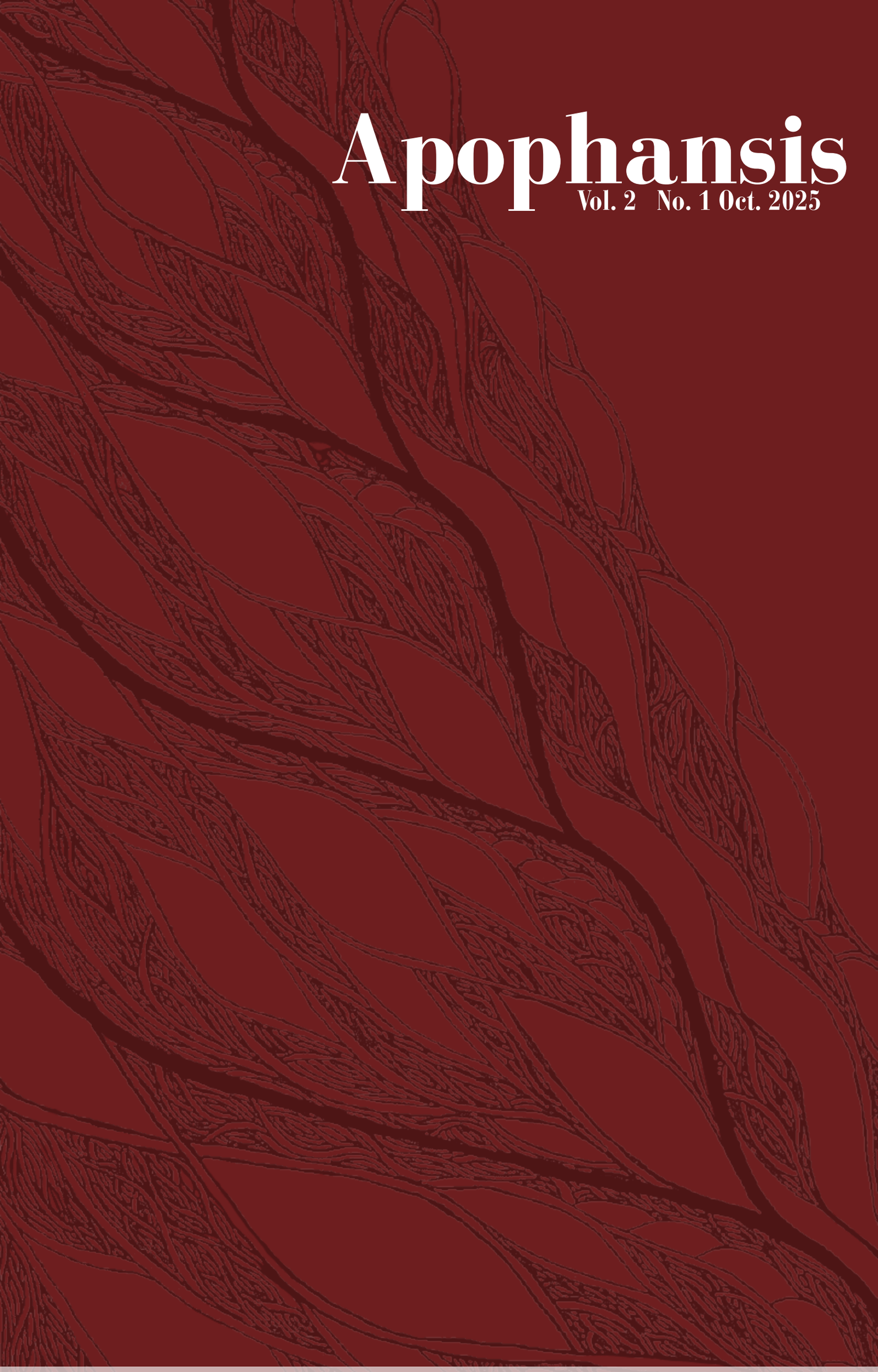

Apophansis

Vol. 2 No. 1 Oct. 2025

The Postgraduate Journal for Academic Philosophy from King's College London

Apophansis

The Postgraduate Journal for Academic Philosophy from King's College London

Contents

Letter from the Editors

We would like to thank Isabella Lucero and Nathanael Dorsch for their tireless efforts to establish *Apophansis* last year. Since they graduated from KCL, it has been our honour to run Apophansis, and we endeavour to live up to the high standard they set. Once again, Andrea Sangiovanni and Zita Toth have been crucial to this publication and we are grateful for their continuing support. We are proud of our first issue as Editors-in-Chief of *Apophansis* and of all the work that our great editors and writers have put into it, without whom none of this would have been possible.

This issue received even more submissions, making it difficult to choose among so many strong contributions, especially as our newly expanded word count left room for fewer pieces. The *Apophansis* Editorial Board carefully read and examined each article, and we selected those we believed had the greatest potential to further academic discussion. We hope you find them as interesting and thought-provoking as we did.

Our first section focuses on the History of Philosophy, featuring contributions from Mounir Freij and Jonathan Anderson. The second section turns to Moral and Political Philosophy, with a stand-alone essay by Rafal Miklas. The third and final section explores questions in Philosophy of Mind and Epistemology, showcasing work from Finlay Thwaite and Alexander McQuibban. We have maintained the same thematic structure as last year to preserve continuity and coherence within the journal. We believe these pieces reflect the diversity of philosophical thought that *Apophansis* seeks to promote, alongside a continued engagement with the past and future of philosophy.

As we bring forth this new issue of *Apophansis*, we find ourselves reflecting once again on what it means to do philosophy today. In a time when intellectual pursuits are too often weighed against their immediate utility, the philosopher's task might appear increasingly fragile. Yet, it is precisely in such times that philosophy can prove its resilience. For to philosophize is not merely to produce knowledge: it is to preserve a space for questioning, for dialogue, and for the slow cultivation of mutual understanding.

We, as young philosophers, take pride in continuing this task, aware of both its vulnerability and its necessity. The articles gathered in this issue bear witness to the diversity of voices and the vitality of thought that still animate our discipline. They remind us that philosophy is not an ornament of culture but one of its driving forces.

In the spirit of Wittgenstein, we might say:

Philosophy is not a theory but an activity. (Tractatus Logico-Philosophicus, §4.112)

It is this activity of thought, questioning, and dialogue that *Apophansis*seeks to celebrate and sustain. Without further preface, we invite you to read, reflect, and perhaps doubt alongside us. May these pages renew the conviction that philosophy's worth lies not only in its past, but also in its future.

Sincerely,
Daniel Ayres and Lorenzo Prota
Editors-in-Chief

Descartes Doctrines of Freedom

Jonathan Anderson

King's College London

1. Introduction

There is a sustained scholarly debate concerning whether Descartes was a libertarian or a compatibilist about freedom of the will.[1] This is perhaps surprising given the centrality of freedom to Descartes' metaphysics and theology. Freedom is central to Descartes' reply to the problem of evil and forms a part of his response to external-world scepticism, and yet his comments on free will are scattered across his philosophical texts and correspondences, often in statements that are unsystematic and apparently inconsistent. It is therefore the duty of historians of philosophy interested in Descartes to piece together his views on the matter for him. I maintain that Descartes' account of free will is compatibilist. In defence of this claim, this essay provides an account of Descartes' theory of freedom based primarily on his *Meditations* and his letters to Mesland, with some additional supporting references to his broader corpus. §1 begins by outlining Descartes' theory of the mind and by giving a *prima facie* case for the compatibilist interpretation of Descartes' thought. I offer textual evidence for his commitments to a doctrine of source freedom and to a doctrine of theological determinism. Since Descartes is at once a believer in free will and a determinist, it follows that he must be a compatibilist. §2 then considers Descartes' apparent commitment to a doctrine of leeway freedom, which has been considered problematic for the compatibilist readings of Descartes. I outline two potential interpretations of Descartes' notion of leeway freedom: as hypothetical or categorical freedom. On the hypothetical freedom interpretation, an agent is free only if she has the ability to do x or not do x, albeit under different initial conditions. By contrast, on the categorical freedom interpretation, an agent is free only if she has the ability to do x or not do x, even given the same initial conditions. In §3, I argue that Descartes is best interpreted as holding to a form of hypothetical freedom, which is consistent with the compatibilist interpretation. In doing so, I argue against some recent libertarian readings from Lilli Alanen and C. P. Ragland which attempt to ascribe to Descartes a doctrine of categorical freedom. Alanen's interpretation conflicts with some of Descartes' central examples of free agency while Ragland's reading is exegetically profligate. Accordingly, §4 concludes in favour of the compatibilist interpretation of Descartes' theory of freedom.

[1] Some terminology. There are at least two ways in which someone might be thought to possess free will. One way is that an agent, whenever he performs an action, has a genuine ability to do otherwise: that the agent has the ability to do x or not do x. To have the ability to do otherwise is to have leeway freedom. The other way is when an agent is the source of his own actions, in such a way that the causal history of the action can be attributed to him. To be the source of one's actions is to have source freedom.

2. Descartes on the Mind, Source Freedom, and Determinism

In order to examine Descartes' account of free will, it will be useful to begin with an account of Descartes' theory of the mind.

Like the scholastics, Descartes' account of the mind in relation to action is two-part, concerning the interaction between the intellect and the will.[2] In Descartes' taxonomy, 'the intellect' (*intellectus*) refers to the faculty of the mind which perceives ideas and propositions. In this sense, it is the faculty of understanding. The intellect perceives propositions that an agent might believe and practical actions that an agent might perform, detecting the extent to which those propositions are true and the extent to which those actions are good. By contrast 'the will' (*voluntas*) refers to that part of the mind which assents to or denies the perceptions of the intellect.[3] The intellect itself does not decide anything that the agent believes or acts upon: the intellect presents the will with propositions and actions, to which the will is then able to assent or deny.

For Descartes, the will is drawn necessarily in this operation towards truth and goodness as perceived by the intellect. This, Descartes claims, is a part of human nature: 'for man, since he finds that the nature of all goodness and truth is already determined by God, and his will cannot tend towards anything else, it is evident that he will embrace what is good and true all the more willingly [...] in proportion as he sees it more clearly'.[4] Nevertheless, even if it is the intellect that perceives truth and goodness, the power of judgement remains firmly with the will. The will does not always choose what is perceived most likely to be true or what is perceived to do the most good.[5] So long as the will continues to perceive a proposition, the inclination of the will remains, and it will be 'impossible' to shake the desire to assent or act according to that perception.[6] Yet Descartes holds that 'the nature of the soul is such that it hardly attends for more than a moment to a single thing'.[7] The will can 'call up before the mind' various perceptions from the intellect, thereby providing the will with a variety of reasons for or against assenting or dissenting to a proposition or action. Thus, even if we have a clear and distinct perception to do A, the will may obtain reasons for doing B instead by turning its attention to B. In such a case, the will retains the ability to suspend its judgement or to even form a contrary judgement to a perception of truth or goodness, so long as there are other perceptions to hand.[8] So considered, Descartes does not hold that intellectual comprehension always determines the volition of the will, though there are times that it might.

With Descartes' theory of the mind in outline, we can turn to consider what Descartes requires in order for the will to act freely. Descartes is absolutely clear *that* the will can act freely, for our freedom is 'very evident by the natural light of experience'.[9] Unfortunately, he is less clear as to *how* the will acts freely. Descartes' comments on the matter are scattered.

[2] Descartes 2003a, p. 46 and 2003b, p. 124; Ragland 2006, p. 61.
[3] Descartes 2003a, p. 47.
[4] Descartes 1985, p. 292; Ragland 2006, p. 78.
[5] Ragland 2006, p. 86.
[6] Descartes 1991, p. 233.
[7] Ibid.
[8] Ibid. p.234.
[9] Descartes 1985, p. 134.

Across his writings, he appears to endorse two different doctrines of freedom: a compatibilist doctrine and a libertarian doctrine.

First, let us take his compatibilist doctrine. This finds clear articulation in Descartes' *Fourth Meditation*. Here, Descartes claims that 'in order to be free [it is not necessary that] I must be capable of moving in either direction; on the contrary, the more I am inclined in one direction the more freely I choose it'.[10] Later in the same passage, he continues: 'if I always saw clearly what is true and what is good, I would never deliberate about what judgement to make or what to choose and thus, although I would obviously be free, I could never be indifferent'.[11] This textual evidence indicates that for Descartes, an agent acts freely when her mind, by virtue of its rational powers, has determined its action to pursue truth and goodness.[12] Therefore, Descartes endorses a doctrine of source freedom, according to which:

S. Some person, S, freely performs some action, x, at a time, t, if S's volition to x at t is determined by a clear and distinct perception that x is good.

In short, an action is free if the mind forms its volition to act according to what it perceives to be true or good.

So understood, this form of freedom is compatibilist. On this analysis, an agent may be causally predetermined to perceive that x is good and consequently assent to x, all the while retaining her freedom.In particular, Descartes has in mind the notion of the agent being predetermined by God. He describes that an agent may act freely even if 'God so disposes [one's] innermost thoughts', because in such a case the perceptions of truth and goodness are still what incline the will.[13] Freedom, so understood, is not incompatible with predetermination.

This aspect of Descartes' conception of source freedom is of fundamental importance when considered against the background of Descartes' wider theoretical commitments. As said, Descartes is certain from his own experience that free will exists, and yet, Descartes is also explicitly a theological determinist. This is clear in his correspondences. In his letter to Princess Elizabeth of October 1645, Descartes expresses that God is 'a supremely perfect being; and he would not be supremely perfect if anything could happen in the world without coming entirely from him'.[14] Here, Descartes is describing that the universe is metaphysically dependent upon God: because God continually sustains creation, nothing can occur other than what is determined by God.[15]

For Descartes, this determination applies not only to the inanimate world but equally to human agency. He writes:

> '[T]he greater we deem the works of God to be, the better we observe the infinity of his power; and the better known this infinity is to us, the more certain we are that *it extends even to the most particular actions of human beings*[. . .] I do not

[10] Descartes 2003a, p. 47.
[11] Ibid.
[12] Ibid.
[13] Ibid. p.47.
[14] Descartes 1991, p. 272.
[15] Wee 2014, p. 194.

> think that you have in mind some change in God's decrees occasioned by actions that depend on our free will. No such change is theologically tenable.[16]

On Descartes' view, it is theologically unacceptable that there could be a change to God's decree that does not come from Him, for this would impinge upon His infinite power. Since this decree applies to human agents as well, human agency cannot change what God has decreed for human action. Hence, Descartes is a theological determinist.

In sum, since Descartes affirms both human freedom and theological determinism, it follows by default that he must be a compatibilist with respect to free agency. (S), therefore, explains the conditions under which predetermined agents can exercise free agency, even given God's predetermination of the universe. As Descartes describes, 'neither divine grace nor natural knowledge ever diminishes freedom; instead, they increase and strengthen it'.[17]

3. Descartes' Doctrine of Leeway Freedom

So far then, we have good reasons in support of the compatibilist reading of Descartes. Unfortunately, the picture is quickly complicated, for in places Descartes appears to insist that a liberty of indifference is a necessary condition for free action.[18] In the *Fourth Meditation*, he claims that freedom of choice 'consists in this alone, that we are able to do or not to do (that is, affirm or deny something, seek or avoid it)'.[19] Meanwhile, in his letter to Mesland of February 9th 1645, Descartes claims that the will has a liberty of indifference, 'not only with respect to those actions to which it is not pushed by any evident reasons [. . .] but also with respect to all other actions'.[20]

It is important to be precise as to Descartes' exact use of 'indifference' here, for the concept of 'indifference' possessed multiple meanings in the early modern period.[21] Descartes employs at least two meanings of the term.[22] The first sort of 'indifference' refers to instances where the intellect does not provide sufficiently clear perceptions to persuade the will to assent one way or the other. In this sense, the will is 'indifferent' to an action because it does not strongly incline towards it.[23] The second meaning of 'indifference' refers to a positive power of the will to determine itself to its actions between (at least) two options.[24] In his letters to Mesland, Descartes makes clear that it is this second sense of 'indifference' that he holds necessary for freedom.[25] So understood, Descartes holds to a doctrine of leeway freedom, according to which:

L. Some person, S, freely performs some action, x, at a time, t, only if S could have performed x or not performed x at t.

[16] Descartes 1991, p. 273. Emphasis added.
[17] Descartes 2003a, p. 47.
[18] Ragland 2006, p. 63.
[19] Descartes 2003a, p. 47.
[20] Descartes 1991, p. 245.
[21] Lennon 2011, p. 578.
[22] Chappell 1994, p. 181-182.
[23] Descartes 1991, p. 245.
[24] Lennon 2011, p. 578.
[25] Descartes 1991, p. 245.

But if so, Descartes' commitment to (S) will be brought into question depending on how (L) is interpreted. Here, C. P. Ragland has distinguished between two possible readings of Descartes' commitment to (L): the hypothetical freedom interpretation and categorical freedom interpretation. On the hypothetical freedom interpretation, an agent is free only if she has the ability to do x or not do x, albeit under different initial conditions. The car can turn left if it is steered left and it can turn right if it is steered right, but turning the car left will not turn the car right, and vice versa. The car has the hypothetical freedom to turn left or right, but the mechanical ability to do so requires a change in the initial conditions. By contrast, on the categorical freedom interpretation, an agent is free only if she has the ability to do x or not do x, even given the same initial conditions. An agent that is free in this way is essentially underdetermined by the initial conditions temporally prior to the action.[26]

At issue then is whether Descartes holds that a free agent necessarily possesses hypothetical or categorical freedom. If he requires only hypothetical freedom, then (L) is in fact consistent with determinism, for a predetermined agent could act differently if she were placed under different initial conditions. However, if Descartes' theory requires that free action exhibits categorical freedom, then Descartes is a kind of leeway libertarian, rendering his view incompatible with determinism and *a fortiori* with the compatibilist interpretation of (S).

4. Hypothetical and Categorical Freedom Interpretations

My own view is that Descartes is committed to a doctrine of hypothetical freedom, for he leaves no logical space in the volitions of free agents for categorical freedom. In defence of this claim, let us suppose that Descartes *does* believe in the categorical freedom of human agency. The question that arises from this assumption is the following: when in the course of a free action does Descartes believe that an agent can exercise this kind of categorical freedom? Or, in other words, at what moment in the exercise of free agency is the operation of the will essentially underdetermined? The challenge for the categorical freedom interpretation lies in the difficulty in finding evidence of there being any place for categorical freedom in Descartes' account.

The simplest answer to this challenge comes from the 'straightforward' libertarian interpretation of Descartes, due to Lilli Alanen.[27] On Alanen's reading, Descartes holds that free agents always possess categorical freedom at the moment of the volition of a free action. So understood, freedom consists in the permanent ability to withhold assent from one's perceptions, even in cases where those perceptions are clear and distinct.

However, such an interpretation quickly finds difficulty. It would require us to revise Descartes' commitment to (S), for it implies that no amount of clear and distinct perception could determine the agent to act. But (S) seems to be a firm commitment in Descartes' account. He holds that the will 'cannot tend toward anything else' but truth and goodness.[28] Thus, if the intellect has perceived a proposition to be true or that course of action good, the only way to shift that assent is through another motive. But this means that, absent any reasons for a contrary judgement, the will *cannot* move in another direction. Thus,

[26] Ralgand 2019, pp. 181-182.
[27] Alanen 2002, pp. 293-294.
[28] Descartes 1991, p. 245.

Descartes holds that our leeway power over our clear and distinct perceptions is conditional: it is *conditional* on there being no countervailing perceptions against the original clear and distinct perception.[29] So Descartes describes, 'it is always open to us to hold back from pursuing a clearly known good, or from admitting a clearly perceived truth, *provided* we consider it a good thing to demonstrate the freedom of our will by so doing'.[30]

That Descartes prefixes the final clause with a conditional conjunction ('provided') is important: it denotes that absent these countervailing perceptions it is simply a part of our nature that we are drawn to assent.[31] In practice, it may be that the mind is indeed presented with a variety of perceptions and therefore the reasons to form judgements are often mixed. However, Descartes does not here suggest that there will always be a countervailing motive to any given clear and distinct perception.[32]

Therefore, in cases of wholly clear and distinct perceptions where there are no countervailing perceptions, no space can be afforded for categorical freedom.[33] The agent can only be said to be hypothetically free, for *if* the agent had different perceptions, she *could have* formed other volitions. In that case, an agent can exercise freedom without being essentially undetermined, and so it follows that categorical freedom is not a necessary condition for free action. Therefore, the evidence suggests that Descartes only requires the presence of hypothetical freedom in cases of perfectly clear and distinct perceptions.

Unperturbed by these considerations, C. P. Ragland has attempted to preserve Descartes' commitment to categorical freedom by attributing to him a diachronic theory of libertarian freedom. On this reading, Descartes does not suppose that free agents possess categorical freedom with respect to all of their actions at the moment of volition. Indeed, Descartes apparently denies this. Rather, Ragland proposes that on Descartes' account an individual still enjoys leeway freedom in cases of determined volitions because that predetermined volition is a result of an earlier volition *which itself enjoyed categorical freedom*. Thus, if S does x at t2, and S's x-ing is a result of a volition which determined that Swould x, S is still free so long as S's volition to x at t2 is the result of some earlier volition at t1which was itself categorically free.[34] Ragland himself provides no examples, but the thought can be illustrated by Daniel Dennett's popular example of Martin Luther.[35] When Luther posted his *Ninety-five Theses* against the Church, in his defence he famously claimed: 'Here I stand; I can do no other'. In Cartesian terms, we may interpret Luther to mean that his actions were necessarily compelled by his clear and distinct perceptions. By his principles, he could not do otherwise and therefore lacked categorical freedom. Nonetheless, on Ragland's proposal, this action is free *derivatively* insofaras it is determined by temporally prior volitions over which Luther *did* possess categorical freedom. In this case, these prior volitions could be Luther's formative decisions as to the kinds of moral and theological teachings he should endorse. For Ragland, if these earlier volitions are categorically free, then, once decided, they form Luther's character and determine some of his subsequent actions. Thus, these

[29] Ibid.
[30] Ibid., Emphasis added.
[31] Ibid., p. 233.
[32] Ibid., p. 245.
[33] Ragland 2019, p. 183.
[34] Ragland 2006, p. 81.
[35] Dennett 1984, pp. 555-556.

later actions are derivatively free, since they are determined by prior volitions which are not themselves causally necessitated.

So understood, leeway freedom can be diachronic: an intellectually determined volition derives freedom from a temporally prior volition that is categorically free. This means that Descartes can be committed to categorical freedom as a necessary condition for free action, *à la* leeway libertarianism, without requiring that all actions involve this genuine indeterminacy at the moment of volition. Moreover, this reading is consistent with Descartes' commitment to (S), with the added proviso that there is a categorically free volition in the causal history of any volition determined by one's clear and distinct perceptions. Ragland's interpretation is thereby able to save Descartes' libertarianism without committing him to the denial of (S); it is therefore more plausible than Alanen's libertarian reading.

