Southwest Philosophical Studies

The Journal of
The New Mexico Texas
Philosophical Society

Volume 43
2021

SOUTHWEST PHILOSOPHICAL STUDIES

http://www.nmwt.org/

Southwest Philosophical Studies, the journal of the New Mexico Texas Philosophical Society, is published annually. General correspondence may be addressed to Justin Bell, editor-in-chief *Southwest Philosophical Studies*; Texas A&M University–Victoria; College of Liberal Arts and Social Sciences; 3007 North Ben Wilson; Victoria, TX 77901. Justin Bell's email address is BellJ1@tamuv.edu.

Submissions: Only papers read at the Annual Meeting of the New Mexico Texas Philosophical Society are eligible for publication in Southwest Philosophical Studies. Manuscripts in any area of serious philosophical interest are blind-refereed for possible acceptance in two phases, the first for formal presentation during the Society's annual spring meeting and the second for publication in the following year's volume of Southwest Philosophical Studies. Contact the Secretary-Treasurer for technical details of manuscript preparation. The call for papers is published on the Society's website at <http://www.nmwt.org/> and sent out by the Secretary through the distribution list. Please contact the Secretary to be added to the list.

ISSN 0885-9310

Selected Proceedings of

The Seventy-First Annual Meeting of The New Mexico Texas Philosophical Society

May 21-22, 2021

The meeting was held online

with the support of Baylor University

EDITOR'S NOTE

COVID-19 presented the society with a number of challenges. Unfortunately, the society had to cancel our 2020 conference. The 2021 conference, of which these are the proceedings, was held online. Those accepted into the 2020 conference were invited to participate in 2021.

I would like to thank the following members of the New Mexico Texas Philosophical Society for their editorial assistance with this volume:

Dave Beisecker
Parish Conkling
Jackson Hoerth
Jacob Mills
Nathan Smith
Sarah Woolwine

I count myself lucky to be a member of a society as supportive as the New Mexico Texas Philosophical society and to have colleagues willing to volunteer their time toward making *Southwest Philosophical Studies* successful.

CONTENTS

Was Socrates the Wisest Man in Athens, and If So, in Which Respects?

Jean-Paul Vessel
New Mexico State University

Presidential Address

1. Introduction

In a well-known passage near the beginning of Plato's *Apology*, Socrates recounts a story about his friend Chaerephon visiting the Oracle at Delphi. Chaerephon asked the oracle whether anyone was wiser than Socrates "and the Pythian replied that no one was wiser" (21a6-7).[1] Much has been written about what the Pythian meant and in precisely which respect (or respects) Socrates might have been regarded by the oracle (or himself) to be at least as wise as any of his contemporaries. Here I want to develop an extensive inventory of the ways in which Plato's Socrates[2] was—or at least might have been—as wise as or even wiser than his peers. I motivate this discussion by developing a puzzle from Plato's *Apology* 21d-23b. I argue that the negative thesis that Socrates' superlative wisdom consists solely in knowing that he does not know anything "beautiful and good" cannot effectively establish the "fact" that no one was wiser than Socrates. I then proceed to canvass a number of Socrates' other features that might justify the Pythian's declaration. Ultimately, I argue that a strong case can be made for the claim that Socrates was in fact the wisest man in Athens.

2. The Oracle's Declaration and Socrates' Response to It

When addressing the jury about how he came to develop the reputation that he has and why many people, some of them prominent political people like Anytus, became comfortable slandering him, Socrates suggests that his possessing a certain type of wisdom explains his reputation and sparked the slander: "What type of wisdom? Human wisdom, perhaps. It may be that I really possess this thing…."

(20d6-7). Socrates then contrasts his human wisdom with the wisdom associated with the sophists Gorgias, Prodicus, Hippias, and Evenus, who "are wise with a wisdom more than human; else I cannot explain it, for I certainly do not possess it, and whoever says I do is lying and speaks to slander me." (20d7-9). Almost immediately after this, Socrates reports the oracle's declaration that no one was wiser than he, precipitating his mission to understand whether the oracle could be right and, if so, in which respects he might be among the wisest people, if not *the* wisest person, of his time.

Socrates then sought out to refute the oracle by seeking out a prominent political person with a reputation of wisdom: If Socrates could find someone wiser than he was, the oracle would be refuted. But when Socrates examined this person, Socrates came to the conclusion that the man "appeared wise to many people and especially to himself, but he was not" (21c5-6). After Socrates attempts to show this person that he is not in fact wise, which results in Socrates becoming the target of public odium of this man and others, Socrates comes to a somewhat puzzling conclusion. In one of the most prominent passages in the *Apology*, Socrates thinks to himself (at 21d3-7):

> "I am wiser than this man; it is likely that neither of us knows anything worthwhile [*kalos kagathos*[3]], but he thinks he knows something when he does not, whereas when I do not know, neither do I think I know; so I am likely to be wiser than he to this small extent, that I do not think I know what I do not know." (Grube translation in Plato 1997)

> "I am wiser than that human. For I daresay that neither of us is aware of anything beautiful-and-good [*kalos kagathos*]; but that one supposes he is aware of something beautiful-and-good, though he isn't aware; whereas I, as indeed I'm not aware, neither suppose that I am. Therefore, I'm at least likely to be wiser than that person at least in just this one small thing: that the things that I'm not aware of I don't suppose that I'm aware of either." (Senn translation in Plato 2019)

There are a few interesting things going on in this passage. First, Socrates concludes that he is wiser than the political man he examined. Second, Socrates suggests that neither of them knows (or is aware of) anything *kalos kagathos* (worthwhile, beautiful-and-good, fine and good, noble and good). Socrates then accuses the political man of thinking that he knows (or is aware of) something *kalos kagathos* when he does not. Finally, Socrates reasons that he is wiser than the political man in at least this small respect: Unlike the political man, Socrates himself does not think he knows what he does not know. Socrates then proceeds to examine another political man regarded even more highly for his wisdom, other politicians, poets, and finally the craftsmen. In each case, he finds the people whom

he examines to be deficient in just the same way: They think themselves to be wise about things outside their areas of expertise—the *kalos kagathos* perhaps and "other most important pursuits" (22d6-7)—but they are not wise about these things.

This suggests that Socrates' (alleged) superlative "human" wisdom can be articulated solely in terms of epistemic humility. Seemingly everyone whom Socrates examines thinks they know things (or are aware of things) that they do not, and this seems especially so when it comes to things *kalos kagathos*. Socrates, on the other hand, claims that he does not know anything (or is aware of anything) *kalos kagathos*, nor does he think that he knows (or is aware of) anything *kalos kagathos*. Understanding precisely what Socrates means to pick out with the phrase *kalos kagathos* is tricky business; however, there are some clues in the *Apology* and elsewhere in the early dialogues that should assist us in understanding what Socrates means with his use of the phrase *kalos kagathos*.

Just before the passage about the oracle in the *Apology*, Socrates discusses the "more than human" (20e2) wisdom attributed to some of the Sophists. Socrates recounts a conversation he had with Callias, who allegedly had "spent more money on Sophists than everybody put together" (20a5). Socrates asks Callias whether he knows of anyone who "is an expert in this kind of excellence, the human and social kind" (20b3-4)[4] and finally asserts: "Certainly I would pride and preen myself if I had this knowledge, but I do not have it, gentlemen" (20c2-3). I think that it is reasonable to suppose that Socrates' use of the phrase *kalos kagathos* at *Apology* 21d is intended to pick out a certain type of arete or excellence: personal and civic virtue.[5] This interpretation meshes well with some of the more prominent themes in the *Apology*, especially Socrates' mission to exhort his fellow citizens to care primarily for the best possible states of their souls (30b1-2) and to "care for virtue" (31b5).[6] With this interpretation in mind, some might be inclined to articulate Socrates' human wisdom solely in terms of epistemic humility, especially with respect to things *kalos kagathos*.

> EH: Socrates does not think he knows things that he does not know. Socrates claims that he does not know anything *kalos kagathos* nor does he think he knows anything *kalos kagathos*.

This attempt to capture the nature of Socrates' human wisdom is naturally extracted from the passages we have been looking at (*Apology* 19e-23b).[7] But I am unpersuaded that it successfully accounts for the entirety of Socrates' human wisdom. Consider this challenge one of my students lodges against it in an extra credit essay on Plato's *Apology*. My student writes:

> So what? Why does this make Socrates so wise? I admit that I don't know anything fine and good. I admit that I really don't know the true nature of

> moral virtue, nor do I know which things are best to pursue in a human life. I guess no one in Ancient Athens was wiser than me either. I'm as wise as Socrates! But I doubt that I'm anywhere near among the wisest here in Las Cruces; in fact, I suspect I'm not even close to as wise as many of my classmates.

This challenge to the claim that Socrates' (perhaps) unparalleled wisdom can be completely captured by EH is straightforward and simple. Many newcomers to philosophy quickly realize how perplexing are the natures of some of our most important moral concepts. And those with intellectual integrity will readily admit that while they have some ideas about these concepts, they cannot be sure that their ideas track the truth. Until these ideas are sufficiently challenged—and until the perplexities about these concepts are resolved in theoretically satisfying ways—we should admit that we do not actually know anything *kalos kagathos*.[8] The upshot of this challenge is that it seems too easy to become as wise as Socrates on this EH account. Those who have never thought seriously and carefully about things *kalos kagathos* can easily admit that they do not really know anything about them. They would thus be as wise as Socrates. And so would those who recognize that their beliefs and ideas about things *kalos kagathos* simply do not meet the standards required for knowledge. But Socrates brought something to Hellas that shocked the entire world. His wisdom cannot be had so easily.

3. Socrates Examines the Craftsmen

One might object to the challenge above by pointing out what might be important differences between Ancient Athenian men and today's young students and citizens. Today's young students have access to philosophers (and other teachers) who can pose serious theoretical challenges to various conceptions of things *kalos kagathos* that today's youth might deem accurate or in some ways theoretically attractive. The Ancient Athenians—it seems—had only Socrates, the first philosopher to investigate moral concepts and principles with his elenctic method.[9] There is also the fact that today's (American) society is radically more heterogeneous than Ancient Athens was. Perhaps the proud, impressive, homogenous Athenian culture bred epistemic confidence—even epistemic hubris—among its citizens regarding things *kalos kagathos*. This is certainly suggested by early passages in the *Apology* and several passages within many of the other early Platonic dialogues. Members of contemporary communities also display similar epistemic confidence and hubris regarding things *kalos kagathos*, but there are many people in today's societies who do not, for a variety of sociological reasons. Perhaps these acute differences between the Ancient Athenians and contemporary societies could be used to mount a defense of an EH account of Socrates' human wisdom. Still, challenges and puzzles remain, some of

which emerge from Socrates' examination of the craftsmen.

Disappointed by the poets, Socrates turns to the craftsmen, whom he claims to have "knowledge of many fine (or beautiful[10]) things" (22d2-3), a certain sort of wisdom to be sure. After examining the craftsmen, Socrates reflects:

> In this I was not mistaken; they knew things I did not know, and to that extent they were wiser than I. But, men of Athens, the good craftsmen seemed to me to have the same fault as the poets: each of them, because of his success at his craft, thought himself very wise in other most important pursuits…. (22d3-7)

Socrates admits here that the craftsmen knew things that he himself did not know, and thus to that extent they were wiser than he was. But then Socrates levels his familiar charge of epistemic hubris against the craftsmen too: Because of his success at his craft,[11] each craftsman "thought himself very wise in other *most important pursuits.*"[12] These *most important pursuits* (or the *greatest things*) should surely be considered things *kalos kagathos*.[13] Socrates concludes that the craftsmen thought themselves to have wisdom about these most important pursuits, but—in fact—they did not.[14] Socrates continues to reflect upon his examination of the craftsmen:

> …and this error of theirs overshadowed the wisdom they had, so that I asked myself, on behalf of the oracle, whether I should prefer to be as I am, with neither their wisdom nor their ignorance, or to have both. The answer I gave myself and the oracle was that it was to my advantage to be as I am. (22e1-5)

Socrates suggests here that the epistemic hubris of the craftsmen regarding the most important pursuits eclipses their craft-oriented wisdom so much so that Socrates—with neither epistemic hubris about the most important things nor beautiful craft-oriented wisdom—emerges as wiser than the craftsman. Socrates' suggestion is both intriguing and puzzling. Moreover, another challenge against an EH account of Socrates' human wisdom can be wrought from it.

It seems that Socrates' ranking of himself over the craftsmen with respect to wisdom is defensible only if *all* of the craftsmen in Ancient Athens were epistemically hubristic regarding things *kalos kagathos* and the *most important pursuits*.[15] Maybe they were; we'll never know. But maybe some of them were not so epistemically hubristic about such things. There are stories—perhaps apocryphal—about Socrates spending a substantial amount of time with various craftsmen. Diogenes Laertius reports that Socrates was friends with Simon—a cobbler—and that "[w]hen Socrates came to his workshop and began to converse, [Simon] used to make notes of all that he could remember."[16] Simon (allegedly)

was a philosopher too. Diogenes Laertius credits him with thirty-three dialogues, including several (seemingly) about things *kalos kagathos*: *Of the Good*, *On the Beautiful*, *On the Just: two dialogues*, *Of Virtue*, *that it cannot be taught*, *Of Courage: three dialogues*, and *Of Honour*, among others. If Diogenes Laertius' report about Simon is accurate, then we can be assured that Simon definitely had a lot of philosophical ideas—maybe he even held positions—about things *kalos kagathos*. But was he epistemically hubristic about these ideas and positions? Maybe. But maybe not.

The titles of some of Simon's (purported) dialogues suggest that he might have been confident about positions on things *kalos kagathos*, perhaps especially: *Of the Good*, *On the Just: two dialogues*, and *Of Virtue*, *that it cannot be taught*. There might even be some stridency in the latter title. But we must also take into consideration that Simon (allegedly) wrote *Socratic* dialogues, not treatises. And we can imagine what Socrates' influence on Simon might have been like.

Consider Plato's early dramatic masterpiece: *Protagoras*. The character Socrates in *Protagoras* seems especially attracted to an intellectualist account of the virtues and argues for it at great length, but also confidently (or ironically?) deploys arguments against the thesis that virtue can be taught. At the end of the dialogue, Socrates articulates the unresolved perplexities that have emerged from his dialectical discussion of the virtues with Protagoras (361a5-c3). Socrates requests that they continue their inquiry with the hope that some theoretical progress can be made (361d1-4), but Protagoras opts to table the discussion for the time being, adding—however—that:

> Indeed, I have told many people that I admire you more than anyone I have met, certainly more than anyone in your generation. And I say that I would not be surprised if you gain among men high repute for wisdom. (361e3-6)

This is high praise from Protagoras.[17] He definitely detects Socrates' wisdom. If Socrates can retain his epistemic humility regarding the true natures of the virtues and whether they can be taught throughout, and even after, his investigation with Protagoras, then so might Simon the Cobbler retain his own epistemic humility upon the completion of his own dialogues. Of course, we'll never know the truth about this. But suppose Simon did, in fact, maintain an attractive form of epistemic humility throughout his philosophical inquiries. Given this supposition, it would appear that Simon is wiser than Socrates. He shares Socrates' epistemic humility regarding things *kalos kagathos* but also possesses valuable craft-specific wisdom about cobbling. Socrates admits as much. Simon could then serve as a counterexample to the Pythian's declaration. But should we conclude that Simon was wiser than Socrates? I would be reluctant to do so. Protagoras seemed to be onto some of Socrates' attractive features. Perhaps those features reflect a sort of

wisdom that outstrips any non-craft-specific wisdom that Simon enjoys. If so, we must reject an EH account of Socrates' superlative wisdom.

Now consider Simon's (completely fictional!) craftsman friend Skepticos. Skepticos is a potter who wanders into Simon's workshop while Simon is engaged in philosophical discourse with Socrates. Skepticos is shell-shocked when he witnesses Socrates and Simon clearly illustrate that popularly endorsed Athenian conceptions of courage, virtue, and honor are mired in inconsistencies. Socrates then turns to Skepticos, asking him whether he knows anything about courage, virtue, and honor. Skepticos immediately realizes that while he has some ideas about those moral concepts, he has no knowledge of them, none whatsoever. And—unlike Simon—Skepticos never attempts to work out accounts of these concepts. Yet he remains skeptical, embodying the type of epistemic humility about things *kalos kagathos* that Socrates claims for himself. If there were a person like Skepticos in Ancient Athens, then Socrates—and perhaps even the oracle—would have to admit that Skepticos (or whoever) is wiser than Socrates. He shares Socrates' wisdom grounded in epistemic humility but also possesses craft-specific wisdom about how to create beautiful pieces of pottery. If we are reluctant to embrace this conclusion, we must reject an EH account of Socrates' human wisdom. There must be something more to the nature of Socrates' wisdom that is responsible for (at least perhaps) his own wisdom enjoying an elevated position above whatever wisdom his peers possessed.

4. Identifying Inconsistencies and Posing Perplexing Philosophical Challenges

In the remaining sections I canvas several other admirable features of Socrates in an attempt to articulate the full extent of Socrates' human wisdom. Some of these features are related to Socrates' epistemic humility in robust ways, others perhaps less so. In any case, my hope is to show that the fusion of these features with Socrates' epistemic humility makes for a more complete and more attractive account of Socrates' human wisdom than can be provided by his epistemic humility alone.

Socrates brought to light that the Athenian positive morality was muddled in serious ways. Socrates pointed out that features of the popularly-endorsed Athenian conceptions of the virtues (and subsequently moral improvement) were unprincipled in some cases and inconsistent in others. And given how important Socrates regarded caring for one's soul or psyche (*Apology* 30a7-b4) and caring for virtue (*Apology* 31b5), these were important discoveries, both for Socrates and so many others. Socrates, however, was not alone in pointing out problematic features of the Athenians' positive morality, and he may not have even been the first. Sophocles displays some of the conceptual tensions within the Athenian positive morality in his plays.[18] But Socrates did more than merely expose some

puzzling features of Athenian positive morality. He proved that popular Athenian beliefs about virtue were inconsistent: He derived contradictions from them.

Socrates also posed serious and puzzling theoretical challenges to commonly embraced features of the moral virtues and thereby injected philosophical perplexity about the nature of morality into the Athenian culture.[19] In Plato's *Laches*, a young Socrates investigates the nature of courage with the esteemed Athenian generals Laches and Nicias. Socrates reveals some of the inconsistencies plaguing Laches' conception of courage and poses perplexing philosophical challenges to Nicias' (Socrates-inspired!) conception of courage and its conceptual relationship to virtue in general. Socrates introduces interesting ideas about courage and virtue into the discussion, but the perplexities are never completely resolved. Similar things can be said about the investigation that Socrates leads into the nature of sophrosune (temperance) with Charmides and Critias in Plato's *Charmides*, or Socrates' discussion of piety with Euthyphro. The ability to identify inconsistencies in Athenian positive morality and make them explicitly known to others requires a certain sort of wisdom, or so it seems to me. And so does the ability to pose perplexing philosophical challenges to popularly endorsed conceptions of the moral virtues. The human wisdom required to ground these abilities is attractive and impressive. Socrates likely enjoyed more of this wisdom than any of his peers. And Socrates used this wisdom to inspire his fellow Athenians to engage in sincere philosophical investigation into the nature of virtue itself.

Notice that this kind of wisdom likely played a key role in developing and justifying Socrates' epistemic humility. If the Athenians' conceptions of the virtues were internally inconsistent, or if they were unable to resolve perplexities at the hearts of these conceptions, then surely they lacked knowledge of the true nature of moral virtue and corresponding things *kalos kagathos*.

5. Operating in the Philosophical Dialectic

Socrates developed his elenctic method with the hope that it would provide a route to progress in ethical and social-political inquiry. As far as we know, no one before him investigated moral matters strictly within the confines of the philosophical dialectic,[20] and Socrates' elenctic method was perhaps the earliest, methodologically fruitful version of the philosophical dialectic in action.[21]

Socrates frustrates but also amazes Protagoras with his dialectical prowess throughout Plato's *Protagoras*. Socrates articulates competing theses about the nature of the virtues and the nature of moral failure before submitting each of them to tremendous dialectical pressure. Socrates seems to make some theoretical progress in various respects, but the dialectical discussion with Protagoras leaves them both perplexed (perhaps) about serious issues. Despite his frustration, Protagoras ends the dialogue by heaping praise upon Socrates (as was noted in

section 3). And earlier in the dialogue, Alcibiades pays homage to Socrates' dialectical abilities: "But when it comes to dialectical discussion and understanding, the give and take of argument, I would be surprised if he [Socrates] yields to anyone" (336c2-4). Socrates' dialectical abilities are impressive and philosophically fruitful too.

Socrates realized that the only route to theoretical improvement regarding insight into moral (and social-political) matters is via the (the very "human") elenctic method. Tough, tricky, and creative dialectical work is required for philosophical progress. And theoretical insight into the nature of moral virtue (and perhaps other things *kalos kagathos*) is required for serious practical moral progress,[22] something Socrates craved not only for himself but for his fellow Athenians and visitors as well.

Socrates came to understand that the targets of virtuous action must enjoy some theoretical justification, and Socrates was (justifiably) confident that any principle, account, position, or argument that could survive dialectical refutation held promise and might even come close to approximating the truth, at least in many cases. Socrates displays this confidence on a number of occasions in the early Platonic dialogues. Socrates' confidence regarding the dialectical challenges he poses for Euthyphro's proposals concerning the nature of piety is clear and apparent. So is his confidence regarding the principles he embraces and uses to justify his position that he should endure his execution rather than escape with Crito. Socrates declares to Crito:

> I cannot, now that this fate has come upon me, discard the arguments I used; they seem to me much the same. I value and respect the same principles as before, and if we have no better arguments to bring up at this moment, be sure that I shall not agree with you, not even if the power of the majority were to frighten us with more bogeys, as if we were children, with threats of incarcerations and executions and confiscation of property. (*Crito* b6-c6)

Socrates expresses similar confidence about the prospects of philosophical success in his discussion with Callicles in *Gorgias* about whether it is worse to commit injustice than to suffer it. Excited about the fruitfulness of a dialectical discussion with Callicles about "the topics of greatest importance" (487b5)—"…that of what a man is supposed to be like, and what he's supposed to devote himself to and how far…" (487e9-488a1)—Socrates declares:

> I know well that if you concur with what my soul believes, then that is the very truth. I realize that a person who is going to put a soul to an adequate test to see whether it lives rightly or not must have three qualities, all of which you have: knowledge, good will, and frankness. I run into many

> people who aren't able to test me because they're not wise like you. Others are wise, but they're not willing to tell me the truth, because they don't care for me the way you do. (486e5-487a6)

Socrates continues:

> If there's any point in our discussions on which you agree with me, then that point will have been adequately put to the test by you and me, and it will not be necessary to put it to any further test, for you'd never have conceded the point through lack of wisdom or excess of shame, and you wouldn't do so by lying to me, either. You are my friend, as you yourself say, too. So, our mutual agreement will really lay hold of truth in the end. (487e1-7)

Socrates even celebrates their mutual success regarding the respective shamefulness of committing injustice and suffering it: "These conclusions, at which we arrived earlier in our previous discussions are, I'd say, held down and bound by arguments of iron and adamantine, even if it's rather rude to say so." (508e6-9).[23] Socrates' dialectical prowess is indisputable. His trust in the dialectical method as the only way to claw towards philosophical truths marks another aspect of Socrates' (perhaps) peerless human wisdom.

