status. This understanding, I believe, is the closest to the foundationalism of common-sense philosophy. However, it entirely begs the question of *why* those beliefs enjoy such a special epistemic status. This is not just philosophically unsatisfactory, but some may even see it as a dogmatic insistence on popular beliefs based on folk psychology.

We can see that each of the three ways to justify the epistemic priority of common-sense beliefs seems controversial. Hence, there appears to be no unproblematic sense in which common-sense beliefs could be thought of as having the epistemic kind of priority. This is not to say that they do not have it or that they never amount to knowledge. The point is that it is something that has yet to be proven. History teaches us this as well, since many widely held beliefs in the past have turned out to be wrong. Therefore, simply claiming that present widely held beliefs are *not* wrong appears rather philosophically naïve. As a result, common-sense foundationalism is far from being universally accepted (to put it mildly), as it raises serious suspicions of circularity or dogmatism.

A supporting reason undermining the idea of epistemic priority of common-sense beliefs may also be found in some recent findings in the field of philosophy of mind. Folk psychology, i.e. common-sense explanations of various psychological phenomena, turn out to be wildly mistaken in many different respects. For instance, Paul Churchland compiles a whole list of issues that folk psychology cannot explain or even fails to address: the nature of mental illness, the faculty of creative imagination, the ground of intelligence, the psychical functions of sleep, and many more (see Churchland 1981: 73 ff.). According to eliminativists, as they are labelled, the main problem is that folk psychology vocabulary is, in its central categories, massively wrong and therefore should be abandoned in favor of a more correct "scientific image", describing those issues with more refined and accurate categories. As Churchland concisely puts it, "our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced"—by science (Churchland 1981: 67). If we draw an analogy between the philosophy of mind and

epistemology, we may come to a suspicion that similar problems may affect at least some of our common-sense beliefs, too. Apart from the fact that they can be factually incorrect, the very categories by which they express and describe the world around us may not be adequate or relevant. Again, this puts one in a position where one should be very cautious with the notion of epistemic priority of common-sense beliefs. Not only their factual, but also their semantic aspects seem to be something yet to be explored.

Conclusion

As far as our inquiry indicates, common-sense beliefs *do* enjoy methodological priority, but they *do not* enjoy epistemic priority. We have found no convincing way to rationally (i.e. non-dogmatically and non-circularly) account for the alleged positive epistemic status of common-sense beliefs. It is possible that the appearance of epistemic priority is caused by psychological features (such as irresistibility) which are related rather to the methodological priority of those beliefs—but for now, this is only a speculation. In any case, if common-sense beliefs do not have epistemic priority, they do not constitute knowledge; or, to be more precise, they do not necessarily constitute knowledge. This consequence is supported by the fact that according to most accounts of knowledge, one necessary condition of knowledge is truth. But since common-sense beliefs are contingent in nature and thus not entirely immune to revision—as the proponents of common sense themselves admit, on occasions the inadequacy of a common-sense belief can be revealed and replaced by a different belief—they cannot be labelled as true by definition. Otherwise, it would have to be the case that in some periods of time, we know something that is not true. This is an *ad absurdum* outcome which shows that common-sense beliefs should not be automatically treated as knowledge. To be clear, it is not to be denied that some common-sense beliefs may actually pose knowledge. But this has to be assessed individually and not just generally declared simply because many people find such beliefs compelling. Good candidates for knowledge will probably be the most ingrained common-sense beliefs, such as that other people exist. But a complex analysis is not my intention here.

As for the methodological priority, we have found no reasons why it should not hold. Some beliefs seem to be more psychologically significant than others, and common-sense beliefs may be considered to be the most significant ones in this respect. Therefore, special attention should be paid to them in theoretical investigations. As a basic norm, theories should not openly contradict such beliefs; they should not imply, for example, that there are no material objects, or that our vision generally deceives us. And if they do, they are expected to explain it adequately and also to show what notable theoretical advantages do we accomplish in exchange. In some domains, a change of the initial

common-sense view already allowed us to better understand and predict different phenomena. The atomic structure of things, the Earth's motions, the viral causes of infections, etc.—all of these theories are not commonsensical in the first place, but science widely proved their usefulness. As previously mentioned, common-sense beliefs are generally revisable, since they are contingent. But any such revision has to be properly motivated and justified, otherwise it may run into a barrier of psychological unacceptability. Thus, theories have to either respect common-sense beliefs, or, when they explain them away, they must do so with enough detail, care, and conviction. For these reasons, common-sense beliefs are thought to be methodologically prior, both in science and philosophy.

To sum up, in this paper we have shown, first, that common-sense beliefs are to be distinguished from mere common beliefs in that they also possess a normative appeal prescribing what (normal) people should believe. Second, that common-sense beliefs are by some thought of as knowledge which serve for philosophical theories as ultimate data that should not be contradicted. Third, that the most critical question concerns the epistemic justification of common-sense beliefs, but providing a satisfactory answer to this question has proved problematic. And fourth, that the priority of common sense in (and over) philosophy seems to be rooted in the distinctive psychological features of common-sense beliefs and thus appears methodological rather than epistemic. If so, common-sense beliefs are not to be conceived as unconditional knowledge, since the real epistemic status of each such belief has to be determined by investigation. But they may be thought of as the starting points of every inquiry, including the one concerning human knowledge. In this sense, they represent the clearest cases of what we *believe* is knowledge, although further philosophical scrutiny may reveal some of those beliefs to be inaccurate.*

References

Alston, W. P. 1996. "A 'Doxastic Practice' Approach to Epistemology." In P. K. Moser. (ed.). *Empirical Knowledge.* Lanham: Rowman and Littlefield: 269–301.

Bergmann, M. 2004. "Epistemic Circularity: Malignant and Benign." *Philosophy and Phenomenological Research* 69 (3): 709–727.