Unfortunately, Ragland's interpretation salvages Descartes' libertarianism at the expense of some fairly substantial exegetical defects. The fundamental issue is that the claim that Descartes believed in derivative freedom is exegetically profligate: by his own admission, Ragland's interpretation of the derivative concept of freedom is an import he brings to Descartes' theory which is never made explicit in Descartes' own texts.[36] The closest Ragland gets to textual evidence is Descartes' letter to Mesland which claims that: 'freedom can be considered in the acts of the will either before they are elicited, or after they are elicited. Considered with respect to the time before they are elicited, it entails indifference in [the sense of having a two-way power]'.[37] This would suggest that there is a two-way indifference in the will prior to the moment of volition. But it must be noted that Descartes is ambiguous as to the nature of the power that is at play here. Descartes does not explicitly claim that this ability 'to do or not to do' is the ability to act or not act *simultaneously*, as is required by categorical freedom.[38] So the passage cited is in fact consistent with the hypothetical freedom interpretation, and therefore the compatibilist reading. Taken by itself then, the cited text does not endorse a notion of categorical freedom as argued for by Ragland.

In any case, what is even less clear is the further claim that for Descartes this earlier freedom is itself constitutive of human freedom in the later deployment of our volitions. In the same passage, Descartes goes on to describe that the freedom of the act *as it is elicited* 'consists simply in [its] ease of operation; and *at that point* freedom, spontaneity and voluntariness are the same thing'.[39] Thus, when describing the freedom of the will as it is exercised, Descartes does *not* refer to a prior form of leeway freedom to explain the basis of its freedom. Rather, Descartes here equates freedom with 'spontaneity' (*spontaneum*), meaning that the will is the source of its own action.[40] The will here has liberty of spontaneity because it is so strongly compelled by its perceptions, ensuring its 'ease of operation'.[41] Here again then, we find Descartes describing the freedom of the will in terms consistent with compatibilism.

If Descartes did believe in the possibility of derivative freedom of the will, we might expect either here or elsewhere to see him attribute the liberty of the will to some earlier categorical

[36] Ibid., p. 34.
[37] Descartes 1991, p. 245.
[38] Collins 2013, p. 224.
[39] Descartes 1991, p. 245. Emphasis added.
[40] Lennon 2015, pp. 71-72.
[41] Descartes 1991, p. 246.

freedom. That he does not is a telling omission for Ragland's interpretation. Thus, his reading goes beyond what is afforded by the textual evidence. Freedom, for Descartes, is not temporally protruded.

In sum then, Descartes' putative commitment to categorical freedom finds no place in his theory of free action. Accordingly, his doctrine of leeway freedom is best understood as a commitment to hypothetical freedom as is consistent with the compatibilist interpretation.

5. Conclusion

In this paper, I defended a compatibilist interpretation of Descartes' theory of freedom. Descartes believed in free agency in cases of determination of the will by clear and distinct perceptions and he believed in theological determinism. I have argued that there is no space for categorical freedom in such an account: libertarian readings of Descartes do not match the textual evidence and import theoretical commitments he does not maintain. By contrast, the compatibilist reading is exegetically parsimonious and best fits with the evidence. It follows that Descartes must be a compatibilist. In particular, he is a source compatibilist. Human freedom is a kind of rational excellence: an agent is free to the extent that she acts on clear perceptions of truth and goodness, and this does not require that she has categorical leeway freedom.

Bibliography

Alanen, L. K. (2002). 'Descartes on the Will and the Power to Do Otherwise'. In H. Lagerlund and M. Yrjönsuuri (Eds.), *Emotions and Choice from Boethius to Descartes.* Dordrecht: Kluwer Academic Publishers.

Chappell, V. (1994). 'Descartes's Compatibilism'. In J. Cottingham (Ed.), *Reason, Will, and Sensation.* Oxford: Clarendon Press.

Collins, B. (2013). 'Adding Substance to the Debate: Descartes on Freedom of the Will'. *Essays in Philosophy*, 14(2), pp. 218-238.

Dennett, D. C. (1984). 'I Could not have Done Otherwise - So What?' *The Journal of Philosophy* 81(10), pp. 553-565.

Descartes, R. (1985). *The Philosophical Writings of Descartes, Volume II.* (J. Cottingham, R. Stoothoff, D. Murdoch, Trans.). Cambridge: Cambridge University Press.

Descartes, R. (1991). *The Philosophical Writings of Descartes, Volume III.* (J. Cottingham, D. Murdoch, R. Stoothoff, A. Kenny, Eds.). Cambridge: Cambridge University Press.

Descartes, R. (2003a). *Meditations.* In D. Clarke (Trans.) Meditations and Other Metaphysical Writings. St Ives: Penguin Books.

Descartes, R. (2003b). *The Principles of Philosophy.* In D. Clarke (Trans.) Meditations and Other Metaphysical Writings. St Ives: Penguin Books.

Lennon, T. M. (2011). 'Descartes and the Seven Senses of Indifference in Early Modern Philosophy'. *Dialogue*, 50, pp. 577-602.

Lennon, T. M. (2015). 'No, Descartes Is Not a Libertarian'. In D. Garber and S. Nadler (Eds.), *Oxford Studies in Early Modern Philosophy, Volume VII.* Oxford University Press.

Ragland, C. P. (2006). 'Is Descartesa Libertarian?' In D. Garber and S. Nadler (Eds.) , *Oxford Studies in Early Modern Philosophy, Volume III.* Oxford University Press.

Ragland, C. P. (2019). 'Descartes on Freedom'. In S. Nadler, T. M. Schmaltz, D. Antoine-Mahut (Eds.), *The Oxford Handbook of Descartes and Cartesianism.* Oxford: Oxford University Press.

Wee, C. (2014). 'The Fourth Meditation: Descartes and libertarian freedom'. In D. Cunning (Ed.), *The Cambridge Companion to Descartes' Meditations.* Cambridge: Cambridge University Press.

The Tyranny of Poetic Pleasure

Mounir Freij

Christ Church, University of Oxford

Abstract

In this essay, I consider Plato's claim that the imitative poet puts a bad constitution in our souls [1]. I argue that although Republic X has been widely discussed, this charge in particular is poorly understood. A popular interpretation which holds that poetry damages our souls because we internalise the psychologies of those we see on stage ignores Plato's own assertion that imitative poetry puts a bad constitution in our soul because of the pleasure we experience from it. Interpreters who do acknowledge this claim are not precise about the nature of poetic pleasure. This is a problem because there is a certain asymmetry in the claims Plato makes about the psychological effects of base pleasures. Whilst he claims that the pleasures of poetry cause a drastic, violent upheaval in our souls, he elsewhere holds that one can experience a plethora of base pleasures over a number of years without the same psychological damage. This means that a complete account of Plato's psychological charge needs to explain what it is about poetic pleasure in particular which accounts for this qualitative difference in effect. This essay aims to fill this gap in the literature by providing a precise account of the reasons why poetic pleasures are uniquely damaging. I examine the unique properties of poetic pleasures through a close reading of key passages in the Philebus, Gorgias, Phaedrus and Republic IX. This provides the basis for my positive account of why they're so psychologically pernicious.

1.Introduction

In Republic X, Plato resumes his discussion of imitative poetry with some damning claims. Imitative poetry is at third remove from the truth [2]; despite being taken for experts, poets know nothing about their subject matter [3]; poetry appeals to the worst element within us [4], and can corrupt even decent people [5].

Perhaps the most intriguing claim, and that which I consider here, is that imitative poetry puts a bad constitution in our souls [6]. In this essay, I argue that although *Republic* X has been widely discussed, this charge in particular is poorly understood. A popular interpretation

[1] Republic 605b.
[2] Ibid. 595e-598e.
[3] Ibid. 598e-602b.
[4] Ibid. 602b-604a.
[5] Ibid. 605c-606d.
[6] Ibid. 605b.

which holds that poetry damages our souls because we internalise the psychologies of those we see on stage ignores Plato's own assertion that imitative poetry puts a bad constitution in our soul because of the *pleasure* we experience from it.

Other accounts do acknowledge this claim, but are not precise about the nature of poetic pleasure. This is a problem because there is a certain asymmetry in the claims Plato makes about the psychological effects of base pleasures. Whilst he claims that the pleasures of poetry cause a drastic, violent upheaval in our souls, he elsewhere holds that one can experience a plethora of base pleasures over a number of years without the same psychological damage. This means that a complete account of Plato's psychological charge needs to explain what it is about poetic pleasure in particular which accounts for this qualitative difference in effect.

This essay aims to fill this gap in the literature by providing a precise account of the reasons why poetic pleasures are uniquely damaging. I examine the unique properties of poetic pleasures through a close reading of key passages in the *Philebus*, *Gorgias*, *Phaedrus* and *Republic* IX. This provides the basis for my positive account of why they're so psychologically pernicious.

This is the plan for the paper. §2 considers a popular interpretation of Plato's psychological charge and, by showing its shortcomings, makes the case for an interpretation which emphasises the role of poetic *pleasure*. §3 continues the case with a close reading of Plato's psychological charge. I consider the asymmetry between the claims Plato makes about the effects of poetic pleasure and those he makes about the effects of other kinds of base pleasures to make the case that a complete account of the charge needs to explain what it is about poetic pleasure in particular that makes it uniquely damaging. I argue that existing accounts which do acknowledge the importance of pleasure are insufficiently precise. §4 constitutes my positive account of the psychological charge. I consider what makes poetic pleasure unique and, on this basis, explain why it is so psychologically pernicious. §5 concludes.

An important preliminary. There is a huge body of literature on the alleged inconsistencies of *Republic* X[7]. Plato has been accused of working with a different conception of imitation to book 1 III's, excluding more poetry than he did in books II-III, and operating with a different psychological apparatus to book IV's. Although this paper does not handle these issues directly, some will rear their heads throughout the discussion.

2. A Popular Interpretation: 'Internalisation'

Here, I consider a widespread interpretation of Plato's charge that imitative poetry puts a bad constitution in our soul [8].[9] My discussion of its shortcomings begins to make the case for the need for an alternative reading of this charge, a case I complete in the following section.

To get this interpretation of the charge on the table, we need to consider what Plato says about the kind of subject matter poets deal in. Poets, Plato famously argues, deal

[7] For some of it, see Annas 1981; Belfiore 1983, 1984; Burnyeat 1999; Greene 1918; Halliwell 1988; Janaway 1995; Marušič 2011; Moss 2007; Nehamas 1982; Singpurwalla 2011; Tate 1928, 1932.

[8] Republic 605b.

[9] This interpretation is endorsed by Burnyeat 1999, p. 322; Kamtekar 2019, p. 619; Lear 1992, pp. 208-9 and Nehamas 2 1982.

with a subject matter which is at third remove from the truth [10].[11] Plato introduces a three-fold distinction between Forms, sensibles, and appearances of sensibles. The works of poets are at third remove from the truth [12] since their subject is the appearance of virtue [13]. They deal with neither the Form of virtue, nor virtuous sensibles but the more ontologically impoverished subject of people who seem virtuous.

What are people who seem virtuous like? Plato describes the type of character poets imitate as 'excitable and multicoloured', aganaktetikon te kai poikilon.[14] In contrast to this is the 'rational and quiet character', phronimon te kai hēsukhion ēthos.[15] These terms mark an important ethical distinction. The word poikilos is particularly pejorative in Plato's writings [16]. It designates an activity or person which is stormy, passionate and emotional. This is exactly what characterises the unjust person [17]. The just person is the very opposite—moderate, sensible, harmonious and unchanging [18]. Most people, in their ignorance of true virtue, mistakenly take these multicoloured and contradictory characters as paragons of virtue, giving poets their subject matter.[19] The appearance of virtue thus turns out to be the very opposite of actual virtue. Poets present vicious characters to their audience, who praise them in the specious belief that they are virtuous.[20]

On what I am calling the 'internalisation' account of Plato's psychological charge against poetry, what makes poetry so dangerous is precisely the kind of character described in the above paragraph. This account claims that by putting ourselves in the shoes of these vicious characters ruled by their appetites, we take on their psychology. Thus, Jonathan Lear argues as follows:

> "'We imaginatively take up the perspective of the characters: even the best of us abandon ourselves and imaginatively take up their feelings ... By pretending to be these characters, we unconsciously shape our characters around them. The mimetic poet, says Plato, sets up a bad constitution in the psyche of each person.[21]

[10] Republic 598e.

[11] Poetry's metaphysical impurity is argued for via an extended comparison with painting. Although this analogy has been the subject of much criticism (eg., Annas 1981, pp. 336-338), not much turns on its success for my purposes, and so I set it to one side here.

[12] Republic 598e.

[13] Ibid. 600e, 599d, 602a.

[14] Ibid. 605a.

[15] Ibid. 604e

[16] Republic 561e, 604e; Laws 704d.

[17] Republic 444b, 445b.

[18] Ibid. 443c-444a.

[19] Cf. Republic 273a, 404e, 557c for the Platonic preference for simplicity over variety in other matters.

[20] This paragraph is indebted to Moss' 2007 discussion of the connection between the metaphysics of mimesis and the nature of the characters Plato thinks poets trade in. Incidentally, understanding the subject of poetic mimesis as appearances of virtue, and hence vice, provides a solution to the vexed problem of whether Plato excludes more poetry in book X than II-III. If in book X Plato is concerned with poetry which copies those who *seem* virtuous but are actually *vicious*, poetry which copies virtuous characters will not be imitative in this, more technical, sense, and so will still be permitted.

[21] Lear 1992, pp. 208-9.

A similar point is made by Myles Burnyeat:

> '[Poetry encourages people] to enter into the viewpoint of emotions which their better judgement, if it were active, would not approve. ... When we share an emotion with a character on stage, we enter (despite our better judgement) the moral outlook from which the emotion springs. [22]

On this account, the reason poetry is so psychologically pernicious is that when we sympathise with the perspective of the characters we watch on stage, we allow ourselves to enter into their perspectives. In the process, our souls are corrupted by imagining ourselves in the position of those we see on stage.

The first thing to note about this account is that for its plausibility it relies, to an extent, on a misrepresentation and oversimplification of tragedy and the Athenian audience's attitude towards it. It is therefore unlikely as an exegesis of Plato's thought given that, as an Athenian, his experience of attending the theatre complete with all its nuance would likely have informed his critique of poetry.[23] First, the account somewhat homogenises the figures we find in the extant tragic canon. Lear's remarks in particular on p. 214 about parricide and incest indicate that he has some of the most notorious tragic figures in mind, figures like Oedipus, Clytemnestra, and Medea. However, it is too straightforward to claim that the tragic poet is someone who 'externalises his appetites' through the characters he portrays.[24] There are plenty of figures who are not appetitive. One feels as though Plato may even have come close to approving of characters like Theseus in Euripides' *Supplices* or Ismene in Sophocles' *Antigone*.

Second, the Athenian audience was surely savvier than this account implies—they did not simply swallow the characters they saw on stage, assuming their vicious psychology. Instead, there was much active critical cultural engagement with tragedy's content and social function. Aristophanes' *Frogs* provides the best extant example of this. During the agōn, the poets discuss the kinds of characters each presented to their audience. Aeschylus claims he made the citizenry braver by putting valiant men on stage (1039-1044) and accuses Euripides of cheapening the genre by introducing immoral characters (771-783). The joke does not rely on the idea that tragedy had no didactic function; far from it—much of the humour of the play rests on the assumption that poets really did act as moral teachers. Rather, the point is that, expressed in the public forum of the theatre, this comment implies that the way in which the Athenian audience interacted with the theatre was less naïve than this interpretation of Plato might suggest. It would be more accurate to say that representations of regicide, parricide, incest etc. invited critical engagement with cultural taboos rather than causing the audience to straightforwardly emulate the figures they saw before them, assuming their vicious psychology.

These are minor points. More important is that Plato himself gives a different reason for why poetry has such devastating psychological effects. At 603c-605a, Plato is discussing

[22] Burnyeat 1999, p. 322.

[23] See Puchner 2010, ch. 1 for discussion of some of the evidence of Plato's personal association with the theatre.

[24] Learn 1992, p. 214.

the kinds of characters present on stage and he stresses that to get a good reputation, poets must represent emotional characters. The discussion shifts focus sharply at 605b to tragedy's effects on its audience's soul. At 605b7, Plato explicitly tells us that poetry changes the soul *by pleasing* its irrational element. The abrupt transition here makes it easy to assume that the same considerations which are at work in the previous section are at work here too: namely, the kinds of figures we find on stage. But making this assumption ignores Plato's own assertion that tragedy is psychologically destructive because of the pleasure it provides, rather than because of its content.

Lear and Burnyeat are surely right to hold that one of the reasons Plato objects to poetry is because its audience misinterprets the characters they see on stage as virtuous role models. But Plato makes his *psychological* charge against poetry not on the basis of tragedy's content but because of the *pleasure* which it engenders in its audience. I consider this passage in much greater detail in the following section when I continue to make the case for an alternative reading of Plato's psychological charge. For now, suffice to say that a more faithful reading of the text requires us to focus on poetry's *pleasure* for our account of the psychological effect it has. I return to the question of the relation of my alternative account to the internalisation reading at the end of §4.

3. Poetic Pleasure and Psychological Damage

Here, I continue to make the case that an alternative reading of Plato's psychological charge is needed, one which emphasises the role of pleasure. Here is the Plato's statement about the psychological effects of poetry, in full.

> a painter, he [the imitative poet] produces work that is inferior with respect to truth and that appeals to a part of the soul that is similarly inferior rather than to the best part ... He arouses (*egeirei*), nourishes (*trephei*) and strengthens (*ischuron poiōn*) this part of the soul and so destroys (*apollusi*) the rational one, in just the way that someone destroys the better sort of citizens when he strengthens the vicious ones and surrenders the city to them. Similarly, we'll say that an imitative poet puts (*empoiein*) a bad constitution (*kakēn politeian*) in the soul of each individual by making images that are far removed from the truth and by gratifying (*charizomenon*) the irrational part.[25]

The point is reiterated shortly after.

> in the case of sex, anger (*thumou*), and all the desires (*epithumētikōn*), pleasures, and pains that we say accompany all our actions, poetic imitation has the very same effect on us. It nurtures (*trephei*) and waters them (*ardousa*) and establishes them as rulers (*archonta*) in us when they ought to wither and be ruled, for that way we'll become better and happier rather than worse and more wretched. [26]

[25] Republic 605a-c.

[26] bid., 606d. The Greek is more explicit than the translation about the effects of poetry being intra-psychological: they take place *en tē psuchē* (606d2).

These rich passages are very close in tone and message and even use some of the same language (eg., *trephei*). There are several important points to make. First, Plato is adamant that imitative poetry is capable of fundamentally reconfiguring our souls: it puts in (*empoiein*) a bad constitution (*kakēn politeian*). It is clear that this "bad constitution" amounts to a soul ruled by the lower elements.[27] In the second passage, the lower parts of the soul are still firmly in view: *epithumētikōn* is an unmistakeable allusion to the appetitive part whilst *thumou* is to Spirit. Moreover, the language of ruling is even more explicit here (*archonta; archesthai*). There are slightly different senses in which a part of the soul may "rule" for Plato.[28] The most explicit statement of the sense relevant to this essay comes during the discussion of different lives in *Republic* IX [29]. Here, Plato suggests that when a part of the soul, X, rules, the soul regularly gives precedence to the ends valued by X.

Second, the nature of this change is presented as thoroughgoing and irreversible. Both points emerge from the incredibly strong language Plato uses. Poetry is characterised as a violent force which destroys (*apollusi*) the rational part of the soul at the expense of the lower element which is aroused, nourished, strengthened and watered. Plato employs the city-soul analogy to reiterate the point—it is like when someone destroys the best citizens (*chariesterous ftheirē*) and hands over the city (*paradidō tēn polin*) to the inferior element. Poetry has a power psychologically analogous to the overthrowing of an established rule—it renders the rational part of our souls impotent at the expense of the irrational part.

Third, the cause of this change is poetic pleasure in particular. The poet puts in a bad constitution by gratifying, *charizomenon*, [30] our irrational part. Plato reiterates this point shortly after [31]—the pleasure the base element of the soul receives in the theatre is directly responsible for the change in our souls. When our base element is nourished, *threpsanta* [32], by this pleasure, it cannot easily be held in check when we are out of the theatre. The point is also implied by the preponderance of pleasure-based language whilst Plato discusses the psychological effects of poetry at 605-6 (eg.,*areskein, charizomenon, chairomen, chairein, apoplēsthēnai, chairon, tēn hēdonēn*). It is clear that understanding poetry's pleasures will be key to getting a grip on why it changes the soul.

In the previous section, we considered the popular interpretation that imitative poetry is psychologically destructive because of a kind of reality effect whereby the audience's entering into the viewpoint of the characters they see on stage amounts to them assuming their psychologies. However, some interpreters have recognised that our account of poetry's psychological effects needs to insist on their pleasures. Iris Murdoch, for instance, writes as

[27] I assume that Plato is working with the same psychological apparatus book IV's and that the bipartite distinction between rationality and non-rationality in book X, maps roughly onto the distinction between reason on one side, and spirit and appetite on the other. (See Moss 2008 for an extensive defence of this assumption and Belfiore 1983, pp. 52-6 and Nehamas 1982, pp. 64-7 for critiques.) However, Plato is ambiguous here about poetry's exact target, eschewing the technical vocabulary of book IV. In my positive argument, I focus on how experiencing poetry leads to appetitive rule; I discuss the effect on spirit in 4.1.

[28] See Klosko 1988 for helpful discussion.

[29] Republic 580d-581e.

[30] Ibid. 605c.

[31] Ibid. 606b; cf. 607a.

[32] Ibid. 606b.

follows:

> 'We may find satisfaction in viewing the misfortunes of others ... Such experiences, for instance in the theatre, are not, as Aristotle later suggested, a purgation of our emotions by pity and fear, but rather a fostering of base impulses of sex and anger and selfish desire which ought to dry up and perish. [33]

Jessica Moss makes a similar point:

> 'Poetry's appearances influence and gratify the nonrational part of the soul, a part that experiences powerful and disruptive pleasure. By gratifying this part of the soul poetry strengthens it; thus the audience's rational thought is crippled, and their souls are harmed. [34]

These interpretations do better justice to Plato's text than the account we considered in the previous section: the reason for the psychological harm poetry is responsible for is the pleasure which it induces in the audience. However, these interpreters do not specify what it is in particular about poetic pleasure which makes it so psychologically pernicious. Moss claims the pleasures are 'strong'.[35] But she does not spell out precisely what it is about poetic pleasures which causes them to have the 'disruptive' effects they do.[36]

This lack of specification is a problem because it leaves unexplained just why it is that poetic pleasure has such an effect on our souls. Plato, as we shall see, holds that other strong pleasures do not affect our souls in this way. A complete account of the charge thus requires us to explain what it is about poetic pleasure in particular which accounts for the qualitative difference in the effect it has on our souls.