This aspect of Socrates' wisdom too is related to his epistemic humility in an intimate way. Recognizing that the only path to philosophical insight is via the dialectic, and realizing how difficult it is to secure anything with certainty via his dialectical method, Socrates came to know how much philosophical truth there was out there yet to be known and how tremendously difficult it would be to grasp those truths with sufficient dialectical justification—a divine task to be sure, perhaps beyond the limited intellectual powers of even the strongest human thinkers, including the likes of Socrates himself. A god-like being, on the other hand, might have direct intellectual insight into the complete structure of the nature of morality, resulting in systematic, comprehensive, certain, and infallible knowledge of the moral realm.[24] Such a being would be able to provide a "full and successful dialectical account"[25] of the nature of moral virtue and all things *kalos kagathos*. Socrates realized that such knowledge was beyond him and perhaps beyond any human being, so he continued to inquire, testing and examining any potentially attractive principle, position, or argument that came his way. Socrates' epistemic humility was hard won and well-earned.[26] This too distinguishes him from most of his peers, if not all of them. Socrates realized that whatever philosophical progress he made was paltry in comparison to the divine philosophical knowledge he craved. His dialectical efforts frequently enjoyed limited successes, but those successes were incomplete in radical ways: Unresolved perplexities lurked everywhere Socrates ventured, and it was Socrates

himself who brought many of those perplexities to light.[27]

6. Character Grounded in Philosophical Commitments

Socrates was an exemplar of someone who leads a philosophical life. He was completely committed to whatever philosophical morsels he discovered or latched onto, that is, whatever resisted refutation in the philosophical dialectic. Socrates committed himself completely to whatever (at least seemingly) theoretically viable features of virtue he could grasp.[28] His behavior never wandered from his philosophical commitments:[29]

> So I disregard the things held in honor by the majority of people, and by practicing truth I really try, to the best of my ability, to be and to live as a very good man, and when I die, to die like that. And I call on all other people as well, as far as I can—and you especially I call on in response to your call—to this way of life, this contest, that I hold to be worth all the other contests in this life. (*Gorgias* 526d4-e4)

Socrates' virtue—or at least the degree to which he exemplified it—was grounded in his philosophical commitments, commitments developed through continual, critical inquiry in the philosophical dialectic. Socrates was a pioneer in this. And this too reflects a sort of wisdom that Socrates enjoyed in ways that his peers did not.

7. Conclusion

I have tried to lay out a strong case in defense of the Pythian's reply that no one was wiser than Socrates. Many philosophers have attempted to articulate Socrates' superlative yet human wisdom solely in terms of epistemic humility, especially about things *kalos kagathos*:

> EH: Socrates does not think he knows things that he does not know. Socrates claims that he does not know anything *kalos kagathos* nor does he think he knows anything *kalos kagathos*.

I readily admit that Socrates was probably peerless when it came to being epistemically humble about philosophical matters. But restricting Socrates' wisdom to just his epistemic humility does him a serious injustice. Socrates was able to identify inconsistencies in the Athenian positive morality and reveal perplexities emerging from philosophical investigations into the moral virtues. It was Socrates who developed the elenctic-dialectical method and recognized that the only path to philosophical (and subsequently moral) progress was via

strenuous, creative, and continual dialectical work—and Socrates worked hard. Sure, these features are intimately related to his epistemic humility; in fact, they likely ground his epistemic humility in a philosophically attractive way. But he was probably unrivaled with respect to these features too. Combining these features with his commitment to become as virtuous as he could while realizing that moral virtue must ultimately be grounded in defensible philosophical commitments makes for a tremendously attractive person, a person who (among his peers) was likely unsurpassed with respect to wisdom, and beautiful forms of wisdom at that. Was Socrates the wisest man in Athens? Probably.

NOTES

1. Unless otherwise noted, translations of the Platonic passages are those provided in Cooper.

2. Hereafter, I will use *Socrates* to refer to Plato's Socrates.

3. Scholars translate this important phrase *kalos kagathos* in a variety of similar ways. Here is a sample: "fine and good" (Harold North Fowler in Plato 1914), "beautiful and good", "admirable and good" (both provided by Scott Senn in Plato 2019), and "noble and good" (Thomas G. West and Grace Starry West in Plato 1984).

4. Here is how Senn translates the passage in Plato 2019: "Who is knowledgeable about the virtue of that sort: about the human's and citizens's virtue?" Hereafter I will refer to Senn's translation of Plato merely as *Senn's translation*.

5. Richard Bett points out that in many of the early aporetic dialogues dedicated to investigations of various virtues and important moral concepts—*Hippias Major*, *Euthyphro*, *Charmides*, *Laches*, *Lysis*, and (the transitional dialogue) *Meno*—Socrates professes ignorance regarding the targets of investigation in these dialogues and "this professed ignorance is not dispelled in the course of the dialogue[s]" (222). Bett also points out that Socrates intimates on several occasions how exceptionally valuable such knowledge would be if one could obtain it. Bett continues on, offering an interpretation roughly the same as my own: "Taken together, these two points suggest that knowledge of the nature of the virtues and other related qualities is, or is at least a prime example of, the wisdom or truly valuable knowledge that Socrates disclaims in the *Apology*." 222

6. I will have a bit more to say in favor of this interpretation when I discuss Socrates' examination of the craftsmen.

7. Cf. Benson 170 and Bett 219.

8. Cf. with the trilogy of early Socratic dialogues *Charmides*, *Laches*, and *Lysis*. The natures of sophrosune (and any conceptual relationships it may have with self-knowledge), courage, virtue, and friendship are investigated. Each investigation ends in perplexity. Even if some theoretical progress is made in these dialogues, knowledge of the concepts under investigation remains elusive, or so it seems. The investigations in *Laches* seem particularly apt when thinking about this. Suppose that after investigating the natures of courage and virtue with Socrates, the generals Laches and Nicias admit that they do not know the nature of courage, nor do they know the nature of virtue in general. In doing so, they too would be just as wise as Socrates on this EH account. But it is doubtful that they

would be. That is why they are eager (Nicias) or willing (Laches) to investigate these matters with a young Socrates.

9. Note, however, that the playwright Sophocles was onto conceptual tensions within the Athenian concept of arete. I touch on this issue below.

10. Senn's translation.

11. Senn's translation: "because of beautifully working out his art".

12. Emphasis is mine. Cf. Senn's translation; again the emphasis is mine: "…each deemed himself to be wisest in the other things too—the *greatest things*."

13. This interpretation is bolstered by attending to the contexts in which Socrates uses the phrases *most important pursuits*, *greatest things*, and *topics of the greatest importance* in Plato's *Gorgias*; see especially 487b5 and 527e1. Also see *Gorgias* 487e-488a, where Socrates expands upon his conception of the "topics of greatest importance" in terms of "that of what a man is supposed to be like, and of what he's supposed to devote himself to and how far…." (487e9-488a1). At *Gorgias* 500c3, Socrates claims that his discussion with Callicles is about "the way we're supposed to live". My interpretation is also consonant with Socrates' renowned declarations at *Apology* 38a.

14. Also see *Gorgias* 452a-c, where Socrates has a doctor, a physical trainer, and a financial expert each claim of his craft that it is "concerned with the greatest good for humankind".

15. Thanks to Mark Walker for pushing me in this direction.

16. Book II, Chapter 13 of *Lives of Eminent Philosophers.*

17. Plato sometimes singles out Protagoras as particularly admirable among the sophists; see, for example, *Meno* 91d2-92a1.

18. In *Antigone* (first performed when Socrates was about thirty years old), poor Clytemnestra is torn between what is morally due to her polis and what is morally due to her family—seemingly competing and incompatible demands—resulting in her tragic, suicidal demise. In *Philoctetes*, Odysseus and Neoptolemus dispute whether Odysseus' plan to deceive Philoctetes constitutes honorable or dishonorable action, especially given that Philoctetes receives Odysseus and Neoptolemus with trust and honesty. No resolution of the dispute is provided; rather, the demi-god Heracles steps in as peacemaker. Alasadair MacIntyre discusses these cases in his chapter "The Virtues in Athens" in *After Virtue*.

19. See Gareth B. Matthew's excellent *Socratic Perplexity and the Nature of Philosophy*.

20. See Guthrie 138.

21. Perhaps Parmenides and Zeno should be credited for being the first philosophers to employ fruitful versions of the philosophical dialectic. Empedocles and Anaxagoras might have employed interesting dialectical strategies too. Even so, none of these Presocratic philosophers dedicated themselves to inquiry into moral matters.

22. Cf. Senn 83, who writes: "… he [Socrates] values philosophizing as a necessary means for getting genuine virtue/wisdom."

23. Near the end of *Gorgias*, Socrates declares:

> But among so many arguments this one alone survives refutation and remains steady: that doing what's unjust is more to be guarded against than suffering it, and that it's not *seeming* to be good but *being* good that a man should take care of more than anything, both in his public and his private life…. (527b3-7)

24. Cf. Cooper, who writes: "Wisdom, then is a permanent, deeply settled, complete grasp of the total truth about human values of all sorts, in all their systematic relationships, primed for ready application to all situations and circumstances of human life." (Cooper, *Pursuits of Wisdom* 46) Cooper continues:

> For Socrates, then, knowledge is the grasp of the truth of some fact or group of facts on the basis of a comprehensive, complete understanding of the whole system of facts, and relationships among them, that constitute some distinct area of intellectual inquiry. In the case of the human good and what is bad for human beings, such knowledge is a grasp of what to do in particular sets of circumstances, where that grasp derives from and depends upon a complete understanding of the whole realm of human values. Wisdom goes beyond that knowledge by requiring that, once acquired, it be so deeply and firmly settled in one's mind that one would be prepared, for all future time, when in normal possession of one's powers, to apply that knowledge, with confidence and demonstrable authority, in any and every circumstance, so as always to do what is right and best, with a complete and fully grounded justification in mind for what one does. (48)

25. Cooper, "Arcesilaus" 182, fn. 24.
26. Thanks to Hal Thorsrud for pressing this point.
27. Cf. Vlastos, who writes:

> [Socratic] knowledge is full of gaps, unanswered questions; it is surrounded and invaded by unresolved perplexity. But this does not trouble Socrates. He does not find it debilitating, but exhilarating…. So if an inquiry should run into *aporia*, he can reckon the exercise not total failure but incomplete success. Nothing has transpired to show that the unfound answer is unfindable, nor yet to invalidate the fragmentary truths unearthed along the way and shake his claim that in their case he does have knowledge$_E$. (19-20)

28. See, for example, *Crito* b6-c6 (quoted above).
29. See especially *Apology* and *Crito* but *Laches* too.

WORKS CITED

Benson, H. *Socratic Wisdom: The Model of Knowledge in Plato's Early Dialogues*. Oxford University Press, 2000.

Bett, Richard. "Socratic Ignorance." *The Cambridge Companion to Socrates.* Edited by Donald R. Morrison, Cambridge University Press, 2011, pp. 215-236.

Cooper, John. "Arcesilaus: Socratic and Sceptic." *Remembering Socrates.* Edited by Lindsay Judson and Vassilis Karasmanis, Oxford University Press, 2006, pp. 169-187.

———. *Pursuits of Wisdom*. Princeton University Press, 2012.

Guthrie, W. K. C. *Socrates*. Cambridge University Press, 1971.

Laertius, Diogenes. *Lives of Eminent Philosophers*. Translated by R.D. Hicks, Harvard University Press, 1972. (First published 1925).

MacIntyre, Alasdair. *After Virtue*. University of Notre Dame Press, 1981.
Matthews, Gareth. *Socratic Perplexity*. New York: Oxford University Press, 1999.
Plato. (1914) *Euthyphro. Apology. Crito. Phaedo. Phaedrus*. Trans. Harold North Fowler. Loeb Classical Library vol. 36. Cambridge, MA: Harvard University Press.
Plato. *Four Texts on Socrates*. Translated by Thomas G. West and Grace Starry West, Cornell University Press, 1984.
———. *Plato: Complete Works*. Edited by John M. Cooper, Hackett Publishing Company, 1997.
———. *The Good, the Just, & the Happy in Four Platonic Dialogues: Socrates' Defense Speech, the Euthyphro, the Crito, & the Meno*. Translated by Scott J. Senn. *Available at http://ssrn.com/abstract=2727511.* (2019)
Senn, Scott. "Ignorance or Irony in Plato's Socrates? A Look Beyond Avowals and Disavowals of Knowledge." *Plato Journal,* vol. 13, 2013, pp. 77-108.
Vlastos, G. "Socrates' Disavowal of Knowledge." *Socratic Studies.* Edited by M. Burnyeat, Cambridge University Press, 1994, pp. 39-66. (Original version published in 1985).

.

The Equal Opportunity to Be a Legislator: Why Randomly Selecting Legislators Is More Democratic Than Electing Them

Eric Shoemaker
Lycoming College

Winner of the Glenn Joy Award

An exciting and recent development in the history of democracy is the use of mini-publics for political decision-making. Mini-publics are assemblies composed of randomly selected citizens and convened for the purpose of arriving at a collective decision through deliberation. Most prominently, mini-publics have been used in Ontario and British Columbia to arrive at proposals for electoral reform which were subsequently put to popular referendums,[1] and in Ireland a mini-public has been convened to deliberate about the repeal of the country's constitutional ban on abortion, which was also subsequently put to a referendum.[2] Additionally, mini-publics have been employed in Belgium, Denmark, France, the Netherlands, Poland, and the US state of Oregon for various public purposes. These applications have been precipitated by the great deal of positive attention which mini-publics have received from political scientists over the past thirty years.[3]

Much has already been said about the epistemic virtues of mini-public deliberation as a method of decision-making. These bodies are cognitively diverse and genuinely deliberative, which are both important for quality group decision-making.[4] These bodies are maximally inclusive in that they resemble the whole public in miniature, and as a result of this these bodies impartially aim at the whole public's interest.[5] These bodies are comparatively incorruptible, with their limited tenure and random selection making them very resistant to illegitimate influences on their decision-making.[6] These epistemic virtues are particularly impressive when compared to those possessed by elected legislative bodies, which do not deliberate

within the legislature and are not very cognitively diverse, do not tend to be very inclusive or impartial, and are comparatively far more susceptible to corruption and illegitimate influence.[7]

As a result of this, several proposals have been created for the replacement of elected legislatures with legislatures composed of randomly selected citizens.[8] The most promising of these proposals, such as Alexander Guerrero's and Terril Bouricius', do away with the generalist nature of elected legislatures and convene many mini-publics, some for setting the legislative agenda and other tasks, and some single-issue legislatures which deliberate and vote about one issue and then disband.

At present, there is an unfortunate disconnect between advocates for political random selection and their detractors. Advocates for political random selection argue persuasively about the epistemic merits of this method of decision-making, and the instrumental value which their preferred procedures have for producing good laws. The most compelling critics of political random selection argue that legislation by random citizens is undemocratic, either because randomly selected representatives are not authorized by the people they represent,[9] or because it makes the opinions of the people not sitting in the randomly selected body irrelevant to the legislative process.[10] Accounts of political random selection's value which give only instrumental reasons to prefer political random selection lack the resources to adequately respond to these objections because they have no force whatsoever in a purely instrumentalist framework.

To respond to its most persuasive critics, political random selection first needs a defense which is grounded in a non-instrumental conception of democracy's value. In this paper I shall demonstrate that on an equality-based conception of democracy's value, political random selection is more democratic than election for selecting legislators.

What is An Equality-Based Theory Of Democracy's Value?

When I say an equality-based theory of democracy's value, I have in mind the sorts of theories put forward by Thomas Christiano, Niko Kolodny,[11] Jaques Rancière, Charles Bietz, and some other democratic theorists.[12] Equality-based theories, while quite diverse in many respects, share some central commitments and themes. Each is committed to these structural principles: (1) it is normatively required that citizens be (in some respect) treated equally by the state, and this puts constraints on what types of institutions and government conduct are normatively acceptable, and (2) these constraints require that states be governed democratically.

These ideas about the nature and requirements of democratic equality are central to many equality-based theories: (1) democratic equality requires that the lawmaking procedure produce compelling justifications for the law that answer to each citizen's interests. (2) Democratic equality requires that decisions be made in a non-hierarchical way that gives each equal standing to contest the law & the authority of

their rulers. (3) Democratic equality requires everyone's judgements about what the law ought to be to be treated equally through the lawmaking procedure. These points are highly interrelated. Each of these three requirements on their own require the process of legislating to be democratic, according to the theorists who posit them.

WHY JUSTIFICATIONS FOR ELECTION BASED ON DEMOCRATIC EQUALITY FAIL

Christiano and Kolodny both offer up justifications for election based in equal treatment. Both are unsuccessful in ways that are instructive for understanding why equality-based theories, as a rule, will fail to justify election.

Per Christiano, in an idealized version of an elective democracy citizens set the *ends* which the elected legislature ought to pursue through the election process, and once in office elected legislators deliberate about the *means* for the achievement of those ends, and enact them. Elected legislators have a responsibility to respect the judgement of their constituents about the appropriate ends to pursue, and it would be a breech of their duty to their constituents to deliberate about these ends and change their mind about them. (Christiano 169-70) Conversely, if an elected representative's judgement about the appropriate means is at odds with the constituents' conceptions, then those constituents have no complaint against their representative (Christiano 170).

Christiano's structure anticipates the worry that political equality is violated by having elected representatives count for far more in the legislative process than ordinary citizens. Per Christiano, legislators qua legislators have no role in that process at all, and they are duty-bound not to use their position of power in ways which undermine the equal voice of all citizens in the process of setting ends.

The main issue with this proposal relates to the role of equality in Christiano's argument. Christiano has not adequately explained why equality entitles citizens to have input in the process of determining who shall sit in the legislature. His process requires that citizens have equal input into the determination of ends, however it is fine for the elected politicians to have a greater influence on the decision about the means than those who elect them—this is why they are expected to deliberate amongst themselves about the means. If all political equality requires is equality in the determination of the ends pursued by legislators, then the actual legislators could just as well be appointed by technocrats or a monarch. Christiano would have to show that the best way to discover the legislators most capable of realizing the public's ends is to have citizens identify them via election. Because these legislators are only permitted to deliberate about means they are not importantly different from the kinds of policy-making bureaucrats which real-world elected legislatures normally delegate regulative tasks to. These bureaucrats do not need to be selected through a democratic process and are not political representatives of the public—we should think the same is true of Christiano's elected legislators.

Christiano's argument establishes that the equality of persons requires equal influence for each citizen over the ends pursued through public policy, but it has failed to establish that they must have an equal influence over who shall sit in the legislature, and thereby has failed to justify election. This is illustrative of a broader struggle which equality-based theories will have with justifying the institution of election: elected officials wield special powers over political processes which ordinary citizens do not, and without a special justification that cannot be provided for election, this is at odds with political equality.

In "Rule Over None II" Kolodny gives a very straightforward argument for election which, unfortunately, is directly contradicted by a far more compelling argument he gives in "Rule Over None I". For Kolodny, democratic governance is valuable because it is an integral component of a society that does not have social inequality, both because it does not involve putting some people in relations of social subordination to others, and because all other social hierarchy is moderated when an egalitarian government can limit and regulate it (Kolodny, "Rule Over None II" 303-7).

Per Kolodny, in order for a government to be democratic, citizens need to have an equal opportunity to influence political decisions—if the opportunities were unequal then that would establish a social hierarchy between citizens. Kolodny, "Rule Over None II" 308-10). Per Kolodny, elections ensure this. However, while under a reasonably idealized election-based democracy, people do have the equal opportunity to vote in elections, it is evidently untrue that people have an equal opportunity to achieve political office, in Kolodny's sense. In fact, under election the equal opportunity to influence political decisions is essentially impossible, because voters vote on the basis of judgements about the merit of candidates. One might reasonably object at this point that with a conventional understanding of the meaning of equal opportunity, as long as our society is constituted such that people have the equal opportunity to achieve the kind of merit which is the basis of the voter's decisions about who to elect, then all citizens do indeed have an equal opportunity to be elected. However, if this was all that political equal opportunity required, then we could, without any harm to democracy, apply broad restrictions on who is permitted to cast a vote. If we lived in an egalitarian society where everyone had the equal opportunity to become a tenured professor, and then restricted the voting franchise to tenured professors, it would still be the case that everyone had an equal opportunity to vote.[13] This is an absurd consequence.

Kolodny also offers an argument that elected political representatives are not socially superior to those who do not have political office, which might save his view. Kolodny's explanation is that the relationship of the citizenry to their representatives is one of delegation. This delegation does not create a social hierarchy with political representatives above their constituents because: representatives exercise powers which they acknowledge belong to "the principal" (that which they represent), on behalf of their interests, under the direction of their

expressed will, and the representative can be replaced if they displease the principal (Kolodny, "Rule Over None II" 317-20).

I agree that, in a reasonably ideal elective democracy, as a representative relates to their principal, they can be a social equal. However the principal of an elected representative is not any one of their constituents. I, as a constituent of my member of parliament, do not have the power to remove my representative if her conduct displeases me, nor does she act on my direction or always for the sake of my interests, nor does she exercise powers which she acknowledges belong to me (the power to vote in parliament doesn't belong to me!). Instead, the principal of a representative is their whole constituency. This should make an important difference to Kolodny, as in "Rule Over None I: What Justifies Democracy?" he objects to the notion that citizens can be collectively in control of the government in a way that we think should matter to individual citizens (Kolodny, "Rule over None I" 209). My representative enjoys a degree of power over political decision making in my country far greater than my own. Again, from this we can see the difficulty equality-based theories have in justifying the unequal access to political office citizens have given their equal access to the right to vote.

Throughout the history of political theory, until relatively recently, elections have been considered to be an aristocratic method of selection, rather than a democratic one. As Aristotle famously said, "[i]t is accepted as democratic when public offices are allocated by lot; and as oligarchic when they are filled by election" (Aristotle 4.1294be). Expressing the same sentiment in a far more modern context, when Rosseau advocated for the election of executive officers he referred to this as an aristocratic element of his form of government (Rosseau 35). The reason for this is that elections select office holders on the basis of their meritorious distinction from and superiority to ordinary citizens, rather than their equivalence to them. The modern impulse in liberal political theory to re-conceptualize elected representatives as one and the same as "the people" and therefor their rule as the rule of "the people" represents a muting of our democratic ambitions. Equality-based theories fail to justify election because elections do not treat citizens as equals.

Democratic Equality And Political Random Selection

In contrast to election, political random selection is the natural fit for equality-based theories. Let's now reconsider the three important ideas about democratic equality which I had earlier mentioned, and compare political random selection to election in terms of the standards set by these ideas.

First, democratic equality requires that the lawmaking procedure produce compelling justifications for the law that answer to each citizen's interests. Here, an approximately ideally constituted elective democracy succeeds. If elections were capable of disciplining politicians into producing compelling justifications for their decisions to their constituents on pain of removal from office (which perhaps they

can be), then the process of legislation produces compelling justifications in terms of each citizens' interests. Political random selection would do so as well, although by a more direct mechanism. While elected representatives are cajoled with institutional carrots and sticks into providing public justifications for their decisions, randomly selected representatives reason based on public interest through the process of deliberation within the legislature. Each randomly selected representative has self-interested reasons to prefer some policies to others which correspond to the self-interested reasons of some fraction of the public, and the whole randomly selected legislature in aggregate has self-interested reasons which correspond to the self-interested reasons of the whole public, because they are a representative sample of the whole public (Stone 138-41). Furthermore, the process of deliberation requires citizens to give reasons that appeal not to their narrow private interest, but to the interests shared by the whole public.[14] The result of this deliberation is that the policies which are adopted will be supported by the reasons given through the deliberation, and these can be publicized. Furthermore, the composition of the legislature and the process of its decision-making itself provides ordinary citizens with compelling reasons to believe that the decisions it arrives at are in the public interest.

Second, democratic equality requires that decisions be made in a non-hierarchical way which gives each citizen equal standing to contest the law & the authority of their rulers. As we have seen, this is one area in which elections are doomed to fail. With an elected legislature, decisions are made in a way which does not give each citizen equal standing to contest the law. By contrast, the power of randomly selected legislators is so circumscribed that they can hardly be considered rulers. When randomly selected and put into a single-issue legislature, they consider one issue, and then disband.[15] They have legislative power, but it is impossible to wield that power in a way that differentially affects the citizens and the body, for as soon as the body's members have altered the law they are no longer legislators. Notably the feature of political random selection which is doing the work here is that those randomly selected sit in single-issue legislatures. Randomly selecting a dictator for life would pose a problem for democratic equality. However, it would be impracticable to combine a single-issue legislative structure with election.