Chisholm, R. M. 1973. *The Problem of the Criterion.* Milwaukee: Marquette University Press.

Chisholm, R. M. 1977. *Theory of Knowledge.* 2nd edition. Englewood Cliffs: Prentice Hall.

Chisholm, R. M. 1982. *The Foundations of Knowledge.* Minneapolis: The University of Minnesota Press.

* This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences, under the number VEGA 1/0258/2019—Realism and Anti-Realism in Contemporary Philosophy of Science.

Churchland, P. M. 1981. "Materialism and the Propositional Attitudes." *The Journal of Philosophy* 78 (2): 67–90.

Fumerton, R. 1995. *Metaepistemology and Skepticism*. Lanham: Rowman and Littlefield.

Greco, J. 2014. "Common Sense in Thomas Reid." *Canadian Journal of Philosophy* 41 (Sup. 1): 142–155.

Hookway, C. 1990. "Common-Sensism and Rational Self-Control." *Noûs* 24 (3): 397–411.

Kant, I. 1950. *Prolegomena to Any Future Metaphysics*. Indianapolis: Bobbs-Merrill.

Lemos, N. 2001. "Commonsensism in Ethics and Epistemology." In M. Steup (ed.). *Knowledge, Truth, and Duty. Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford: Oxford University Press: 204–218.

Lemos, N. 2004. *Common Sense. A Contemporary Defense*. Cambridge: Cambridge University Press.

Moore, G. E. 1922. "Hume's Philosophy." In *Philosophical Studies*. London: Routledge and Kegan Paul: 147–167.

Moore, G. E. 1953. "What Is Philosophy?" In *Some Main Problems of Philosophy*. London: George Allen and Unwin: 1–27.

Moore, G. E. 1959. "A Defence of Common Sense." In *Philosophical Papers*. London: George Allen and Unwin: 32–59.

Moore, G. E. 1993. "Proof of an External World." In *Selected Writings*. Ed. T. Baldwin. London: Routledge: 147–170.

Reid, Thomas. 1823. *An Inquiry into the Human Mind on the Principles of Common Sense*. London: T. Tegg.

Reid, Thomas. 1969. *Essays on the Intellectual Powers of Man*. Cambridge: MIT Press.

Rescher, N. 2005. *Common-Sense: A New Look at an Old Philosophical Tradition*. Milwaukee: Marquette University Press.

Sosa, E. 1996. "Philosophical Scepticism and Epistemic Circularity." In P. K. Moser (ed.). *Empirical Knowledge*. Lanham: Rowman and Littlefield: 303–329.

Sosa, E. 2007. *A Virtue Epistemology. Apt Belief and Reflective Knowledge*. Volume I. Oxford: Clarendon Press.

Croatian Journal of Philosophy
Vol. XXI, No. 62, 2021
https://doi.org/ 10.52685/cjp.21.62.7
Received: June 4, 2020
Accepted: March 3, 2021

The Problem with Impure Infinitism

HUSEIN INUSAH
University of Cape Coast, Cape Coast, Ghana

It is generally believed that pure versions of infinitism face two problems, namely: 1) they are unable to distinguish between potential and actual series of justified reasons because they are defined strictly in terms of relations between beliefs in the series so that every succeeding belief is justified by the belief before it and so on ad infinitum and, 2) they are unable to mark the difference between a set of justified reasons that are connected to truth and one that is not because they are defined strictly in terms of a relation between beliefs in the series of reasons. However, Aikin argues that impure infinitism could surmount these problems without undermining the infinite regress condition because impure infinitism can solve the Modus Ponens Reductio, MPR, argument that threatens pure versions of infinitism. I argue that Aikin does not succeed because his impure infinitism faces some fatal consequences and any attempt to salvage it will undermine the infinite regress of justification.

Keywords: Infinitism; impure infinitism; modus ponens reductio; justification; regress.

1. *Introduction*

Two problems[1] confront what Aikin calls pure versions of infinitism. These are:

IP1: An infinite and non-repeating series of reasons cannot confer justification on a belief or cannot differentiate between an actual and a provisional set of justified reasons. In other words, an infinite series of reasons alone is not sufficient for justification.

[1] There are other problems of infinitism namely, the finite mind objection and the no-starting point objection and certain forms of the reductio problems (Klein 1999: 306ff). I take it that Klein's reply to these objections is convincing and proceed to look at these other two objections.

IP2: There is no way to tell the connection between truth and justification in an infinite series of justified propositions. The point is that an infinite series of reasons alone is arbitrary because it is not properly connected to truth.

IP1 questions our intuition that if there are infinite series of justified propositions of which each justified proposition justifies its predecessor in the series, how is the terminating reason in the series justified? Put succinctly, what justifies the proposition terminating the series of regress of justification? An unjustified proposition cannot do the trick because it will imply one has to construct an infinite array of justified propositions on an arbitrary or unjustified proposition.[2]

IP2 suggests that an infinite series of justified reasons cannot account for the connection between justification and truth. For instance, let us consider a subject, *S*, having two sets of infinite and non-repeating series of justified beliefs supporting *p* and ~ p respectively. Let's suppose that he claims that *p* on the basis of *q*, and *q* on the basis of *r*, and *r* on the basis of *s* and so on. Let's also suppose that she claims, on the same grounds, that ~ *p* on the basis of ~ *q* and ~*q* on the basis of ~ *r* and ~ r on the basis of ~ *s* and so on. The problem is that, on the infinitist account, we lack the requisite resource to differentiate between the series of reasons that have a connection to truth and the one that has not. The assumption is that it is difficult to track the connection between justification and truth if one merely takes into consideration only the logical and the inferential connection between propositions in each set of the series. The standard proposal is that there should be some further conditions that when considered in addition to this inferential relation will mark out the connection between justification and truth. Otherwise, there is no way to adjudicate between these sets of propositions to ascertain which one has a connection to truth.