To get this point on the table, let us consider the asymmetry Plato seems to be committed to concerning the effects of poetic versus other base pleasures. Elsewhere, Plato explicitly claims that other pleasures do not have such detrimental psychological consequences: we can experience them over many years without having our souls corrupted.

Discussing the potential for vicious but clever people to change, Plato claims that all is not lost even after years of enjoying base pleasures.[37]

> 'However, if a nature of this sort had been hammered at from childhood and freed from the bonds of kinship with becoming, which have been fastened to it by feasting, greed, (*edōdais*) and other such pleasures (*toioutōn hēdonais*) and which, like leaden weights, pull its vision downwards—if, being rid of these, it turned to look at true things, then I say that the same soul of the same person would see these most sharply, just as it now does the things it is presently turned towards. [38]

[33] Murdoch 1992, p. 13.
[34] Moss 2007, p. 443. See Ferrari 1989, p. 138 and Halliwell 1988, p. 11 for similar points.
[35] Ibid., p. 441.
[36] Ibid., p. 443.
[37] The pleasures mentioned here are unmistakable references to those of appetite—see Republic, 439d.
[38] MRepublic 519ab.

Notwithstanding the highly metaphorical nature of this passage, its overall message is clear. There are some clear contrasts to draw with the power of poetry's pleasures both in terms of the effects of the pleasures and the possibility of reversing these effects. Plato claims that these pleasures simply make us consumed with the world of appearance; by contrast, poetry's pleasures were said to cause violent turmoil in our souls rendering us ruled by our base elements. Moreover, and crucially, attaining a just soul is still possible even after years of experiencing the smorgasbord of base pleasures. Pruning oneself of appetitive pleasures is a necessary condition on moral improvement but the soul still has the capacity to be turned around to 'look at true things'. In the case of poetry, the violent language of destruction implied that the educated and uneducated alike were powerless to reverse their effects.[39]

Similar points are implied by Plato's discussion of an aspect of the educational program of the guardians and rulers.

> 'We must expose our young people to fears and pleasures, testing them more thoroughly than gold is tested by fire ... anyone who is tested in this way as a child, youth, and adult, and comes out of it untainted, is to be made a ruler as well as a guardian.[40]

Plato does not explicitly tell us that these pleasures are base ones as in the passage above. Yet, given that these are preliminary tests prior to their education in dialectic [41], it is safe to assume that the relevant pleasures are not those of reason accessible to very few but ordinary pleasures, like those of food and drink designed to determine potential rulers' convictions. This is also implied by passages from the *Laws* [42] where men are given wine to test their self-control. By making testing in pleasures a condition on future guardianship, Plato assumes, as in the passage above, that one can quite plausibly experience them without having one's soul violently corrupted.

These passages (and others like them[43]) make it clear that although Plato is suspicious of base pleasures in general[44], he reserves his most damning charge for poetry. It has a power qualitatively unlike other pleasures which, even after lengthy exposure, do not violently and permanently upturn our souls.

The claim here is not that poetry is exceptional in being uniquely capable of causing lack of proper psychic governance. This is evident from the discussion of vicious souls in books VIII-IX. Rather, the claim is that it is unique as far as pleasures go. I return to this point in 4.3, arguing that constitutional change is a complex process, usually emerging from a considerable number of factors, not just pleasures. There, I suggest that experience of poetic pleasure may be an alternative route to developing a tyrannical soul.

The asymmetry between the claims Plato makes for poetry and those he makes for other appetitive pleasures is key to the bite of the question about why poetry has the psychological

[39] There are some exceptions but these are rare (Republic, 605c).
[40] Republic 413e-414a.
[41] Ibid. 503a, e.
[42] Laws 649d, 673e.
[43] See especially Republic 442a; 503a; 612a; Phaedo 83cd.
[44] See Fletcher 2018 for a proposal about how Plato's thoughts on pleasure in different dialogues are unified.

effect it does. It also allows us to understand why an account like Moss' or Murdoch's which does stress that poetry's pleasure is responsible for its psychological effects but is not specific about the nature of this pleasure will be an incomplete explanation of Plato's psychological charge. Plato acknowledges that experience of other "strong" pleasures, like those that come from feasting and drinking, is not incompatible with a just soul, even if experienced over many years. Without a more precise account of the nature of poetry's pleasures in particular, it remains unclear what accounts for their unique psychological effects.

This need to provide a precise account of what makes poetic pleasure unique in order to satisfactorily explain why it puts a bad constitution in the soul is the task I take up in the remainder of the essay.

4. An Account of Why Poetic Pleasure Puts a Bad Constitution in the Soul

Here, I show why the unique nature of poetry's pleasures account for the unique psychological harm they cause. First, I examine what makes poetic pleasure highly unusual. Second, I show why experiencing this kind of pleasure leads to a radical prioritisation of appetite's pleasures in general, a consequence which experience of ordinary base pleasures does not have. Third, I illustrate why this amounts to appetite's rule in the soul.

4.1. The Special Nature of Poetry's Pleasure

There are two important points to make about the special nature of poetry's pleasure.

First, the pleasure the irrational part of the soul receives from poetry is intense, and probably more intense than other kinds of pleasures[45]. Plato does not make this point explicitly in the *Republic* but it emerges clearly from an important passage in the *Philebus*. In his categorisation of different kinds of pleasure, Plato offers weeping at tragedies as an example of pleasure mixed with pains [46][47]. These pleasures, precisely because they are mixed with pain, appear greater and more intense, *hai men hēdonai para to lupēron meizous kai sphodroterai* [48]. Pleasure, when it is juxtaposed with pain, infects a subject's judgement about that very pleasure, making it seem stronger than it really is (*sphodros* almost means 'violent')[49]. Mixed pleasures do not *feel* like mixed pleasures; rather, they feel like extremely strong pleasures. Although Plato does not explicitly locate the mixed pleasures of tragedy in the lower part of the soul in the *Philebus*[50], the discussion here helps to illuminate why Plato insists on the dangers of poetic pleasure in the Republic.

This point about the intensity of poetic pleasures does not yet account for their qualitative difference from other pleasures. Indeed, weeping at tragedies is not the only example offered in the Philebus of mixed pleasures—anger, yearning, mourning, jealousy, envy and love are

[45] I bracket those pleasures (eg. philosophy) of the especially virtuous soul, which few can access.
[46] Philebus 48a.
[47] Cf. Phaedo 60b-c.
[48] Philebus 42b.
[49] Cf. Republic 586b. This is distinct from the criticism of 'false' pleasures at Philebus 37a-40e. Mixed pleasures produce false beliefs; false pleasures are parasitic on false beliefs. See Fletcher 2018 for discussion of this point.
[50] Though note that one of the reasons the pleasures of the lower part of the soul aren't 'true' is because they are mixed 29 with pain (Republic 586b-587a).

mentioned too[51]. We can understand the claim when we take it in combination with the following point about the unique context of the theatre.

The second, and more important point, is that the pleasures of poetry are detached from shame. Plato explicitly draws attention to the conspicuous lack of shame which accompanies watching someone grieve on stage: 'is it right to look at someone behaving in a way that we would consider unworthy and shameful, *mē axioi all'aischunoito an*, and to enjoy and praise it, *chairein te kai epainein*, rather than being disgusted by it?', Socrates asks Glaucon.[52] A few lines later, he reiterates this point: in the theatre, reason thinks 'that there is no shame, *ouden aischron*, involved for it in praising and pitying another man who, in spite of his claim to goodness, grieves excessively'.[53] Part of the problem with poetry, Socrates suggests, is that the pleasures it induces in its audiences are to things we ought properly be ashamed of but are not when we experience them in the theatre.

To appreciate the significance of this point, we need to consider briefly the role that shame plays in Platonic moral psychology. At various points, Plato proposes shame as important for our resisting the lure of appetitive pleasures. For instance, in the *Gorgias*, Callicles propounds hedonism but is forced to recant after realising that the corollary of his position is that the man who scratches an itch forever and the catamite will be happy. He is shamed into admitting that these pleasures are repugnant [54]. In this case, Callicles' shame is responsible for the reservations he feels about his initial position on base pleasure.[55] If it is right to think of shame as paradigmatically self-regarding, then it is clear why one would not experience shame in the theatre when watching someone else do something one would not permit oneself to do. Since the pleasure experienced in the theatre is thirdpersonal—*allotria pathē* [56]—it is unaccompanied by shame. This means that poetry provides appetite with a particularly alluring kind of pleasure since we do not experience the reservations which might arise from experiencing one's pleasure as shameful.[57]

However, the point is deeper than this. To appreciate the full significance of a lack of shame, we need to realise that Plato consistently connects shame to the working of spirit.[58] In *Republic* IV, the very first example offered to distinguish spirit from appetite— the case of Leontius and his ghoulish desire to look at corpses—illustrates this connection:

> had an appetite to look at them but at the same time he was disgusted, *duscherainoi*, and turned away. For a time he struggled with himself, *machoito*, and covered his face, *parakaluptoito*, but, finally, overpowered by the appetite, he pushed his eyes wide open and rush towards the corpses, saying, "Look for yourselves, you evil wretches, take your fill of the beautiful sight, *tou kalou theamatos!*"[59]

[51] Philebus 60b.
[52] Republic 605e.
[53] Ibid., 606b.
[54] Gorgias 494e.
[55] See also Phaedrus 254a; Republic 439e-440a.
[56] Republic 606b.
[57] Cf. Belfiore 1983, p. 61: the poet creates "what appears to be a special circumstance in which ordinary rules do not apply".
[58] For a full defence of this point, see Moss 2005.
[59] Republic 439e-440a.

One of the key words here is *kalos*. But since Leontius' remark is ironic, it immediately brings to mind its opposite, *aischros*. These terms (*kalos* and *aischros*), as Williams (1997) points out, are value-laden. Though they can have a solely aesthetic tone, they're clearly being used here to make ethical judgements. Here, spirit is working against Leontius' appetite's desire to ogle corpses by reacting with shame: Leontius is disgusted (*duscherainoi*) by his appetitive motivations and the result is intra-psychological conflict (*machoito*). Moreover, spirit's, ultimately unsuccessful, ardour is also evident in this example: its reaction is so zealous that it manifests in Leontius' behaviour (*apotrepoi heauton, parakaluptoito*).

The association of spirit with the emotion of shame recurs in the *Phaedrus*. In the palinode, Plato presents us with the famous image of the charioteer with his two horses, an unmistakeable reference to the tripartite soul.[60] Whilst one horse is of noble stock and character, the other is quite the opposite [61]. The bad horse is barely responsive to whips and spurs [62] while the good horse heeds reason's commands [63]. In the presence of a beautiful boy, the bad horse is motivated by the potential for sex but the good horse sides with reason.[64] Its sense of shame, *aidoi biazomenos* [65], prevents it from leaping on him. Again, Plato stresses the ardent reactions of spirit: 'the good one drenches the whole soul in sweat brought on by its shame and horror'.[66] This time, however, it is effective against the power of appetite.

Since shame is spirit's weapon against appetitive pleasure, the impossibility of shame in a theatrical context due to the third-personal nature of poetic pleasures means that spirit is not able to play its usual role.[67] Plato holds that spirit's job in the soul is to guard appetite to prevent it from becoming so filled up with pleasure it attempts to take over the whole soul [68]. Although spirit does not always play this role successfully (as we saw in the case of Leontius), it uses shame to prevent appetite from growing too fat on the pleasures which it experiences. This means that the lack of shame associated with experiencing poetic pleasures is doubly bad. It not only means that they are especially alluring to appetite since we are encouraged to assent to the pleasures without reservation, it also means that, since spirit needs shame to keep appetite in check, spirit is unable to prevent appetite from being directly strengthened by the experienced.[69]

4.2. Poetry's Pleasure and Other Appetitive Pleasures

With these thoughts in place, we are better placed to spell out Plato's psychological charge. To understand the next step, why experiencing poetry—a particular pleasure of the lower part of the soul—leads to a radical prioritisation of appetitive pleasures in general, we

[60] nterestingly the soul here is discarnate which makes it hard to see how it might have appetite. See Gerson (1987) for discussion of this problem.
[61] Phaedrus 246b.
[62] Ibid. 253e.
[63] Ibid. 253d-e.
[64] Though note it is specifically the charioteer who catches sight of the boy (Phaedrus 253e5), a vision which warms the rest of his soul.
[65] Ibid. 254a.
[66] Ibid. 254c.
[67] This point may help to explain the vexed issue of book X's seeming bipartite, rather than tripartite, distinction: the distinction reflects the impotence of spirit in a theatrical context.
[68] Republic 442a.
[69] Plato plays with the notion that poetry is like a drug at 598bd. Cf. Gorgias' *Encomium of Helen* §10-14.

need one more claim.

The pleasure derived from watching poetry is an unstable, and hence addictive kind of pleasure. To see this, we need to turn to Plato's remarks on pleasure in *Republic* IX where he distinguishes between metaphysically pure and impure pleasures and associates these categories with different degrees of satiation.[70] He vividly illustrates the plight of those who have had no experience of pure pleasures:

> aren't filled with that which really is and never taste any stable (*bebaiou*) or pure (*katharas*) pleasure. Instead, they always look down at the ground like cattle, and, with their heads bent over the dinner table, they feed, fatten, and fornicate. To outdo others in these things, they kick and butt them with iron horns and hooves, killing each other, because their desires are insatiable (*aplēstian*). For the part that they're trying to fill is like a vessel full of holes (*oude to stegon*), and neither it nor the things they are trying to fill it with are among the things that are.[71]

In Plato's view, the objects of reason are maximally pure, stable and true [7273]. These qualities enable them to give us genuine and lasting pleasure.

By contrast, metaphysically impure objects like food, drink and crucially for our purposes, poetry[74], offer a kind of pleasure which is not truly satiating. Moreover, they are correlated with a part of the soul which by nature cannot be truly satiated—it is 'like a vessel full of holes'.

Recall from section 2 that Plato holds that the works of poets are at third remove from the truth [75] since they deal with the subject of people who *seem* virtuous. Many commentators have urged that recognising that poetry deals only in appearances of virtue is crucial to understanding Plato's metaphysical (poetry is ontologically impoverished) and epistemological (poets know nothing about their subject matter) charges against poetry. Stephen Halliwell, for instance, puts the point like this: 'poetry and painting have no access to the true, transcendent and unchanging reality which lies beyond appearances. All this, if accepted, necessarily makes it is absurd to attribute deep knowledge of the world to poets or painters'.[76] The point I want to make here is that it is also crucial to understanding the psychological charge against poetry because it helps explain why poetic pleasure is so damaging.

[70] See Frede 1985 and Nussbaum 1986, pp. 146-164 for discussion of how these claims are developed in the Philebus. 41 See Sommerville 2019 for an argument that each of the three arguments in Republic IX is designed to appeal to a different part of the soul and its corresponding political class.

[71] Republic 586ab.

[72] Republic 479a; Phaedo 78d-e; Symposium 211b-d.

[73] See Gosling and Taylor 1982, pp. 112-115 for a discussion of how to square these claims with the earlier claims about 'pure' bodily pleasures like those of smell at 584b.

[74] The situation may be worse with poetry than with food and drink etc. since it is even further from the Forms.

[75] Republic 598e.

[76] Halliwell 1998, p. 7.

Plato argues that the pleasure metaphysically impure objects offer in general is unsatisfying. However, it is clear to see why poetry in particular, when we add this point to what we have just learnt, is uniquely psychologically dangerous. First, it requires a uniquely drastic response to keep ourselves satiated. Poetry's *qualitative* difference explains why experiencing even a plethora of other base pleasures does not call for such a response. Second, by directly strengthening our appetite in a situation in which our spirit is inoperative, it puts appetite in a uniquely powerful position to meet these demands.

One might think that this would just lead us to try and consume more poetry.[77] However, Plato describes how those suffering this plight do not discriminate between different appetitive pleasures in their attempts to satiate themselves: 'they feed, fatten and fornicate'.[78] This explains why poetry leads us to a radical prioritisation of appetite's pleasures *in general*, doing anything we can to try and achieve a semblance of the pleasure poetry offers. Of course, this might also lead to an attempt to keep consuming poetry's pleasures too, but, since they do not keep us satiated for long, we will turn to other appetitive pleasures when poetry is not available.

Moreover, the connection to other pleasures may be stronger than simply wanting or craving them. The image of filling up a vessel full of holes recurs in the *Gorgias* when Socrates tries to warn Callicles of the perils of hedonism. Again, pleasure is associated with the filling of a lack. The man dedicated to hedonism is forced to work day and night at keeping them [the jars] full, or else suffer terribly, *tas eschatas lupoito lupas*'.[79] One of the lessons of this memorable image is that unfulfilled desires can cause dreadful pain. When the initial hit of pleasure from watching poetry resides, we may not only crave other base pleasures, but be forced to seek them out on pain of suffering.

4.2. The Descent to Appetitive Rule

The next point to make is that the Sisyphean labour of gratifying the lower element of the soul with other pleasures amounts to its *de facto* rule in our soul. (In fact, Plato suggests at 580d-581e as I mentioned above that prioritising the ends of a given part of the soul just is what it means for that part of the soul to rule.) Consider how similar the figure just illustrated is to the description of the tyrant in book IX.

> when the other desires … buzz around the drone, nurturing it and making it grow as large as possible, they plant the sting of longing, *pothou kentron*, in it. Then this leader of the soul adopts madness as its bodyguard, *doruphoreitai te hupo manias*, and becomes frenzied, *oistra*.[80]

The tyrant is the paradigmatic example of someone ruled by his appetites. He caters to his insatiable desires by indulging them in the specious belief that he will eventually be satisfied, but this only makes him frenzied and mad. Moreover, Plato implicitly encourages comparison between the effects of imitative poetry and this description. In book IX, we are

[77] Cf. Plato's description of the 'lovers of sights and sounds' (475d).
[78] Republic 586a.
[79] Gorgias 493e-494a.
[80] Republic 573ab.

told that the pleasures of the tyrant's soul are distinguished by their illusory quality. His soul delights in images and shadow paintings (*eskiagraphēmenais*) of true pleasure [81]. This has a devastating psychological impact leaving him enslaved to his desires and wretchedly unhappy. Thus, when Plato makes the painting metaphors of book IX literal in the tenth book, he draws an implicit connection to the life of the tyrant.

The typical evolution of the tyrant is a complex process. It requires a combination of a bad upbringing, lawless activity, and the efforts of enchanters to persuade his lawless desires against his lawful ones [82]. His appetites, so transformed, are multiplied the more they are indulged in. The implicit comparison to the tyrant in book X, could be read as implying an alternative route: Plato seems to be suggesting that poetic imitation ought to be reckoned as potent as the forces which produce the tyrannical soul. Though other base pleasures may have detrimental effects on our souls, they require external forces to affect such a drastic change. By contrast, poetry has this power all by itself.

We now have our answer. Poetry, accompanied by pain and a conspicuous absence of shame, offers the base part of our soul a highly unusual kind of pleasure, and strengthens it in the process. However, this pleasure, since it is correlated with a metaphysically impure object and an insatiable part of the soul, is also ultimately unsatisfying. Since appetite is both hooked and strengthened on the pleasure poetry offers, this leads us to drastically prioritise other appetitive pleasures to try and keep ourselves satisfied and prevent the onset of pain. However, it turns out to be a Sisyphean task to keep this 'leaky' and 'holy' part of the soul satisfied. This labour amounts to prioritising appetite's dictates in our decision-making, letting it take assume rule of the soul.

Let me, in closing this section, briefly return to the question of the relation between my account of Plato's psychological charge which emphasises poetry's pleasure to the internalisation account we considered in §2. There, we saw that interpreters have claimed that poetry has the psychological effects it does because we internalise the vicious psychologies of those we see on stage. I made the case that a more faithful reading of the text requires us to focus on the nature of poetry's *pleasure* in explaining the psychological effects it has. But this alternative reading may be compatible with some of the insights of Lear and Burnyeat, given a different framing. It may, for instance, be the case that the content of poetry—vicious figures masquerading as virtuous role models—is another reason for the pleasure which poetry offers the appetitive part of the soul being so magnified. Taking up such issues, however, is beyond the scope of my discussion here.

5. Conclusion

In this essay, I considered Plato's charge against imitative poetry that it puts a bad constitution in the soul. I argued that this charge was poorly understood not only by those who favour what I called an 'internalisation' account but even those who acknowledge that our account of this charge needs to focus on poetry's pleasures. Since Plato holds that poetry's pleasures are unique in the psychological damage they are capable of, a complete account requires a precise explanation of what it is about poetic pleasure which underwrites this difference.

[81] Ibid. 586b
[82] Ibid. 572d-573b

I have attempted to provide such an account here. I argued that the admixture of pain and absence of shame combine to make poetic pleasure highly unusual. Poetic pleasure is not only experienced intensely, and without reservations, it directly strengthens our appetite since spirit is inoperative in a theatrical context. However, the metaphysical impoverishment of this pleasure ensures that it is highly addictive. Since our appetite is hooked and strengthened on poetic pleasures, we seek out more appetitive pleasures in the endless attempt to keep this part of the soul satisfied. Prioritising the dictates of appetites in this way amounts to letting them take over the rule of the soul.

Bibliography

Annas, J. (1981). *An introduction to Plato's Republic*. New York: Oxford University Press.

Belfiore, E. (1983). "Plato's Greatest Accusation against Poetry". *Canadian Journal of Philosophy*, 13(1), pp. 39-62.

Belfiore, E. (1984) "A Theory of Imitation in Plato's Republic". *Transactions of the American Philological Association*, 114, pp. 121-146.

Burnyeat, M. (1999). "Culture and Society in Plato's Republic". *Tanner Lectures on Human Values*, 20.

Ferrari, G. (1989). "Plato and Poetry". In: Kennedy, G. A. (ed.) *The Cambridge History of Literary Criticism*, Volume 1. Cambridge: Cambridge University Press.

Fletcher, E. (2018). "Two Platonic Criticisms of Pleasure". In: Shapiro, L. (ed.) *Pleasure: A History*. New York: Oxford University Press.

Fletcher, E. (2018). "Plato on Incorrect and Deceptive Pleasures". *Archiv für Geschichte der Philosophie*, 100(4), pp. 379-410.

Frede, D. (1985). "Rumpelstiltskin's Pleasures: True and False Pleasures in Plato's "Philebus"". *Phronesis*, 30(2), pp. 151-180.

Gerson, L. (1987). "A Note on Tripartition and Immortality in Plato". *Apeiron*, 20(1), pp. 81-96.

Gosling, J. C. B., Taylor, C. C. W. (1982). *The Greeks On Pleasure*. Oxford: Clarendon Press.

Greene, W. C. (1918). "Plato's View of Poetry". *Harvard Studies in Classical Philology*, 29, pp. 1-75.

Halliwell, S. (1988). *Plato, Republic 10, with Translation and Commentary*. Warminster: Aris Phillips.

Janaway, C. (1995). *Images of Excellence: Plato's Critique of the Arts*. Oxford: Clarendon Press.