As for equal standing to contest the law, each citizen has an equal chance of sitting in the legislature, which is more than can be said for an elective system. In comparison to some form of direct democracy, it is not at all obvious that an equal chance of effecting the outcome is any less good than an equal piece of the power to affect the outcome. We should not regard an equal chance to sit in the legislature as a deficient substitute for an equal vote in a referendum.[16]

Third, democratic equality requires that everyone's judgements about what the law ought to be, be treated equally through the lawmaking procedure. Election, because of its foundational purposes and aims, falls short of the ideal. Elections select candidates on the basis of their merit as legislators. A process which begins

with the conception of some people as better suited to legislate than others, and then endows those people with powers over the law not possessed by their fellow citizens on that basis, does not treat the judgements of all citizens about what the law ought to be equally. It may treat the judgement of all citizens on the question of who is best suited to be a legislator equally, but this is an essentially distinct question with far less relevance to democratic equality. With political random selection, the judgement of each citizen is treated equally in the legislative process not only vicariously by the fact that each citizen is treated as equally worthy of wielding legislative power, but directly by the fact that, because the legislature constitutes a mini-public, we should expect that the perspectives, knowledge, style of reasoning, and any other constituent components of judgement that one might conceive of, are represented as well in the randomly selected body as it would be possible to represent them in a body of that size.

Democratic equality requires at least that all citizens be equal in the legislative process. Election, which evaluates citizens on their merits, and empowers some citizens to make the law on the basis of merit, cannot treat all citizens as equals in the legislative process. When political random selection empowers a citizen, it does so on the basis of their equal citizenship only. For this reason, political random selection better realizes the democratic ideals emphasized by equality-based theories of democracy's value.

NOTES

1. See Warren and Pearse 9-16, for a discussion of these assemblies.
2. See Farrell et al for discussion of this assembly.
3. One contribution to this research of particular notoriety is the deliberative opinion polls invented and subsequently conducted by James Fishkin, which bring together a random and representative sample of citizens to deliberate about a particular policy issue, with informational input into the deliberation from experts, and culminating in a poll of the group's post-deliberation views on the topic. See Fishkin and Luskin.
4. See Landemore, chapter 4.
5. See Fishkin and Luskin.
6. See Stone, chapter 6.
7. See Guerrero.
8. For such proposals see Guererro, Gastil and Writht, and Bouricius.
9. Most examples of arguments of this type are made rather off-handedly. For a thoroughly developed account, see Dahl 122-5. Pettit also develops an account of this criticism rather thoroughly in "Representation, responsive and indicative."
10. For one very well developed objection of this type, see Lafont.
11. See Kolodny "Rule Over None I", and Kolodny "Rule Over None II".
12. Many other democratic theories arguably fall into this camp, or have elements embedded into their democratic theories which commit them to some of the positions which I describe below. Of these, Dahl and Fishkin are particularly notable here.

13. Here one might think to object that I am here equivocating between the equal opportunity for enfranchisement and the equal opportunity to vote. I can see why this might immediately appear suspicious, however the distinction makes no difference here. If we object to this inference on the basis that, although in my hypothetical everyone would have an equal opportunity to become a tenured professor (by stipulation) it would not be the case that everyone had an equal opportunity to vote because in fact not everyone is a tenured professor, we might similarly object to the claim that even if everyone had an equal opportunity to earn a million dollars not everyone has the opportunity to spend that million dollars on a yacht because in fact not everyone has a million dollars. The conventional understanding of equal opportunity requires that equal opportunity to obtain some means is the same thing as equal opportunity to achieve some end which those means could be used to bring about.

14. This is an important descriptive component of what deliberation is, but also an ideal which can be promoted by facilitators of deliberation through a variety of means. See Landemore, chapter 4 and 6. See Goodin and Spiekermann 132-145. See Carson.

15. That issue being either the agenda for the other single-issue legislatures, or the issue given to them by the agenda-setting assembly.

16. It is true that when distributing a cake, the appropriate distributive principle is equal slices, whereas when distributing a kidney to equally worthy recipients, the appropriate distributive principle is an equal chance. Distributing legislative power is more like distributing the kidney because there are important practical reasons to prefer a smaller legislature, but even putting those practical reasons aside distributing votes is unlike distributing cake because people value them for different reasons. A cake is pleasant to eat, whereas a vote is only valuable for the sake of bringing about one's desired outcome in the legislature. If we allowed everyone to vote on an issue, but instead of tallying the votes drew one vote out randomly and made our decision according to it, it is not obvious that the voters would be treated unfairly, unequally, or that they in any sense had different levels of power over the outcome (although there may be other reasons to prefer the tallying method).

WORKS CITED

Aristotle. *Aristotle's Politics*. Clarendon Press, 1905.

Beitz, Charles R. *Political Equality: An Essay in Democratic Theory*. Princeton University Press, 1989.

Bouricius, Terrill G. "Democracy Through Multi-Body Sortition: Athenian Lessons for the Modern Day." *Journal of Public Deliberation,* vol. 9, no. 1, 2013.

Carson, Lyn. "How to Ensure Deliberation Within a Sortition Chamber." *Legislature by Lot: Transformative Designs for Deliberative Governance*, edited by John Gastil and Erik Olin Wright, Verso Books, 2019.

Christiano, Thomas. *The Rule of the Many: Fundamental Issues in Democratic Theory*. Routledge, 2018.

Dahl, Robert Alan. *After the Revolution?: Authority in a Good Society*. Yale University Press, 1990.

Farrell, David, et al. "When Mini-Publics and Maxi-Publics Coincide: Ireland's National Debate on Abortion." *Journal of Representative Democracy,* vol. 55, no. 1, 2019, pp. 1-19.

Fishkin, James, and Robert Luskin. "Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion." *Acta Politica,* vol. 40, no. 3, 2005, pp. 284-298.

Gastil, John, and Erik Olin Wright. *Legislature by Lot: Transformative Designs for Deliberative Governance*. Verso Books, 2019

Goodin, Robert E., and Kai Spiekermann. *An Epistemic Theory of Democracy.* Oxford University Press, 2018.

Guerrero, Alexander. "Against Elections: The Lottocratic Alternative." *Philosophy and Public Affairs,* vol. 42, no. 2, 2014, pp. 135-178.

Kolodny, Niko. "Rule Over None I: What Justifies Democracy?" *Philosophy and Public Affairs,* vol. 42, no. 3, 2014, pp. 195-229.

———. "Rule Over None II: Social Equality and the Justification of Democracy." *Philosophy and Public Affairs,* vol. 42, no. 4, 2014, pp. 287-336.

Lafont, Cristina. "Deliberation, Participation, and Democratic Legitimacy: Should Deliberative Mini-Publics Shape Public Policy?" *Journal of Political Philosophy,* vol. 23, no. 1, 2015, pp. 40-63.

Landemore, Hélène. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press, 2017.

Pettit, Philip N. "Representation, Responsive and Indicative." *Constellations,* vol. 17, 2010, pp. 426-34.

Rancière, Jacques. *Hatred of Democracy*. Verso Trade, 2014.

Rose, Jonathan. "The Ontario Citizens' Assembly on Electoral Reform." *Canadian Parliamentary Review,* vol. 30, no. 3, 2007, pp. 9-16.

Rosseau, Jean-Jaques. *The Social Contract.* Early Modern Texts, 2017. Available at https://www.earlymoderntexts.com/assets/pdfs/rousseau1762.pdf. Accessed 12 July 2021.

Stone, Peter. *The Luck of the Draw: The Role of Lotteries in Decision Making*. Oxford University Press, 2011.

Warren, Mark and Hilary Pearse. *Designing Deliberative Democracy: The British Columbia Citizens' Assembly.* Cambridge University Press, 2008.

Ectogenesis and Misogyny

Christopher Stratman
University of Nebraska—Lincoln
Southeast Community College

Winner of the Glenn Joy Award

Ectogenesis is an emerging technology that would *inter alia* conceivably allow physicians to transfer a fetus from a human womb to an artificial womb-like environment where gestation continues until it reaches full-term.[1] In recent years, philosophers have wondered whether this technology might provide a direct solution to the problem of abortion.[2] On one approach, the possibility of ectogenesis shows that there is no "moral" right to the death of a fetus, since it could be safely removed from the human womb and gestated in an *ex utero* environment.[3] But even if this is true, it still might be the case that societies should not implement *new* legal restrictions and prohibitions on one's ability to safely obtain a lethal termination of their pregnancy.[4]

Why is this a problem? The significance of this issue can be better appreciated by considering what may seem to be an inconsistent triad of independently plausible statements:

> (A) NO VIOLATON OF AUTONOMY: Given the possibility of ectogenesis, a fetus can be removed from a human womb without violating the pregnant person's right to autonomy or killing it.
>
> (B) NO MORAL RIGHT TO FETAL DEATH: Given the possibility of ectogenesis, there is no "moral" right to the death of the fetus.

> (C) NO NEW LEGAL RESTRICTIONS: The advent of ectogenesis should not lead to the implementation of new legal restrictions on a pregnant person's ability to obtain a legal and safe lethal abortion.

The goal of this paper is to show that these statements are not jointly inconsistent—that is, I shall develop an argument that purports to show that (C) is true, which does not hinge on the negation of either (A) or (B).

In the first part of the paper, I shall explain why the 3 propositions (A), (B), and (C) are independently plausible, and why the possibility of ectogenesis suggests that they are jointly inconsistent. I then go on to argue that these propositions are not inconsistent, and that the reason why they appear inconsistent is because the problem of abortion has been traditionally treated as a conflict between competing rights—the pregnant person's right to bodily autonomy versus the fetus's right to life.[5] But, once we abandon this approach to the problem of abortion, and reject (C) on grounds that abortion laws are misogynistic, then we are free to adopt a new approach to the problem of abortion, which construes it in terms of a psychological struggle between conflicting motivations, goals, desires, and expectations that a human agent experiences as a result of an unwanted pregnancy. In the final part of the paper, I shall consider how the legal questions associated with the possibility of ectogenesis reveal just how deeply one's psychological *reasons* rather than "moral" *rights* are rooted in and cannot be divorced from the social context and systems in which they are embedded.

1

Why should we accept (A)? First, the possibility of ectogenesis shows that the term "abortion" is ambiguous: It could mean *the lethal termination of one's pregnancy*, or *the non-lethal termination of one's pregnancy*. Of course, we tend to assume that an abortion just is the killing of a fetus but this assumption is false. The notion of abortion does not conceptually or logically entail the death of the fetus.[6] This raises an important moral question: Given the possibility of ectogenesis, does a pregnant person's "moral" right to terminate their pregnancy entail that there is also a "moral" right to the death of the fetus?

According to Mathison and Davis, ectogenesis presents the problem of abortion from a novel perspective because the debate will no longer depend on whether a pregnant person's right to bodily autonomy outweighs a fetus' alleged right to life (314).[7] Here is how Mathison and Davis describe this point:

> Some theorist have argued that this technology, especially if made readily available and cost-effective, would essentially conclude the abortion debate, since the two rights commonly thought to be in tension could be jointly exercised: pregnant women will be able to exercise their right of

> autonomy by terminating the pregnancy, and the foetus can be carried to term, thus not being denied its right to life. (314)

To get a better grip on this point, suppose that future medical technology has advanced to the point where fetal transfer surgery is safe and low risk with minor inconveniences to the pregnant person. And suppose that the relevant culture and society is willing and sufficiently capable of supporting hundreds of thousands of fetuses as they develop and grow into children and adults, either in the form of adoptive parents, foster-parents, or a governmental foster-care system that is sufficiently superior to our current system.[8] Given these stipulations, the mere possibility of ectogenesis would clearly be a safe alternative to a lethal abortion.[9] But would this violate a pregnant person's "moral" right to bodily autonomy? To ask the question a bit differently: If a pregnant person were forced to choose a non-lethal form of terminating their pregnancy, would this limit their medical choices so as to count as a violation of their right to bodily autonomy?

To my mind, how we answer these questions depends on whether an abortion just is the termination of a pregnancy, which typically results in the death of the fetus but is not necessarily tied to the death of the fetus. Or whether the death of the fetus is in some philosophically important sense logically or conceptually connected to the termination of one's pregnancy. If an abortion is conceptually distinct from the death of the fetus, then it is not clear why ectogenesis would violate a pregnant person's right to bodily autonomy. After all, ectogenesis does not prevent the pregnant person from choosing to end the pregnancy. The relevant choice, given the possibility of ectogenesis, is not between terminating the fetus or not; it is between terminating the pregnancy or not. And if this is correct, then to justify the claim that a pregnant person's right to autonomy means that they have a right to the death of the fetus, one would first need to show that the relevant choice is not merely a choice to terminate the pregnancy, but a choice to obtain the death of the fetus. But this presupposes that there is a right to the death of the fetus. So, only once one has established that they have a right to the death of the fetus could they reasonably argue that ectogenesis would limit their choices.[10]

Why should we accept (B)? If the fetus is a person, then (B) would immediately follow, for the same reasons that Thomson argued that if the violinist were to survive being unplugged, then you would not have a right to secure his death (66).[11] Presumably, most will grant this much or attempt to remain neutral regarding the moral status of the fetus.[12] But even if we grant that the fetus is not a person, it still might be the case that there is no "moral" right to its death.[13] For instance, Warren argues:

> Indeed, if and when a late-term abortion could be safely performed without killing the fetus, she would have no absolute right to insist on its death (e.g., if others wish to adopt it or to pay for its care), for the same reason

> that she does not have a right to insist that a viable infant be killed. (118)

And if we assume that fetal viability is a good indication of the moral permissibility of a lethal abortion even ignoring fetal personhood, and that ectogenesis shifts the viability of the fetus to be much earlier during gestation, there would be no "moral" right to obtain the death of the fetus once it can be safely removed from the human womb and gestated in an artificial womb. And if this is correct, then to reject (B) one must show that the following conditional is false:

> If a fetus can be safely removed from a human womb without a violation of autonomy and the conditions in the moral community needed to provide care for the fetus exist, then there is no "moral" right to the death of the fetus.

Notice, however, to show that this conditional is false, one must show that there *is* a "moral" right to the death of the fetus, even when the antecedent conditions have all been satisfied. Those who reject (B) do so by appealing to a fundamental "moral" right that is supposed to ground the alleged right to the death of the fetus. For instance, Mathison and Davis consider and reject three arguments for the claim that there is a right to the death of the fetus: (i) The right to avoid parenthood;[14] (ii) The right to genetic privacy;[15] and (iii) The right to property.[16] Each of these purport to show that there is a fundamental "moral" right, which (if it exists) can be deployed in order to establish the negation of (B). While there is some intuitive pull in support of each of these claims, many have found the arguments they are supposed to support problematic.[17]

The general worry in each case is twofold: First, it is doubtful that the alleged more fundamental "moral" right actually exists. Second, even if it is granted that the alleged right exists, it is not clear that it can be used to show that there is a right to the death of the fetus. For our current purpose, I will not rehearse these arguments, objections, and responses here. But it is worth mentioning that, while there may not be any general unanimity amongst philosophers regarding whether the arguments in support of the claim that there is a right to the death of the fetus are sound or unsound, (B) is clearly a controversial position to take. So, one cannot reject (A) simply by assuming the negation of (B). One must demonstrate by sound argumentation that (B) is false. And in what follows I do not need to show that (B) is in fact true, but only that there is *prima facie*, defeasible evidence in support of it.

Why should we accept (C)? The basic idea is this: The problem of abortion has traditionally been understood, roughly, as a conflict between a pregnant person's right to autonomy and the alleged right to life possessed by a fetus. If this approach to the problem of abortion is correct, then our justification for (C) must be grounded in the negation of (A) or (B), or both.

Suppose that securing the death of the fetus was morally permissible before the fetus is viable (currently at approximately 24 weeks) but not afterword, since at the point of viability it could be safely removed from the human womb and gestated in an artificial womb. But in effect, what the possibility of ectogenesis (or more specifically partial ectogenesis via the use of artificial amnion and placenta technology) means is that the point of viability might be construed by some as being far earlier during the gestation of the fetus in the human womb than it is currently. For the sake of argument, let us stipulate that ectogenesis would in fact shift the point of viability from approximately 24 weeks to 14 weeks. If the standard model is correct, then it seems to automatically follow that societies would at the very least be forced to consider whether it would be morally required to implement new legal restrictions and prohibitions on lethal abortions.[18] And it is not difficult to imagine that some would go well beyond a mere question of whether or not societies should implement such technology; many likely would argue that there in fact is an obligation to restrict access to legal, lethal abortions, given the emergence of this reproductive technology. But this approach to (C) is problematic.

Let us grant that ectogenesis will in fact provide *reasons* for fewer lethal abortions. Does this *reason* count as the very same *reason* we would need in order to reject (C)? Even if it is true that partial ectogenesis would give us a reason to think there would be fewer lethal abortions, it is not obviously true that ectogenesis is also a reason to implement new legal restrictions and prohibitions on lethal abortions, since it is not obvious that these will count as the same reason. Indeed, the standard model of the problem of abortion is arguably a product of a social and historical context involving a peculiar juxtaposition of religious/political propaganda and the publications of patriarchal, philosophical and legal departments with their moral and legal scholars—that is, the standard model of the alleged problem of abortion is itself an artifact of a systemic misogynistic social environment.[19] And if this is correct, then the real issue at the core of the abortion debate is intimately entwined to systemic misogyny. Indeed, this fact alone may be sufficient to accept (C).

Consider Manne's ameliorative analysis of misogyny as a structural, social function or mechanism that polices, regulates, and punishes women who fail to conform to the social norms and expectations of patriarchal systems of authority and institutions, which constitute a social context (33-34). On this view, misogyny is not an individual's attitude of hatred held about women—that is, misogyny is not a property of an individual; it is a property of a society.[20] If we assume Manne's (2018) view of Misogyny, this will provide us with the resources to show that abortion laws are inherently misogynistic and, therefore, they are morally problematic. So, we can accept (C) without rejecting (A) or (B) because any law that aims at restricting one's ability to safely obtain a lethal abortion are inherently misogynistic and, therefore, immoral. And this will be true even in the absence of

a fundamental "moral" right to the death of the fetus. In the next section, I will consider how this notion of misogyny can be deployed to argue in support of (C).

2

Consider the Pro-Life Movement (PLM). Proponents of PLM either implicitly or explicitly treat women (and women's bodies) as care givers and baby producers. And according to the very extreme versions of PLM, no deviations from this central trope can ever be allowed. Indeed, an essential feature of this PLM-Dogma is a form of thought police or gaslighting such that one is not even allowed to think for themselves about abortion. The misogyny involved in PLM runs deep, insofar as a woman must be *punished* for resisting or not conforming to the social norms and expectations at the heart of our patriarchal society. Arguably, the misogyny at the core of PLM treats women as though they do not have a mind or genuine intellect—it is a kind of gaslighting and mansplaining, which controls the narrative and version of reality to be recognized.[21] It is in this way that the misogyny at the core of PLM makes abortion laws, generally, immoral.

To my mind, then, the so-called "problem" of abortion, traditionally understood, is not a conflict between competing *rights* between the pregnant person and a fetus. Rather, it is better understood in psychological terms—as an internal and psychologically determined conflict of *reasons* that an agent experiences. Once we abandon the standard model of the alleged problem of abortion, we need not reject (A) or (B) in support of (C). We can, instead, argue in support of (C) in the following way: (1) Given that abortion laws and prohibitions are written, endorsed, promoted, etc., by those who are either directly or indirectly influenced by PLM (i.e., mostly male law-makers), it follows that abortion laws and prohibitions are inherently misogynistic and immoral—that is, abortion laws are grounded in systemic, misogynistic properties of a patriarchal society. (2) If X is inherently misogynistic and immoral, then X should be resisted and rejected. (3) Abortion laws should be resisted and rejected. (4) This suffices to demonstrate that there is good, defeasible evidence in support of (C), which does not depend on rejecting either (A) or (B).

Prior to concluding, I want to consider a further, deeper element involved in the real problem of abortion. I have been assuming that ectogenesis will provide at least one reason for widespread reduction in the overall number of lethal abortions that occur in a society. But this reason is not the same as a reason for new, sweeping implementations of legal restrictions and prohibitions on lethal forms of abortion—that is, the reason for reduction is *not* necessarily the same reason for new abortion laws. Indeed, if what I have argued is correct, if abortion laws are inherently misogynistic, then they cannot be the same. This suggests that we should reconsider how the legal questions associated with the possibility of ectogenesis reveals just how deeply one's psychological *reasons* rather than "moral" *rights* are

rooted in and cannot be divorced from the social context and systems in which they are embedded.

To my mind, the problem of abortion is entwined with an additional philosophical problem: The problem of an agent's "meaningfulness".[22] According to one view, meaningfulness just is the agent's active intellectual or psychological engagement with competing kinds of normative reasons in some form of a decision-making context.[23] If we were to accept this view, we could construe the problem of abortion as being linked to Manne's (2018) understanding of misogyny, insofar as abortion laws aim to close off from the agent the relevant decision-making contexts. Consider, for example, the way that Symons discusses and characterizes this point about what "Meaningfulness" is:

> …meaningfulness is an objective dimension of importance that exists for beings capable of *freely adjudicating between kinds of normative reasons*. This adjudication involves the reality of distinct kinds of value corresponding to distinct kinds of *normative reasons*...One's determination of what kind of value should count as most important in the context of a particular decision; whether one places aesthetic value above moral value, or whether one subordinates prudentially good choices to religious values are examples of how one might constitute one's commitment to an overall conception of meaningfulness. (461)

This approach to what is meant by meaningfulness is not implausible. So, if meaningfulness just is this active intellectual and psychological engagement with competing kinds of normative reasons in a decision-making context, this would help explain why abortion laws are misogynistic and immoral.

In short, the real problem of abortion should be construed in terms of competing psychological *reasons* rather than competing "moral" *rights*. It is an ostensive example of the sorts of competing normative reasons that Symons claims constitutes meaningfulness. Thus, abortion laws prevent one from fully pursuing a life of meaningfulness, which is immoral. This is why, at least in part, we do not need to reject (A) or (B) in order to accept (C).

WORKS CITED

Blackshaw, Bruce P., and Daniel Rodger. "Ectogenesis and the Case Against the Right to the Death of the Foetus." *Bioethics,* vol 33, no. 1, 2019, pp. 76-81.

Cannold, Leslie. "Women, Ectogenesis and Ethical Theory." *Journal Of Applied Philosophy*, vol. 12, no. 1, 1995, pp. 55-64.