While IP1 is usually regarded as a conceptual problem for infinitism, IP2 is construed as an epistemic problem.[3] Although IP1 and IP2 appear as distinct charges against infinitism, closer scrutiny reveals they are quite related to each other. IP1 suggests that an infinite array of justified propositions is not possible because such a series cannot differentiate between provisional and actual series of justified reasons. Any attempt to resolve this problem will require an additional resource that could undermine the infinite regress itself. IP2 rides on the back of IP1 in claiming that if IP1 be resolved, then it must be done in a way that reveals the connection between an infinite series of justified reasons and truth. Whereas Cling expresses pessimism as to how infinitism could resolve these two problems in tandem because to him

[2] This problem is usually credited to J. Post's and J. Cornman's Modus Ponens Reductio argument of infinitism, according to which if an infinite series of reasons is organized in such a way that each member in the series is logically implied by its successor to produce justified beliefs, then no proposition in the series will be actually justified. See Post (1980: 9–10) and Cornman (1977: 290–299). It flows from this that any arbitrary chosen contingent proposition would be justified.

[3] The tags are due to Cling (2004: 107).

any attempt "would undermine the rationale for the regress condition itself" (Cling 2004: 110), Aikin expresses optimism that these problems can be surmounted without undermining the regress condition necessary for justification in the infinitist scheme of things.

Aikin argues that an impure version of infinitism can circumvent both problems if it is couched to accommodate a foundationalist element of non-doxastic support to solve what he calls a Modus Ponens Reductio (MPR) argument against infinitism. He dubbed his version of infinitism 'strong impure infinitism' because his theory appeals to two sources of justification, which he thinks are individually necessary but insufficient for justification. This is unlike pure theories of justification that appeal to only one source of justification. So, an impure infinitist will hold that although the chain of justified reasons must be infinite and non-repeating, other formal principles of justification are legitimate (Aikin 2005: 199; 2008:178; 2011: 73).

Aikin uses justification trees (J-trees) to illustrate his position. J-trees are graphic representations of reasons a subject holds for believing a proposition, say P. He argues that J-trees could have as many branches as possible, but it is necessary that at least one branch in every J-tree proceeds infinitely and that a basic belief is included in every J-tree. These basic beliefs are supposed to be non-doxastic. However, Aikin explains that beliefs on the J-tree are not entirely justified by non-doxastic content only because the justificatory status of these basic beliefs also requires inferential support (Aikin 2005: 200). Hence, he dubbed his theory strong impure infinitism (Aikin 2011: 75); as he notes, impure theories of justification are more "ecumenical" because they permit more than one standard source of justification (Aikin 2011: 73).

I argue that Aikin is not successful in this venture because if any effort is made towards resolving IP1 and 1P2, the result will undermine the infinite regress condition and sway strong impure infinitism from the core moorings of epistemic infinitism. I mount two defences in favour of my thesis. First, I peruse Aikin's classification of meta-epistemic theories of both foundationalism and epistemic infinitism and argue that other options within the infinitist's ken could generate a strong impure infinitism without necessarily invoking a foundational source of justification within a chain of infinite reasons. Second, I argue that although the MPR argument against pure infinitism succeeds on one account, on another account, it fails against pure versions of infinitism.

2. *A Misleading Taxonomy*

I shall devote this section to rehearse Aikin's analyses of various strands of meta-theories of justification. He notes four possible combinations of meta-epistemic theories for foundationalism and epistemic infinitism. I shall present only that of infinitism here for brevity purposes but may draw on that of the foundationalism for illustration purposes.

Option 1: Strong, pure epistemic infinitism: the view that only infinite and non-repeating series of reasons is necessary and sufficient for justification.

Option 2: Strong impure epistemic infinitism: the view that an infinite and non-repeating series of reasons is necessary but insufficient for justification, thus other sources of justification are necessary.

Option 3: Weak impure epistemic infinitism: the view that an infinite and non-repeating series of reasons could yield justification but are not necessary for justification.

Option 3 initially appears as a non-starter until recently Fredrik Herzberg defended a version of it.[4] Given Aikin's account of infinitism, he favours option 2, an impure version of infinitism, which he says is analogous to strong impure foundationalism because it requires basic beliefs in addition to an infinite series of justification for a justification producing regress (Aikin 2008: 177ff). Furthermore, he assigns Klein and Fantl to option 1 because, according to him, they require that an infinite and non-repeating series of reasons as both necessary and sufficient conditions for justification producing regress (Aikin 2008: 177–178). I have a strong reservation concerning the foregoing classification, not on Fantl (because Fantl has clarified Aikin's view of him in the review of the latter's 2011 book)[5] but on Klein. The fact that Klein is a pure infinitist is contestable though in some portions of his papers he sometimes alludes to the thesis that infinitism is committed to the idea that the structure of justificatory reasons is infinite and non-repeating.[6] To be fair to Klein, he takes an infinite and non-repeating series of reasons to be a necessary but insufficient condition for a belief being justified for a subject; the propositions in the chain of reasons must also be available to the person (Klein 1999: 312). There must be another source of justification in addition to the infinite series of reasons for a proposition to be justified for a person on Klein's model of infinitism. This way of looking at the matter sits pretty well with option 2, the idea that infinite and non-repeating series of reasons though necessary is insufficient for justification; so other sources of support is legitimate. Thus, by Aikin's approach, Klein's brand of pure infinitism is impure.[7]

[4] F. Herzberg has developed an account of coherentism and infinitism that synthesises the presence and strength of inferential support or connections with probabilistic consistency to generate a thesis he dubbed the dialectics of infinitism and coherentism. See Herzberg (2014).

[5] See footnote of (Aikin 2008: 181). Moreover, Fantl's version of infinitism does not specifically draw on the regress, so it is difficult to begin to pitch his brand of infinitism onto any of the options although Aikin does that. Meanwhile, Fantl has clarified this misconception in his review of Aikin 2011. See Fantl (2012).