Kamtekar, R. (2019). "Plato on Education and Art". In: *Fine, G. (ed.) The Oxford Handbook of Plato, second edition*. Oxford: Oxford University Press.

Klosko, G. (1988). "The "Rule" of Reason in Plato's Psychology". *History of Philosophy Quarterly*, 5(4), pp. 341-356.

Lear, J. (1992). "Inside and outside the "Republic"". *Phronesis*, 37(2), pp. 184-215.

Marušič, J. (2011). "Poets and mimesis in the Republic". In: Destrée, P. Herrmann, F-G. (eds.) *Plato and the poets*. Leiden: Brill.

Moss, J. (2007). "What is imitative poetry and why is it bad?" In: Ferrari, G. R. F. (ed.) *The Cambridge Companion to Plato's Republic*. Cambridge: Cambridge University Press.

Moss, J. (2005). "Shame, Pleasure, and the Divided Soul". In: Sedley, D. (ed.) *Oxford Studies in Ancient Philosophy*, 29, pp. 137-170.

Moss, J. "Appearances and Calculations: Plato's Division of the Soul". In: Inwood, B. (ed.) *Oxford Studies in Ancient Philosophy*, 34, pp. 35-68.

Murdoch, I. (1992). *Metaphysics as a Guide to Morals*. London: Chatto Windus.

Nehamas, A. (1982). "Plato on Imitation and Poetry". In: Moravcsik, J. M. E. Temko, P. (eds.) *Plato on Beauty, Wisdom and the Arts*. Totowa: Rowman Littlefield.

Nussbaum, M. (1986). *The fragility of goodness: luck and ethics in Greek tragedy and philosophy*. Cambridge: Cambridge University Press.

Plato. *Republic*. (1992). Translated by G.M.A Grube; revised by C.D.C. Reeve. Indianapolis: Hackett Publishing Company.

Plato. *Gorgias*. (1994). Translated by R. Waterfield. Oxford: Oxford University Press.

Plato. *Phaedrus*. (1995) Indianapolis :Hackett.

Puchner, M. (2010). *The Drama of Ideas: Platonic Provocations in Theatre and Philosophy*. Oxford: Oxford University Press.

Singpurwalla, R. (2011). "Soul division and mimesis in Republic X". In: Destrée, P. Herrmann, F-G. (eds.) *Plato and the poets*. Leiden: Brill.

Sommerville, B. (2019). "Pleasure and the divided soul in Plato's Republic book 9". *The Classical Quarterly*, 69(1), pp. 147-66.

Tate, J. (1928). "'Imitation' in Plato's Republic". *The Classical Quarterly*, 22(1), pp. 16-23.

Tate, J. (1932). "Plato and Imitation"". *The Classical Quarterly*, 26(3/4), pp. 161-169.

Williams, B. (1997). "Plato against the Immoralist". In: Höffe, O. (ed.) *Platon: Politeia*. Berlin: Akademie Verlag.

Sexual and Nonsexual Autonomy in Cases of Deception into Sex

Rafal Miklas

London School of Economics and Political Science

Abstract

In this paper I challenge the view that all instances of deceiving someone into sex (DIS) are seriously morally wrong. While I admit that DIS can be morally problematic, I argue that its seriousness does not result directly from the fact that it pertains to sex. Rejecting Tom Dougherty's claim that all DIS violates sexual autonomy, I introduce a distinction between sexual and nonsexual sphere of personal autonomy. I argue that not all deceptions undermine one's sexual autonomy and, thus, should not be categorised as distinctly sexual wrongs. This approach explains why some intuitively trivial deceptions, e.g. about diet or education should not be equated with offences such as impersonating one's spouse or disguising sexual assault as medical procedure. In defending this distinction, I address objections from sexual moralism, redundancy, and the risk of excusing sexual abuse. My account preserves the seriousness of violating sexual autonomy while offering a more coherent moral framework for assessing DIS.

1. Introduction

"I never sleep with anyone on the first date", whispered Jake into Jenny's ear as he unbuttoned her blouse. The night before he said – and did – the same to Alex and two evenings ago to Richard. Each of Jake's partners had clearly stated that they were not looking for one-night stands and did not want to have sex with someone who regularly hooked up with strangers. In this paper I will attempt to show that, although Jake's behaviour may be morally wrong, perhaps even 'seriously' so, its gravity does not distinctly result from its sexual objective. I will argue that it is *not always* seriously morally wrong to *deceive another person into having sex* (DIS) and that whether DIS is *seriously* wrong must be assessed independently from the fact that it involves sex. My argument distinguishes between sexual and nonsexual spheres of personal autonomy and, consequently, between sexual and nonsexual reasons for action. I argue that not all instances of DIS violate one's sexual autonomy, even if the deception pertains to a deal-breaker for that person. I begin by highlighting conflicting intuitions about the wrongness of DIS. I then reject Tom Dougherty's view that all DIS is seriously wrong and argue for a distinction between sexual and nonsexual reasons for sex. I show that adopting this distinction allows for a nuanced assessment of how serious a case of DIS is. I also argue that it helps to avoid the problems associated with Dougherty's view, such as its consent-centrism and the risk of undermining the perception

of severity of sexual offences overall. Finally, I respond to objections and motivate my conclusion that an intuitively non-absurd account of sexual offence is necessary for ensuring that sexual consent is taken seriously.

2. The Lenient Thesis and Conflicting Intuitions

The debate over the wrongness of DIS reveals two conflicting intuitions. Initially, misinforming someone about a potential deal-breaker for sex appears wrong since it precludes morally valid consent. On the other hand, it seems implausible to label minor deceptions, such as about being vegan, as serious sexual offences. Furthermore, some matters of deceit are widely viewed as more morally significant than others. One heavily scrutinised area is the nature of the sexual act itself. It seems non-controversial to claim that more people would classify as serious wrongdoing cases of lying about the type of the intended sexual act, e.g. protected vs unprotected, than deceiving someone about one's professional prospects. Moreover, the numerical identity of a sexual partner – the "*who* they are" – is considered to be more morally important than the sum of their personal characteristics – the "*how* they are". For example, impersonating someone known personally to the victim is usually perceived as more seriously wrong than misrepresenting one's religious convictions. The UK's Sexual Offences Act 2003 explicitly states that consent is vitiated in cases where: 'the defendant intentionally deceived the complainant as to the nature or purpose of the relevant act;' and where 'the defendant intentionally induced the complainant to consent to the relevant act by impersonating a person known personally to the complainant.'[1] These legal provisions render the sexual act non- consensual only when the deceit occurs in two specific areas – the nature of the act and the impersonation of an acquaintance. Thus, it appears intuitive that at least some cases of DIS are not seriously wrong, however blurry the boundary between them is.

Tom Dougherty captures the above intuition in the Lenient Thesis (LT): 'It is only a minor wrong to deceive another person into sex by misleading her or him about certain personal features such as natural hair color, occupation, or romantic intentions'[2]. However, he rejects it and argues that DIS is always seriously wrong since any deception may vitiate morally valid consent. His argument can be summarised as follows [3]:

A: *'Having sex with someone, while lacking her morally valid consent, is seriously wrong'*[4]

A1: *Non-consensual sex is seriously morally wrong since it violates one's sexual autonomy.*

A2: *Non-consensual sex violates one's sexual autonomy since one does not validly consent to the sexual act.*

B: *Deceiving another person into sex involves having sex with that person, while lacking her morally valid consent.*[5]

[1] Sexual Offences Act 2003, s. 76.
[2] Dougherty 2013, p. 718.
[3] In the below summary I intentionally omit Dougherty's final premise, the "Argument from a Substantive Account of Consent", since Dougherty himself admits that his conclusion can be reached without accepting this final premise and it does not challenge the argument which I am presenting.
[4] Dougherty 2013, p. 720.
[5] ibid.

B1: *DIS vitiates/precludes morally valid consent since all matters subject to DIS may constitute a meaningful part of a sexual encounter.*

B2: *DIS about meaningful matters of a sexual encounter vitiates morally valid consent since morally valid consent must be informed – the consent giver must know what they are consenting to – and the deceit means they are not informed.*

B3: *The mere awareness of the risk that the matter of deception may be a deal-breaker to the consent giver means that the consent receiver cannot reasonably believe to have the deceived morally valid consent.*

C: *Therefore, deceiving someone into sex is seriously wrong*[6]

To illustrate, consider the following example:

The Devout Vegan

Malcolm, a vegan, meets Jenny, a chef. He tells her that he wants nothing to do with people who eat or even touch meat. Hearing that, Jenny omits that her work involves meat, so as not to ruin her sexual advances. Malcolm and Jenny begin flirting and they proceed to engage in sex which they both find satisfying.

According to Dougherty's account, Jenny acted seriously wrong, she had sex with Malcolm without his morally valid consent (B2). Jenny's behaviour was seriously wrong since it violated Malcolm's sexual autonomy (A1). However, although Jenny wronged Malcolm, it appears far-fetched to accuse her of sexual assault. Dougherty denies this intuition, arguing that had Malcolm known the truth, he would not have had sex with Jenny and, thus, her deceit vitiates his consent (B2). Dougherty's conclusion is based on rejecting what he calls 'sexual moralism, the position that there are morally significant and insignificant reasons to have sex.[7] He argues that all lies can potentially vitiate consent for sex since all matters can be deal-breakers for the consent giver. Dougherty effectively denies the possibility of objectively distinguishing between core and peripheral features of a sexual encounter and, consequently, rejects such distinction as grounds for differentiating between minor and serious moral wrongness of DIS. He emphasises that such distinction 'rests on an objectionably moralised conception of sex'.[8]

Dougherty's point is that without committing to such 'moralised' conception of sex, whether some seemingly trivial feature really is trivial is necessarily subjective. He argues that in the absence of moralism about sex, every feature of the sexual encounter may be core to the deceived party (be a deal-breaker for that person). As I will argue, however, accepting Dougherty's argument in this shape would have unwelcome implications for the overall framework of protecting sexual autonomy. First, his account seems to place too much weight on consent, overlooking serious harms that can result from deception about matters that are not deal-breakers, as well as cases of morally troubling but consensual sex. Second, by treating all cases of DIS as seriously wrong, his view risks being overextensive. It places intuitively minor wrongs, such as lying about one's diet, on the same plane as clear sexual offences, which may undermine the perceived severity of the latter. What appears to be needed is a way which explains why some cases of DIS are more serious than others, without

[6] ibid.
[7] Ibid., p. 730.
[8] ibid.

collapsing them all into the same category. I will now move to outlining an approach which I believe does just that and effectively reconciles the conflicting intuitions. I will do so by arguing against premise B1 of Dougherty's argument and showing that while all matters subject to DIS may be a meaningful part of a sexual encounter, only some of them are constitutive of the sexual act itself.

3. The Sexual-Nonsexual Distinction and Defending The Lenient Thesis

Even accepting Dougherty's rejection of "sexual moralism" does not mean that DIS is always seriously wrong. One approach to relieving the intuitive tension is to distinguish various spheres of personal autonomy. From this, it is natural to assume various kinds of reasons for performing some act: having a R-reason to do Z amounts to expressing your R-autonomy to do Z. For example, I may have a culinary (R-) reason to eat pasta over pizza (Z) since, I find pasta tastier. By acting on this reason, I am *expressing my culinary* (R-) *autonomy*. In other words, the expression of my culinary autonomy is captured in my decision to eat pasta rather than pizza. Accepting this assumption yields a distinction between *sexual* and *nonsexual* spheres of personal autonomy and, consequently, *sexual*, and *nonsexual* reasons for having sex. This distinction highlights that one can be wronged sexually and non-sexually, depending which autonomy sphere is violated. This intuition supports the following argument:

D1: *DIS about nonsexual features of a sexual encounter does not violate X's sexualautonomy.*

D1.1: *There are sexual and nonsexual features of a sexual encounter.*

D1.2: *Sexual and nonsexual features of a sexual encounter weigh into X's decision about having sex in different ways.*

D1.2.1: *Sexual features determine X's sexual decision to have sex.*

D1.2.2: *Nonsexual features can determine X's all-things-considered decision to have sex.*

D1.3: *Expression of X's sexual autonomy is captured by X's sexual decision.*

D2: *The distinct seriousness of harm from non-consensual sex results from the violation of X's sexual autonomy*

E1: *Therefore, DIS is not always seriously morally wrong.*

E1.1: *The severity of the wrongness of DIS depends on the matter of deceit.*

E1.2: *In cases where DIS is about nonsexual features, its severity is determined by considerations independent from the sexual features.*

This argument claims that DIS is not always seriously wrong by virtue of being a distinctly sexual offence. One can deceive someone into sex they would not engage in had they had known about some nonsexual feature, e.g. their partner's political inclinations, without wronging them sexually. To illustrate this, consider another example:

The Catholic Burger Eater

On Friday, Rachel offers to make Tom a meat cheeseburger. She knows Tom is fond of burgers and enjoys them. However, he is also a devout Catholic who refrains from meat on Fridays. Desperate to cook the burger and have Tom eat it, Rachel deceives him that it is Thursday. Tom, convinced he is committing no sin, eats the burger and enjoys it.

Tom was clearly wronged by Rachel. Her deceit resulted in a violation of his autonomy. However, I hope that the example extracts the intuition that it was not the consumption itself which violated it, but rather the context in which this consumption took place and Rachel wronged Tom by situating the act in this context. Let us differentiate between Tom's: (i) culinary autonomy – freedom to decide what one eats – and (ii) religious autonomy – freedom of belief and religious expression. Rachel's deceit violates (ii), but *not* (i). She does not coerce Tom into eating nor lies about the burger's meat content. She deceives Tom about his *non-culinary* reasons to eat the burger. Since she does not deceive him about the *culinary* reasons, she does not violate his *culinary* autonomy and, consequently, does not *wrong him culinarily*.[9] Applying this logic to the Devout Vegan clarifies that, while Jenny wrongs Malcolm by remaining silent about her employment, she does not wrong him *sexually*. She still commits a moral wrongdoing, however, not a distinctly *sexual* one. Thus, the severity of Jenny's DIS cannot be directly attributed to the sexual nature of this DIS and must be established on grounds independent from such sexual nature.

The above result can also account for cases of consensual, yet morally troubling sexual relations. Consequently, adopting the sexual-nonsexual distinction allows for more nuance which solves one of the key problems with Dougherty's account, namely its consent-centrism. While Dougherty's approach is restrictive – it labels all cases of DIS as seriously morally wrong – his framework appears objectionably consent-centric. The severity of DIS, on his account, derives from its capacity to vitiate morally valid consent (premise B1), thereby violating sexual autonomy. By locating DIS's severity in such manner, Dougherty places consent at the centre of the moral evaluation of sexual conduct. This approach has, however, been increasingly criticised for legitimising abusive and exploitative, yet formally consensual, sexual practices.[10] As Robin West points out, it does not follow that because non-consensual sex is seriously wrong, consensual sex is not.[11] While Dougherty's theory remains technically compatible with cases in which consensual sex is morally problematic, this possibility is neither clearly acknowledged nor developed. Furthermore, by defining DIS as deceit concerning a dealbreaker, Dougherty approach seems to omit the harms and violations associated with deceit which *does not* concern a deal-breaker. To illustrate, consider the following example:

The Matrimonial Seduction

Jackson and Sebastian have been dating for several weeks. As their relationship develops, Jackson becomes increasingly emotionally invested and begins to develop feelings for Sebastian. Sensing this, Sebastian wants to initiate sex and does so by promising that he intends

[9] I should clarify that by Tom's *culinary autonomy*, I refer specifically to his food preferences guided by taste and proclivity alone. While his overall food choices may reflect both, his culinary preferences, and religious convictions, I take his culinary autonomy to be expressed in choosing what he enjoys or finds satisfying. I am grateful to the Editor for encouraging this clarification.

[10] Fischel 2019.

[11] West 1996.

to marry Jackson one day. In truth, Sebastian has no such intention, and he says it only to encourage Jackson to sleep with him. However, even if Jackson knew Sebastian was not serious about marriage, he would still have chosen to have sex with him.

In this example, Sebastian's deceit does not concern a deal-breaker, Jackson would still choose to sleep with him even if he knew the truth, yet it nonetheless appears morally problematic. For instance, had Jackson known the truth, he might have adjusted his emotional attitude and expectations and contextualised the sexual act accordingly. Since he is being deceived, however, he may interpret Sebastian's sexual advances as indicative of lasting feeling and a future marriage proposal. Depending on how invested he becomes in this expectation, Sebastian's deceit may produce serious psychological and emotional harm which, in turn, can render the deceit seriously wrong. Dougherty's account remains insensitive to such cases: since the deceit did not concern a deal-breaker, it falls outside his definition of seriously wrongful DIS. The sexual-nonsexual distinction, by contrast, removes the locus of severity from the mere fact that the deceit concerns sex into situates it within a broader contextual framework. As such, if Sebastian's deceit results in serious psychological or emotional implications for Jackson, then it may *still* constitute a *serious* moral transgression, even if it does not pertain to a deal-breaker. Conversely, even if Jackson would not choose to sleep with Sebastian had he known Sebastian's true intentions, Sebastian's DIS may still be only a minor wrong, as it pertained to Jackson's nonsexual reason for engaging in sex. The above example demonstrates the explanatory flexibility of the sexual-nonsexual distinction, and its capacity to better approximate the complex morality of deceptive sexual encounters.

3.1. Is The Sexual-Nonsexual Distinction Moralistic?

One could argue that my distinction mimics the differentiation between the core and peripheral features of sex, which Dougherty rejects as resting on sexual moralism. After all, what is the difference between a nonsexual feature of a sexual encounter and a morally irrelevant feature of such encounter? In defining DIS, he claims that: 'Since each person is an essential part of the sexual encounter, one is deceived about the sexual encounter by deception about the other person', effectively presupposing that all personal characteristics count as features of sex.[12] Accepting this presupposition would be incompatible with the claim that some personal characteristics count as nonsexual features of a sexual encounter and, consequently, undermine the LT.

My approach, however, does not rest on distinguishing between morally relevant and irrelevant features of a sexual encounter, which is the view Dougherty opposes. Rather, it distinguishes between *sexual* and *nonsexual* features of a sexual encounter which give rise to the *reasons* pertaining to the sex itself (sexual) and those concerning its context (nonsexual). There appears to be nothing moralistic about recognising that the sexual act is defined by a limited set of bodily and experiential features such as the numerical identity of participants, their actions, and the nature of physical contact. Other features, while potentially important to the broader meaning or context of the act, simply do not fall within its constitutive structure. Some features are simply not sexual, and they do not become so merely because someone treats them as decisive for sexual decision-making. Consider the

[12] Dougherty 2013, p. 719.

case of Jordan, who refuses to sleep with supporters of a rival football club. If, because of deceit, he has sex with someone who supports that club, it may be understandable for him to feel misled or betrayed, but it would be implausible to say that his strictly *sexual* autonomy was violated, or that the football club allegiance became a sexual feature of the act. This example highlights the point that the features which define a sexual act are not endlessly extendable – they concern the immediate, embodied structure of the act – who is involved, what is performed and how. Distinguishing between sexual and nonsexual features of the act is thus not moralistic, but a matter of conceptual discipline.

Personal features such as political inclination are not inherently a sexual feature since they do not appear to impact the immediate decision to have sex – they do not preclude the expression of sexual autonomy. Even if a subsequent discovery of deceit pertaining to a nonsexual personal feature would render the sex all-things-considered undesirable, this *immediate*, sexual decision remains unchallenged. In the context of the Devout Vegan example, even if Jenny's purported veganism was a major factor in Malcolm's sexual attraction to her, her deceit *did not nullify this attraction.* The same would be true even if Malcolm was attracted to Jenny *because* of her supposed veganism – for instance, if he found veganism sexually exciting. Although his attraction was based on a false belief, the sexual reason (motivation) it produced still expressed Malcolm's sexual autonomy, just as it would have if the belief had been true. There appears to be no relevant sexual difference between attraction based on fact and attraction based on a false belief. After all, people often feel sexual desire in response to impressions that may not correspond to reality. In fact, the phenomenon of sexual fantasy shows that desire can be shaped even by things we *know* are not real. The fact that someone's attraction is influenced by a false belief does not, by itself, mean that their sexual autonomy was undermined. While Jenny certainly wrongs Malcolm by remaining silent about her employment, the seriousness of her deceit cannot be attributed solely to the fact that it was used to induce Malcolm into sex. Just as in the Catholic Burger Eater example, this seriousness must be established on grounds independent from the sexual features. For example, Malcolm may suffer a serious psychological trauma associated with unintentionally abandoning his stringent vegan principles. However, such wrong is not a distinctly *sexual* wrong, so the potential severity of the moral wrongness associated with Jenny's DIS cannot be directly attributed to the sexual nature of this DIS. The matter of deceit may still be a valid reason to refrain from sex but if the DIS about it violates a nonsexual sphere of autonomy, e.g. religious or moral then its seriousness does not inherently follow from the fact that it pertains to sex. Thus, the sexual-nonsexual distinction avoids the moralism objection.

3.2. Is The Distinction Stable?

One could argue that the sexual-nonsexual distinction may falter in complex, real-life cases when it remains unclear whether a given feature of the encounter is part of the sexual act itself. This perceived instability might prevent the distinction from tracking our moral intuitions and leave it open to misuse by those seeking to minimise or obscure genuine harms. However, this worry goes too far. Conceptual complexity does not mean we cannot reliably distinguish between features which are internal to the sexual act itself, and those which belong to its broader context. Even in complicated cases, we can ask a clear question: was

the thing that someone was deceived about part of what happened during the sex itself – the partners, the actions, the feelings in the moment – or was it something external, such as a fact about the person's past or their beliefs? This approach does not ignore harm. Rather, it helps to identify what *kind* of harm occurred, and whether it was a distinctly sexual transgression of another kind of wrong. To illustrate, consider the following example:

Stolen Valour

Taylor is deeply attracted to military veterans. He finds their courage and discipline sexually exciting and says he would only ever sleep with someone who had served in combat. During a conversation at a party, Mark mentions that he did a few tours and was honourably discharged, allowing Taylor to assume that he is a veteran. In truth, Mark has never seen combat and was dishonourably discharged for refusing an order. Taylor and Mark have sex that evening. Later, Taylor discovers the truth and feels deceived, even violated.

While Taylor's feelings are valid and understandable, it does not follow that Mark's deceit concerned a sexual feature of the act. The physical structure, mutual awareness and the numerical identity of the partners remained unchanged. Mark's past, while significant to Taylor, is a contextual feature – not part of the act's internal structure. Even though Taylor's sexual attraction was based on Mark's purported past, its falsity did not nullify it. Taylor would be equally attracted to Mark had he actually served in combat or not. If, however, Mark pretended to be someone Taylor knew, an old acquaintance for example, then the deceit would classify as concerning a sexual feature of the encounter. This is because in such case, Mark would misrepresent his numerical identity – he would claim to be another distinct agent altogether. This would, in turn, make the sexual act entirely non-consensual since consent is given to *a particular agent*. It is worth stressing that under the sexual-nonsexual distinction, Mark's deception can still be seriously wrong. It is just not enough that it pertains to sex to demonstrate its severity. This example demonstrates that the sexual-nonsexual distinction can be coherently maintained even in morally complex scenarios, without collapsing into conceptual ambiguity or risking excusing sexual transgressions.