Colgrove, Nicolas. "Defining 'Abortion': A Call for Clarity." Theoretical Medicine and Bioethics (2):137-175 (2025)

Di Stefano, Lydia, Catherine Mills, Andrew Watkins, and Dominic Wilkinson. "Ectogestation Ethics: The Implications of Artificially Extending Gestation for

Viability, Newborn Resuscitation and Abortion." *Bioethics,* vol. 34, no. 4, 2020, pp. 371-384.
Gillon, Raanan. "Is There a 'New Ethics of Abortion'?" *Journal Of Medical Ethics*, vol. 27, Supplement 2, 2001, pp. ii5-ii9.
Hendricks, Perry. "There is No Right to the Death of the Fetus." *Bioethics*, vol. 32, no. 6, 2018, pp. 395-397.
Horn, Claire. "Ectogenesis is for Feminists: Reclaiming Artificial Wombs from Antiabortion Discourse." *Catalyst: Feminism, Theory, Technoscience,* vol. 6, no. 1, 2020.
Kaczor, Christopher. "Ectogenesis and a Right to the Death of the Prenatal Human Being: A Reply to Räsänen." *Bioethics*, vol. 32, no. 9, 2018, pp. 634-638.
Kauppinen, Antti. "Meaningfulness and Time." *Philosophy And Phenomenological Research*, vol. 84, no. 2, 2011, pp. 345-377.
Kingma, Elselijn, and Suki Finn. "Neonatal Incubator or Artificial Womb? Distinguishing Ectogestation and Ectogenesis using the Metaphysics of Pregnancy." *Bioethics,* vol. 34, no. 4, 2020, pp. 354-363.
Manne, Kate. *Down Girl: The Logic of Misogyny*. Oxford University Press, 2018.
———. *Entitled: How Male Privilege Hurts Women*. Penguin Books, 2020.
Mathison, Eric, and Jeremy Davis. "Is There a Right to the Death of the Foetus." *Bioethics,* vol. 31, no. 4, 2017, pp. 313-320.
Overall, Christine. "Rethinking Abortion, Ectogenesis, and Fetal Death. *Journal Of Social Philosophy*, vol. 46, no. 1, 2015, pp. 126-140.
Räsänen, Joona. "Ectogenesis, Abortion and a Right to the Death of the Fetus." *Bioethics*, vol. 31, no. 9, 2017, pp. 697-702.
Romanis, Elizabeth Chloe. "Artificial Womb Technology and Clinical Translation: Innovative Treatment or Medical Research?" *Bioethics*, vol. 34, no. 4, 2019, pp. 392-402.
Simonstein, Frida, and Michal Mashiach–Eizenberg. "The Artificial Womb: A Pilot Study Considering People's Views on the Artificial Womb and Ectogenesis in Israel." *Cambridge Quarterly of Healthcare Ethics* 18, no. 1 (2009): 87-94.
Stratman, Christopher M. "Ectogestation and the Problem of Abortion." *Philosophy & Technology,* vol. 34, no. 4, 2021, pp. 683-700.
Stratman, Christopher M. "Ectogestation and the Good Samaritan Argument." *Journal of Law and the Biosciences*, 2023, pp. 1–26.
Stratman, Christopher M. "Ectogestation, In Vitro Fertilization, and the Abortion Debate." Monash Bioethics Review (forthcoming).
Symons, John. "Meaningfulness and Kinds of Normative Reasons." *Philosophia,* vol 49. no. 1, 2021, pp. 459-471.
Warren, Mary Anne. "On the Moral and Legal Status of Abortion." *The Problem of Abortion,* 2nd ed., edited by Joel Feinberg, Wadsworth, 1984.
Wolf, Susan. "Happiness and Meaning: Two Aspects of the Good Life." *Social Philosophy And Policy*, vol. 14, no. 1, 1997, pp. 207-225.
———. *Meaning in Life and Why it Matters*. Princeton University Press, 2010.

NOTES

1. There is a distinction between what we can call "full-ectogenesis" and "partial-ectogenesis." The former refers to the entire fetal gestation occurring independently of a human womb. The later refers to an ectogestation technology that allows a fetus to be surgically removed from a human womb and transferred to an artificial womb where it will continue to be gestated until it reaches full term. For the purpose of this paper, I shall focus only on the notion of partial-ectogenesis or "ectogestation" as it relates to the problem of abortion. It is also worth mentioning that this technology is importantly different from incubator technology currently used in the NICU, which serves the purpose of rescuing the viable fetus in some sense. Ectogestation is better understood in terms of artificial amnion and placenta technology, which allows for the continued development of the vital organs like the lungs, heart, and liver. While the argument to be defended in what follows does assume that the possibility of ectogestation means that a fetus will be viable earlier during pregnancy, in an attempt to remain as neutral as possible, I will not specify a specific point during gestation where the fetus would "become" viable.
2. The literature is growing, but for a representative selection of those who find ectogenesis to be a plausible solution see for e.g., Mathison and Davis; Kaczor Hendricks; and Blackshaw and Rodger.
3. But see e.g., Cannold and Ovverall for philosophers who holds that there is a right to the death of the fetus; and see e.g., Räsänen for an example of someone who argues that there is a collective right to the death of the fetus that is possessed by both of the biological parents.
4. The different between lethal and non-lethal abortions will need to be made clear. But the general idea is just that one should be legally permitted to terminate their pregnancy in a lethal way.
5. For a discussion of this point, see e.g., Gillon. Additionally, I have defended similar arguments in Stratman (2023) and in a forthcoming article in Monash Bioethics Review.
6. This is a crucial but overlooked point. For further discussion, see e.g., Mathison and Davis 313-314. However, not everyone agrees. For a discussion of why one might believe that our concept of abortion necessarily requires the death of the fetus, see e.g., Colgrove (2025).
7. Mathison and Davis directly address this claim in footnote 2 (313-314).
8. Considerations of space preclude a needed discussion of all that would be needed for this proposed hypothetical scenario to be actualized, especially in the United States. But it will be sufficient for our current purpose to simply stipulate that this situation is not impossible, even if it is unlikely.
9. Arguably, ectogenesis technology calls into question the metaphysical and moral significance of the human womb and the artificial womb. Given considerations of space, I cannot discuss the multiple issues involved here. But, for further discussion, see e.g., Di Stefano et. al, Romanis, Simonstein and Mashiach-Eizenberg, Kingma and Finn, and Horn.
10. There may still be ways to object here, but given considerations of space, I shall move on to consider why we cannot very easily reject (B).
11. See e.g., Thomson's discussion of the Famous Violinist case (48-52)
12. See e.g., Mathison and Davis 314.
13. For a discussion of such an argument, see e.g., Stratman (2021).
14. See e.g., Mathison and Davis 314-316.

15. See e.g., Mathison and Davis 316-317.
16. See e.g., Mathison and Davis 317-319.
17. Considerations of space preclude further detailed discussion of these arguments.
18. But consider e.g., Di Stefano et.al; see also Romanis.
19. In what follows, I shall assume Manne's understanding of misogyny from *Down Girl*.
20. Manne (*Down Girl*) develops this ameliorative analysis of the concept misogyny in painstaking detail. I am assume that Manne sufficiently develops this analysis in order to deploy it in support of (C), but one may doubt whether such an analysis can be adequately carried out for my current appropriative purposes. This is a reasonable worry but one that I do not have the space to investigate here. Future work, however, should pay close attention to objections to Manne's understanding of misogyny and my deployment of this concept.
21. For discussion of these points concerning misogyny as it relates to intellectual entitlement, see e.g., Manne's *Entitled*
22. There is a long history and well developed literature on the philosophy of "meaningfulness" or a meaningful life, which I will not comment on here because these views diverge in numerous ways. For example, Wolf ("Happiness and Meaning" and *Meaning in Life*) focuses on the notion of proper fulfillment arising from worthy love; and Kauppinen focuses on how achievement relates to our understanding of meaningfulness.
23. For further discussion of this approach see Symons. This view of meaningfulness takes it a kind of subjective experience whereby an agent actively engages in an intellectual or psychological episode involving competing kinds of normative reasons that constitute a decision-making context of the right sort. I will assume but not argue for Symons understanding of meaningfulness.

Why Knowledge Might Not Entail Belief

Pranav Ambardekar
FLAME University

I

The entailment thesis holds that knowledge entails belief. Most epistemologists take it to be obvious.[1] Now, there are two ways to argue against the entailment thesis: first, one can argue for a particular account of knowledge or belief such that the entailment thesis does not hold; second, one can provide an intuitive example of knowledge without belief. The second strategy was employed by Colin Radford.[2] In Radford's proposed counterexample, Jean—a French Canadian—is being quizzed by Tom on English history. The former earnestly avows that he knows nothing about English history. Jean requires prompts to answer questions, and even when he answers them he does so with hesitation. Jean is fairly certain that his answers are wrong. At the end, Jean gets five out of ten questions right. One of the questions he is asked is: when did Queen Elizabeth die? Jean answers (E): 'Queen Elizabeth died in 1603.' When told by Tom that his answer is right, Jean reports that he feels he might have "picked that up on a Shakespeare course or somewhere."

Many take Radford's claim *à propos* his example to be the following: Jean knows (E) but does not believe it.[3] But Radford does not dwell on establishing the absence of belief, presumably because he takes Jean's behavior to be revealing. Likewise, no sustained defence of the presence of knowledge is offered: at a point Radford says we should think Jean knows because he remembers.

In light of all this, one plausible way to interpret a crucial part of Radford's argumentative strategy is to take him to be pushing the following move: an agent can know p without believing it when i) he has evidence that is sufficient to justify p, and ii) he is uncertain about the epistemic status of the evidence he possesses and p; the thought then is that we can infer disbelief from uncertainty, and knowledge from the possession of sufficient evidence. Interpreted this way, Radford's argumentative strategy faces several problems.

First, in the case Radford proposes, it's not clear that the agent possesses sufficient evidence to know p.[4] Even if Radford's original case is improved upon to deliver the intuition that the agent has sufficient evidence for knowledge, his move still faces—and this is the second problem—the following dilemma: if the agent's uncertainty about the evidence he possesses and p is well grounded i.e. there is some reason that casts doubt on the proposition in question, thus acting as a defeater, then the agent does not really have sufficient evidence for knowledge on the other hand, if the agent's uncertainty about the evidence he possesses and p is not well grounded (i.e., there is no reason whatsoever for the agent to be uncertain about or disbelieve p) then the agent is irrational. Such irrationality might be a symptom of some deeper problem in the agent's cognitive apparatus, in which case it seems implausible to credit such an agent with knowledge. Either way, there is no knowledge.

The third problem is even more devastating: recently Jonathan Schaffer and David Rose have argued that even if there is knowledge in the kind of case that Radford proposes, there will always be dispositional belief.[5] They make their point without offering a specific account of dispositional belief. The case they discuss is a slightly modified version of Radford's original case: in their case it is clear that the agent (Kate) possesses sufficient evidence, however, her recall capacities are compromised due to fear or stress. When asked about Queen Elizabeth's death, Kate answers correctly. Surely, Schaffer and Rose think, Kate does not *merely* guess in such cases. She gets the right answer because the information she possesses is unconsciously guiding her behavior. In the absence of any other plausible explanation for Kate's behavior, an unconscious mechanism guiding her behavior has to be posited. They admit that there is no occurrent belief in Kate's case, as Kate does not consciously give assent to (E); nevertheless, they contend that there is dispositional belief due to the unconscious mechanism. Thus, Schaffer and Rose's argument takes the form of an inference to the best explanation hypothesis. So the most that Radford would have shown is that in the kind of cases he describes, the agents do not have occurrent belief. However, as Schaffer and Rose emphasize, that is not the epistemologically relevant notion of belief.

In this paper, by providing an improved version of Radford's proposed counterexample, I will show that Radford's strategy of providing an intuitive case of knowledge without belief can be salvaged from the aforementioned problems. Recall Radford's central move: the introduction of disbelief through uncertainty,

even when the agent purportedly has sufficient evidence for knowledge. Instead of introducing uncertainty to get outright disbelief, I specify the agent's cognitive motivations differently: the agents in my proposed counterexample care more about avoiding false belief than believing in what is true. As we shall see later, this move allows me to avoid both the dilemma and the Schaffer and Rose kind of response. Ultimately, I will argue that there are cases where knowledge is followed by something (e.g., a mental state) that falls short of full belief (and not disbelief).

Here is the plan. In section II, I will begin by presenting what I take to be an improved version of Radford's counterexample to the entailment thesis. The crucial innovation has to do with a particular kind of agent: *cautious believers*. I will make explicit and briefly motivate the claims that I am committed to in introducing such agents. In section III, I will show how my case avoids Schaffer and Rose's response and the dilemma mentioned before. I will conclude by addressing an objection to my proposal, showing how the agent in my case satisfies a version of the basing requirement for knowledge, and stating what the main takeaway of this paper is.

II

Consider the case of Saraswati:

Saraswati. The recommended textbooks Saraswati had read for her history examination did not explicitly mention the year of Queen Elizabeth's death. Instead, the author who had written the books mentioned in one of the books that: 'Queen Elizabeth died in the early sixteen hundreds, in the year that the "Father of New France" sailed to Canada'; in another book on world history the author wrote: 'Samuel de Champlain sailed to Canada a year before England had concluded the Treaty of London with Spain (1604).' When Saraswati reads the question, "When did Queen Elizabeth die?" she feels a little annoyed. She says to herself, "Why wasn't the year explicitly stated in those textbooks!" She entertains several years as possible answers but then entertains the following inference to the best explanation hypothesis in her mind:

> P1: Queen Elizabeth died in the early sixteen hundreds, in the year that the "Father of New France" sailed to Canada. [at T1: entertains the proposition and consciously gives assent to it]
>
> P2: Samuel de Champlain sailed to Canada a year before England had concluded the Treaty of London with Spain (1604). [at T2: entertains and assents]
>
> P3: Engraved on the grand statue of Samuel de Champlain in Quebec is

the word "Père." [at T3: entertains and assents] {background knowledge}

From P1 through P3:
P4: "The Father of New France" is Samuel de Champlain. [at T4: entertains but does not assent]

From P1, P2 and P4:
(E): Therefore, Queen Elizabeth died in 1603. [at T5: entertains but does not assent]

Saraswati sees an entailment relationship between premises P1 through P4 and (E) but she suspends judgment about P4 and thus (E). Since there is just a minute remaining and she has to write down something, Saraswati writes down "1603" as the answer: she is, overall, more confident in (E) than not. After the exam, she rushes to a public library and borrows all the books which contain historical research on the early seventeenth century. When she reads that all twelve of them say that "The Father of New France" is Samuel de Champlain and that Queen Elizabeth died in 1603, she smiles and consciously gives assent to both P4 and (E).

Note: here are some premises, such that if they were all in Saraswati's pool of evidence, her evidence would be sufficient (and not just barely sufficient to justify) to justify P4 and (E):

P5: There are many statues across Canada paying tribute to Samuel de Champlain.

P6: Part of the territory of what was "New France" now belongs to Canada.

P7: Samuel de Champlain was French.

P8: Samuel de Champlain made the first accurate map of what is today's Canada's east coast.

And had she read a sentence explicitly mentioning the identity of the two names "Father of New France" and "Samuel de Champlain" in a respected book on seventeenth century history, she would have overwhelming evidence for P4 and thus (E).

Saraswati is what I call a cautious believer in a forced choice situation. What features characterize such an agent in such a situation? First, she cares more about avoiding false belief than acquiring true belief. In his classic essay entitled "The Will to Believe," William James pointed out that, broadly speaking, human beings can be motivated, in various degrees, by two kinds of epistemic goals: first,

believing what is true and second, not believing what is false.[6] The second goal can be achieved with respect to any given issue not just via outright disbelief, but also via suspending judgment; whereas, to achieve the first goal suspending judgment will not do. Now, my claim is that it is plausible to hold that there can be cases like Saraswati's where one of these two cognitive motivations is predominant. This is an empirical claim—and although I do not offer an argument in favour of it here, I take it that Saraswati's case is as intuitively familiar as Radford's original case and this serves as a kind of *prima facie* justification.

Second, she has the ability to assess the epistemic status of her evidence, i.e. what the evidence she has really amounts to. There are two pictures here, each compatible with the case I have offered: the ability she has will depend on the *kind* of cautious believer she is. On one picture—where a cautious believer is a virtuous epistemic agent—the agent has the ability to tell if the evidence she has for a given proposition p is barely sufficient for p. A virtuous epistemic agent, as Hume once said, proportions their beliefs to the evidence: so if the evidence for a certain proposition p is weak and insufficient for knowledge, a virtuous epistemic agent will not form the belief that p.

Now a brief but important detour. I hold that any given agent's pool of evidence regarding any given proposition p must fall under one of the four rubrics: insufficient evidence, barely sufficient evidence, sufficient evidence, and overwhelming evidence. Whatever one's preferred view about the nature of evidence is i.e. whether evidence is always propositional, what it means to *possess* evidence, and so on, if a component of that view is that evidence admits to degrees of strength, then I think that one should follow me in making the kind of fourfold classification made above. Instead of just making coarse-grained distinctions like weak and strong evidence for a particular proposition, we can make more fine-grained distinctions.

Back to virtuous epistemic agents: I take it that Saraswati counts as a virtuous epistemic agent because she suspends belief about (E)—recall that she does not give conscious assent to either P4 or (E) —as the evidence she has at her disposal is right on the knife's edge. Although, intuitively, it is sufficient for knowledge it is only barely sufficient. Any less and she would not count as knowing (E). Might more evidence in favour of (E) come her way, so as to make it sufficient for knowledge, she will *not* suspend judgment about (E). This makes her virtuous, espitemically speaking.

On the second picture—where a cautious believer has a personality quirk—the agent has the ability to tell if the evidence she has for a given proposition p is *not* overwhelmingly in favour of it. Saraswati might be so passionate about history that she only consciously assents to claims that she knows have overwhelming evidence for them. Until a time when she is in possession of overwhelming evidence for (E), Saraswati suspends judgment and thereby does not give conscious assent to (E). Someone might object here that an agent who only

consciously assents to propositions when the evidence is overwhelmingly in favour of them is irrational. This might be concerning since—as the second horn of the dilemma I presented before goes—this irrationality might be so central to the agent's cognitive system that the agent could not be credited with knowledge. But this objection succeeds only if the kind of personality quirk I mentioned is pathological across domains of knowledge i.e., if with respect to any kind of knowledge the agent suspends judgment when the evidence is not overwhelming. On my picture, the agent only suspends judgment with respect to propositions that fall under one domain of knowledge (e.g., historical knowledge).

The plausibility of these pictures hinges on the following claim: agents can possess the general ability to assess the epistemic status of the evidence at their disposal. Why think the claim above is true? Again, a *prima facie* defence: we cannot make sense of much of human learning without positing such an ability. Reflection on what one has learned (the information one has acquired) and what it really amounts to, allows us not only to tell others about which domains of knowledge we are proficient or deficient in, but it also helps us selectively explore within and without domains of knowledge as per our goals. This ability might also be the causal basis of our confidence in asserting knowledge claims. In a nutshell, we are not mere computers which store and process information—we are capable of being aware (even if we are not always aware) of what that information really amounts to. Now, if it is plausible to hold that humans in general have such an ability then I think it is plausible to hold that some gifted humans might wield this ability better than others: they can tell if the evidence they have at their disposal is barely sufficient for, or if it is not overwhelmingly in favour of, some proposition p.

I now turn to how the counterexample offered in this section avoids the problems faced by Radford's original counterexample.

III

Recall Schaffer and Rose's response to Radford. On their view, Kate still believes in (E) because it is implausible to hold that she *merely* guessed: she got the answer right due to the unconscious mechanism guiding her behavior. The positing of dispositional belief serves well as a causal explanation of Kate's behavior. So why not think the same in Saraswati's case? In Saraswati's case, I contend that there is no need for the kind of explanation Schaffer and Rose are offering to account for her actions: there is a stepwise conscious process—laid out in terms of an inference to the best explanation hypothesis—that goes on in Saraswati's mind. It is plausible to hold that it is this conscious process that is causally responsible for Saraswati's proximate behavior: getting the answer right.

To this an objector might say: you seem to assume that the relevant disposition only manifests if the unconscious mechanism is in play; why not think instead that

Saraswati's conscious process is itself a manifestation of the disposition?

My response is that what counts as a manifestation of the disposition will depend on the account of dispositional belief at play. The account at play here seems to be the following one: whatever else the account might hold, it holds that responding to a question as to whether p in a forced choice situation is sufficient for dispositional belief. I reject such an account of dispositional belief since it would fail to take seriously an agent's behavior that conflicts with p. For instance, in Saraswati's case, it's clear that she does not have cognitive dispositions that accord with p: she fails to consciously assent to (E). Moreover, it could be that she lacks other relevant dispositions: like the disposition to place a bet on whether (E), the disposition to publicly affirm that (E) after the exam in front of friends, and other behavioral dispositions.

Saraswati's case is a little complicated: she has some dispositions that accord with p, and others that do not. Nevertheless, I take it that even on more attractive liberal dispositionalist accounts of belief—like the one defended by Eric Schwitzgebel—Saraswati does not count as believing (E).[7] While it also follows from the account that she does not lack belief entirely, that is not a problem for someone arguing against the entailment thesis. For knowledge will be followed by something that falls short of full belief and so the entailment does not hold.[8]

We can now see how my counterexample avoids the dilemma faced by Radford's. In introducing uncertainty to get outright disbelief, Radford invited the charge that the agent is irrational not to believe if the evidence is sufficient for knowledge. But with Saraswati, there is no outright disbelief: the agent merely suspends judgment about p, while acting in ways that both accord and conflict with p. If you will, Saraswati is in a mental state that is distinct both from belief and disbelief. And if any of the two kinds of pictures I have presented are plausible, Saraswati is not irrational for neither virtuous epistemic agents nor agents with minor personality quirks are irrational.

If Saraswati's pool of evidence is sufficient - even if *barely* so - to justify (E), then *were she* to form the belief that (E), she would be propositionally justified in doing so. Many philosophers think that knowledge is intimately connected with the notion of doxastic justification, which comes apart from propositional justification. The thought is that a belief must be held not only because evidence supports it, but also for that same reason and not some bad reason. Now, of course Saraswati will fail to be doxastically justified in this way: she does not form the belief that (E), and so the *belief* cannot be properly based on her evidence. She does, however, satisfy what could be seen as a version of the basing requirement: her behavioral dispositions that accord with p are causally linked to a *conscious* episode involving reasoning (an inference to the best explanation). And this is exactly the kind of thing that her belief—were she to have formed it—should have been based on were it to be counted as properly based. Thus, we have stronger reason to think that knowledge is present in my counterexample, as opposed to

Radford's.[9]

The main take-away from this paper is the following: caution need not destroy knowledge. I think that Radford was right to think that there is a way to specify an agent's psychology such that belief would not follow knowledge. However, he was wrong in thinking that throwing uncertainty into the picture would do the job. Caution, I have suggested, is a more robust candidate to do the job. Thus, if the counterexample offered in this paper is *prima facie* plausible, then I take it that the Radfordian project is still alive. The next step would be to show that *all* (save belief, of course) the plausible pre-conditions for knowledge can be satisfied in cases involving cautious believers in forced choice situations.

NOTES

1. For a *Wittgensteinian* pushback against the entailment thesis, see chapters 4 and 5 in Peter Hacker's *The Intellectual Powers.*

2. Even his title heralds his argumentative strategy. Colin Radford calls his paper "Knowledge—By Examples."

3. See for instance the introduction in Blake Myers-Schulz and Eric Schwitzgebel's "Knowing that P without Believing that P."

4. For a brief review of responses to Colin Radford's purported counterexample, see Jonathan Ichikawa and Matthias Steup's "The Analysis of Knowledge."

5. See Jonathan Schaffer and David Rose's "Knowledge Entails Dispositional Belief."

6. William James, "The Will to Believe," reprinted in *The Will to Believe and Other Essays in Popular Philosophy* (New York: Longmans, Green and Co., 1897).

7. See Eric Schwitzgebel's "A Phenomenal, Dispositional Account of Belief."

8. For a discussion on the possibility of 'in-between believing,' see Eric Schwitzgebel's "Acting Contrary to our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief."

Also, Saraswati ought to know: she is, after all, the goddess of knowledge.

WORKS CITED

Hacker, Peter Michael Stephan. *The Intellectual powers: A Study of Human Nature*. John Wiley & Sons, 2013.

Ichikawa, Jonathan Jenkins; Steup, Matthias, "The Analysis of Knowledge", *The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>.

James, William. *The Will to Believe and Other Essays in Popular Philosophy* Longmans, Green and Co., 1897.

Myers-Schulz, Blake, and Eric Schwitzgebel. "Knowing that P without believing that P." *Noûs,* vol. 47, no. 2, 2013, pp. 371-384.

Radford, Colin. "Knowledge: by examples." *Analysis,* vol. 27 no. 1, 1966, pp. 1-11.

Rose, David, and Jonathan Schaffer. "Knowledge entails dispositional

belief." *Philosophical Studies,* vol. 166, 2013, pp. 19-50.
Schwitzgebel, Eric. "Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief." *Pacific Philosophical Quarterly,* vol. 91, no. 4, 2010, pp. 531-553.
Schwitzgebel, Eric. "A phenomenal, dispositional account of belief." *Noûs,* vol. 36, no. 2, 2002, pp. 249-275.