[6] See Klein (1999: 297)

[7] One can also mount the same argument using the emergentist approach where Klein notes that there is an emergentist intervention in his approach of infinitism where justification emerges as the series of reasons widens (See Klein 2007: 8 and 2008: 494).

Indeed, Aikin acknowledges Klein's idea of subjective availability of beliefs as a strategy for circumventing some reductio arguments but indicates that such intervention is different from his own (Aikin's) view. He avers the following, "my difference here is not with Klein's notion of belief, but with what supports those beliefs and in what way they may be dispositionally available to the subject" (Aikin 2005: 200). But this difference is so trivial to insulate Klein's brand of infinitism from occupying option 2. From the account offered above, what renders a theory impure is not the meta-justificatory challenge that it is likely to generate but whether it attracts individually necessary conditions as jointly sufficient for justification conferring regress.

Even if we stretch the argument, Klein's brand of infinitism could as well generate a meta-justificatory regress. For instance, if say a subject, S, utters the following expression: P: "I have hands" and proceeds to argue that the reason for P is subjectively available to her in the form of a dispositional belief, Q and that Q tentatively terminates the regress. A persistent interlocutor could demur by asking S why he thinks that Q can serve reasons for another proposition; or what makes Q assume that privileged epistemic status of rendering support to another proposition in the chain of reasons and so on. Here, the interlocutor could hold the infinitist accountable to their standards of ensuring epistemic responsibility. If the infinitist fails to answer this meta-justificatory challenge, then she endorses tacit foundationalism.

On the other hand, if she does, then she is a meta-justificatory infinitist. In the former case, the subject could be said to be an impure infinitist who is also a foundationalist at the meta-justificatory level. In the latter case, the subject is a thorough-going infinitist.

The lesson drawn from the above analysis is that Aikin's basis for rejecting Klein's version of meta-epistemic regress is not hinged on the reason that Klein's brand of infinitism is not impure. Instead, he rejects the meta-justificatory challenge that Klein's brand of infinitism could generate. But as noted, this should not be a basis for assuming that Klein promotes pure infinitism. Thus, Klein could be said to promote strong impure infinitism because, on the one hand, he allows that infinite and non-repeating series of reasons is necessary for justification and, on the other hand, he allows that subjective availability of reasons is necessary for justification. Both conditions are jointly sufficient for justification conferring regress.

But this does not immediately put the matter to rest. There is something clumsy about the label "strong impure infinitism." The reason is that pure theories of justification are exclusivist and, as Aikin notes, are committed "to the exclusivity of one source or formal structure of justification" (Aikin 2008: 175). Impure theories are not exclusivist and accommodate other sources of justification aside from their main structure of justification.

A theory is strong if the condition for a justification conferral regress is necessary but insufficient for justification. So for instance,

strong foundationalism will be the view that basic beliefs are necessary for a justification conferral regress. Weak foundationalism will be the view that basic beliefs are not necessary for justification because other sources of justification may be legitimate. On this construal, pure theories of justification are strong because of their commitment to only one source of support in the structure. In a similar vein, weak theories are impure because they allow other sources of justification amongst propositions in the structure aside from the relevantly main source.

What about strong impure infinitism? The answer appears simple when looked at from Aikin's construal of infinitism. But this generates some unpalatable consequence as well. Aikin argues that his version of infinitism is strong because it is committed to the idea that non-inferential series of reasons is necessary and productive of justification; yet it is impure because another source of justification is legitimate. It is refreshing to understand Aikin's strategy because he argues that his theory is not a dialectics of infinitism and foundationalism but an integrationist approach that integrates a non-doxastic component of foundationalism. He remarks: … "this insight that drives foundationalism can be incorporated and appreciated in most meta-epistemic theories, and it certainly can work here" (Aikin 2005: 199).

The non-doxastic states, according to Aikin, "have justificatory purport only in the context of inferentially rich support, but that condition does not mitigate their own independent, non-inferential justification" (Aikin 2005: 200). This means that the non-doxastic support is incorporated not for justificatory purposes but something else because the infinitist structure of justification remains intact with the introduction of the non-doxastic support. This "something" else according to Aikin ensures that there is a connection between infinite and non-repeating series of reasons in a chain, on the one hand, and the external world, on the other, to mitigate the trouble of the isolation objection (the idea that justification structures that rely on only the relations between reasons as the criteria of justification isolate the belief system from the world because beliefs are not justified in virtue of relations beliefs and the external world). So, unlike just being a mixed theory of justification, Aikin's theory is also strong in that it allows that the infinite regress and the foundationalist non-doxastic support are individually necessary for justification. As Aikin himself notes "what impure theorists do with the piecemeal of intuitive cases of knowledge is cobble together a systematic view of knowledge that allows a variety of sources of justification" (Aikin 2008: 176). But this is not what his impure theory does. Although it does "cobble together" other sources of justification, only one source of justification is salient regarding the function apportioned to the non-doxastic support in the infinite structure of justification. So Aikin's impure view, in a sense, is not the same as Haack's foundherentism, which combines foundationalism and coherentism (Haack 1993); BonJour's weak foundationalism, which combines observational

requirement into a coherentist structure (Bonjour 1985); or Herzberg, coherentism and infinitism which synthesises infinitism and coherentism (Herzberg 2014).

So, what does this bear on the prospects of Aikin's strong impure infinitism? The picture that we are presented with is that we have a brand of infinitism which focuses on the infinitist mode of justification (fitting pretty well into option 1) whose positive epistemic duty is to connect the entire chain of infinite reasons to the external world through the integration of non-doxastic support even though the regress condition undermines the non-inferential status of this non-doxastic input. So, there is an initial triumph over IP1; after all, it is not too stringently a problem, we can introduce another condition of justification as necessary without undermining the infinite regress of reasons.