3.3. Is The Distinction Redundant?

A further objection could be that the tension between the LT and Dougherty's conclusion can be resolved without the sexual-nonsexual distinction, thus rendering the distinction redundant. One such strategy, proposed by Campbell Brown in his 2019 "Sex crimes and misdemeanours"[13] points out that the wrongness resulting from DIS may differ in degree.

Brown argues that since it is plausible to admit that deal-breakers vary in strength, then it is natural to assume that the moral wrongness resulting from deceiving someone about them is similarly variable. The strength of some principle as a deal-breaker for any individual can be, in turn, assessed by the degree of risk aversity exhibited by this individual's behaviour towards the possibility of trespassing this principle. Brown gives the example of a vegan deciding whether to eat a meal potentially containing animal products. The greater the risk of compromising their vegan ideal they are willing to take by eating the meal, the weaker the dealbreaker this vegan ideal is for them. According to Brown, some deal-breakers will be so weak that it will no longer be plausible to associate them with serious moral wrongness.

[13] Brown 2020.

Applying his logic to the Devout Vegan example from before, the fact that in any sexual courtship there is a reasonable chance of the other person not being completely honest about their life means that Malcolm the vegan is willing to risk having sex with a meat-eater. And the greater the risk he is willing to take, the weaker the deal-breaker in form of his partner's diet is. As such, Brown's analysis shows that there are cases where DIS is not seriously morally wrong and remains compatible with Dougherty's subjective view about the relevant features of a sexual act, rendering the sexual-nonsexual distinction redundant.

Brown's argument, however, and other arguments employing a similar framework of variability in wrongness or voluntariness[14][15][16][17], all appear to miss the mark. They may succeed in showing that some cases of DIS are *more seriously* morally wrong than others, but this does not amount to demonstrating that some instances of DIS are *not seriously* morally wrong. It may still be the case that *all* cases of DIS are seriously morally wrong, even if they pertain to very weak deal-breakers, just by virtue of violating one's sexual autonomy. While it seems admissible that sexual autonomy can be violated to a greater or lesser extent (as Brown argues), it may still follow that all violations of sexual autonomy are serious moral wrongs. Furthermore, this appears to be Dougherty's intuition evident in his argument for attributing the wrongness of sexual violation to lack of consent rather than actual harm resulting from the sex itself.[18] On Dougherty's account, even if the act had no way of directly harming the victim, e.g. in case of covertly sexually assaulting a comatose patient, the act would still be seriously morally wrong. In other words, its moral wrongness results from the fact that it is a distinctly *sexual offence* and, as such, it is seriously wrong by definition. Therefore, contrary to Brown's assumption that in some instances: "the degree of wrongness might be so small that we would not regard the deception as "seriously" wrong"[19], even the most minor cases of DIS would still be seriously morally wrong by virtue of being distinctly sexual offences, i.e. offences violating someone's sexual autonomy. Distinguishing between sexual and nonsexual reasons for having sex, on the other hand, is free from this difficulty and demonstrates how not all cases of DIS bear the hallmarks of a distinctly sexual offence. Therefore, the sexual-nonsexual distinction is not redundant.

3.4. Is The Distinction Harmful and Why Do We Need It?

Finally, one may argue that defending cases of deceiving someone into sex amounts to excusing sexual abuse and that accepting my argument would do nothing more than exonerate sexual offenders. I accept this criticism as a valid worry and one which must be addressed. However, it should be noted that my argument only defends DIS cases which we intuitively think are *not* cases of sexual assault. Moreover, I believe that resolving the tension between the common intuition captured by the LT and Dougherty's conclusion is necessary for creating a useful and coherent theoretical framework for protecting sexual autonomy. Otherwise, we run the risk of trivialising the critical matters of sexual consent and autonomy. Accepting Dougherty's conclusion would entail that someone who lied about

[14] ibid.
[15] Feinberg 1986.
[16] Archard 1999.
[17] Manson 2017.
[18] Dougherty 2013, p. 727.
[19] Brown 2020, p. 14.

their natural hair colour or having a small tattoo under their armpit ought to be labelled a sexual offender. And labelling DIS about such matters as sexual offences detracts from the perception of severity of sexual offence overall. Moreover, it creates an unhealthy reality of having to guess whether someone is a sexual offender in the strict sense, e.g. they committed rape or just the technical, unintuitive sense, e.g. they were dishonest about what they usually have for lunch. Therefore, I see the sexualnonsexual distinction as aiding the advancement towards an inclusive, coherent, and actionable framework about sexual autonomy and its protection.

4.Conclusion

In this essay, I have explored whether it is always seriously morally wrong to deceive a person into having sex, or whether the moral gravity of doing so depends on the matter of the deception. My analysis has highlighted the conflicting intuition about DIS: while some deceptions intuitively seem to constitute serious moral wrongs, others appear too trivial to be labelled as serious sexual offences. This tension raises important questions about the nature of personal autonomy and the criteria for serious moral wrongdoing in sexual contexts. Tom Dougherty's argument rejects this common intuition and claims that all cases of DIS are seriously morally wrong by virtue of inherently violating one's sexual autonomy. Dougherty rejects the distinction between core and peripheral features of a sexual encounter, arguing that all features can be dealbreakers for the consent giver and that drawing such distinctions rests on an objectionable moralism about sex. However, by introducing the distinction between sexual and nonsexual spheres of personal autonomy, I have challenged his conclusion. I argued that individuals can have various kinds of reasons – both sexual and nonsexual – for engaging in sexual activity, and that these reasons correspond to different spheres of autonomy. I have shown that deception about nonsexual features of a sexual encounter does not necessarily violate one's sexual autonomy and, therefore, may not constitute a serious moral wrongdoing in the same way that deception about sexual features would. The sexual-nonsexual distinction also avoids the moralism and redundancy objection. By acknowledging that not all features of a person or a situation are part of the sexual encounter itself, the distinction recognises the complexity of personal autonomy without undermining the seriousness of sexual offences or excusing immoral behaviour. Instead, it provides a framework which distinguishes between different kinds of wrongs, thus ensuring that serious sexual offences are not trivialised by being conflated with minor deceptions. Furthermore, it allows for a more nuanced assessment of the seriousness of cases of DIS and escapes the consent-centrism which renders Dougherty's approach insensitive to cases of morally problematic consensual sex. In conclusion, it is not always seriously morally wrong to deceive a person into having sex. Rather, the moral gravity of DIS depends on what the deception is about. Deceptions which violate one's sexual autonomy – such as those concerning the nature of the sexual act or the numerical identity of the sexual partner – are wrong by definition and constitute distinctly sexual offences. However, deceptions about nonsexual matters, while still problematic, potentially even seriously, do not have to carry the same moral weight. Recognising this distinction appears essential for developing a coherent and actionable framework for protecting sexual autonomy and ensuring that consent is taken seriously in all its dimensions.

Bibliography

Archard, D. (1999). "Sexual Consent". *Philosophical Quarterly*, 49 (197), pp. 556-557.

Brown, C. (2020). "Sex crimes and misdemeanours". *Philosophical Studies*, 177 (5), pp. 1363- 1379.

Dougherty, T. (2013). "Sex, Lies, and Consent". *Ethics*, 123 (4), pp. 717-744.

Feinberg, J. (1986). "Victims' excuses: The case of fraudulently procured consent". *Ethics*, 96 (2), pp. 330-345.

Fischel J. J. (2019). *Screw Consent: A better politics of sexual justice*. University of California Press.

Manson, N. C. (2017). "How Not to Think about the Ethics of Deceiving into Sex". *Ethics*, 127 (2), pp. 415-429.

West, R. (1996). "A comment on consent, sex, and rape". *Legal Theory*, 2(3), pp. 233–251.

Consciousness Doesn't Emerge—It Endures: A Defence of the Continuity Argument

Finlay Thwaite

King's College London

Abstract

This paper defends a continuity-based argument for panpsychism: the view that consciousness is a fundamental feature of reality, present either in basic physical entities themselves, in their constituents, or in the larger systems they compose—rather than something that emerges abruptly at complex biological scales. I argue that if phenomenal consciousness is truly irreducible—neither deducible from, nor identical with, structural or functional facts—then invoking its abrupt emergence at some threshold of complexity becomes metaphysically suspect. Drawing on the scientific preference for continuity over abrupt leaps, I show that rejecting panpsychism demands accepting one of three costly alternatives: (1) metaphysical vagueness, where the presence of consciousness is indeterminate; (2) radical emergence, where consciousness appears suddenly from non-conscious matter; or (3) an implausibly fine-grained threshold, where consciousness switches on at an exact point in physical complexity. The paper begins by outlining the motivations for panpsychism, particularly the explanatory gap left by physicalist accounts of consciousness. I then introduce the core conceptual framework—clarifying continuity, irreducibility, emergence, vagueness, and a minimalist conception of consciousness that avoids anthropomorphic assumptions. With these terms in place, I formulate the continuity argument in a structured form and critically examine the three principal strategies for resisting it. While not a proof of panpsychism, the argument significantly narrows the space of viable alternatives. What remains is not a speculative indulgence, but a default view arising from a principled process of elimination.

1. Introduction

Panpsychism maintains that consciousness is inextricably linked to all physical entities. More precisely, it claims that any given object must satisfy at least one of the following conditions: (i) it possesses a unified consciousness of its own; (ii) it is composed of constituents that are each conscious in their own right; or (iii) it participates as a component within a larger conscious system. Human and many non-human animals exemplify the first category, enjoying a single, integrated point of view. Most panpsychists extend this status to the fundamental building-blocks of matter—quarks, electrons, and the like[1]. Everyday inorganic

[1] Consciousness, as understood here, does not entail the capacities for thought, abstraction, self-awareness, etc. These are treated as complex manifestations of consciousness that arise only in sophisticated systems. Thus, panpsychism does not imply that particles engage in thinking or possess self-consciousness.

composites such as tables, chairs, or sand-heaps would,on this picture, qualify only in virtue of condition (ii): they inherit whatever experiential spark their microscopic parts possess. By contrast, the neurons and sub-structures inside an already-conscious brain satisfy condition (iii), contributing to but not independently duplicating the subject-level point of view.

Panpsychism, then, navigates a middle path between the difficulties besetting its two traditional rivals. Physicalist theories face an apparent *epistemic gap*, canonised by the knowledge argument[2] and conceivability thought experiments[3] . Alternatively, dualist models struggle with the problem of mental causation[4] . Panpsychism claims to evade both pitfalls, thereby offering a fresh framework for explaining how mind relates to matter without invoking brute emergence or causal disconnect. It avoids physicalist reductionism or eliminativism while placating dualist intuitions— insisting that the mind is not merely the brain, and that consciousness is irreducible.

While panpsychism circumvents brute emergence by treating consciousness as fundamental, it is not without its own explanatory burdens. Chief among these is the combination problem—how simple experiential units yield unified minds—and the question of why specific physical configurations correspond to particular qualitative feels. In this sense, panpsychism repositions rather than resolves the explanatory gap: it avoids the question of how consciousness arises from nonconscious matter, but leaves open why phenomenality has the structure it does. Still, its central appeal lies in what it avoids: the metaphysical leap from wholly non-conscious matter to conscious life.

The argument I will be considering in this paper is the so-called continuity argument, which leverages science's reluctance to accept abrupt ontological leaps. Call this the No-Radical-Emergence principle[5]

I contend that anyone rejecting panpsychism must accept at least one of three costly alternatives: (1) *metaphysical vagueness*, where the presence of consciousness is indeterminate; (2) *radical emergence*, where consciousness appears suddenly from non-conscious matter; or (3) an *implausibly fine-grained threshold*, where consciousness switches on at an exact point in physical complexity. The following sections examine each of these options. .

I first introduce key conceptual clarifications before outlining panpsychism and its core motivations. I also pre-empt the implausibility objection by clarifying how minimal experience might be conceived: not as human-like awareness, but as an extremely stripped-down phenomenal presence, lacking content, selfhood, or complexity. I then present the continuity argument and examine three possible ways to reject it: metaphysical vagueness, radical emergence, and the invocation of a precise threshold. Each is found wanting.

While not a proof of panpsychism, the argument significantly narrows the space of viable alternatives. The resulting view—minimal though it may be—emerges not from speculative preference but from a principled process of elimination.

2. Key Terms and Theoretical Considerations

[2] Jackson, 1982, 1986.
[3] Kripke, 1980; Chalmers, 1996.
[4] Kim, 1988.
[5] Robinson 2024, p. 152.

I will now begin by defining and commenting on a few terms. What follows is not a neutral conceptual map but a minimal set of framing assumptions adopted for the sake of the argument. Some readers may reject these premises, but my aim is not to defend them in full here—only to examine what follows if they are granted. I introduce terms like "consciousness" and "continuity" with specific meanings, mindful that they are contested. These definitions are not arbitrary: each is clarified where relevant and justified in context.

2.1. Consciousness

Now, by "consciousness", I refer to what many contemporary philosophers mean by "phenomenal consciousness": the *what-it's-likeness* that is concretely present whenever we see colours, think thoughts, feel pain or fear, sense amusement or anxiety, etc., as conceptualised by Nagel (1974). It is the subjective dimension of qualitative character, private in the straightforward sense that only the being undergoing it has direct acquaintance with how it feels from inside. Our own experiences—vision, pain, emotion—serve as prime examples, illustrating that we can imagine a vast range of possible phenomenal forms, from the very simple or primitive to the intricately complex. Here, *consciousness* is understood in the strictly phenomenal sense: the possession of subjective experience, a felt *what-it-is-likeness* that marks the difference between mere function and genuine awareness.

The co-referential use of "experience", "awareness", "subjectivity", "seeming", "mind", and "the mental"[6] in this paper rests on the assumption that what marks out *consciousness* is the presence of phenomenal character—that there is *something-it-is-like* for a system. While different traditions nuance these terms differently (e.g., "awareness" may connote attention; "subjectivity" may imply selfhood), I treat them here as pointing to a common metaphysical core: the occurrence of first-person phenomenality.

This core feature—*what-it-is-like-to-be-something*—is what the continuity argument concerns. Since the argument does not depend on finer distinctions among modes of awareness, degrees of access, or conceptual self-understanding, the terms are used interchangeably to refer to the mere presence of phenomenal feel[7]. This reflects a minimalist phenomenological realism: where experience exists at all, that "seeming" is the phenomenon under discussion. No further structure or richness is required.

To put it more rhetorically, no matter how confused one could be about the nature of reality or the veridicality of one's mental states, there is one thing one cannot be confused about: that something *seems* to be happening. That "seeming" is what I hold to be the fact of consciousness. I take it as a self-evident and undeniable fact that consciousness in this sense exists, as is warranted by my experience of writing this and your experience of reading this.

[6] I do not claim these terms are always equivalent across all philosophical contexts; however, for the purposes of this argument, what matters is the possession of phenomenal presence—that there is something it is like for a system to exist. All these terms are intended here to refer to this common core phenomenon: the existence of subjective, phenomenal feel. They are treated as co-referential unless otherwise specified.

[7] This approach echoes the strategy used by Thomas Nagel (1974), Galen Strawson (2008), and others who take phenomenal consciousness to be self-intimating and unitary at the minimal level—i.e., distinguishable in structure or content, but not in kind. From this perspective, terminological diversity reflects linguistic emphasis, not ontological multiplicity.

Thus, I adopt a thoroughly realist stance about consciousness: we know precisely what consciousness is by *having* it—at least in the case of reflexive humans. Non-human conscious beings, by contrast, need not have any explicit knowledge of their consciousness for it to be real nonetheless. This applies not just to individual experiences (we know exactly what red looks like or pain feels like simply by undergoing them) but also at the general level: once we have experience in any sense, we grasp what "consciousness" is broadly, because its presence is a form of direct acquaintance[8] . The phenomenon that *anything seems like anything* is the phenomenon of consciousness. It is in this sense by which I will employ the concept of "consciousness".

2.2. Vagueness

These reflections feed directly into the debate over vagueness in consciousness. One essential point is that consciousness occupies a distinct ontological status unlike that of non-conscious entities. An inanimate physical object (bracketing concerns about mereology)—like a rock or a cloud—simply *is* or *is not*, irrespective of whether anyone is aware of it. Experiences, by contrast, must be subjectively manifest; they depend on having a "for-whom" aspect in order to exist at all. This difference underpins the claim, developed soon, that consciousness cannot be vague, because experiences never occupy an indeterminate zone between "manifest" and "not manifest". Vagueness is associated with borderline cases. Different theories interpret borderline cases in different ways:

Linguistic: The borderline arises from our language (e.g. we know exactly which object "Rob" picks out, but we do not know if "is short" applies, because our semantics is imprecise).

Epistemic: Our incomplete knowledge of an otherwise sharp boundary (e.g., how many centimetres define "short") leaves statements about Rob's shortness unclear, though there may be a real fact of the matter unknown to us.

Metaphysical: The world itself may contain indeterminate states of affairs (e.g. a "cloud" with fuzzy boundaries).

All such theories, however, assume borderline instances—that the phenomenon in question can be neither clearly present nor clearly absent. The argument to come maintains that consciousness cannot fit the description of metaphysical vagueness. This is the sense in which I will use "vague/vagueness".

2.3. Emergence

There are broadly two kinds of emergence: *weak* and *strong*. *Weak* emergence is when a highlevel phenomenon arises from a low-level domain but is in principle *reducible* to truths and principles governing the lower-level domain. *Strong* emergence is when a high-level

[8] An analogy can be made to DNA. Each piece of DNA has the instructions for the entire organism. The partwhole relationship is unusual in this sense, as the part has information about the nature of the whole. Similarly, each quale tells you something about the nature of the whole—consciousness—in the sense that the mere fact of the *seeming* of a quale is the same *seeming* that permeates each possible quale.

phenomenon arises from a low-level domain but is in principle *irreducible* to truths and principles governing the lower-level domain.

Radical emergence is subtly distinct form of strong emergence. Radical emergence goes further than strong emergence by suggesting that some emergent phenomena are so fundamentally different from the lower-level domain that they cannot be derived or explained within that domain. A property radically emerges when it is ontologically novel and appears without supervenience or systematic bridge relations to the lower level, thereby introducing an entirely new causal category that requires additional fundamental principles. Consciousness, as I will argue, if it emerges, must emerge as a radically emergent phenomenon. By contrast, if consciousness does not emerge at a specific level of complexity but rather is basic—in some extremely minimal sense—then no additional laws need to strongly emerge consciousness from zero.

2.4. Continuity

By "continuity", I mean the principle that when a system's underlying parameters vary, the system's states change in a manner that can be traced through a chain of intermediate configurations, without invoking an abrupt introduction of qualitatively new kinds of entity or property in a single step. It does not rule out sharp phase transitions in the ordinary physical sense; it simply requires that any apparent "jump" be reducible (at least in principle) to finer-grained steps or underlying mechanisms rather than positing a *sui generis* category that appears *ex nihilo*. Such continuity dovetails with modern physics and evolutionary biology's emphasis on incremental changes rather than abrupt leaps. Nature, broadly speaking, tends to evolve without sudden metaphysical breaks, and this continuity suggests that consciousness too must arise—or endure—incrementally, rather than appearing *ex nihilo* at a sharp threshold of complexity.

For this to be problematic for consciousness uniquely, it needs to have the property of being *irreducible*. A property P is irreducible to the extent that (i) it cannot be identified with, deduced from, or explained away by any set of lower-level, non-P facts, and (ii) the explanatory gap between P and those lower-level facts is not merely the result of current ignorance but is in-principle uncrossable if one restricts oneself to the vocabulary and ontology of the lower level. Neuro-functional descriptions specify information-processing or causal roles but leave open why and how such roles *feel* like anything. Physics (fields, mass–energy, etc.) capture structure and dynamics but they are silent on qualitative, first-person appearance. I therefore assume the ontological claim that conscious experience is a *sui generis* feature of reality; it is not just a complicated arrangement of nonexperiential stuff.

Thus, if consciousness is both real and irreducible, introducing it only at advanced biological complexity demands a "switch-on" moment if it does not exist at lower and lower levels of organismic and physical complexity. These "switch-on" moments may *prima facie* seem acceptable, but this kind of radical emergence is viewed with suspicion.

3. Panpsychism and Its Motivations

The initial motivation for panpsychism is the persistent shortfall of mainstream physicalist theories in attempting to account for subjective phenomenal experience using purely structural properties of matter. Mainstream physicalist accounts aim to explain conscious-

ness in terms of cognitive architecture, information integration, or access mechanisms. Yet these models often address correlates of consciousness without accounting for phenomenal feel itself. Functionalism identifies mental states by their causal role, but this leaves open the question of why certain roles feel like anything at all. Global Workspace Theory (GWT)[9] and Higher-Order theories[10] offer accounts of consciousness as informational availability or meta-representational access, but such frameworks describe the *structure of awareness* without explaining its *existence*. Functionalism, GWT, and Higher-Order theories typically assume that consciousness is reducible to structural or functional features. However, if one accepts that phenomenal consciousness is irreducible, then these theories face a renewed difficulty: they must specify a threshold at which subjective experience first appears, and explain why that point—rather than any other—marks the ontological transition from nonexperience to experience.

Integrated Information Theory (IIT)[11] and Orchestrated Objective Reduction (Orch OR)[12] attempt more radical departures. IIT quantifies consciousness via integrated causal structure (Φ), while Orch OR posits quantum-level collapse events with experiential character. Both face challenges. IIT attempts to ground consciousness in integrated causal structure, assigning experience to any system with non-zero Φ. While this moves toward panpsychist conclusions that align with the continuity argument, IIT still faces a core challenge: it quantifies structure but does not explain why integration should entail phenomenality, rather than simply co-occur with it. The presence of Φ explains informational unity, but not subjective presence. Meanwhile, Orch OR appeals to unverified quantum-collapse mechanisms within microtubules[13], resting on physical assumptions—such as noncomputable influences on space-time geometry[14]—that remain outside the consensus of contemporary physics[15].

These views, while mature attempts to procure a physical explanation for consciousness, share a core difficulty: none explain how or why structural arrangements should entail subjective experience, rather than merely correlate with it.

This persistent shortfall invites a choice: either (i) consciousness is not real, or (ii) it is real but cannot be satisfactorily reduced to structural facts. For those who accept the reality of experience, this forces a re-evaluation of the assumption that consciousness must emerge from non-conscious matter.