"We Know You Better Than You Know Yourself": Aesthetic Taste and Recommender Systems

Kenneth L. Brewer
Montclair State University

Several years ago, walking out of a screening of *The Hangover* ("Absurd, Raunchy, Irreverent" according to Netflix), I said to my friend, "Well, that wasn't very good." "What are you talking about?" she replied. "You laughed the entire movie." While it is possible that I might have laughed a lot and still not liked the movie (perhaps the laughs were what we call "cheap laughs"), I was startled. If I had been alone, I would have walked out of *The Hangover* firmly believing that I had not enjoyed it, when apparently, I had enjoyed it tremendously. I like to think that I know what I like, but this seemingly trivial incident made me wonder how much I really understand my own taste.

Naturally, I turned to philosophy for answers to the worrisome question of whether I am in fact a fan of absurd, raunchy, irreverent bromance comedies. It turns out that my befuddlement about what I like and do not like makes me completely typical, at least according to a number of philosophers working in aesthetics who have pored over a considerable amount of data from the behavioral and brain sciences.[1] The most thorough account so far is Kevin Melchionne's "On the Old Saw 'I know nothing about art but I know what I like.'"[2] It is entirely likely, Melchionne argues, that I do not know what I like or why I actually like what I believe that I like. The unreliability of introspection seriously undermines our knowledge about our own taste, and Melchionne attacks what he views as our complacency about knowing what we like. Melchionne maintains that self-knowledge about what we like and why are necessary conditions for a valuable aesthetic life. Absent these things, we can certainly have aesthetic pleasures, but

we cannot cultivate them, as in the absence of self-knowledge "the pursuit of aesthetic satisfaction would be no more than a fishing expedition" (132). Presumably, we want more efficiency from our aesthetic experiences. We also like to understand our taste as consistent, and to see some sort of coherent narrative in our aesthetic lives.[3]

To my knowledge, Melchionne's argument has not been convincingly answered. My approach will be to accept his well-supported attack on the unreliability of introspection, and focus on one possible solution to this problem.[4] Melchionne rejects skepticism about aesthetic self-knowledge, instead recommending a "weak fallibilism" that "leave[s] room for reliable aesthetic self-knowledge while recognizing that our aesthetic responses often remain as opaque as the rest of conscious life" ("On the Old Saw…" 140). Is there any way, however, to make my aesthetic responses less opaque? I will explore a decidedly new-fangled method for better understanding one's own taste: recommender systems that rely on algorithms to predict whether or not a user will find a certain item satisfying.[5] We are familiar with these from Netflix, Amazon, Spotify, and so on. There is a vast technical literature in computer science-related fields on how to improve the performance of these systems.[6] Conversely, there is a fairly substantial amount of discussion in what we might loosely term "cultural studies" about the possible effects of such systems on individuals and societies; these effects are usually seen as pernicious.[7] Even though most of these systems are well into their second decade, philosophers have had relatively little to say about them.[8] I will discuss the possibility that technologically sophisticated recommender algorithms such as Netflix can enhance our knowledge of our own taste precisely by improving our capacity to introspect about what we like and why.

Melchionne begins by pointing out that the infallibility of introspection has pretty much been taken for granted in philosophical aesthetics, which has been more concerned with how "individuals with self-transparent taste disagree with others equally aware of their preferences" (133). Melchionne draws on a number of empirical studies to suggest "that ignorance or error about our inner states is common," such as Timothy Wilson's *Strangers to Ourselves* (134). Nor is Melchionne inclined to accept the view that while this ignorance may be true of many areas of my emotional life, aesthetic experience is special, as John Dewey argues. In fact, Melchionne notes, that we go to such elaborate lengths to avoid distractions during our aesthetic experiences (headphones, darkened movie theaters) suggests that we are well-aware of the fragility of aesthetic experience and aesthetic attention (135).

Citing John Lambie and Anthony Marcel on "emotion unawareness," Melchionne argues that there are three ways we might get knowing what we like wrong: we might have aesthetic responses without second-order awareness of them; we might have aesthetic responses with inaccurate second-order awareness; and we might give "false attributions of the causes of aesthetic response" (135).

With the first, aesthetic responses without second-order awareness, Melchionne is particularly interested in cases in which belief about my "aesthetic personality," or "biographical taste" ("I like French New Wave films") leads me to simply ignore moment-to-moment aesthetic responses that conflict with those attitudes. Thus, if I thought before *The Hangover*, "I'm not really the kind of person who enjoys raunchy comedies: I much prefer the mid-period works of Ingmar Bergman," then I would simply ignore evidence—and it was apparently considerable—that I am precisely the kind of person who likes raunchy comedies.

With the second possibility, I am aware of my affective state but may mis-categorize it. For example, with an avant-garde work of art, I misunderstand my confusion as dislike. Worse, this problem apparently cannot be solved by simply focusing more intensely on my introspective states. An interesting example of this problem was found by T. D. Wilson and J. W. Schooler, who conducted a study on pleasure in listening to music. Subjects were asked to listen to a piece of music, and either given no instructions, instructions to try to be happy, instructions to monitor their happiness, or instructions to try to be both happy and monitor their happiness.

Subjects who monitored their happiness were significantly less happy while listening to the piece than subjects in the control group who were given no instructions. Conclusion: "Vigilantly monitoring one's ongoing hedonic experience can undermine one's ability to actually gain happiness" (Wilson and Schooler 191).[9]

Finally, and perhaps most significantly, we tend to believe that taking the "aesthetic attitude" gives us more insight into what we feel. I regard aesthetic experience as a highly self- conscious activity in which I can give clear explanations for why I'm feeling what I'm feeling because I pay such close attention to the experience. In daily life, I may not know why I'm angry or sad, but with a work of art, I expect myself to be able to explain the source of these emotions. But this very effort to attribute reasons to what we feel leads, Melchionne argues, to distortions, what social psychologists term the "introspection effect." Figuring out why I like Cézanne (why looking at a Cézanne painting causes certain mental states that I classify as "pleasurable") is not at all straightforward, and I often end up relying on critical clichés I have picked up at some point in my life (Melchionne, "On the Old Saw…" 139).

So how might recommender systems deal with these three problems of reliable aesthetic self-knowledge? A couple of initial points. I am going to focus on Netflix because I believe its algorithm offers the most interesting attempts, so far, to solve these problems. At the same time, I am aware that Netflix is a profit-making enterprise. We do not know, for example, exactly how its algorithm works (it is proprietary) and this lack of transparency is obviously not ideal.[10] Moreover, there is general agreement that Netflix's algorithm promotes shows that Netflix has itself produced and favors television shows (which create longer user engagement)

over films.[11] However my aim is to explore the basic aesthetic issues these systems raise, not engage in social critique.

Briefly, it helps to have some idea of how these systems work.[12] Most of them typically use some form of *collaborative filtering*, either user-based on item-based. Based on my own ratings or behavior, I am paired up with similar users, and this similarity is used to predict my rating for a film I have not seen but similar users have seen. In an item-based system, the algorithm isolates items that tend to be consumed together (peanut butter and jelly, or Tom Cruise and Liam Neeson action movies) and makes recommendations accordingly. There are also *content-based* systems that mark features of what a user consumes and then recommends similar items based on these features: this is the "if you liked X you'll love Y" appeal present in film trailers from way back, updated into an algorithm. Most recommender systems use a combination of these and are termed *hybrid* systems, as all of them have significant limitations on their own.[13]

So the first problem Melchionne describes: I ignore my own responses because of my preconceptions about the kind of aesthetic person that I am. I do not notice the pleasure I am taking in *The Hangover* because I see myself as more of an Ingmar Bergman film sort of person. My sense of my biographical taste does not jive with my actual aesthetic experiences: my revealed preferences are quite different from my stated preferences.[14] If you go on Movielens, a research project on film recommendations based at the University of Minnesota, you are asked to rate as many movies as possible on a 1-5 scale.[15] According to the research Melchionne describes, my ratings will not tend to be very accurate. I will rate movies highly that I believe I am supposed to have liked, not even aware that I actually did not enjoy them.

While Netflix still allows a thumbs-up or a thumbs-down, this metric can be relatively crude most likely because ratings do not really matter that much for the reasons just explained. So Netflix's algorithm guru Todd Yellin has argued that I will probably give Al Gore's *An Inconvenient Truth* five stars because I think it is an important film, but it turns out that I actually watch *Paul Blart: Mall Cop 2* several more times a year than I watch the documentary.[16] As far as Netflix is concerned, it is pretty obvious which film I actually prefer, so it no longer cares to compile ratings I might give them. When Netflix shifted over to a streaming service, it began to intensively monitor user behavior with the interface, termed "data-behaviorism" by Sarah Arnold. As Arnold puts it, the user, for Netflix, is no longer an individual but a set of data points. Our behavior tells Netflix what we like, not our (suspect) ratings (Arnold 50). In rating films, we may think we know what we like, but from our behavior, Netflix actually does know what we like, and then offers a prediction: Film X is a "95% match" to our preferences. These preferences are drawn from my behavior, not from other data about me. In fact, Yellin argues, much of my biography is not relevant to my taste: my gender, class, and so on do not matter.[17]

Freed of the need to rely on users' (suspect) ratings, Netflix thus addresses the second problem of how monitoring my hedonic states actually undermines my self-knowledge about them. As Sherry Irvin notes, data from the behavioral and brain sciences indicates that if I know that I have to rate an item, my likes and dislikes typically change significantly (38). With Netflix, if all of my behavior matters, then on some level, none of it really matters in terms of affecting my behavior. Without having to monitor my hedonic state because I'm constantly thinking, "Is this film 3.5 stars or only 3?", I may be better aware of what my hedonic state actually is.

The third problem Melchionne discusses is my confident belief that I can give clear reasons for why the work of art caused certain feelings: "I felt nervous watching *Psycho* and this was due to Hitchcock's use of overhead shots." This may very well be the case, Melchionne argues; it may also just be something I learned about Hitchcock in Film Studies 101, and I'm just repeating it.[18] As Irvin explains, this is the problem of "coarse-grainedness": "Aesthetically relevant aspects of a work fail to have the expected effect on people's judgments" (39). What is really affecting me in *Psycho* is Marion Crane's moral struggle, but I overlook this, instead attributing my feelings to the usual sources.

Netflix tries to deal with the coarse-grainedness problem with "tags." Described in a hagiographic 2014 article in *The Atlantic*, Alexis C. Madrigal's "How Netflix Reverse-Engineered Hollywood," Netflix developed a 36-page manual for employees who work at tagging movies with descriptors. Netflix has broken down the dozen or so film genres we are familiar with (drama, comedy, action, horror, etc.) into 76,897 micro-genres. You may think of yourself as someone who enjoys horror, but what you actually enjoy are "Cult Evil Kid Horror Movies."

These are Kendall L. Walton's "Categories of Art" on steroids.[19] Is a movie funny? Not good enough. How is it funny? Is it witty? Gross-out? Quirky? Drily witty? And so on.

Madrigal's homage to Netflix's genius specifically makes a claim about introspection: "It's not just that Netflix can show you things you might like, but that it can tell you *what* kinds of things those are. It is, in its own weird way, a tool for introspection." Madrigal does not explain how precisely Netflix's tags function this way, but Emily Lawrence has suggested that Netflix has the capacity to make us go "Huh" about our taste, which should lead to self-reflection about what we actually like and why we like it.[20] Essentially, the tags function as a form of aesthetic education, and the next time I am enjoying a movie, I can presumably go beyond critical clichés ("I like Hitchcock because he was the master of suspense") to "I like *Rear Window* because it is cerebral and features a strong female lead."

If we lack self-knowledge about our own tastes, then we can expect inconsistency in our choices. What do recommender systems do about this problem, raised in its classic form in Ted Cohen's "On Consistency in One's

Personal Aesthetics"? Cohen describes his confusion at his wife's seemingly inconsistent tastes. In particular, he is troubled by her love the music of Béla Bartók and his dislike of the music of Aaron Copeland, specifically because she enjoys Bartók's use of folk music and *dislikes* Copeland for his use of folk music idioms (107).

Typically, Cohen argues, when we like an object x for reason R, we expect to like another object y that also features R.

What do we do? Cohen offers two basic possibilities, and I suggest that recommender systems have some relevance to both of these (122). The first is that I simply have not looked hard enough at why I like object x or dislike object y. My reasons are shallow ones. With recommender systems, tags should help alleviate this problem. I enjoy the films of Éric Rohmer, and I suspect that I will also enjoy the films of Jean-Luc Godard, a fellow French New Wave director, as they are similar in significant ways. I do not enjoy Godard's films. However, I do enjoy the films of French New Wave director Claude Chabrol. It turns out--when I examine the tags--that what I actually like are films with strong female leads, which Chabrol's and Rohmer's films have in common. If, as Melchionne argues, Cohen's wife's problem is not inconsistency but instead a vague understanding of her own taste ("opacity"), the tags should/might make her understanding less vague (Melchionne, "On the Old Saw…" 133).

A second possibility is that I do not have a "style," as Cohen puts it. My tastes are simply not consistent, and while there is educational value, he notes, in attempting to clarify my taste to myself, it may just be that there is no there there (Cohen 122). On some level, recommender systems accept inconsistency as a fact about our taste. While they need to make successful predictions to stay in business, the systems are flexible enough to accommodate what seem to be gross inconsistencies in ratings.[21] Netflix does not care if you like Bartók and dislike Copeland; it just needs to be able to successfully predict if you are going to like Phillip Glass and the system is designed to figure this out even if Cohen cannot.[22]

Melchionne concludes that "[b]y embracing anti-introspectivism, we move toward a more naturalistic view of taste, with more room for confusion, indifference, and transience in our aesthetic preferences" ("On the Old Saw…" 140). Given his earlier call for "weak fallibilism," the point is that we dethrone introspection from its position of privilege rather than rejecting it entirely as a means to understanding our own taste, though one could certainly use recommender systems to avoid introspection entirely. Netflix is confident that I would like this, so I must have liked it.[23] Dominic McIver Lopes recommends that "it is prudent to turn to data about appreciative practice gathered through empirical methods that do not rely on subjective reports" (34).[24] With its "data behaviorism," Netflix and other recommender systems seem to be carrying out, at least in part, this kind of program; these sites do not just offer opportunities for aesthetic pleasure, but also ways of thinking about aesthetic practices that are

highly relevant to recent discussions in philosophical aesthetics, and not just in the area of aesthetic self-knowledge.

Netflix is 95% confident that I would like *The Hangover*.

NOTES

1. Robson suggests that most of us, most of the time, are entirely too confident about believing that we know what we like.
2. In addition to Melchionne, see also Irvin, Lopes, and Robson for detailed discussions of the fallibility of aesthetic introspection.
3. On the importance of consistency in one's personal taste, see Cohen.
4. Though she does not discuss Melchionne's arguments, Irvin explores similar research and is a bit more optimistic about the reliability of introspection with aesthetic experience, suggesting that aesthetic "experts" are less prone to errors in this regard than laypeople (41).
5. Shriver offers a standard definition: "Recommender systems filter information to help users make decisions when lacking personal experience or knowledge or when the set of choices is overwhelmingly large. In general, they take in data about users' past preferences for items and aggregate it to predict a user's preference for unknown items, often presenting a list of the items most likely to be preferred" (1).
6. Aggarawal offers a clear and thorough overview.
7. To be fair, these are partly in response to popular media Odes to the Geniuses of Silicon Valley; Madrigal (2014) is a typical example. Lawrence offers a clearly argued, philosophically-informed account that reaches some well-earned pessimistic conclusions about the loss of personal autonomy recommender systems, especially Netflix, might encourage.
8. Melchionne notes that they must work pretty well, given user satisfaction ("On The Old Saw…" 140). In "Aesthetic Choice" he suggests that we "keep the recommendations of the streaming firms at arm's length" (298), as these firms are concerned first about their own interests, not ours. This is fairly mild stuff compared to Frankfurt School-influenced cultural studies folks, who argue that an arm's length is nowhere near far enough from these algorithms that are undermining our taste and destroying our capacity to think critically. If Melchionne is Jeff Lebowski from *The Big Lebowski*, saying to Netflix, "That's just like, your opinion, man," the cultural studies theorists are Dr. Bennell at the end of *Invasion of the Body Snatchers*, screaming that the pod people are coming for you.
9 . Melchionne discusses this research as well (see "On the Old Saw…" 138).
10. Ariellii is especially critical of Netflix's lack of transparency with its algorithm, arguing that this kind of corporate secrecy is anti-democratic (89). He also argues that Netflix aims at increasing consumption more than it aims at increasing introspection on the part of its users (79).
11. See Alexander 90.
12. Morris provides a good overview (451).
13. Bobadilla has a good discussion.
14. Lawrence 359. Arnold terms this *user expression* versus *user behavior* and notes that Netflix sees former as a poor guide to taste (58).

15. See Arielli (84) on Netflix's shift from what he agrees are the flawed star ratings that Movielens still uses, as does Amazon.

16. Arnold discusses Yellin's claims (Arnold 58).

17. Gray quotes Yellin claiming in an interview that Netflix has put traditional markers of identity "in the garbage heap" (Gary 80).

18. Lawrence: "Netflix essentially mediates the common struggle to articulate why you like what you like by providing a brief account of what features you enjoy" (360).

19. Lawrence has a brief discussion of Walton in relation to Netflix's tags (360). This is obviously an area of rich interest that I will explore further in a subsequent paper.

20. Lawrence(359. See also Arielli 78-79. In theory, of course. I am not defending the actually existing tags as necessarily good. Lawrence offers a very compelling critique of their weaknesses.

21. See Alexander's discussion of Netflix's assumption that we are "consistently inconsistent" (84).

22. Cohen also discusses a third possibility, a "deep truth" that consistent aesthetic principles are simply not possible to find (122). This is a larger issue than I can tackle in this essay. What is also interesting about Cohen's example is that it is about *dislike*. Why does his wife dislike Copeland's music when she likes music by similar artists? Here, recommender systems have run into problems. As journalist Tom Vanderbilt explains in his *You May Also Like: Taste in an Age of Endless Choice*, these algorithms have had a very difficult time explaining dislike. Your shirt tells me what kind of shirts you like, Vanderbilt notes; it does not tell me what kind of shirts you dislike (89). It seems that disliking is a lot more complicated than liking; according to Vanderbilt, the music service Pandora experimented with allowing users to give a reason for a thumbs-down on a particular piece of music, and the system was overwhelmed with explanations (111). Granted, it is perhaps because Mr. Cohen is pressing her, but his wife has much more to say in Cohen's article about why she dislikes Copeland than why she likes Bartók. This is, apparently, typical, and recommender systems, Vanderbilt notes, are very interested in cracking the problem of dislike.

23. See Ross for a discussion about fears of a loss of autonomy when we heed the advice of critics instead of thinking for ourselves. Obviously, recommender systems raise similar concerns. Ross briefly mentions recommender systems, and intriguingly suggests that we look to "public" rather than "inward" sources to preserve the authenticity of our aesthetic responses, but she does not offer any detailed analysis of how recommender systems might function in this context (367; 370). Arielli questions the mind/outside of mind model, suggesting that recommender systems operate as a kind of extended mind (78). On the extended mind, see Clark. This is another topic worth exploring that is beyond the scope of this essay.

24. He recommends "the methods of the social sciences" (Lopes 39).

WORKS CITED

Aggarwal, C. C. *Recommender Systems: The Textbook*. Springer, 2016.

Alexander, Neta. "Catered to Your Future Self: Netflix's 'Predictive Personalization' and the Mathematization of Taste." *The Netflix Effect: Technology and Entertainment in the 21st Century,* edited by Kevin McDonald and Daniel Smith-Rowsey, Bloomsbury,

2016, pp. 81-98.
Arielli, Emanuele. "Taste and the Algorithm." *Italian Journal of Aesthetics,* vol. 46, 2018, pp. 77-97.
Arnold, Sarah, (2016) "Netflix and the Myth of Choice/Participation/Autonomy." *The Netflix Effect: Technology and Entertainment in the 21st Century,* edited by Kevin McDonald and Daniel Smith-Rowsey, Bloomsbury, 2016, pp. 49-62.
Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez. "Recommender Systems Survey." *Knowledge-Based Systems,* vol. 46, 2013, pp. 109-32.
Clark, Andy. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press, 2008.
Cohen, Ted. "On Consistency in One's Personal Aesthetics." *Aesthetics and Ethics: Essays at the Intersection,* edited by Jerome Levinson, Cornell University Press, 1999, pp. 106-125.
Gray, Jonathan. "Reviving Audience Studies." *Critical Studies in Media Communication,* vol. 34, no. 1, 2016, pp. 79-84.
Irvin, Sherry. "Is Aesthetic Experience Possible?" *Aesthetics and the Sciences of the Mind,* edited by Greg Currie, Matthew Kieran, Aaron Meskin, and John Robson, Oxford University Press, 2014, pp. 37-56.
Lawrence, Emily. "Everything is a Recommendation: Netflix, Altgenres and the Construction of Taste." *Knowledge Organization,* vol. 42, no. 5, 2015, pp. 358-64.
Lopes, Dominic Iver. "Feckless Reason." *Aesthetics and the Sciences of the Mind,* edited by Greg Currie, Matthew Kieran, Aaron Meskin, and John Robson, Oxford University Press, 2014, pp. 21-36.
Madrigal, Alexis C. How Netflix Reverse-Engineered Hollywood. *The Atlantic,* 2 Jan 2014, www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/.
Melchionne, Kevin. "Aesthetic Choice." *British Journal of Aesthetics,* vol. 57, no. 3, pp. 283-298.
———. "On the old saw 'I know nothing about art but I know what I like.'" *Journal of Aesthetics and Art Criticism,* vol. 68, no. 2, 2010, pp. 131-41.
Morris, Jeremy Wade. "Curation by Code: Infomediaries and the Data-Mining of Taste."*European Journal of Cultural Studies,* vol. 18, no. 4-5, 2015, pp. 446-463.
Robson, Jon. "Aesthetic Autonomy and Self-Aggrandizement." *Royal Institute of Philosophy Supplement,* vol. 75, 2014, pp. 3-28.
Ross, Stephanie. "Comparing and Sharing Taste: Reflections on Critical Advice." *Journal of Aesthetics and Art Criticism*, vol. 70, no. 4, 2012, pp. 363-71.
Shriver, David. "Assessing the Quality and Stability of Recommender Systems." *Computer Science and Engineering: Theses, Dissertations, and Student Research.* 2018, digitalcommons.unl.edu/computerscidiss/147.
Vanderbilt, Tom. *You May Also Like: Taste in an Age of Endless Choice*. Vintage, 2017.
Walton, Kendall L. (1970) "Categories of Art." *The Philosophical Review,* vol. 79, no. 3, 1970, pp. 334-367.
Wilson, T.D. and J. W. Schooler. "Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions." *Journal of Personality and Psychology,* vol. 60, no. 2, 1991, pp. 181-92.

Rational Resolve as Magnification: A Response to Holton

Daniel Grasso
Saint Louis University

In his paper "Rational Resolve," Richard Holton raises two questions about resolutions: first, the descriptive question of how they work, and second, the normative question of whether it is rational to persist in them. Holton answers the latter, but he leaves the former question alone, assuming that resolutions work. However, central to Holton's account of rational resolve is the idea that resolutions "entrench" a previous rational decision, that any rationality in obedience to resolutions gets its rationality from some previous source. Because Holton assumes the descriptive account of this previous source, the origins of his argument remain blurry. The aim of this paper is to outline the foundation Holton assumes, and to give a descriptive account of how and why resolutions work.

In order to give this account, I will argue that Holton is wrong to claim that resolutions add no extra features of rationality and will instead propose that resolutions add a multiplication or magnification feature, one dependent on the reason the resolution is based upon. I will argue that this extra feature of resolution works by an agent knowingly entering into a test of will—a test of will that I will show cannot be escaped.