What this means is that strong impure infinitism appears to miss its appropriate tag and qualifies to be placed in option 1 because, by the foregoing explication, it allows only inferential justification to be justification producing; the non-doxastic element only plays an epistemic role of allowing that an infinitist structure of justification possesses the relevant connection with the external world. In a similar vein, it qualifies to be slotted in option 2 because it introduces a further element in addition to the main infinitist condition of justification, the infinitist structure. The consequence appears to be that if Aikin's version of infinitism can be slotted in option 1, then there could be an exciting twist to the story and this would be explored in what follows.

3. *The Modus Ponens Reduction*

From Aikin's point of view, impure infinitism has a dialectical advantage over pure infinitism because the latter cannot solve the modus ponens reductio. The modus ponens reductio, MPR, according to Aikin, is invoked to adjudicate between series of justified reasons that are truth-conducive and those that are not. The argument is that an infinite chain of reasons is arbitrary when considered as a mere relation between beliefs. This is because there is no clear manner to adjudicate between a chain that is truth-conducive and one that is not. He remarks:

> The MPR …was simply that for any proposition, one could construct an infinite series of logically valid modus ponens inferences that support the proposition, and it turns out that a formally identical series can be constructed for the proposition's negation. The lesson was supposed to be that an infinite series of inferences cannot themselves distinguish between true and false propositions. (Aikin 2011: 105)

From Aikin's point of view, to circumvent this problem, basic beliefs are needed to distinguish between justified chains of beliefs that are truth-conducive and those that are arbitrary. Hence, *basic beliefs don't serve as regress enders* but only play an epistemic role of marking out an infinite chain of reasons that is truth-conducive from one that is not (Aikin

2005: 199; 2008: 183; 2011: 105). His strategy hinges on the assumption that basic beliefs possess non-doxastic support and this support does not mitigate their inferential role. On this construal, the reason why basic beliefs do not end the regress is that they only play the role of perceptual experience to indicate why a series of justified reasons is connected to truth: sensory experience provides our only acquaintance with the empirical world, so it provides the framework for our beliefs to match external reality. Hence, beliefs grounded in experience are more likely to be true because they map empirical reality.

The modus ponens reductio is at the heart of IP2, and the way Aikin resolves IP2 is by integrating a basic belief with non-doxastic support within an infinite series of inferential support. Aikin may be said to be right here. But he appears to argue that what he calls pure theories of infinitism, like Klein's and Fantl's versions, lack the resources to solve the modus ponens reductio. This is where I think Aikin is mistaken because Klein's strategy of invoking dispositionally available belief is to resolve the modus ponens reductio, albeit not exactly the way Aikin resolves it. Although Klein does not reply directly to the MPR argument against infinitism, one may draw on his strategy of dispositionally available belief to resolve this problem. Klein has replied John Post's reductio argument that if every proposition is justified by its successor, then for any contingent proposition, p, one can construct an instance of an infinite chain of regress. His reply was to suggest that the propositions in the chain must also be available to the subject. According to Klein, this way of looking at the issue does not place any constraint on the regress (Klein 1999: 312). This same strategy can be invoked to resolve the MPR. I am not holding brief for Klein, but I think that Klein has anticipated that any brand of infinitism will be vulnerable to the MPR if every contingent proposition and its negation are justified for the same person at the same time. But it will be fine if for a subject, S, the proposition in the chain terminating in p is available and for another subject S* (where S and S* are identical at different times or where S and S* are not identical), the proposition terminating in ~p is subjectively available in the form of dispositional belief. This will not raise any problem because S's set of beliefs that are dispositionally available to her will be different from S*'s and there is no constraint on why p shouldn't be justified for S and ~p be justified for S*. Once we grant this explanation some credibility, then Aikin's argument that Klein's version of infinitism cannot adjudicate between a series of infinite reasons and its negation is not compelling. So, Klein's version of infinitism avoids the unpalatable situation where a chain of infinite and non-repeating series of reasons will support both p and ~p.

It may be objected that this way of resolving the MPR is not an adequate account for addressing IP2. I concede. So, Aikin might be right after all that Klein's versions of infinitism lead to IP2. But it will seem erroneous to assume that Klein's version is vulnerable to MPR; it is

rather susceptible to the isolation objection problem. Put succinctly; it suffers from an epistemic problem of linking an infinite and non-repeating chain of reasons to the external world. The MPR and the isolation objection problem are not *necessarily* conjoined challenges against epistemic infinitism although there is a point of interdependence. An infinitist can resolve the MPR in another appropriate way and still may be troubled with the isolation objection problem as seen in Klein's intervention. But once the isolation objection problem is resolved, it is fixed in tandem with the MPR. The problem with Klein's version of infinitism is that it solves the MPR but falls shy at solving the isolation objection problem.

On the above showing, both Klein's and Aikin's positions appear to have chalked some initial success in the way they circumvent the IP1. The crux of IP1 is that one cannot introduce an additional source of justification in the regress of reasons without undermining the regress condition itself—meaning that one cannot introduce another source of justification in the chain of reasons and still be said of espousing infinitism. The challenge is that any theory that does so loses its moorings with epistemic infinitism. But have not Klein and Aikin show that one can achieve this feat and still be an infinitist?

Now, what is the prospect of both Klein's and Aikin's versions of infinitism on the account of IP2? I have already argued that Klein only resolves one part of the MPR and fails to resolve the other—the part that throws up the isolation objection problem. I will, therefore, focus only on Aikin's impure infinitism and how it attempts to circumvent IP2.

One of the remarkable tractions about impure infinitism as Aikin notes is that it "has the dialectical advantage over pure infinitism, because it can address the challenge of the 'modus ponens reductio'" (Aikin 2008: 184). As noted earlier, once the isolation objection problem against infinitism is resolved, it also caters for the MPR. So, Aikin could be right that he has successfully resolved IP2. But the worry is that Aikin resolves IP2 in a way that makes his impure infinitism vulnerable to IP1.