Panpsychism enters at precisely this juncture. By rejecting the notion that consciousness must arise out of wholly inert matter, it aims to reposition the explanatory gap[16] between physical processes and subjective experience—often called the "hard problem" of consciousness[17]. Rather than confronting the sudden appearance of mind in an otherwise mute universe, panpsychism suggests that consciousness was never absent—that it is intrinsic

[9] Baars, 2005; Dehaene, Kerszberg and Changeux, 1998.
[10] Rosenthal and Weisberg, 2008.
[11] Albantakis et al., 2023.
[12] Hameroff and Penrose, 2014; Hameroff, 2020.
[13] Hameroff, 2012.
[14] Hameroff, 2012.
[15] Tegmark, 2000.
[16] Levine, 1983.
[17] Chalmers, 1995.

to matter at the most fundamental level[18].

The chief objective, then, is to bridge—or even dissolve—that explanatory gap. Rather than confront an apparent gulf between physical descriptions and the *what-it's-like* of conscious states, panpsychists argue that *if* matter is fundamentally non-conscious, then it is highly improbable for there to be a leap to the mental. Instead, as consciousness is always present—according to the panpsychist—in however rudimentary a form, it averts the initial explanatory gap from the outset.

3.1. Biological Continuity

Panpsychism's most compelling feature is its commitment to natural continuity. Science generally depicts nature as lacking sharp, fundamental breaks[19]. In biology, claims that require "ontological leaps" are viewed with suspicion unless compellingly justified[20] . Biology indeed records what one may describe as "leaps"—for instance, in the Cambrian explosion, or abiogenesis. However, these leaps typically reflect a rapid diversification building on pre-existing genetic or ecological substrates, rather than a wholly new metaphysical property. Hence, even these leaps seldom equate to emergent consciousness from inert matter.

This suspicion of abrupt metaphysical change is captured clearly in the following evolutionary continuity argument, which—though informal—helps illustrate why drawing a hard line between conscious and non-conscious systems may be conceptually misguided:

••

- **P1.** Humans, much of the time, are undoubtedly conscious.
- **P2.** Humans evolved from simpler life forms, which themselves evolved from single-celled organisms, and ultimately from non-living matter.
- **P3.** There is no *prima facie* conceptual limit to the simplicity of conscious experience.
- **P4.** If there is continuity in physical and biological evolution, it is difficult to define a precise point in history where consciousness could have suddenly emerged.
- **P5.** Any attempt to draw a line between conscious and non-conscious entities within the evolutionary chain would require justification for the exact point where continuity breaks down.
- **P6.** There is no identifiable point along the continuum of organismic complexity—the phylogenetic or ontogenetic chain—where one could justify a definitively drawn line between conscious organisms and non-conscious organisms.
- **C.** Therefore, consciousness is likely continuous in nature, extending to simpler forms of life and non-living matter.

[18] Chalmers, 1995; Nagel, 1979; Strawson, 2008.
[19] Hendry, 2010; Leibniz 1687/1989a, as cited in Leibniz 1702/1989b, p.297.
[20] Gould, 1980; Walker Davies, 2013.

This continuity is not assumed dogmatically. Absent a robust mechanism and compelling evidence for any ontological leap, it seems more parsimonious to posit incremental transitions than to postulate sudden ontologically novel appearances. Of course, parsimony alone does not conclusively exclude such abrupt changes if strong future proof arises—indeed, should incontrovertible data reveal a genuine discontinuity for consciousness, it would override key aspects of the continuity argument. While I rely on this preference for incremental continuity, I concede that science could discover an exceptional case. Any exceptional case that demonstrates strongly or radically emergent phenomena would certainly influence our intuitions in important ways. Yet, absent direct evidence for a purely non-conscious-to-conscious threshold, the continuity stance remains more plausible. If science or philosophy were to confirm a true abrupt event for consciousness, it would be an extraordinary result requiring robust justification. However, such data have not surfaced thus far.

Critics might object that certain traits (e.g., flight or photosynthesis) do indeed arise midstream in evolutionary history[21]. However, panpsychists emphasise that consciousness is not an ordinary trait but a phenomenal property—one that defies reductive explanation in purely functional or behavioural terms[22], because these roles admit of borderline cases. Flight and photosynthesis can be captured biologically, but the "feel" of experience is more elusive. If there is no clear boundary marking "experiencing matter" versus "non-experiencing matter," then continuity arguments imply that consciousness did not abruptly emerge from nothing; rather, at least some rudimentary form of it was present all along.

From a continuity perspective, we treat abrupt leaps as an *extraordinary claim* requiring extraordinary evidence. Hence, the continuity stance is more parsimonious until data mandate otherwise. This is one reason I propose the second and third prongs of my trilemma: it pinpoints that radically emergent consciousness (i.e. an abrupt leap) cannot be presumed without strong justification. For instance, if a single morphological trait truly appeared from nowhere, that might be an instance of genuine radical emergence. But typically, science *does* uncover incremental precursors. Hence, my trilemma highlights how, empirically, these leaps are not well-founded.

If we restrict consciousness strictly to advanced central nervous systems and posit that consciousness is wholly absent in "lower" stages of ontogeny, phylogeny, and taxonomy, then we face the question of where and when (not to mention *why*) does mind spontaneously appear? Given that consciousness seems an all-or-nothing phenomenon (I will expand on this point later), this is a hard question to answer. Panpsychism's resolution is to dismiss the question and claim that a faint spark of subjectivity might run all the way down the evolutionary ladder, so no single generation or morphological stage marks the abrupt switch from zero to nonzero conscious awareness.

Panpsychism thus taps into a dissatisfaction with "magic steps" in nature. This is the crux of what motivates the continuity argument. I develop this more formally in Section 5.

4. The Immediate Objection

Before continuing, we ought to address something important. At this point, many find

[21] Blankenship, 2010; Carroll, 2001, pp. 1104; 1107; Hunter, 2007
[22] Strawson, 2008, p.20.

the theory hard to take seriously. The most immediate objection to panpsychism is that it appears to lead to bizarre conclusions, namely the possibility that electrons, quarks, or fields might be or contain consciousness. This is counterintuitive for many people, in part because it contradicts a widespread common-sense or hierarchical outlook on which entities can plausibly be conscious. Roelofs and Buchanan (2018) label this the *Great Chain of Being* (GCOB) intuition, a conceptual ladder:

- **Level 1:** Humans (definitively conscious).
- **Level 2:** Higher animals (likely conscious).
- **Level 3:** Lower animals/insects (questionably conscious).
- **Level 4:** Plants/other non-sentient life (assumed non-conscious).
- **Level 5:** Inanimate objects/fundamental particles (taken to be non-conscious).

According to this scheme, consciousness is strongly correlated with complexity and with observable capacities for sophisticated behaviour. As such, it is inconceivable on the GCOB view that something as small and simple as a quark might have experience. Consequently, panpsychists are accused of flouting deeply entrenched intuitions about where consciousness resides. We can summarise this implausibility objection as follows:

P0.P0.

1. Panpsychism maintains that fundamental entities, such as quarks, have consciousness (even if extremely rudimentary).
2. This claim offends an intuition-laden hierarchy of consciousness, which places fundamental particles at the bottom, lacking even the possibility of experience.

C. Therefore, panpsychism is wildly implausible, defying commonsense views about who or what has consciousness.

4.1. Responding To The Implausibility Objection

There are various replies available to the panpsychist, but it is crucial to clarify intuitions about consciousness before addressing broader arguments. It can be argued that our negative intuitions about consciousness in unfamiliar or non-human contexts can be seriously misleading. In hat regard, I emphasise that our intuitions about consciousness are not built for detecting it in an absolute sense, but rather to recognise a specific type of consciousness—one with familiar characteristics, like human or certain animal experience. These positive intuitions (e.g., the sense that humans and certain animals are conscious) tend to be reliable because they align with consciousness as we recognise it; however, negative intuitions about non-consciousness (e.g., vegetative humans, lower animals, and even plants) are more questionable[23]. We struggle to understand alien experiences because they differ fundamentally from our own, limiting our ability to recognise or comprehend them.

[23] Owen, 2013; Trewavas, 2021.

Although we are directly acquainted only with our own distinctive modes of consciousness, there might well be experiential phenomena we cannot conceive from our vantage—perhaps the most rudimentary versions of it, or forms that predate our own biological lineage. We are epistemically limited by our biology: each subject samples only its own style of experience and can at best infer alien phenomenologies.

As such, we may fail to detect forms of consciousness that are unfamiliar or outside our perceptual range. This is an asymmetry: positive intuitions of consciousness are generally reliable, whereas negative intuitions of non-consciousness are unreliable. Thus, we lack a universal mechanism to detect consciousness simpliciter. Consequently, both non-conscious entities (if there are such things) and alien conscious beings can escape our intuitive grasp, making negative intuitions about consciousness unreliable indicators of its absence. The GCOB intuition, while a useful guide, now appears to reflect how closely a being's consciousness mirrors our own rather than indicating consciousness simpliciter.

In other words, what we might label as the "implausibility" of assigning consciousness to a quark is shaped by the intuitive assumption that consciousness must resemble our own in complexity, behavioural signature, or biological structure. If, however, consciousness could take an unfamiliar form—an ultra-simple felt presence with no sense data or sense of self—then the standard "that's impossible" reaction is not so obviously correct.

To flesh out what such an "ultra-simple felt presence" might entail, I briefly propose the following thought experiment, which traces the conceptual path from rich experience to its most stripped-down possible form.

4.2. Conceiving Minimal Consciousness

It first bears mentioning that truth is not a democracy and that the prevalence of certain intuitions does not inherently confirm their truth. Since intuitions are appearances within consciousness, speaking personally should not be dismissed as improper, and I will do so in this section. When an idea is genuinely conceivable, internally consistent, and not rooted in deception or misunderstanding, it deserves at least provisional consideration as a candidate for truth.

As such, one way to support the idea that even a fundamental particle could have consciousness is to earnestly consider the possibility of extremely minimal or contentless experience. To illustrate this, I propose a thought experiment that begins with losing each of the standard senses— hearing, sight, smell, taste, and touch. Even after removing those channels, there would remain an internal awareness of thoughts or emotions. Now imagine going further still, entering the "stillness" states reported by contemplatives or psychonauts, where one's sense of self, time, and space largely or completely disappear[24]. At no point in this process does consciousness itself obviously blink out; instead, it seems that something like a bare "space of awareness" lingers, even when most or all external and internal stimuli have been stripped away. I am inclined to think that, even in such an extreme state, one would still be conscious—just conscious of "almost nothing", or perhaps actually nothing. There would remain a bare space of awareness, though devoid of specific objects or sensations.

[24] Griffiths et al., 2008, p.19; Milliere, 2020.

By showing that we can at least conceive (and perhaps briefly attain) a largely or entirely content-free mode of consciousness, the notion that consciousness necessarily requires elaborate, human-like mental operations is challenged. If a faint "core sense" of being could persist in us despite drastic sensory or cognitive reductions, then perhaps a similar spark of awareness can exist—albeit in a much simpler form—in a fundamental particle. This image approximates how I conceive the phenomenality that might reside in something as simple as a quark. In that sense, imagining an "empty" or "contentless" consciousness, however minuscule, makes the notion of minimal consciousness marginally more natural or at least less outright dismissible. The idea may feel counterintuitive, but it loosens the implicit assumption that consciousness must always involve rich phenomenology. It importantly sidesteps the objection that attributing consciousness to fundamental particles is self-evidently absurd: if we grant that consciousness can be stripped down to little more than sheer presence and that there is no *prima facie* conceptual limit to phenomenal simplicity, there is less reason to dismiss outright the hypothesis that matter at the smallest scales might harbour a kernel of experience.

Naturally, some may find my analogy implausible, while others might see it as perfectly coherent. Either way, whether one labels such a scenario absurd or feasible has little bearing on its metaphysical merits—because, in the end, we each treat our fundamental intuitions differently.

5. Formulating the Continuity Argument

Now I have introduced defined my key terms and attempted to dismiss the *prima facie* charge of implausibility, let me restate the continuity argument formally:

- **(P1)** Consciousness is irreducible and real (the "explanatory gap" stands).
- **(P2)** Radical, strong emergence is scientifically repugnant; modern physics and biology prefer incremental developments of pre-existing properties,
- **(P2.1)** and this should guide our intuitions when thinking about the natural world.
- **(P3)** Consciousness, by its nature, cannot be partially present in an ontological sense—there is always a minimal subjectivity or none. It cannot be vague.
- **(P4)** If consciousness is wholly absent in simpler matter, we face a metaphysical leap at an arbitrary or implausibly fine-grained threshold, which reintroduces radical emergence.
- **C:** Hence, consciousness must exist at all levels, but typically in extremely minimal or content-poor forms.

Thus, if we question the "sharpness" of consciousness, we open the door to partial or fuzzy states. If we disclaim the norm against radical leaps, we can accept strong emergence. If we allow a pinpoint threshold, we adopt a *prima facie* implausibly fine-grained "switch-on". The next sections assess these three potential escapes.

6. The Trilemma of Rejecting Continuity

6.1 Vagueness

One possible response to the continuity argument is to deny that consciousness is a binary property. On this view, consciousness might gradually emerge—starting off as "faint" or "partial"—allowing for a smooth ramp from non-experience to experience without requiring a hard threshold. If consciousness can be ontologically vague, then there may be no precise moment at which it emerges, and no need to posit an abrupt ontological leap.

But this manoeuvre faces a fundamental challenge: consciousness, unlike many other properties, appears to resist vagueness in this metaphysical sense. To be conscious is to have a subjective point of view—to feel like something from the inside. That subjective presence either exists or it doesn't; to have any subjective vantage at all is already to have full, if minimal, experience. There are no "borderline cases" of having a vantage point. The idea that one might be "half-aware" in the same way that one might be "sort of tall" misunderstands the nature of phenomenal presence. While the contents of experience can be vague, dim, or confusing, the existence of experience itself is not a matter of degree. The presence/absence of phenomenal experience is introspectively bivalent, leaving no room for borderline instantiation.

If by "consciousness" you mean "awake", then it seems consciousness can be *linguistically* vague, as it depends on what you mean by "awake". If by "awake" you mean "some such specific conditions involving norepinephrine, the thalamus, . . . ", then it seems like consciousness may be *epistemically* vague, as it's hard to confirm when, exactly, someone is awake. If, instead, by "consciousness" you mean "something-that-it-is-like—some seeming experience—in the simplest possible sense", then it cannot be *metaphysically* vague. It is a yes or a no. True or false. Manifest or not manifest.

Whichever account of vagueness one prefers, each presupposes the possibility of an indeterminate instance—something that is *sort-of* awake, *sort-of* short, etc. But phenomenal consciousness is self-intimating: the very having of it settles the fact of its presence. An experience is either present or absent; there is no ontic intermediate state with an indeterminate "for-whom-ness", whether it is metaphysically unclear about *whether* anything is happening. As such, the phenomenon of consciousness does not supply the kind of borderline cases that the standard accounts need in order to get vagueness off the ground.

Consequently, this sharpness of consciousness is central to the panpsychist's objection. If consciousness is present at all—however faintly—it is fully present as a mode of being. Illusions, near-sleep states, or cases of low awareness may confuse *what* is experienced, but they do not blur the *fact* that something is experienced. We may doubt whether we saw or felt something, but we cannot inhabit a state that is metaphysically halfway between consciousness and non-consciousness. The vantage point does not flicker into half-being; it either is or is not.

Some emergentist views invoke "dim consciousness" as a transitional state. But even this admits a sharp threshold between *no* experience and *some*—between absolute absence and minimal presence; ϕ_0 is ontologically different from ϕ. Any instance of consciousness, however simple or weak, remains metaphysically distinct from its total absence. If any phenomenally present feeling is present, it is *fully* present, however low in intensity. And

once one accepts that experience—however minimal—must be minimally *present*, the case for continuity resurfaces intact: one can vary the complexity, richness, or integration of conscious states, but not the binary fact of their presence.

Certainly, people can have unclear or incomplete *knowledge about* whether they perceived something, as in near-sleep confusion or illusions. This uncertainty, however, reflects our awareness *of* experience, not whether the experience itself was metaphysically borderline. If the vantage was ever there, it was simply "there", if not, it was absent. We do not get "*half-real*" phenomenology.

In short, phenomenal consciousness admits degrees of *intensity*, but not degrees of *existence*. There may be gradients *within* experience, but no gradient *into* it. Therefore, accepting the metaphysical vagueness of consciousness, as I have defined it, is incoherent.

6.2 Radical Emergence

One way to salvage consciousness as being sharp but not fundamental is to accept that consciousness arises out of nowhere once matter hits a certain complexity. Perhaps there is a specific point—whether neural density, information integration, or structural intricacy—at which a previously non-conscious system suddenly "switches on". Proponents of this stance may respond to the continuity argument by saying, in effect, "Yes, consciousness is simply a novel phenomenon—this is how it works, and it needs no deeper explanation".

However, observe that no other domain in physics or biology actually demonstrates a lawless, spontaneous appearance of wholly new properties. Where so-called emergent phenomena appear (in chemistry, fluid dynamics, superconductivity, etc.), they typically remain weakly emergent, traceable back to simpler interactions. By contrast, positing radical emergence in consciousness contradicts the usual incremental approach of physical science.

Moreover, if we assume consciousness is an on/off phenomenon (see 6.1), then jumping from zero to nonzero experience is an especially stark shift. It also leaves the explanatory gap intact, simply calling it "brute" or "unbridgeable"—which is philosophically unsatisfying. Thus, while radical emergence might avoid the panpsychist notion that consciousness is fundamental, it often appears ad hoc and out of step with standard scientific reasoning. Cynically, it can appear as a last-resort posture, rather than a truly coherent framework.

Radical emergence (**RE**), in more formal terms, implies that thoroughly non-conscious stuff (**NC**) gives rise to something categorically different: conscious stuff (**C**). This is "radical" because it posits that subjective experience, with its first-person "what-it's-like-ness", emerges from matter that is entirely void of such a quality. Concretely:

- **C:** Conscious stuff.
- **NC:** Non-conscious stuff.
- **RE:** Radical emergence.

I argue that "NC produces C through RE" faces a genuine puzzle:

1. We know C definitely exists—at least at the biological level of adult humans.

2. We are not certain that NC exists (there cannot be direct observation of "stuff" wholly devoid of experience).

3. Even if one presupposes NC's existence, there must be some mechanism for it to spawn C from itself.

4. If RE is unexplained, it undermines the initial positing of NC.

An immediate issue for **NC + RE** is that it appears empirically indistinguishable from a "universal C" (panpsychist) approach. Physics itself is neutral about whether the underlying nature of matter is experiential or not: it only describes structures, relations, and mathematically modelled laws. Terms denoting the "inner" or "categorical" reality can be drastically reinterpreted without clashing with the data. Because both "all-conscious" and "non-conscious with a sudden C-leap" remain consistent with observed physics, it is unclear what justifies choosing **NC** theories at all.

This undercuts a key presumption often used to favour **NC + RE**: that panpsychism is somehow "less scientific" or "nonphysical". In fact, physics places no ontological constraint on whether its formal structures are instantiated in non-experiential or proto-experiential terms. That is, the empirical success of physics does not adjudicate between a universe that is qualitatively mute at the fundamental level and one whose base constituents possess minimal subjectivity. Both are equally compatible with the equations. Consequently, rejecting panpsychism on the grounds that it contradicts "the physical" presumes a vastly thin "physical" mean.

Despite feeling safe, this security turns out to be illusory. Definitions of the physical range from current physical theory (which is incomplete), to whatever future physics posits (which renders the term trivially inclusive and may well include experiential properties), to anything that is causal or spatiotemporal (which fails to rule out experiential properties if they are causal, and rules out theories of quantum gravity, which would clearly be physical theories[25]). In this sense, **NC + RE** does not gain theoretical credibility by appealing to "the physical"—it merely assumes that radical emergence is the only naturalistic way to explain consciousness.

But if panpsychism and **NC + RE** are both empirically compatible with physics, and if panpsychism offers a smoother ontological continuity without appealing to unexplained leaps, then it is unclear why **NC + RE** should be preferred. The burden of proof thus shifts: those who reject panpsychism need to explain why consciousness must be absent at simpler levels despite the lack of any empirical warrant.

Defenders of radical emergence might then argue that, even if physics cannot decide what the "physical" is, their theory is theoretically better. **NC** theory advocates often retort that panpsychism faces its own stumbling block: the *combination problem*. How could trillions of tiny micro-experiences combine to produce one unified subject, such as a single human mind? This problem truly does pose a major challenge to most panpsychist and cosmopsychist theories, and no final solution to it yet exists. My aim in this chapter, however, is not

[25] See Belfaqhi et al., 2025; Lee, 2023.

to weigh competing metaphysical costs, but to show that rejecting panpsychism in favour of radical emergence is itself empirically and theoretically problematic—regardless of who one accounts for the mental features required in any proposed RE. If such a combination is conceptually impossible, some argue that it is simpler for consciousness to radically emerge from an **NC** base, rather than result from the fusion of countless microexperiences.

Nonetheless, it remains deeply mysterious how a purely **NC** world flips "on" into **C** at any threshold—radical leaps do not clarify matters. The genesis of the problem is that singling out consciousness for a unique "unbridgeable gap" is suspiciously *ad hoc* when every other phenomenon finds a place in nature's continuous evolution. This is why I take **NC + RE** to violate basic naturalist methodology.

Considering the above, we now have some options. If we accept experiential realism—that consciousness is a real, irreducible datum—then we face a bifurcation: either (1) consciousness radically emerges from non-conscious matter, violating this methodological norm, or (2) consciousness is fundamental, built into the basic architecture of reality. Given the absence of any mechanistic or explanatory account of how radical emergence might work in the case of consciousness, and given that it contravenes the prevailing logic of scientific continuity, the second option appears more coherent.

Indeed, **NC + RE** accounts have yet to show how structural or functional properties alone could give rise to consciousness. This continuing explanatory shortfall reinforces the plausibility of panpsychist or proto-experiential views—not because they solve all problems, but because they preserve ontological continuity without positing unexplained metaphysical leaps.

In this light, radical emergence appears less like a theoretical insight and more like a placeholder for ignorance. It functions not as a model, but as a stipulation—a way of saying, "consciousness just happens," whilst waiting for science to provide a well-articulated mechanism for the only known instance of an assumed radically emergent property.

Perhaps this is fine. I disagree.

6.3 Fine-Grained Threshold

Now, if one accepts that consciousness is indeed sharp, and yet maintains still that it strongly emerges at a certain level of complexity, then this threshold requires an extremely well-articulated explanation. While this "fine-grained threshold" problem is being presented as distinct, it is in fact a special case of radical emergence. It assumes that consciousness is wholly absent up to a specific structural or functional point, at which it emerges fully and discontinuously. As such, it inherits all the metaphysical burden of radical emergence—namely, a novel ontological category appearing *ex nihilo* from a non-conscious base.

However, I treat fine-grainedness as a separate horn of the trilemma for two reasons. First, it is often presented as a more palatable or biologically plausible form of radical emergence, appealing to complexity metrics or computational thresholds. Second, it faces a unique challenge of arbitrariness: the difficulty of justifying why any specific degree of complexity should mark the boundary between non-experience and experience. Thus, while dependent on the logic of radical emergence, the fine-grained view introduces its own explanatory burden

and warrants independent scrutiny.