HOLTON AND THE WORK OF RESOLUTIONS

In investigating the rationality of resolutions, Holton first considers the nature of practical rationality, and then pivots to the bootstrapping problem. The bootstrapping problem surmises that if resolutions are just a special kind of intention, then they should not create extra reasons to act that we would not have

otherwise (Holton 513). If I have no reason to jump in a lake then certainly intending or resolving to jump in a lake will not furnish some new reason to do so. Holton sees two ways out of the bootstrapping problem: either resolutions *do* create extra reasons or we embrace a two-tier account of resolutions. Holton ultimately rejects that resolutions are able to provide extra reasons on their own. Instead, he endorses a two-tier method; if resolving to V is rational, it is because it was rational to decide to V in the *first place* and the rationality of the original decision is *conferred* onto the resolution. Holton claims that on the two-tier account resolutions do not *create* reasons but merely *entrench* the original rationality through a transfer principle (515). Holton's account is therefore dependent on what I will explore: an account of resolutions and why they work in the first place.

To push back on Holton's entrenchment account, we may ask: why do resolutions entrench reasons? Why do they not merely rest on the surface, blown away with other intentions in the face of foreseen temptation when it may be rational for our judgement to shift (Holton 507)? Faced with the enticing Sirens, Odysseus prevented his own suicidal action by tying himself to the mast. Odysseus was held by rope, but with what do resolutions bind us? Holton's entrenchment presumes some sort of *extra* strength or durability is present. The rationality from the original intention is somehow more protected than if it were *not* entrenched. So, how can a resolution help protect the original intention more than one non-resolved, especially if Holton denies resolutions create any extra reasons?

RESOLUTIONS AS SELF-PROMISES

If a resolution *did* furnish extra reasons, then the bootstrapping problem would be solved and there would be no real mystery. Consider something we do think creates extra reasons—promises. If I promise you I will jump in a lake then we do think that I have created a reason, even though I had none before. What if it is argued that resolutions are simply *self*-promises? This would address the first horn of the dilemma; anything we resolve (promise ourselves) to do would immediately become rational. This seems wrong; so, we attend to self-promises next.

Labeling resolutions as self-promises is enticing, yet the structure of a self-promise is inherently impotent, which would leave resolutions just as impotent to entrench rationality. Unlike a standard promise between two parties, within a self-promise a single indiviudal plays both roles: promisor and promisee. In a standard promise, if the promisee releases the promisor the reason for acting is terminated. The promise can be settled by fulfilling the promise or by being released. So, if I am both the promisee and the promisor, how can a self-promise bind me in the face of temptation? If resolutions are merely self-promises, what is to stop me (the promisor) from releasing myself (the promisee) from my promise?

If resolutions were self-promises we would expect broken resolutions to have

no impact on us, akin to the way I am no longer impacted by a settled promise. Yet, somehow resolutions remain entrenched and continue to hold psychological sway over us in a way that settled promises do not. It seems where we can be released from promises we can only *break* resolutions. A broken resolution is *not* the same as a settled self-promise.[1] I suggest that resolutions, unlike self-promises, have what I will call a "one-way feature" that, once lowered, cannot be lifted off. This one-way feature will provide the descriptive account of *why* resolutions work, the description that grounds Holton's normative account and provides the explanation for entrenchment.

In order to describe how this entrenching one-way feature works, I must first show that while resolutions do not *create* extra reasons, they do add something. This something may be vaguely described for the time being as some *magnification* of mental grief, a feeling of dissonance. I will spend the rest of this paper explaining *how* and *why* resolutions generate this extra feature.

HOW RESOLUTIONS WORK: A DEPENDENT EXTRA FEATURE

Holton was tempted to recognize some extra feature of resolution, but he ultimately rejected it. I will argue that the force of some extra feature was felt because it exists. As argued above, the extra feature that makes a resolution's entrenchment resilient is not the *creative* feature of promises.[2] However, it is a subtle feature, one that is present in every resolution but its effect not always noticeable due to its dependence on the original reasoning. Consider three people: Alvin, Beatrice, and Christine.

Alvin wakes up and desires a salad for lunch. The whim strikes him, and he resolves to order one at lunch. When lunch arrives, Alvin realizes his appetite has changed and that a hamburger is preferred; he orders the hamburger instead. What work has this resolution done? Since Alvin made his lunch resolution based on nothing but incidental whim and resolutions do not *create* reasons, this resolution is empty. Alvin will not feel any dissonance; he will not feel any *extra* weight due to his resolution. The obstinate resolver aside, it is more likely that Alvin will brush off the idea as exactly what it was—a whim.

Beatrice is in a different situation. Beatrice has been told by her doctor that her cholesterol is dangerously high. If Beatrice wishes to live a longer and an overall better life, she is told she must eat a salad each day for lunch and avoid red meat. Beatrice believes this assessment and does want to live a longer and overall better life; at a distance from temptation, she understands and agrees she must eat the salad, even though she loathes it. Beatrice wakes up and, considering this diagnosis, resolves to eat a salad for lunch, knowing that she will be tempted to order her habitual hamburger instead. Beatrice arrives at lunch and breaks her resolution. The resolution itself was not strong enough to hold Beatrice back. Will Beatrice feel dissonance? In this case, because the original reason to order the salad

was strong, the resolution *will* add to his dissonance. Not only will Beatrice feel the weight of knowingly harming her health, she will also feel the *extra* weight of breaking her resolution, of being weak willed.

Christine is in the identical situation as Beatrice but with one crucial difference. Christine does *not* resolve to eat salad for lunch. Just like Beatrice, Christine believes the doctor and thinks it is best to eat the salads, but, unlike Beatrice, she has not resolved to order a salad.[3] Unlike Beatrice, Christine is in a vague middle ground, knowing what she should do, believing why she needs to do it, and desiring the results of the action yet still avoiding committing to it. Avoiding taking a stance, Christine is in a position akin to Harry Frankfurt's "wanton," the essential characteristic being "that he does not care about his will (68)."[4] When Christine orders a hamburger instead of the salad, she will feel and think *different* than Beatrice. While Christine may still understand what they are doing is irrational, they will not have broken a resolution. They will not feel the extra weight of weakness that comes with explicitly failing a test of will; instead, they are more likely to continue thinking: "Yes, I should get the salad, and when I *really* want to I will." Unlike Beatrice, Christine is poised to remain in self-deceit. Without resolution, Christine can begin to bury their head in the sand, moving further from rational action.

Between our characters we can examine the subtle extra feature of resolutions. Resolutions are not some powerful extra reason that overwhelms the agent, for Beatrice failed at her task just like the unresolved Christine. Yet, there must be some additional feature within resolve or else Beatrice and Christine would have an identical psychological response. The answer lies in case Alvin. Even though Alvin resolved just like Beatrice, we don't think Alvin *should* feel any dissonance—Alvin's *resolution* was null because their *rationality* was null. Alvin's resolution was made on a whim with no real reason to eat the salad, the same act Beatrice had *very good* reason to do. The extra feature of a resolution is not a *creative* one, one that creates reasons *ex nihilo*; it is a multiplication or a magnification feature, one *dependent* on the reason it is employed. Alvin's resolution was ineffective because his reason was empty; it was zero, and, no matter how large the multiplier, the result will always be zero. No matter how powerful the microscope, nothing will be seen if there is no specimen on the slide.

This extra feature of resolutions is not a silver bullet, one that immediately steels the will against temptation; instead, because of its dependence on rationality, it helps transfer the strength of the original reason that is determined through rational deliberation—just like Holton suggested. This account, seeks to meet rationality and the resolutions that flow from it at the most basic level— the deliberative process made in the face of reality.[5]

If the original reason for making the resolution is believed rational by the agent, resolution is "effective." If it is not rational then the resolution is still present, but its effect is not seen because of its dependence on the original

rationality.[6] Likewise, the stronger the reason the more effective the resolution. *By "effective" I do not mean successful*. The effect of a resolution is not to infallibly steel the will so that the action is achieved; the will is vital, but it is separate from the feature of resolutions I am discussing here.[7] Instead, effective here means that the resolution is activated, that it does have an effect on the situation.

Resolutions have a function akin to sight; they orient the agent, clarify their problem, and hold it in magnified focus, setting the stage for the agent in calling down a challenge on themselves to achieve this clear task. By magnifying the rationality of the original decision, resolutions clarify and multiply the motivational force of the original rationality, this extra clarification and force enlarge the dissonance felt through failure. Like Holton said, this is an entrenchment, a transfer of knowledge and rationality, but the trench's depth reflects the rationality of the original resolve. Now we can turn to the final question: why don't rational judgment shifts in the face of temptation release us from the extra feature of resolutions; in short, *why* do resolutions work?

WHY RESOLUTIONS WORK: A ONE-WAY FEATURE OF KNOWLEDGE

If what I have argued is correct, then resolutions really do add an extra feature, a kind of amplification of mental dissonance which is perpetrated through magnification of the original rationality. We have all broken resolutions and felt a sense of guilt, not only about the failure to do what was rational but about the failure *itself*. If I resolve to quit smoking every day and fail every day, after a month or two, I am more likely to quit *resolving* than to quit smoking. If I must be weak willed then at least let me be honest about it; why keep resolving when all it makes me is a hypocrite?

I feel this way exactly because the resolution has played its role well. Each time I resolve to stop smoking I enter into a test of will; each time I smoke I fail the test. I quit resolving instead of smoking because I no longer wish to be a failure or at least to so clearly see myself as one. If I continue to smoke, knowing it is irrational to do so, but do not resolve to stop, then I have not entered into a test of will; therefore, I cannot fail it. I may be slowly ruining my health and *implicitly* weak willed but at least I am not *explicitly* weak willed. Like Christine in the example above, I may still feel or even *know* smoking is irrational but until I declare I am *really* trying I can live in an uncommitted vagueness—one where I try without resolving, one I can hide in when I fail. At the very least, a position of non-resolution is a set up for self-deceit and wantonness, a position that resolution fights against.

Resolutions pull us from vagueness and this non-committal attitude. They magnify and clarify the situation and the rational decision made to address it. They acknowledge the problem, decide on a solution, hold it in focus, set clear and definite rules of the test, and, finally, get our agreement to enter into this test of

will. We metaphorically "see" in better definition and sharpness; a vision that makes our commitment unequivocal instead of amorphous and vague. This may still seem the *how* of the work of resolutions, but it is the nature of how resolutions work that explain *why* they work. The one-way feature of resolutions, the reason that a resolution cannot be escaped like a self-promise, is that of knowledge. Resolutions work by an agent's knowledge of entering into a test of will, one that is based in rationality.

To describe this one-way feature of knowledge, consider a typical resolution. If I resolve to quit smoking,[8] I am entering into a type of test. There are only so many ways to exit this test: I can quit smoking and pass; I can smoke and fail; or I can quit the test, claiming it was bunk—that my judgment was off earlier or that it is more rational now. Unless there was a substantial change to justify the judgment shift, the test of resolution remains.[9] Pass, fail, or quit, we are still affected by our rational resolutions because they show us our failings, and we know it; we ask them to judge us and they do. We cannot be released from them like a self-promise for the reasons shown previously, and we cannot worm out of them through a foreseen judgement shift for this is exactly why we asked them to judge us in the first place. But if we were to try, where would we look?

Consider the most ambiguous exit from the test of resolution—that of quitting. This seems to be the best way out; after all, why can't I quit a test I set up and I am grading? If I am the test-taker and the test administrator, why not tear up the test and walk out? The answer is that you can. You can claim the test was bunk and walk away, but I argue that you cannot "forget" your weakness, the action of quitting a resolution you knew to be rational. In normal cases, a proper rational resolution, one that fulfills the characteristics I outlined above, will not allow an agent to dive totally back into self-deceit without dissonance or shame.[10] The act of resolving clarifies your reasoning and endorses it; to then decry the test because you no longer like it is to try to escape back to vagueness, to a place where deceit is easier.[11]

I am not claiming that self-deceit is impossible after a resolution. I am claiming that in standard cases an agent cannot immediately and indefinitely pretend that the resolution was irrational to avoid the dissonance of the one-way extra feature of a resolution.[12] This self-deceit may certainly take place over a longer period of time. I may convince myself that my smoking is not so bad after all, but it will not be instantaneous to dodge a feeling of dissonance at the knowledge of my weakness of will. There may be other types of forgetfulness that complicate this account, but I will have to leave these objections for a different place.[13] The test of resolve is powered by the one-way feature of knowledge; it is a trap that once we enter into, we have no escape. We have only to pass, fail, or know we were too afraid to find out which of the two we'd score—its own admittance of failure. Like the proverbial apple, the price of knowledge is that it is not so easily forgotten.

So, resolutions are entrenchments, but in order to explain why the

entrenchment works we have embraced a subtle extra feature of resolutions, a feature dependent on the particular situation and its rationality. This feature clarifies and magnifies the rationality of our original resolve, multiplying the justificatory and motivational force it finds there. This multiplication factor amplifies dissonance when we fail in our resolve due to the extra clarity and force it brings. I have described this extra factor as a one-way feature; one-way because once it is entered into it cannot be so easily escaped. This one-way feature is plausible because resolutions are powered and sustained through knowledge—the knowledge that we are asking ourselves to judge ourselves—our resolutions acting as our own indelible rulers, mirrors reflecting acts we'd rather not see, instruments in the arduous call to know thyself.

NOTES

1. One may say we can *quit* resolutions but I will show how this is not an escape route later on.
2. It is more likely closer to what Holton had in mind when he mentioned extra reasons "even though these are not reasons that the agent will consider" when in the face of temptation. Holton turns from here to what he calls unreflective dispositions. If there is an extra feature of resolution it is that it gives us a reason to be unreflective and arational.
3. Note that this is not a case of someone who judges that a short life lived on red meat is better than a long life lived on vegetables.
4. Like the wanton, Christine does not want to decide her will. She wants to remain in a place characterized by indecision and lack of second order will. This may not be the case for Christine currently, but a position of non-resolution certainly sets Christine up for wantonness.
5. This account meets the standard that Holton says extra features of resolve don't—that they are not descriptively accurate. Holton was concerned that though we prefer to be resolute rather than weak, this preference is not strong enough to outweigh temptation in which it would be rational to persist. I agree and this description can account for this, allowing for both dissonance in failure and yet failure all the same. It also addresses many difficulties surrounding resolutions: why are they effective for some but not others; why do they work in certain situations but not others; why do they seem to be on a sliding scale of importance, working only sometimes?
6. It is also plausible that a resolution made without a reason is not a resolution at all but simply a whim or passing fancy. But this is a peripheral contention and can be put aside for now.
7. There is an easy conflation of resolutions as I am discussing and resolve as will power. I offer a quick metaphor. If you were driving in a downpour, the side of resolutions I am investigating would be the headlights and the windshield wipers, but the will is the transmission of the car. Resolutions help you see where you are going, but no matter how clear the sight, without the transmission (resolution in the sense of willpower) you go nowhere. The interaction of the will and resolutions is worthy of much examination elsewhere. Some of this examination is taken up by Holton in "How is Strength of Will Possible?"

8. Let's assume I consider this the rational and overall best decision.

9. See above for Holton's rules of thumb for legitimate judgment shifts. In this case it may look like a massive medical reversal that smoking cigarettes is actually beneficial for your health. If health concerns were the main reason for quitting, the resolution could be quit without any feeling of dissonance.

10. A weaker claim would be no *rational* agent could escape; I think even non-totally deluded irrational agents will still not escape scot-free. When they return to their rationality later on the feeling will be waiting for them. I'll discuss some more extreme irrational cases below.

11. This is assuming a case where there is no legitimate judgment shift. In this case the change of mind is most likely from the exact temptation that the resolution was made to avoid.

12. Normal cases of self-deceit in the face of a rational resolution are not immediate. To bury one's head in the sand of deceit is a very real possibility, but it is a process and not done in one swift act. There may be abnormal cases of agent's particularly practiced at self-deceit but I set them aside. It is more likely that an agent is skilled in rationalization; in which case they may slip the bounds. In fact, the study of when it is rational to stop rationalizing is the aim of Holton's "Rational Resolve." But here we would transition from a descriptive account of resolutions to an account of when they are rational and tricks of how to persist in resolutions—something outside our purview. Yet, the weaker claim of no rational agent escaping can remain intact with Holton's account of rational arationality.

13. There may be a version of legitimate forgetfulness, not one of cognizant self-deceit, but of perhaps busyness. Consider Jeremy from Christine Korsgaard's footnote story in *The Normativity of Instrumental Reason.* Jeremy makes a positive resolution like: "I resolve to study for my exam tonight." Restless at the moment, he goes for a walk. He stumbles into a friend who invites him for a drink; Jeremy can afford the time for one so he goes. Jeremy is then bounced from place to place until the night is gone and he hasn't studied. Has Jeremy forgotten his resolution in a way that escapes the test of resolution? I'll follow Korsgaard in suggesting this a problem of an inactive will. I imagine that when the resolution is remembered, when Jeremy awakes the next day, he will find not only the consequences of not studying waiting for him but also an extra dissonance at being so weak willed as to *not even remember* his resolution. I suspect the answer lies in how resolutions interact with the will, that if an agent is so forgetful of resolutions then the problem rests in their will power, not in the structure of resolutions. However, this type of forgetfulness of a resolution does add an interesting case worthy of longer treatment elsewhere.

WORKS CITED

Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy,* vol. 68, no. 1, 1971, pp. 5-20.

Holton, Richard. "Rational Resolve." *The Philosophical Review,* vol. 113, no. 4, 2004, pp. 507-535.

Korsgaard, Christine. "The Normativity of Instrumental Reason." *Ethics and Practical Reason*, edited by Garrett Cullity & Berys Gaut, Clarendon Press, 1997, pp. 215-254.

The Dual Role of Kant's Imagination and Nature as Art

Jackson Hoerth
Tyler Junior College

The status of Kantian aesthetics remains contentious despite renewed attention in recent decades. Of particular concern is the value of aesthetic experience for the judging subject. While answers to this topic vary, one consistent feature is the position that aesthetics is valuable only insofar as it reveals something about the subject's capacities, be it in relation to cognition or some other interpretive capacity.[1] The full value of aesthetic experience, I argue, is not only in relation to the subject; rather, it also reveals a fundamental order in nature independent of our capacities to structure sensibility according to our laws. Nature as independently ordered, or 'nature as art,'[2] achieves the full value of aesthetics within Kant's system.

Recognizing an independent order in nature is critical because the goal of aesthetics for Kant is to provide a bridge between sensible nature and our rational ends. Aesthetics is the proposed way for humanity to become aware of this connection, as if there is an order in nature that makes it amenable to our rational ends. Therefore, I aim to provide an answer to the problem of bridging this gap by demonstrating how aesthetics provides a way for the judging subject to glimpse the order already given in nature. I argue that this can only be achieved by a particular view of the faculty of imagination as it works in the shaping of experience. This particular role is what I call the 'dual role of the imagination.' It represents the imagination's capacity as not only responsible for the sensible content of a judgment, but also its capacity to bring the form, or order, of that judgment as well. This function of the imagination is critical to recognizing the

order already in nature, thereby bridging the gap between nature and reason, and completing the task set for aesthetics within Kantian philosophy.

THE IMAGINATION IN THE SCHEMATISM

The 'Schematism' is offered as a solution to the problem of the actual application of a concept of the understanding to the sensible manifold. Schema are proposed as a 'third thing' that is amenable to both sensibility and the understanding.[3] To summarize the solution: schemata contain the sensible conditions for the application of the concept to the object of representation. Kant reiterates a point from the Deduction, namely that sensible intuitions need to be shaped into an amenable form for our conceptual apparatus. The difference being that in the Schematism we are given schemata as the actual features that bridge the gap between represented objects and our concepts. Kant also indicates a procedure, schematism, in which the understanding operates upon a schema of a represented object. Based on the arguments provided in the Deduction, schemata occupy a similar place as the faculty of the imagination, which mediates between the sensible manifold and the understanding in grounding the possibilities of cognition. The difference between the sections is the possibility of applying concepts to the sensible manifold and their actual application to real objects.[4]

That Kant intends the imagination to fulfill a critical role in the actual application of concepts is evinced when Kant specifically names the schema as "always only a product of the imagination."[5] In the Schematism there is a role for the imagination in the actual application of concept to sensible object. I contend that this role actually goes beyond the picture of the imagination as a merely mediating faculty.[6] That is, even in the Schematism, we can view the imagination in a form creating capacity. In claiming that schemata are the product of the imagination, Kant is pointing towards a productive power of the imagination in its ability to form the order required to make sensible experience translatable to our concepts of the understanding. Another way to put this is that the imagination's schemata make sensibility governed by rules without the schemata themselves being rules, which can only be concepts of the understanding.[7] Kant indicates that a requirement for the actual application of concepts to sensible objects is harmony, or a 'condition' for harmony between the object and our concept.[8] With that harmony offered between sensibility and concept, experience is possible. The faculty responsible for offering the instances of this harmony is the imagination because schemata, as products of the imagination, are the bridge in the first *Critique* for the real application of a concepts to a sensible object. It is in this way that the Schematism represents a formal power of the imagination, albeit a capacity that in the first *Critique* is guided by a concept, or a rule of the understanding.

REFLECTIVE JUDGMENT AND NATURE AS ART

The Schematism is not the final word on either the capacity of the imagination or the power of judgment. The third *Critique* demonstrates this through Kant's position that judgment itself demonstrates a rule or lawfulness separate from cognition. This rule, Kant writes, "would have to be the concept of a purposiveness of nature in behalf of our faculty for cognizing it."[9] That is, the power of judgment requires that sensibility conform to our conceptual abilities.

Judgment giving its rule to itself assumes this conformity and seeks it in order to ground sensible experience. What is key here is that judgment assumes a lawfulness that is not the result of a rule of the understanding, even in cognition. However, the difficulty, as in the Schematism, is that the process of cognition veils this power of judgment by directing it through the understanding's laws.[10] As we will see, judgment is capable of its own lawfulness in its interaction with nature; and, this lawfulness independent of a conceptual rule will yield a different experience with nature.

While Kant does not explicitly address the imagination in reference to this power of judgment in the First Introduction, I argue that this principle of judgment is ultimately based on the dual role of the imagination. I argue this by carrying over the first *Critique*'s connection between the imagination and judgment. That is, the imagination not only demonstrates the transcendental conditions for uniting sense and concept in the Deduction, but also provides their actual unity in the Schematism. The real conformity of sensibility to our concepts involves the imagination recognizing the points of harmony and then providing the form for their homogeneity. With the principle of judgment in the First Introduction, Kant simply reveals the assumption judgment makes independent of cognition, namely that nature is purposive towards our cognitive faculties. This principle of judgment is the same recognition made by the imagination in the Schematism, only now asserted as a principle of judgment. Therefore, we can view the application of this principle of nature's purposiveness as an act of the imagination's productive dual role. The difference is simply the lack of a guiding concept of the understanding. Now, the imagination through the principle of judgment is capable of shaping a form based on this harmony that remains independent of cognition.

Differentiating this process of judgment from the Schematism, we say that in the new case judgment is reflective, rather than determinate. The difference being that determinate judgment places the particular under a universal, as we see in the process of cognition.