Recall that with impure infinitism, Aikin notes that the non-doxastic support introduced into the chain of infinite reasons does not end the regress. In the same vein, the regress does not mitigate the independence of the non-inferential justification of this non-doxastic support. So, the regress is nonetheless infinite despite the non-inferential status of the non-doxastic support– a plausible scheme to ensure impure infinitism gets caught up in option 2. Any theory of infinitism deserving of the name, however assorted, must be a 'strong' infinitism (because it regards non-inferentiality as necessary for justification) otherwise it loses anchorage with the core tenets of epistemic infinitism. Aikin's does this manoeuvre remarkably well except his solution generates an unpalatable consequence for his brand of infinitism. The problem is that Aikin does not only advocate for the incorporation of non-doxastic

support in an infinite series of justificatory reasons but also require it in addition to the infinite reasons to confer justification. That is to say, Aikin seems to say that the infinite and non-repeating set of reasons and the non-doxastic input are individually necessary conditions and both combined are necessary and sufficient conditions for justification. The foregoing, *a fortiori*, grounds his strong impure theory of justification.

This is what Aikin says about a justification conferring chain of infinite reasons: "Only infinite chains of reasons that integrate standing non-inferential input can confer justification" (Aikin 2014: 32). Henceforth, I shall refer to this as the justificatory conferring master statement for strong impure infinitism (JAA). Now, JAA seems to generate some unpalatable implications for impure infinitism.[8] First, the explanation Aikin offers to buttress JAA runs contrary to his initial reasons for incorporating the non-doxastic support into an infinite series of reasons. Aikin writes:

> Take these sort of arguments on analogy to the alternate systems arguments against coherentism: if formal relations between beliefs constitute justification, then couldn't one stipulate any series of beliefs with those formal relations, and thereby have justification? I have proposed one answer to this objection elsewhere, holding that other sources of non-inferential justification must work alongside these chains of reasons, which distinguish acceptable from unacceptable chains (and so, empirical evidence can determine which chain of reasons confers justification, and which one is merely hung in the air). (Aikin 2014: 32)

Aikin refers to two of his works (Aikin 2008 and 2009) and argues that he made a similar argument in those works. This is quite misleading because, in those works, Aikin had stated that the non-inferential support is meant to adjudicate between beliefs that are truth-conducive and those that are not in an infinite series of reasons. After all, the non-doxastic support does not end the regress (I have made this point earlier). How then does an infinite regress with non-doxastic support that does not end an infinite regress confer justification? How is justification conferred in an infinite series of reasons that incorporates non-doxastic support which does not end the regress of reasons?

Second, JAA appears to concede to IP2, but a concession to IP2 seems fatal to Aikin's infinitism because it undermines the regress of infinite justification and damages impure infinitism as a brand of infinitism. Aside from the concession to IP2, the idea itself raises questions because the infinite series of reasons and the non-doxastic support alone cannot confer justification without some other requirements counting in favour of a justification affording regress. What is the way out of this confusion? Three proposals are considered, and each is seen to deal a deadly blow to impure infinitism.

[8] Oakley noted one of these worries in his work. Refer to Oakley (2017).

P1: The infinite regress, the non-doxastic support and a sort of transmission requirement may be considered as individually necessary and jointly sufficient condition for justification conferral.

P2: The infinite regress, the non-doxastic support and a kind of emergentist criteria are individually necessary and jointly sufficient for a justification affording regress.

P3: Each of the propositions in the infinite regress of reasons contains non-doxastic support for complete justification.

None of these proposals seems plausible. For instance, P1 will make the regress of reasons redundant because justification would be conferred to the other propositions in the series of reasons via the transmission requirement which could undermine the regress of justification. For instance, if the propositions in the chain, p, q, r, and t are justified because the series is infinite and the proposition p at the head of the series integrates a non-doxastic support, then there cannot be any justification conferral unless justification is transmitted from p to the other beliefs in the series. Failure to endorse the transmission requirement in this context will amount to evading IP1. Alternatively, an endorsement of the transmission requirement will amount to a concession to IP1 which undermines the very regress of justification the infinitist hopes to promote.

In a similar vein, P2 will undermine the infinite regress of reasons. The assumption is that if the transmission conception is denied, then the impure infinitist can subscribe to the emergentist conception to prevent the theory from lapsing into foundationalism. Indeed, Aikin has opted for an emergentist view where he endorses a kind of "strong, synchronic, *emergentist* infinitism" (Aikin 2011: 72). So, let us see how the emergentist conception works with impure infinitism. Notice that the non-doxastic support is not a regress ender, so if there are series of propositions where each succeeding proposition is justified because its predecessor is, even where one of the propositions in the series integrates the non-doxastic support, the series will nonetheless run afoul of IP2. Giving that we deny the transmission conception, then the emergentist conception is an appropriate alternative. So, let us assume that each of the reasons in the series, p, q, r, s, and t are justified because justification emerges in the series as the series lengthens. This way, each proposition in the series starting from p could be independently justified and considered as basic propositions.[9] So the infinite regress of reasons is made redundant by this series of basic beliefs. The very idea of an impeded regress of justification seems to me inimical to any version of infinitist theory of justification.

Finally, P3 seems to me a non-starter because all the propositions in the series cannot integrate a non-doxastic element. As Oakley succinct-

[9] Oakley invokes a similar argument against Klein's emergentist conception and concludes that Klein's version of infinitism collapses into a kind of foundationalism.

ly remarks "it is impossible that each of the infinite chains of reasons should be matched at every point by an infinite number of non-doxastic input" (Oakley 2017: 17). The result is that since the non-doxastic support is not transmissible, only the proposition that incorporates the non-doxastic support will circumvent the MPR; others in the series cannot because they have not been "infested" with the non-doxastic support. So, impure infinitism fails to answer the MPR. The alternative is to assume the transmission requirement which I have pointed out is not feasible.