In principle, one might posit a specific measure N of neural or functional complexity, so that N^1 yields no consciousness, whereas N^2 suddenly yields some. Although this may appear contrived, it could be compared to phase transitions in physics, which occur at exact points, for instance when a material changes from liquid to solid.

Some have argued that these phase transitions—the shift of a substance from one physical state to another—could be examples of "truly [ontologically] emergent properties"[26] and thus undermine ontological reductionism[27]. The central issue lies in the fact that statistical mechanics, when applied to actual, finite systems, does not fully account for phenomena such as boiling. Instead, it depends on the idealisation of an infinite system, which "seems to play an ineliminable explanatory role."[28] In other words, a pot of water should not boil unless it contains infinitely many particles—yet no real system is infinite. This discrepancy potentially indicates that genuinely emergent properties may exist, thus challenging the argument against radical emergence.

However, typical phase changes in physics do not introduce brand-new fundamental properties: they give rise to new macro-behaviours that remain derivable from underlying laws, even if presently unknown. By contrast, threshold-based consciousness would be something irreducible that is not logically implied by micro-physics—a form of radical emergence masquerading as an "exact boundary".

Hence, if consciousness is truly all-or-nothing, then a "switch-on" at a morphological threshold seems deeply puzzling. Can adding one neuron or one small cluster of cells flip a system from absolute non-experience to some-experience? Is it credible that *Caenorhabditis elegans*, with its precisely 302 neurons and extensive nervous system[29], might lack all experience, while a slightly more complex organism like the *Drosophila melanogaster* larva with 3,016 neurons[30] possesses it? Or must we wait until the adult fruit fly's nervous system reaches over 140,000 neurons[31] before subjective experience is permitted? And if even the adult fruit fly is not enough, then what is—an ant? A mouse? A human infant? Are we to believe that only once the adult human brain reaches somewhere between 61 and 99 billion neurons[32], consciousness at last emerges? Is 60 billion not enough? These implications stretch credibility. It is precisely this sort of abrupt ontological leap that the continuity argument challenges: no single micro-step should give rise to a radically new kind of property like consciousness.

If, on the other hand, we maintain that consciousness "fades in", there must still be a point—from zero to non-zero—where it crosses some boundary. That boundary again becomes a fine-grained threshold, undermining the idea of a smooth ramp. Alternatively, if one says the system never truly crosses from zero, implying consciousness is always non-zero,

[26] Liu, 1999, p. 92.
[27] Bangu, 2015, p. 323; Lebowitz, 1999, p. S346; Prigogine, 1997, p. 45.
[28] Bangu, 2015, p. 323.
[29] Soares et al., 2017.
[30] Winding et al., 2023.
[31] Shiu et al., 2024.
[32] Goriely, 2025.

one circles back to the panpsychist position.

If the continuity argument aims to exclude abrupt leaps altogether, then positing an ultra-fine threshold still amounts to a radical leap, merely relabelled as a problem for "future science" rather than current metaphysics. The problem of radical emergence persists, and the continuity objection—that it is both ad hoc and scientifically suspect—remains intact. In effect, positing a fine-grained threshold does not escape the continuity critique.

Conclusion

The continuity argument does not prove panpsychism. But it does place heavy pressure on any attempt to deny it. If consciousness is neither vague, nor able to radically emerge from wholly non-conscious matter, nor plausibly tied to an exact structural threshold, then we are left with little option but to regard it as fundamental—however minimal its early expression may be.

This does not mean we must embrace a full-blown panpsychist ontology. It means only that the alternatives, when examined closely, carry metaphysical costs that many find unacceptable. None of these sit easily with the methodological norms that structure modern science. If we uphold those norms, continuity forces our hand.

Rejecting these three "escape hatches"—vague consciousness, radical emergence, or micro-threshold solutions—deflates the argument for threshold-based emergence in consciousness. Consequently:

(a) We deny that consciousness can just appear fully from zero,

(b) we deny that it exists half-formed or indefinite, and

(c) we deny the plausibility of an ultra-fine but genuine switch-on boundary.

This leaves us with a simpler, if stranger, proposal: that consciousness was never absent in the first place.

To say that consciousness does not emerge is not to say that it is everywhere in any familiar or human-like sense. It is to say that some form of experience—however dim, minimal, empty, or unfamiliar—may be stitched into the very structure of the world. In the absence of decisive reasons to believe in sudden ontological leaps, the simplest assumption may be that what endures did not come from nowhere.

If panpsychism still feels counterintuitive, that's no longer enough to dismiss it. Intuition is not argument, and discomfort is not disproof. What matters is coherence. And coherence lies not in emergence, but in endurance.

Bibliography

Albantakis, L., Barbosa, L. R., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Zaeemzadeh, A.,

Boly, M., Juel, B. E., Sasai, S., Fujii, K., Imhonopi, D., Hendren, J., Lang, J. P. and Tononi, G. (2023). "Integrated information theory (IIT 4.0): formulating the properties of phenomenal existence in physical terms". *PLOS Computational Biology*, 19 (10), pp. e1011465.

Baars, B. J. (2005). "Global workspace theory of consciousness: toward a cognitive neuroscience of human experience". *Progress in Brain Research*, 150, pp. 45–53.

Bangu, S. (2015). "Why does water boil? Fictions in scientific explanation". In: Dieks, D. et al. (eds.). *Recent Developments in the Philosophy of Science*: EPSA13 Helsinki. Cham: Springer, pp. 322–323.

Belfaqih, I. H., Bojowald, M., Brahma, S. and Duque, E. I. (2025). "Lessons for loop quantum gravity from emergent modified gravity". *Physical Review D*, 111 (8), pp. 086027.

Blankenship, R. E. (2010). "Early evolution of photosynthesis". *Plant Physiology*, 154 (2), pp. 434– 438.

Carroll, S. B. (2001). "Chance and necessity: the evolution of morphological complexity and diversity". *Nature*, 409 (6823), pp. 1102–1109.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chalmers, D. J. (1995). "The puzzle of conscious experience". *Scientific American*, 273 (6), pp. 80–86.

Dehaene, S., Kerszberg, M. and Changeux, J.-P. (1998). "A neuronal model of a global workspace in effortful cognitive tasks". *Proceedings of the National Academy of Sciences*, 95 (24), pp. 14529–14534.

Goriely, A. (2025). "Eighty-six billion and counting: do we know the number of neurons in the human brain?" *Brain*, 148 (3), pp. 689–691.

Gould, S. J. (1980). "Is a new and general theory of evolution emerging?" *Paleobiology*, 6 (1), pp. 119–130.

Griffiths, R., Richards, W., Johnson, M., McCann, U. and Jesse, R. (2008). "Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later". *Journal of Psychopharmacology*, 22 (6), pp. 621–632.

Hameroff, S. (1998). "Quantum computation in brain microtubules? The Penrose-Hameroff 'Orch OR' model of consciousness". *Philosophical Transactions of the Royal Society A*, 356, pp. 1869–1896.

Hameroff, S. (2012). "How quantum brain biology can rescue conscious free will". *Frontiers in Integrative Neuroscience*, 6, pp. 1–17.

Hameroff, S. and Penrose, R. (2014). "Consciousness in the universe". *Physics of Life Reviews*, 11 (1), pp. 39–78.

Hendry, R. F. (2010). "Ontological reduction and molecular structure". *Studies in History and Philosophy of Modern Physics*, 41 (2), pp. 183–191.

Hunter, P. (2007). "The nature of flight". *EMBO* Reports, 8 (9), pp. 811–813.

Jackson, F. (1982). "Epiphenomenal qualia". *The Philosophical Quarterly*, 32 (127), pp. 127–136.

Jackson, F. (1986). "What Mary didn't know". *The Journal of Philosophy*, 83 (5), pp. 291–295.

Lebowitz, J. L. (1999). "Statistical mechanics: a selective review of two central issues". *Reviews of Modern Physics*, 71, pp. S346–S347.

Leibniz, G. W. (1989a). "Letter on a general principle useful in explaining the laws of nature through a consideration of divine wisdom". In: Ariew, R. and Garber, D. (eds.). *Leibniz: Philosophical Essays.*Indianapolis: Hackett, pp. 351–354. (Original work published 1687).

Leibniz, G. W. (1989b). *Philosophical Essays*. Indianapolis: Hackett Publishing Company. (Original work published 1702).

Lee, S.-S. (2023). "Massless graviton in a model of quantum gravity with emergent spacetime". *Physical Review D*, 108 (2), pp. 024054.

Liu, C. (1999). "Explaining the emergence of cooperative phenomena". *Philosophy of Science*, 66, pp. S92–S106.

Milliere, R. (2020). "The varieties of selflessness". *Philosophy and the Mind Sciences*, 1 (1), pp. 1–41.

Nagel, T. (1974). "What is it like to be a bat?" *The Philosophical Review*, 83 (4), pp. 435–450.

Nagel, T. (1979). *Mortal Questions*. Cambridge: Cambridge University Press

Owen, A. M. (2013). "Detecting consciousness: a unique role for neuroimaging". *Annual Review of Psychology*, 64 (1), pp. 109–133.

Prigogine, I. (1997). *The End of Certainty*. New York: The Free Press.

Roelofs, L. and Buchanan, J. (2018). "Panpsychism, intuitions, and the great chain of being". *Philosophical Studies*, 176 (11), pp. 2991–3017.

Rosenthal, D. and Weisberg, J. (2008). "Higher-order theories of consciousness". *Scholarpedia*, 3 (5), p. 4407.

Shiu, P. K., Sterne, G. R., Spiller, N., Franconville, R., Sandoval, A., Zhou, J., Simha, N., Kang, C. H., Yu, S., Kim, J. S., Dorkenwald, S., et al. (2024). "A Drosophila computational brain model reveals sensorimotor processing". *Nature*, 634 (8032), pp. 210–219.

Soares, F. A., Fagundez, D. A. and Avila, D. S. (2017). "Neurodegeneration induced by metals in Caenorhabditis elegans". In: Aschner, M. and Costa, L. (eds.). *Neurotoxicity of Metals (Advances in Neurobiology, vol. 18)*. Cham: Springer, pp. 331–352.

Strawson, G. (2008a). "Realistic monism: why physicalism entails panpsychism". In: Strawson, G. (ed.). *Real Materialism*. Oxford: Oxford University Press, pp. 53–74.

Strawson, G. (2008b). *Real Materialism and Other Essays*. Oxford: Clarendon Press.

Tegmark, M. (2000). "Importance of quantum decoherence in brain processes". *Physical Review E*, 61 (4), pp. 4194–4206.

Trewavas, A. (2021). "Awareness and integrated information theory identify plant meristems as sites of conscious activity". *Protoplasma*, 258 (3), pp. 673–679.

Walker, S. I. and Davies, P. C. W. (2013). "The algorithmic origins of life". *Journal of the Royal Society Interface*, 10 (79), pp. 20120869.

Winding, M., Pedigo, B. D., Barnes, C. L., Patsolic, H. G., Park, Y., Kazimiers, T., Fushiki, A., Andrade, I. V., Khandelwal, A., Valdes-Aleman, J., et al. (2023). "The connectome of an insect brain". *Science*, 379 (6636), Article add9330.

Doubt Wide Open
How (Meta-Theoretic) Epistemic Possibility Sustains Scepticism while Grounding Rational Inquiry

Alexander McQuibban

University of St. Andrews

Abstract

Despite how they're often advertised, even the most radical forms of externalism fail to close the door on scepticism. This is because the sceptical scenario always remains epistemically possible. Not only is this possibility stipulated, it is guaranteed by the (meta-theoretic) uncertainty of any epistemological framework used to rule it out. Externalists may try to redefine knowledge to exclude sceptical possibilities, but for all anyone knows, externalism could be false. Thankfully, that scepticism cannot be entirely stamped out is neither surprising, nor particularly worrying: the same uncertainty which dooms any absolute rejection of scepticism is also reflective of an epistemic humility which is necessary for any rational inquiry. Part I presents a formal reconstruction of Cartesian scepticism and highlights how the epistemic possibility of the sceptical hypothesis holding frustrates standard externalist attempts at dissolving scepticism. Part II considers more radical approaches, showing that they-too-fail, succumbing to the meta-theoretic possibility that the epistemic concepts to which they appeal may be mistaken. Part III reframes epistemic humility not as a concession to scepticism but as a necessary condition for rational inquiry. While scepticism cannot be refuted in a way that guarantees its falsity, it can be disarmed, so long as we pragmatically understand any claim to knowledge as being conditional on the falsity of certain undefeated defeaters-at the very least, on the falsity of the sceptical hypothesis.

1. Introduction

Any epistemological theory oriented at knowledge must answer sceptical challenges. The most basic involve Cartesian-inspired thought experiments which posit experientially-identical scenarios which differ in their factual content, undermining subjective claims to knowledge. Given their setup, it is often thought that externalists have an advantage responding to them.

I reject this claim. Externalists (and internalists alike) must not only argue that it is possible for us to be in the non-sceptical scenario but that we *really are* in the non-sceptical scenario-or, at least, that we are fully justified in taking it that we are *not* in the sceptical scenario. This, however, is never guaranteed. The epistemic possibility of the sceptical scenario always remains open, and, even when an epistemic thesis (e.g. a brand of externalism

or internalism) is engineered with the express purpose ofrendering the non-sceptical scenario epistemically *im*possible, then an appeal to the 'meta-theoretic' epistemic possibility that said thesis could itself be mistaken is enough to revive this epistemic possibility. That being said, we are not thereby doomed to a sceptical life. The very notion of epistemic possibility is what allows for rational inquiry among epistemically humble agents *despite* the very real possibility that the sceptical scenario obtains. This consideration allows us to effectively disarm scepticism, endorsing a more pragmatic conditional conception of knowledge.

To substantiate the above, I first introduce Cartesian scepticism, usual externalist responses to it, and a rejoinder from the mere epistemic possibility of the sceptical hypothesis holding. I then consider a radical externalist response but argue another form of epistemic possibility-the meta-theoretic possibility that the externalist has the wrong concepts of justification/knowledge-is enough to defeat this argument. This open meta-theoretic possibility forces the externalist to adopt what I take to be a distinctly unappealing attitude-epistemic arrogance-should they continue to reject the epistemic possibility of the sceptical hypothesis. Finally, I highlight the importance of the opposite attitude-epistemic humility-in grounding rational inquiry and practically disarming scepticism. This is followed by a brief discussion which offers a summative conclusion, emphasises the paper's contributions, and proposes avenues for future theorising.

2. Opening the Door to Doubt: Securing Cartesian Scepticism against Externalism

2.1. An Overview of Cartesian Scepticism

Although various sceptical challenges exist, the classical Cartesian challenge is probably the most popular. This involves juxtaposition of two scenarios identical in all aspects, except for the fact that the external world differs between both. Agents perceive the external world to be a certain way, and yet, in the sceptical scenario, the external world is not in fact the way it looks. In the starkest formulations, it is because the external world simply does not exist or has been manipulated in such a way that the agent cannot trust any of their judgements (at the very least with regards to judgements that rely on perceptual experience). These scenarios can include the hypothesis that the agent is dreaming, being tricked by an evil demon[1] or is a brain-in-a-vat.[2].

For my purposes, these thought experiments are identical in the epistemic challenge they pose. I propose the following rough formalisation with an agent (**S**), warrant (W) — i.e. whatever justifies belief ina certain proposition — sceptical hypothesis (**SH**), non-sceptical hypothesis (**SH**), and propositions (P) about the external world which are true given SH but false given SH:

(P1) SH and $\sim SH$ are indistinguishable to S (by stipulation).

(P2) $SH \Rightarrow P$ and $\sim SH \Rightarrow \sim P$ (by stipulation).

(P3) $P \Leftrightarrow \sim SH$ (from P2).

(P4) $S : WP \Rightarrow S : W \sim SH$ (from P3, given closure of warrant under logical entailment).

[1] Both due to Descartes, 2017.
[2] Due originally to Putnam, 1999.

(P5) If S is equally (un)warranted in believing two contradictory hypotheses, S lacks a positive warrant for either.

(P6) S is equally (un)warranted in believing SH and $\sim SH$ (from P1).

(P7) $S :\sim W\text{-}SH$ and $S :\sim W\sim SH$ (from P6 and P5).

(P8) $S :\sim W\text{-}SH$ (from P7 by conjunction elimination).

(C) $S :\sim WP$ (from P4 and P8).

In essence, it is stipulated that the sceptical and non-sceptical scenario are indistinguishable to an agent **S** (P1), and that propositions about the external world which are true in the latter are false in the former (P2). Thus, these propositions are true if and only if **S** is, indeed, in the non-sceptical scenario (P3). Because of this, for a person to be justified in believing any such propositions they must also be justified in believing that they are, indeed, in the non-sceptical scenario (P4). This is a consequence of warrants being closed uner logical entailment — that is to say, that if one is warranted in believing a propoisition, they must be warranted in believing the other propositions that are logically entailed by that proposition holding true. Although such a closure principle is not universally uncontroversial, it seems eminently reasonable in this case. After all, how can one be justified in believing any proposition which is inconsistent with scepticism if they are not justified in believing that the world isn't mired by scepticsm. Of course, if one is positively warranted in believing that they *are* in the sceptical scenario, then they cannot be warranted in believing propositions which are inconsistent with it either. Yet, it seems that one is equally (un)warranted in believing they are in the sceptical scenario as they are in believing they are in the non-sceptical scenario (P6). This is due to the very stipulation of sceptical hypotheses (P1): for all a subject 'can tell' they could be in either scenario. Nothing which is accessible to them gives them any indication one way or the other. Their experiences are entirely compatible with both, even if in the sceptical scenario their experiences are fundamentally misleading. If an agent cannot be said to be warranted in believing one hypothesis any more than they can be said to be warranted in believing the opposite, then it would seem that they lack a positive warrant for either (P5). Thus even if an agent is not warranted in believing they're in the sceptical scenario, they also lack any warrant to believe they are in the non-sceptical scenario (P7 P8). Since any warranted belief in an external world proposition inconsistent with the sceptical scenario *would* confer a warrant that one isn't in the sceptical scenario (P4), it must be that there are no such propositions which an agent is warranted in believing (C). Ultimately, if one cannot have *any* warranted beliefs about the external world, they can scarcely be said to possess any *knowledge* about the external world.

2.2. Externalism and Cartesian Scepticism

Whereas the very design of Cartesian scepticism makes it difficult for internalists to deny that one is as warranted in believing the sceptical hypothesis as they are warranted in believing that they're in the non-sceptical hypothesis — though many still argue this[3] — externalist denials of this are, by comparison, quite simple.

[3] See Vogel, 2009.

Intuitively, one might think that given the internal/experiential indistinguishability of **SH** and **SH**, **S** cannot be any more warranted in believing one than the other; so, it is entirely consistent with **S**'s evidence to believe in **SH**. This, however, assumes that evidence and so warrants are to be construed in an internalistic fashion. If warrants are externally-construed, e.g. according to safety[4], sensitivity[5], causal history[6], or some other form of reliability, then the sceptic's argument is no longer as obvious. After all, if warrants depend on the external world, and the external world is precisely what differs from **SH** to **SH**, then it is not obvious that **S** is equally warranted in believing both. Not only this, but warrants are generally taken to be factive in externalist theories and so, if we are able to generate a warrant for **S** to believe in **SH**, **S** even has knowledge of **SH**.

While it is outside the remit of this paper to survey each type of externalist warrant, it must be noted that simply plugging in *any* externally-construed warrant is hardly sufficient for defeating sceptical challenges; and there is always the worry that warrants which *are* seemingly capable of defeating sceptical challenges are vulnerable to accusations that they are arbitrary, ad-hoc, or that they do not exclusively reflect cases of genuine justification/knowledge. For example, if reliability is taken to confer warrant, it needs to be construed so as to avoid or otherwise justify methods which regularly generate truths but seemingly do so 'in the wrong way' — e.g. clairvoyance[7] — as well as account for semi-sceptical worlds — e.g. ones in which otherwise reliable methods are on rare occasions subject to scepticism-generating trickery. Insofar as warrants are generated according to safety and sensitivity conditions, they must rely on a modal analysis which excludes sceptical and similarly semi-sceptical worlds from the neighbourhood of relevantly-close possible worlds. This exclusion must, of course, be justified, especially in light of the fact that sceptical hypotheses are generally stipulated such that agents cannot be sure whether the sceptical scenario is *merely* possible or, indeed, *actual*. Insofar as warrants rely on appeals to causal history, these appeals must explain to what degree (and how exactly) justified beliefs reflect whatever it is that caused them and unjustified beliefs fail to do this.

2.3. Externalism and Epistemic Possibility

Nevertheless, let us just assume that externalists *are* able to establish appealing externally-construed warrants that guarantee that if **SH** holds then **S** is warranted in believing **SH**.

If the externalist theory we adopt is reliabilist in nature, there is some method (say, perception) which is reliable if **SH** holds, thus warranting **S**'s belief in some propositions P (and so **SH**). If the theory depends on safety or sensitivity, the modal architecture of the world given SH is such that **S** has some safe/sensitive beliefs **P** (and thus a warrant to believe **SH**). If warrants are causally-construed then, given **SH**, the world is such that **S** has some beliefs **P** which are 'correctly-caused' by an external world, warranting belief in **SH**. In all cases, given **SH**, **S** seems to have warranted beliefs which are reliably-formed, safe, truth-sensitive or correctly-caused. It would seem that the sceptical challenge is met, given **SH**, **S**: **WP**, **W SH** and presumably **WSH**.

However, if **SH** does not hold and **SH** does, then none of this is true; in this scenario,

[4] See Sosa, 2009.
[5] See Nozick, 1981.
[6] See Armstrong, 1973.
[7] See BonJour, 1980 .

S's beliefs cannot be externally-warranted because the external world just is not such to allow for reliability, safety, sensitivity, or 'correct causation'. Crucially, however it does not seem that **S** can tell for certain whether **SH** or **SH** holds and so if the right conditions obtain. Therein lies the failure of externalist responses to sceptical challenges. Arguing that we *could* have knowledge *if* we *were* in the good case grants little comfort if it is possible that we are in the bad case. To undermine the whole epistemic enterprise, the sceptic need only secure a draw. Externalists must therefore argue that **SH** is epistemically impossible, but the epistemic possibility of **SH** cannot easily be dismissed. In fact, the stipulated similarity of **SH** and **SH** seems to suggest that for **SH** to be possible **SH** would also have to be possible.