Reflection, however, seeks a universal based on the particular. Reflective judgment, without the guidance of a determinate concept and thereby exercising the concept of nature's purposiveness, provides a different experience of sensible nature. Kant explains the distinction as follows:

> The reflecting power of judgment thus proceeds with given appearances, in order to bring them under empirical concepts of determinate natural things, *not schematically*, but **technically**, not as it were merely mechanically, like an instrument, but **artistically**, in accordance with the general but at the same time indeterminate principle of a purposive arrangement of nature in a system, as it were for the benefit of our power of judgment, in the suitability of its particular laws (about which the understanding has nothing to say) for the possibility of experience as a system[.][11]

Kant differentiates between a view of nature that is schematic and mechanical and one that is technical and artistic. The schematic-mechanical view is the result of sensible nature conforming to the rules of the understanding. It is nature in accordance with *our* order. However, there is a technical-artistic view as well, which is the result of the reflecting power of judgment. This is the view of nature that assumes nature's purposiveness, the harmony between it and our cognitive capacities, but without a concept present. In this way, it is the exercise of the imagination's dual role capacity. Like in the Schematism, the imagination recognizes harmony between sensible nature and a system of order; however, unlike earlier, this harmony is independent of any concept of the understanding. Without the guidance of the concept, the form of its amenability to nature is still recognized and produced. And it is in this reflective activity that we can see the pure power of the imagination's ability to shape not only sensible content, but also the form that unifies the natural order with the human order. This is ultimately that value of aesthetic experience that reveals the power of the imagination to bring nature's independent ordering to our awareness. [12]

Response to the Division of the Imagination

Without the imagination's dual role, we risk concealing the bridge between the natural order and the human order. However, the imagination as a form-bringing faculty is not the only manner in which the imagination has been interpreted. One approach is to take the imagination's schematism capacities as fundamentally separate from the imagination's role in aesthetic judgment, which this approach claims is to provide an interpretive basis for meaning in the realm of politics, history, and culture. I will call this the 'interpretive' view.[13] The interpretive view takes the imagination's role in the schematism as too limiting for the free activity that it demonstrates in aesthetic reflection. Providing the schema that allows for the application of a concept to a sensible object, the imagination serves the understanding in its function to order sensible nature.[14] Therefore, this view argues, any function of the imagination in the Schematism is subsumed by the understanding and not fit for aesthetic consideration. Instead, the capacities of the

imagination demonstrated in the third *Critique* are fully separate from those given in the Schematism because Kant intends the imagination in aesthetic judgment to demonstrate a freedom that, by the interpretive view, is not amenable with the more limited faculty of the first *Critique*.[15]

I argue that this division between the Schematism and aesthetic judgment is too strong. In the Schematism the imagination, admittedly, operates in service of the understanding by providing the form that makes sensible material amenable for conceptual application. However, in the case of reflective judgment, a concept is not necessarily required as the motivating proof of a judgment. A reflective judgment might have a concept present, as in determining that a beautiful rose is, in fact, a rose. But this is not the same as saying that the imagination's formal operations are necessitated by this concept. Rather, aesthetic judgments are differentiated from empirical judgments in that the imagination recognizes another order, independent of our understanding, and shapes the necessary harmonious form for our judgments based on this recognition. The difference being that the imagination is not guided by a rule of the understanding, but a systematic order in nature. Thus, what the imagination brings to aesthetic judgment is not the same as the Schematism; however, it does demonstrate the same *capacity*. Therefore, the strong division of the interpretive view threatens to miss a key feature of aesthetic judgment, nature's independent order, which is only recognizable if we take the imagination as dual role faculty, bringing both form and content to aesthetic judgment.

THE VALUE OF AESTHETICS

In my final section I would like to go back to nature as art, grounded by the dual role imagination, and its relation to the importance of aesthetics within the Kantian system. I claim that the artistic view of nature is the key to bridging the gap between nature and reason and that the artistic view can only be fully recognized through the dual role of the imagination. Therefore, fully realizing the value that Kant intends for his aesthetics to have within his systematic philosophy depends on adopting the dual role imagination.

In the Introduction of the third *Critique*, Kant claims that reflective judgment is the way in which the gap between nature and reason can be bridged. Now, we can say nature separated from reason is mechanical-schematic nature. What is needed is a bridge that is both grounded in nature and capable of touching reason. Aesthetic judgment, as a form of reflective judgment, fills that role. "The reflecting power of judgment, given its nature," according to Kant, "could not undertake to **classify** the whole of nature according to its empirical differences if it did not presuppose that nature itself **specifies** its transcendental laws in accordance with some principle."[16] Nature cannot be experienced holistically, given its empirical diversity, unless it is assumed that nature itself specifies its own lawfulness

independent of our conceptual ordering. It is the power of judgment that takes the aggregate of nature and forms a systematic picture that makes categorization possible. But, as Kant points out, "such a classification is not a common experience, but an artistic one, nature, to the extent that it is thought of as specifying itself in accordance with such a principle, is also regarded as **art**[.]"[17] This is what Kant terms as the 'purposiveness of nature,' and the only way in which we can glimpse anything based on nature's purposiveness is through this artistic, not mechanistic, view of nature. As nature, it remains rooted in sensibility, yet it also demonstrates, albeit indeterminately to us, its own lawfulness, extending the bridge farther out into the gap between nature and reason. Ultimately, this points to the value of aesthetics for Kant. It is part of how we operate given the assumption that nature operates as a system independent of concepts.

The imagination's dual role strengthens this claim and returns to an issue raised at the beginning of the paper—the overemphasis on cognition in Kantian aesthetics. The imagination's dual role brings into question the power of cognition upon aesthetic reflection. After all, the imagination in its dual capacity provides both the sensible content and the formal ordering for aesthetic judgments. The concept of the understanding, while present, only requires that the form of the imagination remain amenable to its own rule. That is, the concept of the understanding is not the force of proof for aesthetic reflection.[18] My position allows for the weakening of the understanding for aesthetic judgment because the imagination, in its dual role, calls upon the ordered unity it discovers in nature – an ordering that is beyond concepts of the understanding, since it cannot interact with nature without the form provided by imagination, schematic or otherwise. That is, the imagination, in its interaction with certain particulars in nature, recognizes a real feature of its systematic ordering. This is what I take to be the power behind my argument and where it is differentiated from competing views of the imagination.[19] Rather than merely reflective upon our own cognitive faculty in general, aesthetic reflection through the dual role of the imagination also reveals a fundamental feature of nature's own lawful ordering; however, it cannot be demonstrated as a specific law, but only more generally as lawful.

While this claim runs against some statements that appear in the third *Critique*, I believe we can read some of the problematic passages in a more permissible manner, opening room for the benefits of aesthetics given the imagination's dual role. First, when discussing the purposiveness of nature Kant states, "the end is not posited in the object at all, but strictly in the subject and indeed in its mere capacity for reflecting."[20] This might appear to rule out my interpretation since I am asserting that aesthetic reflection, which assumes the purposiveness of nature, reveals a glimpse of nature's order. First, the imagination's form does not discover nature's order within an object or experience, by this I mean an intuition that has been schematized and applied a concept. Rather, the imagination, prior to any schematization based on a concept of the understanding, recognizes an order in the

sensible presentation of nature as a whole, represented by the particular, and thereby shapes the form to capture this order. So, the systematicity is not posited to an object, but to the whole of nature, artistically conceived.[21] However, my key position is that the form provided through the dual role of the imagination is not determinate in the same sense as reason or the understanding. Yet, it does reveal an order to nature as a whole, revealed through the particular that exceeds the capacity of cognition alone.

CONCLUSION

Aesthetics is thus not fully revealed through its connection with cognition. Its full account can only be given through its relationship with nature's purposiveness and its connection with reason. While the link to reason extends beyond the immediate scope of this project, I have demonstrated the grounds for a renewed emphasis on the connection between nature's purposiveness and the imagination's role in providing the form in aesthetic judgment.

Without the consideration of the sensible aspects of aesthetic reflection, we risk missing the imagination's unique power to bring both sensible content and form. Further, this unique capacity of the imagination is the key to the artistic view of nature, which is vital to realizing the role of aesthetics as an avenue for making possible our human and moral ends within the realm of nature. Artistic nature, guided by the principle of purposiveness, is experienced as amenable for our systems of order applied by reason and the understanding. This amenability is the value of aesthetics; however, it is recognized through the dual role of the imagination guided by the ordering of nature, productive in the crafting of forms, and ultimately harmonious with the understanding.

NOTES

1. In this concern, I find myself in agreement with Lara Ostaric. In "The Free Harmony of the Faculties and the Primacy of Imagination in Kant's Aesthetic Judgment," her related criticism of Kant scholarship states, "Most Kant commenters focus on the epistemological import of Kant's claim and, hence, argue that the question of what it means for aesthetic judgment to have its own a priori principle can best be answered by exploring the connection between aesthetic judgment and cognition" (Ostaric 1376).

2. Wording taken from the 'First Introduction' of Kant's third *Critique*, the *Critique of the Power of Judgment*.

3. Kant at (A139/B178-A140/B179) in the *Critique of Pure Reason* states: "[P]ure concepts *a priori*, in addition to the function of the understanding in the category, must also contain *a priori* formal conditions of sensibility (namely of the inner sense) that contain the general condition under which alone the category can be applied to any object. We will call this formal and pure condition of the sensibility, to which the use of the concept of the understanding is restricted, the **schema** of this concept of the understanding,

and we will call the procedure of the understanding with these schemata the **schematism** of the pure understanding."

4. I share this view with Sarah Gibbons. In *Kant's Theory of Imagination,* she writes, "The Schematism turns our focus away from the nature and role of the categories as forms of unity in judgment to the problem of *how* it is possible to apply them to material that is sensibly intuited" (Gibbons 53).

5. Kant, KRV A139/B178-A140/B179.

6. This picture of the imagination is largely due to Kant's changes to the Deduction in the B-Edition. However, the role of the imagination in the Schematism remains consistent in both versions. Therefore, it is reasonable to assume Kant intended the imagination in the Schematism to be compatible across editions.

7. Kant's statement is as follows: [I]n addition to the rule... which is given in the pure concept of the understanding, it can at the same time indicate *a priori* the case to which the rule ought to be applied... it must at the same time offer a general but sufficient characterization of the conditions under which objects in harmony with those concepts can be given, for otherwise they would be without all content, and thus mere logical forms and not pure concepts of the understanding. See KDU A135-136/B174-175.

8. Gibbons points out the relationship between the concepts and the given just before the Schematism. She writes, "Schematism, the, specifies not (conceptual) rules, but the conditions for the recognition of *instances*; it does so by specifying the conditions under which the (spatio-temporal) *given* is 'in harmony' with the categories" (Gibbons 61).

9. Kant, KdU 20:202.

10. At A141/B180-1 of the Schematism, Kant claims that the form of the schematism of the understanding and appearances can be unveiled only with difficulty. More notably, he writes "We can say only this much: the **image** is a product of the empirical faculty of the productive imagination, the **schema** of sensible concepts (such as figures in space) is a product... of pure *a priori* imagination[.]" The function of the imagination in the schematism remains hidden through its necessary connection with a concept of the understanding.

11. Kant, KdU 20:213-214. Italics added.

12. A potential objection here is that by providing a formal aspect of experience, the imagination demonstrates a particular law strictly reserved for reason and the understanding. The form provided by the imagination in aesthetic judgment is not determinate in the same manner as the law provided in cognition. Rather than legislating order upon sensibility, the imagination is receptive to an independent order found in particular instances of nature and shapes this order to be harmonious with our own manner of ordering sensibility. In this way, the imagination is not legislative in the same sense as reason or the understanding. Further, it does not exclude the understanding as a necessary component for cognition. Instead, the harmony between the imagination's lawfulness and the understanding in aesthetic judgment point beyond the experience of nature as a mechanism and to an artistic view of nature. To attain the experience of the artistic view of nature, however, we must have a way to differentiate what is given by the rule of the understanding and what is expressed in the lawfulness of the imagination. Ostaric alternatively describes the outcome of aesthetic judgment as, "grasping that the lawfulness of the imagination is consistent with the discursive demands of the understanding... and, moreover, that the connections of the imagination move well beyond those demands"

(Ostaric 1394).

13. See Rudolph Makkreel's *Interpretation and Imagination in Kant* as a key example of this approach.

14. Makkreel considers the schemata of the imagination in this context as 'semantical rules' for the 'grammatical rules' provided by the understanding in concept application (Makkreel 41).

15. We see evidence of this strong division in Makkreel. He writes, "the extent to which the conditions of the first *Critique* can be transferred to the third *Critique* is limited by the different functions assigned to the imagination in its aesthetic setting" (Makkreel 49). The functions of the imagination in aesthetic judgment make comparison from objective experience to aesthetic experience difficult, by Makkreel's interpretation.

16. Kant, KdU 20:215.

17. Ibid.

18. Here I am in agreement with Ostaric that "the process of schematization and the rule that governs the and orders the manifold in aesthetic judgment is the imagination's own achievement" (1377).

19. Gibbons, for example, while allowing for some dual capacity of the imagination, will claim the forms of the imagination are expressions of "the reciprocity between conceptual thought and (human) intuition and of the appropriateness, or fittedness, of each to the other" (Gibbons 58). In addition, Ostaric maintains the connection between the connections of the imagination as "purposive, not for our cognition, but for our cognitive faculty of the understanding" (Ostaric 1395).

20. Kant, KdU 20:216.

21. Passages that support my more permissive reading of nature viewed artistically are, for example, when Kant writes it is "permissible for us to apply such a special concept as that of purposiveness to nature and its lawfulness, although it cannot of course be an objective concept of nature, but is rather derived merely from the subjective relation of nature to a faculty of the mind" (Kant, KdU 20:218). Kant links nature to the faculty of the mind, not, for example, the general faculty of cognition to another faculty of the mind. This suggests that Kant wants to maintain the connection between aesthetic reflection and nature, rather than move the connection to only subjective connections between the faculties.

WORKS CITED

Gibbons, Sarah L. *Kant's Theory of Imagination: Bridging Gaps in Judgement and Experience*. Oxford University Press, 1994.

Makkreel, Rudolf A. *Imagination and Interpretation in Kant: The Hermeneutical Import of the Critique of Judgment*. University of Chicago Press, 1990.

Ostaric, Lara. "The Free Harmony of the Faculties and the Primacy of Imagination in Kant's Aesthetic Judgment." *European Journal of Philosophy* vol. 25, no. 4, 2017, pp. 1376-1410.

Buying Luxuries and Saving Lives

Leonard Kahn
Loyola University New Orleans

> Since luxury may be considered either as innocent or blamable, one may be surprized by the preposterous opinions, which have been entertained concerning it.
>
> -David Hume (189)

> There are certain things that I treat myself to. For example, I sometimes take a helicopter to travel to my farm, but I don't live a luxurious lifestyle.
>
> -Jeff Koons, commenting on his *Luxury and Degradation* series (quoted in Neuendorf)

Philosophers like to begin by defining their terms, but the definition of "luxury goods" is somewhat elusive. Of course, a typical economics textbook definition reads something like the following: *X is a luxury good if and only if it has high income elasticity of demand.* In other words, a luxury good is a good such that the rate at which one demands this good rises faster than the rate at which one's income increases (Mankiw 90). But Hume's remark that "Luxury is a word of uncertain signification," seems a wiser response, so I will not attempt to define "luxury goods" in this paper. In lieu of a definition, I will simply say that I use the expression to denote those goods that tend to cluster at the far end of this spectrum. Examples include (but are not limited to) tourbillon watches, Bentley Mulsannes, bespoke suits from shops on London's Savile Road, high-end jewelry, penthouse apartments along New York's Central Park, mansions, yachts, and works of fine art by masters of the craft.

It certainly seems morally impermissible to purchase and enjoy luxury goods for oneself if one could instead use that money to save or to improve significantly the lives of others. This appears true even if one has no personal connection to these people. Peter Singer's shallow pond thought experiment (which I will, hereafter, refer to as "Shallow Pond") is a well-known illustration of this idea. Yet might there be circumstances in which one is morally obligated to purchase a luxury good instead of saving others now—and precisely because of the greater amount lives that one might save by purchasing it? In this paper, I argue that there are, and at least some of us might actually be in those circumstances. In the process, I try to clarify what importance, if any, the fact that something counts as a luxury good has on our moral intuitions in cases like these.

Let me begin with a brief reminder of Shallow Pond. Singer asks us to imagine the following situation:

> On your way to work, you pass a small pond. On hot days, children sometimes play in the pond, which is only about knee-deep. The weather's cool today, though, and the hour is early, so you are surprised to see a child splashing about in the pond. As you get closer, you see that it is a very young child, just a toddler, who is flailing about, unable to stay upright or walk out of the pond. You look for the parents or babysitter, but there is no one else around. The child is unable to keep her head above the water for more than a few seconds at a time. If you don't wade in and pull her out, she seems likely to drown. Wading in is easy and safe, but you will ruin the new shoes you bought only a few days ago, and get your suit wet and muddy. By the time you hand the child over to someone responsible for her, and change your clothes, you'll be late for work. What should you do? (Singer, *The Life You Save* 3)

The answer to the question posed at the end of Shallow Pond is obvious. You should wade into the pond and save the child from drowning, even though doing so ruins your shoes, dirties your clothes, and makes you late for work. The lesson of Shallow Pond seems to be that we have a standing obligation to prevent great suffering and even death as a result of a lack of these goods, if we can do so without giving up anything that is equally important.

Though few if any of us will find ourselves in the circumstances of Shallow Pond, Singer puts the lesson of the thought experiment to work. Obviously, great suffering and even death as a result of a lack of such goods as water, food, shelter, and basic medical care are bad. Singer contends that we can prevent such evils by donating to highly effective charities, which include such GiveWell mainstains as the Against Malaria Foundation, the Malaria Consortium, Helen Keller International, New Incentives, and the Schistosomiasis Control Initiative.[1] Moreover, Singer maintains that we can donate in ways that involve little sacrifice

on our part, sacrifice that is more-or-less on a par with the costs associated with saving a drowning child in Shallow Pond. Singer concludes that "if you do not donate to effective charities, you are doing something wrong" (Singer, *The Life You Save* 15-16).

In a moment, I will tweak the parameters of Shallow Pond in order to explore further some of our ethical intuitions. But, before I do so, I want to underline the way that Singer often contrasts donating to highly effective charities with using money in order to purchase and enjoy luxury goods.

> [F]or conspicuous waste of money and resources it is hard to beat a luxury yacht. As *Business Insider* reported in 2017, 'It has become normal for the world's wealthiest individuals to drop millions, even billions, on lavish superyachts." Billionaires compete to be the owner of the largest private yacht—a title held at the moment by Sheikh Khalifa bin Zayed Al Nahyan, the Emir of Abu Dhabi and owner of Azzam, which at 180 meters long, edged out the previous largest, Eclipse, owned by the Russian billionaire Roman Abramovich. Azzam is estimated to have cost $400 million. It has accommodation for 36 guests. (Singer, *The Life You Save* 10)

Nor is Singer's focus on contrasting, on the one hand, helping others with, on the other hand, purchasing luxury goods unique to him. For instance, Peter Unger often makes the comparison between saving lives and enjoying luxury goods (24, 52, 124). I return to Unger later in this paper. Though Singer's tone is measured in the passage above, it is hard not to feel indignation—and more—toward the super-rich for wallowing in luxury, when even paltry amounts of restraint on their part could save and improve many lives.

However, it is not only the luxury goods of the super-rich at which Singer takes aim. In the same breath, he points out that members of the North American and European middle class spend thousands and even tens of thousands of dollars on watches by Rolex, Patek Philippe, and Omega, despite the fact they are no more accurate than a $50 quartz timepiece. And, if most of us are not high-end chronophiles, it is likely that we sometimes spend money on other luxury goods, money that is quite literally a matter of life and death for others.

Singer occasionally makes remarks that are aimed at softening this blow, at least to some extent. He writes, "Of course, we all have our little indulgences. I am not arguing that every luxury is wrong" (*Ethics in the Real World* 310-312). At first, this claim might seem inconsistent with the lesson of Shallow Pond. To the extent that we are trading off luxuries for ourselves with the lives of others, it is difficult to accept Singer's interpretation of his own argument. The indignation that goes around, comes around, at least for many of us—myself included.

So what gives? Perhaps the most plausible reason to accept Singer's interpretation is rooted in our own less-than-heroic qualities as moral beings. Here

the thought is that by compromising and allowing ourselves certain luxury goods now and again, we'll be more likely, over the course of our entire lives, to be more charitable than if we tried to be unremittingly austere and spartan.[2] I will return to this point later in the paper.

At the moment, I want to consider a variation of Shallow Pond that will add some interesting complications to our thinking about the relationship between our duty to help others in dire need and our possession and enjoyment of luxury goods. In this thought experiment, you again pass by a knee-deep pond in which a youngster is drowning, and no one else is nearby to help him. But there are some differences. In this case, you are not on your way to work. Rather:

> You are on your way to an auction at which a painting that you are certain is Leonardo's lost *Salavator Mundi* is being sold. For reasons that are beyond your control, you have no time to spare, and if you do not arrive on time, you will be unable to purchase the painting. Just before arriving at your destination, you see a child drowning in a shallow pond. You can either proceed and purchase the painting or save the child, but you cannot do both. What should you do?

Call this thought experiment Painting-1.

Before I discuss the answer to the question at the end of Painting-1, I need to provide two pieces of context. First, it is worth noting that, in fact, Leonardo's lost *Salavator Mundi* really was sold at auction only 15 years ago (in, as it so happens, my current home town). Its new owners, who realized what was up for bid, paid a little more than $9,000 for it. In contrast, the *Salavator Mundi* resold for $469,700,000 in 2017.[3] Here is the second piece of context for Painting-1. If you stop and fish the child from the pond, you will have saved one life, and good on you for it. But, before concluding that this is the best course of action, consider your opportunities if you purchase the painting instead. Let us say that, during your lifetime, you will have it restored and hung in your house, where only your friends and you will enjoy it. However, you will also direct that the painting be sold after your death with the proceeds going to the world's most effective charities. Of course, estimates vary concerning the marginal cost of saving a human life (understood in terms of, say, a certain period of quality adjusted life years) in the poorest parts of the world. GiveWell claims that "As of November 2016, the median estimate of our top charities' cost-effectiveness ranged from ~$900 to ~$7,000 per equivalent life saved."[4] The fact that the range of uncertainty covers an order of magnitude underlines the fact that these estimates can be difficult to make, but we should also keep in mind the fact that the marginal cost of saving a life will differ considerably because of market fluctuations and perturbations, including, for example, world demand for nylon, which is necessary for making mosquito nets. Nevertheless, $2,500 is a reasonable figure (Singer, *The Most Good*

and *The Life You Save* as well as MacAskill).

At that rate, the sale of the *Salavator Mundi* would result in about 188,000 lives saved. Hence, it seems that your path is as clear as it is disturbing: You should let the child die and purchase the luxury good. Of course, Singer sometimes calls out the extraordinary sums paid for works of art, lamenting "the capacity of the art world to co-opt works or art, irrespective of their creators' aims" which can "turn them into items of consumption for the rich" (*Ethics in the Real World* 179-181 and *The Life You Save* 124). But if art can be co-opted as luxury consumption, it is at least possible that the grotesque proceeds of this process can be put to good use—in this *extraordinarily* good use.

Yet even if you find the answer I have proposed to the question at the end of Painting-1 plausible, you might also think it is just a typical philosopher's fanciful edge-case, which can—and perhaps should—be humored and then ignored. But I wish to push back on this response in two ways. First, consider a related thought experiment that I will call Painting-2:

> You are at an auction and about to make a bid of about $9,000 to purchase a painting that you are certain is Leonardo's lost *Salavator Mundi*. However, you could instead donate that money (the $9,000) to the Against Malaria Foundation. If you make the donation rather than purchase the painting, it is reasonable to expect your action will save the lives of 3 or 4 people. What should you do?

It seems to me that the person who purchased the *Salavator Mundi* was actually in the circumstances of Painting-2. Of course, I am not claiming that the purchaser *knew* that she was in these circumstances, though she might have. My concern here is with the rightness or wrongness of action, not with its praiseworthiness or blameworthiness (Parfit 31-37). Moreover, these circumstances are closely analogous to those of Painting-1, in which you can choose between either saving a low number of lives by forgoing the purchase of the painting, or you can purchase the painting and save far more lives. As a result, if the purchaser had the charitable intentions described in Painting-1, then it would be right for her to forgo the present donation to the Against Malaria Foundation in order to save vastly more lives later. I find it equally compelling that, if the purchaser lacked any such charitable intentions, then it would be wrong of her to decline to donate the money at the cost of her private enrichment. It is worth noting that it would have been a simple matter for the person who purchased the *Salavator Mundi* to reveal the nature of the painting. This work of art need not be lost again if it were not purchased at a bargain-basement price, and the world need be no worse off aesthetically as a result.