As it can be seen from the foregoing analysis, neither of these proposals appears plausible enough to salvage impure infinitism from possessing justification conferring status without losing its moorings with infinitism. On this account, one notes that only two options are available to Aikin: either he admits that strong impure infinitism is justification affording or he does not. Suppose he admits that strong impure infinitism is justification affording. In that case, the JAA is incoherent because JAA alone is an insufficient condition for a justification affording regress (this is based on the assumption indicated earlier that non-inferential support is not a regress ender). On the other hand, if it is not justification affording, then JAA with another requirement will be jointly necessary and sufficient for a justification affording regress, which may undermine the regress of justification. From all indications, by JAA Aikin requires that strong impure infinitism will yield a justification affording regress. The overall implication is that he succumbs in part to the challenge in IP1 that impure infinitism cannot generate a justification affording regress without undermining the regress of justification. Thus, if Aikin requires a justification affording regress for strong impure infinitism, he must admit that such intervention will undermine the infinitist regress of justification on the one hand, and on the other hand, he must accept that impure infinitism does not solve the MPR argument. It appears to me there is no way out of this problem.

4. *Conclusion*

The difficulty that beset Aikin's version of infinitism is apparent from his endorsement of non-doxastic support as a non-regress ender so that the theory may not stray out of the theoretical parameters of infinitism. But this problem generates more difficulties as an attempt to salvage epistemic infinitism from this clutter commits him to either undermining the regress of justification or evading the MPR argument. Aikin's master thesis is that the inferential chain with non-doxastic support is capable of conferring justification to a series of infinite and non-repeating series of reasons. This, however, is a mistake because Aikin notes that the non-doxastic support does not end the regress. So, strong impure infinitism does not possess a justification conferring property after all. Alternatively, to ensure that an inferential chain with non-doxastic support can confer justification in a series of reasons, one may

need to introduce another requirement in addition to the inferential chain and the doxastic support for a justification conferral regress. But this, as seen in the foregoing narrative, will undermine the regress of justification. My argument is a vindication of Cling's argument that it is not possible for one to be an impure infinitist of the type Aikin envisages without undermining the infinitist regress of reasons.

References

Aikin, S. F. 2005. "Who is afraid of epistemology's regress problem?" *Philosophical Studies* 126 (2): 191–217.

Aikin, S. F. 2008. "Meta-epistemology and the varieties of epistemic infinitism." *Synthese* 163 (2): 175–185.

Aikin, S. F. 2011. *Epistemology and the Regress Problem*. London: Routledge.

Aikin, S. F. 2014. "Knowing better, cognitive command, and epistemic infinitism." In J. Turri and P. Klein (eds.). *Ad infinitum: New Essays on epistemic infinitism*. Oxford: Oxford University Press: 19–36.

Bonjour, L. 1985. *The structure of empirical justification*. Cambridge: Harvard University Press.

Cling, A. D. 2004. "The trouble with infinitism." *Synthese* 138 (1): 101–123.

Cornman, J. 1977. "Foundational versus non-foundational theories of empirical justification." *American Philosophical Quarterly* 14: 287–297.

Haack, S. 1993. *Evidence and inquiry: A pragmatist reconstruction of epistemology*. New York: Prometheus Books.

Herzberg, F. 2014. "The dialectics of infinitism and coherentism: Inferential justification vs holism and coherence." *Synthese* 191 (4): 701–723.

Fantl, J. 2012. "Review of Scott F. Aikin, epistemology and the regress problem." *International Journal for the Study of Skepticism* 2: 157–60.

Klein, P. 1999. "Human knowledge and infinite regress of reasons." *Nous* 13: 297–325.

Oakley, T. 2017. "The reductio argument against epistemic infinitism." S*ynthese*. doi.org/10.1007/s11229-017-1629-z

Post, J. 1980. "Infinite regress of justification and explanation." *Philosophical Studies* 38: 31–52.

Croatian Journal of Philosophy
Vol. XXI, No. 62, 2021

Book Review

J. David Velleman, On Being Me: A Personal Invitation to Philosophy*, Princeton and Oxford: Princeton University Press, 2020, xiii + 91 pp.*

David Velleman's short and engaging invitation to philosophy consists of, as he describes it, "dispatches from an examined life" (xiii): concise and subjective reports about the nature of personhood, or, rather, about what it is like to be me. These reports are organized in seven chapters, which are nicely illustrated with drawings made by Emily C. Bernstein.

The first chapter, "Being Glad I Was Born," has ten sentences, one page, one drawing, and a meaningful thought about coming to exist being a chancy event, which seems as something that does not bother Velleman much, as opposed to the possibility of not going on existing. He leaves it as such, as he moves on to the second chapter, "Wanting to Go On," where he discusses what is usually called the problem of personal identity over time. As he exists, he wants to continue, but not as a soul or a thing. He wants to continue as a subject of experiences, thoughts and feelings. He wants to be able to refer to his *I* in any passing moment without being obliged to assume that it is this very person, David Velleman, who is continuing to exist. The personal pronoun "I" is as it were a hook which is being casted forwards and backwards in time thanks to memory and anticipation according to which persons organize their existence.

The cornerstone of the third chapter, "Fearing the End," is the experience of passing. Is our experience of the passage of time real? Velleman thinks it is not; rather than moving in time, he is extended in it. Thus, it is different segments, rather than the whole of him that occupies different days in his life, in a way in which the string of cars make up a train. "So the end of me is not a cliff toward which I am constantly hurtling; it's merely a segment of me with nothing beyond it in time, just as there is nothing of me above the crown of my head or below the soles of my feet" (34). This does not mean, however, that he does not mind having an endpoint, or final segment, in time: after all, the meaning and value of his present experience consists in desires and emotions about the future and the past.