Crispin Wright gives a similar argument to the effect that the second-order sceptical challenge — that **SH** *could* be true — is all the sceptic needs and that externalist responses cannot satisfyingly defeat this.[8] He illustrates this primarily with regards to a safety-based externalist response. Ultimately, to defeat the first-order challenge, one must secure that it be possible that **S** have a warrant in **SH** — precisely, that "it it is not be ruled out" that **W SH**. For this belief to be safe, worlds which are sufficiently close to **S**'s own must also be non-sceptical, disallowing warrant in the sceptical hypothesis — **WSH**. But this rationale works equally well for generating that **S** has a safe sceptical belief in the bad case. In the bad case, **S** secures that it be possible that **S** have a warrant for **SH** — precisely, that "it is not to be ruled out" that **WSH**. Close possible worlds which are all similarly sceptical mean **S** does not have a warrant for believing that the sceptical hypothesis does not obtain — *W SH*. Thus, safe beliefs for both **SH** and **SH** can be generated in both cases with warrants for their negation also being denied in each respective case. Yet, it still remains to be seen which scenario we are *actually* in to determine which safe belief *we* hold.

In essence, it is all well and good to claim that if the sceptical scenario does not obtain, then we have reliable, safe, sensitive, and 'well-caused' beliefs, but the very question which we are trying to determine is whether the sceptical scenario obtains. The externalist response to scepticism begs the question. To the question of whether we know anything, the externalist answers: 'if we are in a position to know, then we know.' But this fails to provide any independent argument for thinking that we are indeed in a position to know. And absent complete certainty that we are in such a position, that **SH** holds and not **SH**, we are forced to accept from a neutral standpoint that **SH** *could* hold. The open epistemic possibility of the sceptical scenario obtaining is arguably enough to undermine any knowledge we might have even if the sceptical hypothesis is ultimately false.

3. Doubt Wide Open: Meta-Theoretical Epistemic Possibility and Exposing the Limits of Radical Externalism

3.1. Radical Externalism to the rescue?

It could be contended that the above criticism fails to take externalism seriously. The epistemic possibility of **SH** is mistakenly allowed because we have snuck in internally-construed warrants. For **SH** to be epistemically possible, it must be true that "for all **S** *knows*, **SH** could be true". However, "for all **S** can *tell*, **SH** could be true" is only equivalent to this for-

[8] Wright 2008, pp. 509-517.

mulation if warrants and so knowledge are internalistically-construed according to something like 'seemings' or 'tellings', exactly the kind of perceptual evidence against which Cartesian scepticism is designed. Similarly, Matt Jope argues that Wright's "it is not to be ruled out that **SH**" only holds if we construe warrant internalistically.[9] If instead we take a thoroughgoing factive externalist account, e.g. one by which knowledge equals evidence ($K \equiv E$) *à la* Tim Wiliamson (1997), then agents in the good case have a body of evidence which includes myriad pieces of knowledge about the external world leaving **SH** epistemically possible but **SH** epistemically impossible. Moreover, agents in the bad case have no evidence at all and so are not justified even in believing **SH**.

This Jope argues is "by stipulation", although one could plausibly stipulate a sceptical scenario where this is less obvious, undermining his argument.[10] We might, for example, consider 'semi-sceptical' scenarios engineered such that the agent only has one piece of knowledge but has no warrant, externally-construed or otherwise, to believe (and so know) anything else. If full-blown scepticism demands resolution, a 'lonely fact' scenario likely does too — it is no great consolation for the anti-sceptic to only be able to know one solitary fact. And, yet, *prima facie*, such scenarios are easily imagined: take an agent **S** who occupies a world where only one feature of their perceptual experience — say, a certain smell — is genuine. To keep externalists happy, we might say that all other features of their perceptual experience are simulated and are not reflective of any 'real' fact about the world, whereas the lonely smell is 'correctly' caused under some relevant externalist conception. Externalists might still argue that a properly externalist construal of evidence and warrant (especially as factive) diagnoses this situation properly. Here, **S** 'knows' this lonely fact but no other, and in knowing all the (first-order perceptual) facts that are to be known, they are perhaps even positively warranted in believing the further (second-order deductive) fact that they are in a semi-sceptical scenario. Of course, this response must still contend with the worry that the knowledge it secures is no secure knowledge at all given that it seems that **S** seems no more justified in believing that they are in the 'lonely fact' world than they are in believing that they are in any other 'N fact(s)' world in which any (completely arbitrary) number of facts about the real world are actually knowledge-apt, or arguably, than they are in believing that they are, in fact, in the '0 fact' world set out in the sceptical hypothesis. Nevertheless, we might consider tweaking our scenario even further such that this response appears all the more defective. While the intricacies of these tweaks are firmly outside the remit of this essay, especially given the fact the following sub-section on meta-theoretic epistemic possibility should be enough to deal with any externalist rejoinders to sceptical, semi-sceptical or similar epistemically-worrying scenarios, they might, for example, include agents who are induced to have troublingly indeterminate though still knowledge-apt lonely or near-lonely beliefs such as the belief that "at least one other of their beliefs is (or merely could be) true", the belief that "it is possible that none, some, or all of their beliefs are true", the belief that "the sceptical scenario *might* hold", or the belief "that scepticism is (or merely could be) philosophically defensible". These strategies might also be conceived of in a subtly different way: they can all be reinterpreted as New Evil Demon-like scenarios involving 'minimal epistemic twin worlds' which are exactly the same as the actual non-sceptical world

[9] Jope 2021, pp. 43–50.
[10] Ibid.

(assuming that it is indeed non-sceptical) but where only the relevant belief or set of beliefs are correctly caused, leaving all other beliefs to register as technically unsuitable for constituting knowledge under strict externalist analysis. Either way, it does not seem obvious that there aren't any 'sceptical-adjacent' cases which include the sort of evidence which would warrant some kind of knowledge, perhaps even the knowledge that one is in such a case, but which would still present the epistemic limbo sufficient for the sceptic to establish a suitably repugnant epistemic conclusion. It is even less obvious that one can deny that such cases exist simply "by stipulation", as Jope does.

Of course, the kind of radical externalist response which Jope has to rely on is theorised precisely to resolve these and all other 'problematic' cases by including within (and arguably excluding from) the body of evidence exactly what is required to correctly diagnose the 'epistemic' state of the world despite agents in these cases and seemingly our actual world *appearing* none the wiser. While this might secure an anti-sceptical conclusion by its own radically externalist merits, it again fails to take scepticism seriously, leading us to doubt whether these merits should be considered merits at all. This response only tells us that *if* **S** is in the good case **S** is not justified in thinking that **S** is in the bad case, and that if **S** is in the bad case she cannot come to know it, but the response gives us no clarity on whether *we* are in fact in the good or the bad case. If anything, factive anti-sceptical externalist accounts like $K \equiv E$ run into a kind of absurdity. They argue that *if* **S** is in the good case, **S** is not only justified in thinking so, **S** actually *knows* it. However, it seems intuitively true that one cannot actually *know* if they are in the good case. At the very least, I am positively uncertain that I am in the good case. So, it would seem that I must be in the bad case, although I am allegedly not justified in believing this either. Perhaps this is just illustrative of a failure to recognise one's own knowledge — Williamson[11] and other like-minded externalists like Amia Srinivasan[12] do, after all, dispute luminosity and the principle that possessing knowledge entails knowledge that one possesses said knowledge (the KK-principle)— but it , at least, strike us as odd that we possess the certain knowledge that we are indeed in the good case despite debates about scepticism raging on. The prospect that one could be in a position to consistently fail to recognise one's own knowledge, such that to them and even to others — no matter how rational we otherwise take everyone involved to be — they do not seem to have knowledge at all, *should*, in any case, strike us as deeply worrisome.

3.2. Meta-Theoretic Epistemic Possibility

Notwithstanding this, there remains a worry for the radical externalist even if she believes that she has proven that it is epistemically impossible for **S** to be in the bad case if warrants are 'correctly' externally-constructed. This worry appeals to another kind of epistemic possibility, namely the meta-theoretic epistemic possibility that she could be wrong about justification or knowledge, i.e. that the true' or, at least, most philosophically useful concepts related to justification and knowledge — the ones she thinks she is appealing to with her radical externalism — *ought* actually to be construed in some non-externalist fashion.[13]For the sake of simplicity, I characterise both the possible 'falsity' and philosophical

[11] Williamson, 2000.

[12] Srinivasan, 2015.

[13] Risberg (2023) offers a similar concept: "meta-scepticism", the worry that any particular conception of knowledge could be 'wrong'.

disutility/irrelevance of externalism in terms of 'truth'.

Following the "for all **S** knows. . . " formulation, this meta-theoretic epistemic possibility can be stated thusly: "for all **S** knows, externalism could be false." Now even if externalism is 'true', it is not obvious that it is not the case that "for all **S** knows, externalism could be false." After all, it may be true that it is raining in Sydney, but for all *I* know, it could be that this is not the case. If externalism is 'false', then it is trivially true, that "for all **S** knows, externalism could be false." There is thus, a relevant disanalogy, between the epistemic possibility of scepticism which the radical externalist denies and the meta-theoretic epistemic possibility of externalism being false. For the radical externalist response to work, they must argue that if externalism is 'true' then it is epistemically impossible for **S** that it could be false, else the epistemic possibility that externalism is false allows for the epistemic possibility that externalist responses to scepticism are wrong and so that **SH** is again epistemically possible.[14]

Even if we manage to fix this analogy in the radical externalists' favour, we are left with our previous worries. It may *not* be the case that "for all **S** knows, externalism could be false" if externalism is true (**E**) but it *is* the case that "for all **S** knows, externalism could be false " if externalism is indeed false (**E**). Yet, there is presumably no way to verify beyond any doubt whether *we* are in a world where **E** or **E** holds. What's worse is that appeals to externalism seem all the more dubious to resolve this worry, since here externalism is the very thing at stake.

3.3. Meta-Theoretic Epistemic Possibility Reformulated

This thinking leads to an important insight. The radical externalist might argue that as long as epistemic possibility is defined by the "for all **S** knows. . . " formulation they can interpret "knows" externalistically and so dismiss both scepticism and this meta-theoretic scepticism. But it is clear that this move should not be allowed if we want epistemic possibility to pull its philosophical weight. The thought is if the very possibility of knowledge or any particular analysis of knowledge is at stake, then we cannot smuggle in this possibility or any particular analysis in an argument which, using these as premises, seeks to prove them. This is akin to Wright's[15] analysis of transmission failure. Consider the following Dretske-style syllogism[16]:

(P1) The wine in this glass looks like red wine. (Assumption)

(P2) The wine in this glass is red wine. (Inferred from P1)

(C1) The wine in this glass is not white wine dyed to look like red wine. (Deduced from P2)

If we accept all the premises, then, the argument runs smoothly with no failure in transmission. However, the inference made in P2 although a rather reliable-seeming inference

[14] Of course, this does not mean all lines of objection to scepticism are misguided, but if the arguments presented in this paper succeed, the onus is on the anti-sceptic to make their case, a case which seems all the more hopeless in light of the meta-theoretic epistemic considerations which this paper advances.

[15] Wright, 2003.

[16] cf. Dretske, 2013.

is clearly only a materially justified deduction if we already assume C1 is true. The negation of C1 acts as possible defeater; if it is not true, then the link between P1 and P2 is severed. Epistemic possibility when interpreted non-neutrally acts the same way. The externalist interprets the "for all S knows" formulations with "knows" operating in an externalist anti-sceptical way to generate general conclusions about the validity of externalism and anti-scepticism themselves. In fact, the actual conclusions generated are trivial ones that are in a sense conditional on themselves, the same way C1 is conditional on C1 holding. The upshot is that if the epistemic possibility we are judging has an effect on the first half of the formulation we use to judge it, i.e. the "for all **S** knows", then we must interpret "for all **S** knows" in a way consistent with withholding judgement on that epistemic possibility. As such the epistemic possibility of scepticism and the meta-theoretic epistemic possibility of externalism being false cannot be interpreted using solely an anti-sceptical or externalistic conception of knowledge. Withholding this judgement, it is clear that both remain open epistemic possibilities, which as we have seen is all the sceptic needs to claim some sort of victory.

Ultimately, radical externalists can only claim to have secured the epistemic impossibility of externalism's falsehood and of the sceptical scenario on the epistemically arrogant grounds that they are so certain of both that they could not even imagine themselves to be wrong. While some externalists may be so inclined, this reasoning is not likely to sway many opponents who would do well to flag the existence of sustained, sensible, and seemingly serious philosophical debate both on the topic of externalism's merits and of scepticism as clear contra-indications against treating either this way.

4. Closing the Door to Doubt: Epistemic Humility, Rational Inquiry, and Disarming the Sceptical Hypothesis

4.1. The Importance of Epistemic Humility to Rational Inquiry

There is a greater lesson to be learnt from all of this, however. Epistemic humility is key to all pursuits of knowledge through rational inquiry.

At the very least, classical logic seems to presuppose an initial position of epistemic humility. This is evident from cases of transmission failure. The methods of classical logic only generate knowledge from valid arguments if these argument's premises can be substantiated independently from the very conclusion they purport to prove. For any competent deduction to take place, we must theoretically withhold judgement of — i.e. be in a position of epistemic humility towards — the matter being deduced. In practice, we may already confidently believe the conclusions of a given argument — e.g. because it has been proven elsewhere — but the only way for the argument itself to properly transmit knowledge is if the argument itself is posed in this epistemically humble way.

Admittedly, this argument from logical transmission alone establishes only a narrow, context-specific kind of epistemic humility — one necessary for warrant acquisition in deductive arguments. However, this limited notion strongly motivates, even if not strictly entailing, the adoption or recognition of epistemic humility as an *epistemic* virtue, more broadly.

In fact, a rather natural generalisation of the argument to cover epistemic humility as necessary (or at least conductive) to *any and all* rational inquiry, in general, can be made.

The acquisition of knowledge requires epistemic humility towards said knowledge, at least prior to its acquisition — else one would either not be in a position to acquire it or already be in possession of it. This also applies to any serious debate or discussion surrounding a piece of 'knowledge' one claims to possess; if a participant is not epistemically humble with regards to this tentative knowledge, then there is no point in arguing against them for they are not open to the epistemic possibility of its negation.

Epistemic humility is simply a virtue. Sharon Ryan motivates it as a distinctly epistemic virtue insofar as it is "an awareness of, and a proper response to, our epistemic limitations."[17] If as rational albeit non-ideal agents, we ought to reason with certain epistemic tools (e.g. evidence, reasoning capacities, and empirical methods), then epistemic humility is simply doing justice to any limitations these tools might exhibit — these tools always exhibiting *some* limitations given non-ideal circumstances. Insofar as we might have certain moral duties to reason correctly, this epistemic humility also takes on a distinctly ethical character.

4.2. Disarming the Sceptical Hypothesis

That the same epistemic humility which grounds rational inquiry ensures the success of sceptical challenges might be alarming. Thankfully, epistemic humility is also the key to disarming the sceptical hypothesis. **SH** only poses any threat to epistemically arrogant conceptions of knowledge, that is to say conceptions of knowledge by which knowledge has a certain and epistemically unassailable quality; but it poses no threat to epistemically humble pragmatic conceptions of knowledge.

Of course, the ideal aim and theoretical subject matter of epistemology is intuitively the former. Knowledge, by its opposition to concepts like a 'hunch', 'guess' or 'bet' and by its frequent association to truth, must arguably reflect this certain and unassailable quality. After all, it seems perfectly correct to say that if S truly *knows* that **P** then it cannot be the case that **P**.

That being said, this kind of knowledge cannot be the *practical* aim of epistemology. When non-ideal agents in a potentially sceptical world engage in rational inquiry, they typically take, or at least *should* take, the outcome of said rational inquiry to be a distinctly limited type of knowledge — call it, '$knowledge_p$'. $Knowledge_p$ is conditional on certain unsubstantiated, perhaps even unsubstantiable, axioms. Concluding that I have a hand on the basis that it *seems* I do is only true if there is an external world that matches my seemings. Concluding that someone genuinely feels hurt on the basis that they display certain behaviour is only true if that behaviour is actually associated with the correct mental/physical state. Every piece of $Knowledge_p$ we acquire, no matter how basic, has an undefeated (perhaps yet-to-be-defeated) defeating condition. This means that using any $Knowledge_p$ to prove the very conditions on which it is based entails a kind of transmission failure.

To the extent that any such conditions are taken to be necessary for rational inquiry to proceed, however, they are what calls metaphysical 'cornerstones' or 'heavyweights' which we are rationally entitled to believe in even if we cannot be fully certain of them.[18][19] Whether something is a metaphysical heavyweight might depend on the context in which the possession of 'knowledge' is asserted. In the context of (mainstream) geological sciences, for

[17] Ryan 2019, p. 132.
[18] Wright 2004.
[19] Moretti Wright, 2023 .

example, the rejection of the Omphalos Hypothesis — the creationist thesis that the world was created a mere several thousand years ago, but in such a way that it appears 'old' consistent with the mainstream evolutionary and cosmological consensus — might be considered a metaphysical heavyweight. In the empirical sciences more generally, the same might be true of the principle that induction is a reliable method. The rejection of scepticism, however, is a uniquely context-independent metaphysical heavyweight insofar as *any* knowledge presumably exhibits the sceptical hypothesis as a defeating condition.

This tells us that appeals to any practically attainable $knowledge_p$ must be understood to be conditional at the very least on the negation of the sceptical hypothesis.[20] But this is no revolutionary conclusion. It strikes me as perfectly acceptable that whenever someone claims to 'know' something they only claim (or, at least, *ought* only claim) to $Know_p$ it conditional on the assumption that the sceptical hypothesis (and any other undisclosed defeaters) do not hold. Of course, the sceptical hypothesis is not thereby disproved — far from it — but it is surely disarmed for practical purposes.

5. Discussion

In conclusion, externalist responses, no matter how radical, stand no chance against sceptical challenges. The success of Cartesian scepticism and arguably any sound form of scepticism, more generally, is secured by the indissoluble epistemic possibility that the sceptical hypothesis might indeed hold. Any claimed eradication of this epistemic possibility arguably betrays an underlying epistemic arrogance which contrasts sharply with the epistemic humility required to appreciate scepticism as a sensible philosophical position which poses a serious challenge to knowledge. Thankfully, the same epistemic humility which opens the door to sceptical doubt is also what grounds all rational inquiry and is ultimately what allows us, for practical purposes, to close that very same door — or at least leave it ever so slightly ajar. Scepticism will forever undermine the quest for certain knowledge, but this is no problem once it is appreciated that it is only $knowledge_p$, knowledge conditional on the sceptical hypothesis being false, which we must and can ever seek.

Though the rejection of externalist responses to scepticism is not in itself novel, this paper's identification of such responses with cases of transmission failure and epistemic arrogance, as well as its focus on meta-theoretic possibility in undermining them are. Similarly, while pragmatic (re-)interpretations of knowledge in response to sceptical challenges is routine discourse in the conceptual engineering literature, I diverge from this discourse in maintaining that knowledge, as an ideal, should retain its certain and unassailable quality while still endorsing a pragmatic interpretation as a separate (though related concept) to account for everyday knowledge claims. In this respect, I endorse a very real albeit palatable scepticism; I view knowledge as *in principle* unattainable while neutralising the worry that is usually associated with this outlook. Crucially, in identifying epistemic humility as both the very same thing which sustains sceptical challenges and which grounds all rational inquiry, I take the recognition of uncertainty and the appreciation of scepticism to exhibit a distinctly positive epistemic value.

Of course, there is still work to be done to fully flesh out and extend the thesis I have laid

[20] Schiffer (2004) offers a similar pragmatic reconception of justification/knowledge.

out in the preceding passages. In particular, the notion of meta-theoretic epistemic possibility would likely benefit from further formalisation as would the brief sketches I have offered of different sceptical and semi-sceptical' scenarios which I believe are even less amenable to externalist resolution than standard Cartesian scepticism. The value I give epistemic humility and my attitudes towards knowledge as a concept might deserve further justification and/or scrutiny as well. Nevertheless, I believe the paper's main claims — that rejections of scepticism either fail to take scepticism seriously or require adopting an epistemically arrogant attitude, and that taking scepticism seriously and adopting an epistemically humble attitude promote rather than undermine the pursuit of the kind of knowledge which fallible humans can ever hope to possess — hold firmly.

Bibliography

Armstrong, D. (1973) *Belief, Truth and Knowledge*. London, UK: Cambridge University Press.

Bonjour, L. (1980) "Externalist Theories of Empirical Knowledge". *Midwest Studies in Philosophy*, 5(1), pp. 53–73.

Descartes, R. (2017) *Meditations On First Philosophy: With selections from the Objections and Replies*. 2nd ed. Cottingham, J. and Williams, B. (eds.). Cambridge, UK: Cambridge University Press.

Dretske, F. (2013) "Is Knowledge Closed Under Known Entailment? The Case Against Closure". In: Steup, M. and Turri, J. (eds.). *Contemporary Debates in Epistemology*. Oxford, UK: Blackwell, pp. 13–26.

Jope, M. (2021) "On the Alleged Instability of Externalist Anti-skepticism". *The Journal of Philosophy*, 118(1), pp. 43–50.

Moretti, L. and Wright, C. (2023) "Epistemic Entitlement, Epistemic Risk and Leaching". *Philosophy and Phenomenological Research*, 106(3), pp. 566–580.

Nozick, R. (1981) *Philosophical Explanations*. Cambridge, MA: Belknap Press of Harvard University Press.

Putnam, H. (1999). "Brains in a Vat". In: DeRose, K. and Warfield, T.A. (eds.). *Skepticism: A Contemporary Reader*. Oxford, UK: Oxford University Press.

Risberg, O. (2023) "Meta-Skepticism". *Philosophy and Phenomenological Research*, 106(3), pp. 541–565.

Ryan, S. (2019) "Epistemic Humility, Defeat, and a Defense of Moderate Skepticism". In: Fitelson, B., Borges, R. and Braden, C. (eds.). *Themes from Klein: Knowledge, Scepticism, and Justification.* Springer International Publishing, pp. 129–143.

Schiffer, S. (2004) "Skepticism and the Vagaries of Justified Belief". *Philosophical Studies*, 119(1/2), pp. 161–184.

Sosa, E. (1999) "How to Defeat Opposition to Moore". *Philosophical Perspectives*, 13, pp. 141–153.

Srinivasan, A. (2015) "Are We Luminous?". *Philosophy and Phenomenological Research*, 90(2), pp. 294–319.

Vogel, J. (2008). "Internalist Responses to Skepticism". In: Greco, J. (ed.). *The Oxford Handbook of Skepticism*. Oxford, UK: Oxford University Press.

Williamson, T. (1997) "Knowledge as Evidence". *Mind*, 106(424), pp. 717–741.

Williamson, T. (2000) *Knowledge and its Limits*. New York, NY: Oxford University Press.

Wright, C. (2003). "Some Reflections on the Acquisition of Warrant by Inference". In: Nuccetelli, S. (ed.). *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, MA: MIT Press, pp. 57–77.

Wright, C. (2004) "Warrant for Nothing (and Foundations for Free)?". *Aristotelian Society Supplementary Volume*, 78(1), pp. 167–212.

Wright, C. (2008) "Internal-External: Doxastic Norms and the Defusing of Skeptical Paradox". *The Journal of Philosophy*, 105(9), pp. 501–517.