What might have been if he had chosen ballet career instead of career in philosophy is in the focus of the forth chapter, "Regretting What Might Have Been." Velleman draws attention to different kinds of regret that are connected to omitted possibilities and asks whether the person who would have chosen a different life path would be him. His answer is negative: the

ballet career could not have been *his* past. "There is no point in comparing my story with other stories that, as of now, couldn't have been mine" (43); consequently, there is no point in regretting for what he might have been if he had made a different choice in the past.

The fifth chapter, "Aspiring to Authorship," and the sixth chapter, "Making Things Happen," address the issues of agency and free will. Velleman is considering whether determinism is a threat to personal identity and responsibility, and argues that his choices, desires and deeds are the result of his authorship and self-reflection. "Certified authorship" as a consequence of self-reflection is a key element of responsibility, with which he can "surprise even God" (45). The last chapter, "Wanting to Be Loved," goes further in declaring the value of introspection. The introspection by itself is not enough: he wants to be loved by others. Yet he wants to be loved for *being him*—not for being David Velleman or for having particular qualities or accomplishments, but for his deepest self that he meets and greets in silence, that has motives, does things, knows what it will do, and that is present throughout all changes.

The nature of Velleman's reports make the book readable, although the effort to riddle out the implications is solely the task of the reader. That task is imposed by the author's ability to say more with fever words, as well as by his personal style, which makes an impression that he is writing as he is thinking along the way. The good thing is that he is using an ordinary language while presenting a variety of difficult issues concerning personhood. The bad thing is that ordinary language is being used as a perfect cover for arguments in disguise. "There are no arguments," says Velleman in the "Preface" (xii); yet, the book is organized as a philosophical stream of consciousness, or a stream of consciousness by a philosopher, so that the sense of arguments lurking behind what are supposed to be personal explorations is inevitable (especially for readers acquainted with Velleman's writings). The motive for writing an invitation to philosophy in this fashion is observable if we take a closer look at the "Preface," and connect it with the author's scattered comments throughout the book. Velleman says that he realized that in his lengthy career, which has lasted for forty years, he was striving to make his arguments convincible. Now, he has written a philosophical book which is less convincible and more personal. As he points out, even though his personal side was always present in his arguments, it was there in the impersonal tone. "Most of these observations and speculations have appeared in my academic publications, sometimes supported by more or less formal arguments. For many years I thought the arguments were meant to convince the reader, but then I found myself oddly unconcerned when few if any were readers convinced. I finally realized that I have all along been reporting on personal explorations, composing dispatches from an examined life" (12–13). I would not say that this book is a career retrospective; however, there is some appeal to it. Velleman writes about what he was thinking about and what he was professionally doing for the last four decades in a more leisurely manner than he used to do throughout his career.

Velleman conceives of his book as an invitation to philosophy. Even though it does not have "*répondez s'il vous plaît*" included, the reader is

supposed either to accept the invitation or to reject it. As with other kinds of invitation, the response will depend, among other things, on expectations. Those who expect an easy to follow introduction to philosophy will remain unsatisfied. Being a person is, as Velleman notes, "a convoluted sort of thing to be" (p. xi), and some parts of the book are perhaps too difficult to follow for a beginner, who is, I assume, the principal recipient of the invitation. Besides, some of the thoughts developed in the book will be more easily understood by those who are already familiar with Velleman's views on the topics discussed here. For these reasons, some will probably waver about accepting the invitation.[1]

ANA GRGIĆ
Institute of Philosophy, Zagreb, Croatia

[1] This work has been supported by Croatian Science Foundation under the projects "Young Researchers' Career Development Project—Training New Doctoral Students" (DOK-2020-01-4459) and "Harm, Intentions and Responsibility" (UIP-2017-05-4308).

Croatian Journal of Philosophy is published three times a year.
It publishes original scientific papers in the field of philosophy.

Croatian Journal of Philosophy is indexed in
The Philosopher's Index, PhilPapers, Scopus, ERIH PLUS
and in *Arts & Humanities Citation Index* (*Web of Science*).

Payment may be made by bank transfer

SWIFT PBZGHR2X
IBAN HR4723400091100096268

Croatian Journal of Philosophy is published with the support
of the Ministry of Science and Education
of the Republic of Croatia.

Instructions for Contributors

All submissions should be sent to the e-mail: cjp@ifzg.hr. Submissions must be in English and formatted to be double-spaced with suitably wide margins, an A4 page size, and automatic page numbering.

Articles are normally no longer than 8,000 words of main text including bibliography. The Journal will consider longer papers, but, once these limits are exceeded, authors should bear in mind the editorial policy that the acceptance bar raises with increasing length.

Manuscripts should be compiled in the following order: cover page; title; abstract (not exceeding 200 words); keywords (3 to 6); main text; appendices (as appropriate); references.

All the authors of a paper should include their full names, affiliations, postal addresses, telephone and fax numbers and email addresses on the cover page of the manuscript. If a paper is co-written, one author should be identified as the Corresponding Author. The cover page must be submitted as a separate document. All submitted manuscripts must be prepared for blind review, with revealing acknowledgements and self-identifying references removed.

Sources are cited in the text by the author's last name, the publication date of the work cited, and a page number if needed, e.g. (Barber 2007: 324). Full details appear in the reference list in which the year of publication appears immediately after the author's name:

Barber, A. 2007. "Linguistic Structure and the Brain." *Croatian Journal of Philosophy* 21 (7): 317–341.

Williamson, T. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell.

The publication of a manuscript in the Croatian Journal of Philosophy is expected to follow standards of ethical behavior for all parties involved in the publishing process: authors, editors, and reviewers. The journal follows the principles of the Committee on Publication Ethics (https://publicationethics.org/resources/flowcharts).

ISSN 1333-1108
9 771333 110001