Suppose that you agree that Painting-1 and Painting-2 are sufficiently analogous and, therefore, that the former is not quite as fanciful as it might have

appeared. Nevertheless, you might still think cases like Painting-1 and Painting-2 are sufficiently rare and should be politely dismissed. You certainly would not be wrong that the two cases are unlikely, though that is true of Shallow Pond and many other philosophical thought experiments as well (see Dennett). But my point here, as in Shallow Pond, is to try to put the lesson to work. I grant that many luxury goods are remarkably poor investments. For example, a bespoke suit made by a brilliant Savile Row tailor—glorious as it is—will be close to worthless after a decade of use. The extravagant luxury goods that Singer discusses—superyachts and such—are essentially bonfires of cash. Indeed, their owners likely know this and likely count on others knowing it too, since it is meant to enhance their social status by showing that they are not only rich, but so wealthy they can afford to set money ablaze (see Wang and Vladas as well as Kastanakis and Balabanis).

Yet not all luxury goods are likely to depreciate like these. For example, many brands of high-end chronographs—especially vintage examples—tend to appreciate with age, including those mentioned earlier by Rolex, Patek Philippe, and Omega. Ironically, one might do more good in the world by ignoring Singer's advice not to buy one of these luxury watches, even though doing so means we can't use the same resources to help others now, precisely because we will be able to save more lives later.

Obviously, my purpose here is not to walk the reader through a series of wise investments in luxury goods, something which I would be especially poorly equipped to do in any event. If philosophers have any wisdom to offer, it is not in this corner of the world. Rather, I want to try to say why doing something this like this might make sense. Recall Singer's claim that "[W]e all have our little indulgences. I am not arguing that every luxury is wrong." I said a moment ago that the most plausible interpretation of Singer's claim is that by compromising and allowing ourselves certain luxury goods, we'll be more likely, over the course of our lives, to be more charitable. Suppose that I am right about this interpretation and that certain luxury goods are likely to appreciate over time. Such goods could be used carefully, and the proceeds from their sale after our deaths could save more lives than we could have saved if we took the austere path and denied ourselves any luxuries at all.

Of course, circumstances like these are rather quotidian when compared to those in Painting-1 and Painting-2, but for that reason some of us might be in them right now. If we are, it seems like the right thing to do is buy the luxury goods, even though we can save lives right now if we don't make these purchases. Yet, if this conclusion is not as disturbing as that in Painting-1, it still leaves me uneasy.

One way to bring out that sense of unease is to consider another kind of luxury good that can - and often does - appreciate, namely: a certain class of automobile. This fact also carries with it a modicum of irony. In a thought experiment that is almost as famous as Singer's Shallow Pond, Peter Unger puts his readers in the position of 70-year-old Robert R. Roberts, who is known to one-and-all as "Bob."

In this thought experiment, Bob must choose between saving the life of a child and saving his Bugatti, into which he has poured all of his savings and on which he is relying for a comfortable retirement. If this seems unwise, Unger casually notes that the Bugatti can be "expected to appreciate at over 20% per year" (135). If so, it *is* a fantastic investment in a luxury good. As Unger frames the matter, Bob must choose between his comfortable retirement, which will result from his selling the car, and the child's life. Unger concludes that, if Bob chooses to save the Bugatti rather than the child, then his action would be "monstrous" (136). It is hard not to agree.

Nevertheless, if Bob's intention were to sell the vehicle and donate the money to a highly effective charity, he could save far more lives than that of a single child. Writing 25 years ago, Unger imagines Bob buying the vehicle "for the bargain price of just under $3 million" (135). In 2019, the most expensive Bugatti ever sold went for $19 million.[5] Its sale would provide enough money to save 7,600 lives, using our earlier estimate of $2,500 per life. Even if the imagined worth of the vehicle is off by a factor of 7, Bob would still be saving 1,000 times as many lives by choosing the Bugatti over the child. By analogy with Painting-1 and Painting-2, it would seem that Bob should do the "monstrous" thing and allow the child to die, if by doing so he will save thousands later.

However, if this is the right thing to do, it is also, as I said above about Painting-1, disturbing. And, as I have already noted, this quality manifests itself as unease in the more realistic cases of buying luxury goods too. What explains this? Clearly, many hypotheses suggest themselves. But the subject of this paper is luxury, so I want to conclude by considering this very possibility. More exactly, I want to ask whether the reason I feel this sense of disturbance in some cases and unease in others is a result of the fact that I imagine us enjoying luxury goods, even though the result is a greater savings of lives.

In order to test this idea, consider a final thought experiment, which I will call Copper Mine.

> You are on your way to an auction at which a piece of land in the Rocky Mountains is up for bid. On this piece of land is an old copper mine. For reasons that are beyond your control, you have no time to spare, and if you do not arrive on time, you will be unable to purchase the piece of land. Just before arriving at your destination, you see a child drowning in a shallow pond. You can either proceed and purchase the land or save the child, but you cannot do both. What should you do?

Before answering, note that in Copper Mine, as in Painting-1, only you know the real economic value of what is at stake. Everyone else believes that the copper mine has been exhausted, but you know that it contains vast quantities of the metal, which has increased in value per unit by 2 orders of magnitude over the last 60

years and is likely to continue to appreciate rapidly because of its use in advanced technology. You will be able to purchase the land for $2,500, even though it is worth tens of millions. Moreover, you will leave the land untouched during your lifetime and direct that it be sold (with a full disclosure of its value) after your death with the proceeds going to GiveWell, where it will be used to save thousands of lives.

My reaction to Copper Mine is much like my reaction to Painting-1 and Painting-2. On the one hand, it seems to me that you should allow the child to drown, and, on the other hand, this conclusion chills me. But note that this response is informative. Copper mines are not luxury goods, and they are unlikely to be the kind of thing that one gains an intrinsic pleasure from owning. That suggests that the claim that my feelings of ambivalence about Painting-1 and Painting-2 are not a function of the fact that luxury goods were involved. Whatever it is that bothers me about these cases is not to be found in the subject of luxury goods.[6]

NOTES

1. GiveWell's top charities change somewhat from year-to-year. See "Our Top Charities." *GiveWell*, www.givewell.org/charities/top-charities.

2. This interpretation involves coming down squarely on the possibilism side of the actualism/possibilism debate. While I think that is more than likely how one should come down, I cannot defend that claim here. See Goldman as well as Jackson and Pargetter.

3. Reports of the painting's precise sales price differ slightly but all are astronomically high. See Holland.

4. See "Cost-Effectiveness." *GiveWell*, www.givewell.org/how-we-work/our-criteria/cost-effectiveness.

5. See Valdes-Dapena. Of course, this only provides us with an upper bound for the worth of the vehicle, but we do not require a great deal of precision here to make the point at hand. It is worth noting that this price is more than an order of magnitude lower than would be expected if Bob's vehicle were appreciating at 20% per year. If there is a lesson to be learned, then it is to take investment advice from philosophers with a grain of salt!

6. My considerable thanks to Drew Chastain, Rebecca Farinas, Everett Fulmer, Celeste Harvey, Cassaundra Hill, Marisa Jurczyk, Kimberly Kahn, Joel MacClellan, Tzofit Ofengenden, Jennifer Rothschild, Jack Stetter, and Caner Turan for helpful comments, criticisms, and conversation. My apologies to any and all whose contributions I have forgotten to acknowledge here.

WORKS CITED

Dennett, Daniel C. *Intuition Pumps and Other Tools for Thinking*. WW Norton & Company, 2013.

Goldman, Holly S. "Dated Rightness and Moral Imperfection." *The Philosophical Review,* vol. 85, no. 4, 1976, pp. 449-487.

Holland, Oscar. "Rare Da Vinci Painting Smashes World Records with $450 Million

Sale." *CNN*, Cable News Network, 16 Nov. 2017, edition.cnn.com/style/article/da-vinci-salvator-mundi-sale-christies/index.html.

Hume, David. "Of Refinement in the Arts." In *David Hume on Morals, Politics, and Society*. Yale University Press, 2018, pp. 187-195.

Jackson, Frank and Robert Pargetter. "Oughts, Options, and Actualism." *The Philosophical Review,* vol. 95, no. 2, 1986, pp. 233-255.

Kastanakis, Minas N., and George Balabanis. "Explaining Variation in Conspicuous Luxury Consumption: An Individual Differences' Perspective." *Journal of Business Research,* vol. 67 no. 10, 2014, pp. 2147-2154.

MacAskill, William. *Doing Good better: How Effective Altruism can Help You Help Others, Do Work that Matters, and Make Smarter Choices about Giving Back*. Penguin, 2016.

Mankiw, N.Gregory. *Principles of microeconomics.* Cengage Learning, 2014.

Neuendorf, Henri. "Jeff Koons claims his art is anti-consumerist." *Artnet Nnews*, 1 Oct. 2015, news.artnet.com/exhibitions/jeff-koons-market-interview-336644, 2015.

Parfit, Derek. *Reasons and Persons*. Clarendon, 1984.

Singer, Peter. *Ethics in the Real World*. Princeton University Press, 2016.

———. *The Life You Can Save: How to do Your Part to End World Poverty*. thelifeyoucansave.org, 2019.

———. *The Most Good You Can Do*. Yale University Press, 2015.

Unger, Peter K. *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press, 1996.

Valdes-Dapena, Peter. "At Almost $19 Million, this Bugatti is the Most Expensive New Car ever Sold." *CNN*, Cable News Network, 5 Mar. 2019, www.cnn.com/2019/03/05/success/bugatti-la-voiture-noire/index.html.

Wang, Yajin, and Vladas Griskevicius. "Conspicuous Consumption, Relationships, and Rivals: Women's Luxury Products as Signals to Other Women." *Journal of Consumer Research,* vol. 40, no. 5, 2014, pp. 834-854.

How the "As If" Becomes True: Fictions from Vaihinger to Appiah and Beyond

Zachary Simpson
University of Science and Arts of Oklahoma

In the past fifty years, and, in particular, in the work of David Lewis, Kwame Anthony Appiah, and Mark Alfano, there has been considerable attention paid to the epistemic and practical significance of fictions. This is not a coincidence, as our world has seen an explosion of fictional entities, from string theory to credit default swaps to fake news. While most of the recent work on fictions and fictionalism has attempted to address its potential truth or epistemic status, today I would like to focus on a slightly different question: how a practical fiction, self-consciously created, eventually becomes a "fact." This is a question left unanswered by various thinkers, and yet, if fictions are to be of practical significance, I would offer that it is *the* question. Not whether a fiction is true, but, rather, the process whereby a fiction *becomes* true. Today I will argue that, for fictions to become true, they must be plausible, repeated, shared communally, and have physical correlates.

As Appiah, Arthur Fine, and others have noted, the focus on fictions began most ostensibly with the late 19th-century Kant scholar, Hans Vaihinger. For Vaihinger, our minds are "inventive" (v, 12), constantly searching for a "way about more easily in this world" (15). A good Kantian, Vaihinger knows full well that our reason-seeking minds cannot match the world as-is; indeed, the best we can do, he argues, is to reason "analogically" (29), in the subjunctive mode of the "as if" [*als ob*]. Such analogies are, at best, practical expedients, ones

which we know to be as such. As he states, "the true fiction, formulated in a strictly scientific manner, is always accompanied by the *consciousness* that the fictional idea, the fictional assumption, has no real validity" (80).

Vaihinger makes two distinctions that are of benefit to our discussion. First, Vaihinger makes a careful distinction between fictions and hypotheses. The distinction is purely empirical. Fictions have explanatory power and are regulative, but cannot, in practice, be proven. Hypotheses, on the other hand, can be demonstrated (49, 88). Or, alternatively stated, a fiction is a useful artifice, whereas a hypothesis is potentially reflective of reality. This leads to the second observation, that, whereas hypotheses eventually become a conclusion if proven, a fiction, on the other hand, holds only a mediating role—once it has been demonstrated or given light to a particular phenomenon, it goes away. Like the middle term in a syllogism, a fiction vanishes once it becomes true for an agent (111). Again, as Vaihinger offers, "We cannot accept the ordinary view that thought is an end in itself. Thought serves as a means of communication with others and its individual mechanisms must all be regarded as mechanical expedients" (102).

The ethical implications of Vaihinger's position are perhaps best shown through his use of a rather Kantian example, that of freedom. Owing to his reading of the First Critique, Vaihinger doubts the ontological status of freedom. And yet we must act "as if" we were free and accountable for our actions (47, 258). The fictional "as if" has performative force here—it allows us to think of ourselves as bound to certain norms and laws, even if we must remain ultimately agnostic about their source. Given the noumenal status of freedom, then, the idea can only be described as a "practical fiction" (48), an expedient (89, 99). Because freedom can never be proven, it must remain as a "mere idea," a fiction which has value but not explanatory force (260).

This perspective, Vaihinger offers, is the mature one, where we do not mistake our fictional constructs for reality itself. It allows us a richer picture of action and of the world. And, indeed, as Arthur Fine notes, Vaihinger eventually finds fictions everywhere, from ethics to atoms to calculus to theology. The modern world, in Vaihinger's estimation, is filled with fictions that allow us to paint a deeper picture both of moral activity and of what we know.

Vaihinger's work has, over the past hundred years, laid largely dormant. Yet it occasionally returns, as in the case of Kwame Anthony Appiah's *As If*, where he makes the case, once again, for what he calls "idealizations" as "useful untruths" (KAA, xii). Whereas Vaihinger saw fictions as neither true nor untrue, Appiah makes the careful argument that many idealizations are both untrue and, yet, useful. These idealizations often contradict our simple picture of the world, and, thus, we have to partition our thinking into truth (which may not be of use), and untruth (which may be of use). As he states, "In believing that it is *as if* something is so, I dispose myself to act in a certain way, but only in certain contexts and for

certain purposes. In that context and for those purposes I will do what I would have done if I had just straight out believed it" (22).

Appiah spends considerable time discussing the utility of ideas that are, ostensibly, false. An idealization is potentially quite false, and at best a simplification. And yet such idealizations, as in the case of consciousness studies or quantum physics, are necessary. Hence, they seem to occupy a liminal epistemological space. As Appiah offers, "But being sorta true is not, alas, a way of being true—it is a special way of being false." (45). These idealizations are valuable not because they are demonstrable, but by the pictures they give us of the world and how they allow us to act. Appiah spends much of the rest of his book discussing the applicability of idealizations to consciousness and ethics.

Both Vaihinger and Appiah, then, consistently demonstrate the utility of fictions/idealizations. And, in the case of Appiah, I think we can read some of his other work through this lens. And yet neither can demonstrate *how* we make these idealizations real in our minds. In the case of Vaihinger, he is simply silent on the issue. Appiah at least recognizes this aporia, but does not offer much in the way of explanation, offering that "showing *that* it does work doesn't explain *why*" (36) and "Idealization works here [in the instance of correlating brain states and consciousness], then, for reasons we do not understand" (49). Thus, I think we should ask the question: How do we entertain concepts that we know to be false? Or, how do we keep fictions in mind long enough for them to *become* true?

I will return to Appiah once again, but to perhaps gain some clarity regarding the performative force of fictions, I'd like to call upon Mark Alfano's work in virtue ethics, specifically his *Character as Moral Fiction.* While Alfano spends considerable time discussing the epistemological status of fictions and countering competing moral theories, at the heart of his book lies a simple idea: that, if a person is treated as having certain desirable moral characteristics, they will eventually come to adopt those same characteristics (Alfano 82-3). This leads to his most notable concept, that of "factitious" virtue, a portmanteau of "factual" and "fictitious," whereby ethical behavior is claimed for an agent without them yet possessing that behavior. Like a placebo effect (83) or a self- fulfilling prophecy (88), an agent performs the truth of a prior claim.

Alfano marshals considerable social-scientific evidence in support of his thesis, which, I will claim, stands on three key points. First, as Alfano argues (91), the attributions made of an agent must at least be plausible. If I want to claim goodness for Suzy, she must at least be capable of goodness and tending towards a particular kind of good behavior. Secondly, such labeling is, by its very nature, interpersonal (92, 96). The social dimension of character development allows agents to follow norms (97) and to fulfill certain roles deemed to be socially beneficial. As Alfano states, "Aristotle thought that people became courageous by acting courageously; I contend that they become courageous (or near enough) by being called courageous" (102). Third, virtue labeling must also be repeated over

time. Alfano notes many social strategies for creating virtues, such as training, habituation, and the arts (177-9). All are signified by repetition and multiple modes of cognitive reinforcement.

Taken together, these three qualities form, for Alfano, a kind of looping effect, whereby a label, when plausible, given socially, and repeated over time, eventually comes to be real. It is not real at its utterance. It is real in the future. For the agent, they must, at least unconsciously, act "as if" the label itself is real. These three features seem to be sufficient for generating factitiously virtuous behavior.

While Alfano does not further discuss the cognitive mechanisms at work in the agent receiving such utterances as real, I think they are of critical importance here and are a necessary supplement to the framework laid out by Alfano. And I think that some aesthetic theory helps to shed light on how an agent allows for social and repetitive practices to work, at least on a psychological level. In particular, recent theoretical work on make-believe and play may allow us to see the mechanisms at work here.

In this regard, I will make two points that bear out our ability to receive, and make real, ethical utterances. First, as Kendall Walton, R.M. Sainsbury, and even Appiah make clear, what is not needed with respect to fiction is "suspension of disbelief." Rather, when we receive something we perceive as fictional, we actually *believe* in such a fiction in rather distinctive ways. Walton, in multiple books, famously calls these emotions "quasi-emotions." In an example that Walton uses to great effect, when watching a movie I actually do experience fear, hate, disgust, etc. (e.g., Walton, *In Other Shoes* 256). While, along with Sainsbury (20), I disagree that these emotions are "quasi" felt—we really do feel them as real—I would offer that Walton gives a compelling account of fictional belief. In short, in fiction our belief is automatic (if not involuntary), felt as real, and represented in our minds. These beliefs are pre-reflective and unconscious. I do not attempt to conjure fear when watching a movie. It simply happens (see Walton, *In Other Shoes* 262, Sainsbury 12, and Casey 30ff). Moreover, these spontaneous imaginings are self-enclosed: they occur in their own world and according to their own logic. I am not surprised, for example, when Wolverine's claws can slice through a Humvee. This is because, as Edward Casey notes, imaginings tend to have their own space and time, an internal schema or "state of affairs" that pertains solely to those imaginings (see Casey 42, 51), And this aesthetic is autonomous: each imagining has its own logic. The logic of X-Men is not the logic of Avengers, and vice versa. In short, when rapt in aesthetic appreciation, we immediately conjure both a world in which such fictions occur and the emotions appropriate to those fictions.

This recognition is conditioned by a second, and equivalent, observation, regarding the nature of aesthetic experience. For, in the midst of experiencing a fiction, our minds also simultaneously partition fictional worlds from the real world in which we live. This, too, is spontaneous. Walton refers to this as our

psychological mechanisms running "off line" (Walton, *In Other Shoes* 280). Appiah notes, similarly, that we abandon "the normal consequences of that belief" (108) or "its normal consequences" (110). Finally, Sainsbury dubs this our "within-the-fiction state," one which is distinctive from our normal way of reckoning with the world (15-16).

The common feature of each of these observations is the notion of a disunified mind, or of varying psychological schema that refer to particular kinds of phenomena. This is not, as some may infer, a form of Type 1 and Type 2 processing. Rather, as Floyd Merrell recommends, we can better think of aesthetic reception as an act of "framing," whereby certain fictions exist within a particular frame or schema that allows them to create sense (Merrell, Chs. 1 and 2). These frames have a peculiar phenomenological status. We employ them pre-cognitively, but, when asked, we are always aware of the fictionality of a particular frame. As Merrell states, we never say, "In this fiction with respect to which I have suspended my disbelief…" (22). Because of the way in which we slip easily into fictional schema, we can oscillate seamlessly between fictional worlds and real worlds, somehow making careful and autonomic psychological distinctions between the two states (cf. 24-6).

Taken collectively, I'd like to now summarize the features of what I believe allows for the "as if" to become real in our minds:

1. Plausibility, either ontologically or narratively;
2. Intersubjectivity;
3. Repetition;
4. Belief, a perception or event felt as real;
5. The normal psychological process of distinguishing the "as if" from the real.

Taken together, I would offer that these constitute what many have called "make-believe." To "make-believe" is to engage with something as if it is true, to give it truth value on a conscious level. The above criteria, I would offer, are sufficient for making believe and for, on a more extended level, rendering the world of the "as if" true in a more-than-fictional sense.

To employ a familiar example, think of children's games of make-believe. When my daughter pretends to be a fairy princess, she actively believes in fairy princesses while also segregating that belief from other bodily and psychological schema. If such play is intersubjective and repeated—as it often is—it can become even more real to my daughter and her friends.

I am not arguing, of course, that fairy princesses can become real to my daughter. They rarely do. And this is principally because they lack plausibility and even a potential narrative or ontological match with the real world. And yet, on a very basic level, even this example demonstrates the positive role that make-

believe plays in allowing us to enact scenarios without suffering their real consequences. This is, of course, the notion of play-as-rehearsal, made by primatologists, psychologists, and even Appiah (see Appiah 105-6). It is also at the root of Appiah's notion of idealization and heuristics, whereby we can model systems and ideas "as if" they were real, imagining them and even enacting such models in a controlled way.

With respect to ethics, however, something more is asked. If we are to treat an individual as virtuous, or a collective *as if* they are equal, or a system *as if* it is just, then all five of the above conditions must be met. For an individual to eventually manifest a certain virtue, they must not only be labelled as such repeatedly, but they themselves must also cognitively entertain such a label as potentially true, and, moreover, believe it to be true *long enough for it to become so.* A fiction becomes real if we collectively treat it as real for a sufficient period of time.

Arguably, then, the boundary between fiction and reality is, as many of us already experience it, quite fuzzy (see, e.g., Merrell 36). Fictions can *become* true, at least socially, if they have traction with reality, are repeated, and allow for collective belief. Such fictions need not be correspondently true at the moment of their utterance or thought; rather, they are potentially true if certain conditions are met.

This account of virtue labeling and aesthetic experience, I would offer, fills in a critical lacuna in the work of fictionalists like Vaihinger, Appiah, and Alfano, as well as aesthetic theorists like Walton. (It may also have some bearing on epistemological discussions of fictionalism.) Fictionalists are right to say, that, normatively, fictions have value. But they leave open the ways in which fictions ultimately become real. And aesthetic theorists like Walton give powerful accounts of imagination and make-believe, but do not account for the ways in which such make-believe can become even more real in the minds of individuals and collectives. By offering an explanation for how ideas become real in individuals and groups, particularly ethical and cultural concepts, we can better understand the mechanisms behind ethical labeling, cultural acceptance of clearly arbitrarily defined cultural norms (like sports rules), and even social acceptance of certain memes and norms of discourse. What is required, for each, is a synergism between plausibility, community, repetition, belief, and framing.

WORKS CITED

Alfano, Mark. *Character as Moral Fiction*. Cambridge University Press, 2013.
Appiah, Kwame Anthony. *As If: Idealization and Ideals*. Harvard University Press, 2019.
Casey, Edward. *Imagining: A Phenomenological Study.* Indiana University Press, 1976.
Merrell, Floyd. *Pararealities: The Nature of our Fictions and How we Know Them.* Philadelphia, PA: John Benjamins Publishing, 1983.
R.M. Sainsbury, *Fictions and Fictionalism.* Routledge, 2010.

Walton, Kendall. *In Other Shoes: Music, Metaphor, Empathy, Existence.* Oxford University Press, 2015.

———. *Mimesis as Make-Believe: On the Foundations of the Representational Arts.* Harvard University Press, 1990.

Vaihinger, Hans. *The Philosophy of “As If”: A System of the Theoretical, Practical, and Religious Fictions of Mankind.* Translated by C.K. Ogden, Martino Fine Books, 2009.

Rehabilitating Agent-Regret

Robert B. Tierney
University of Houston

I. Introduction

In this paper I will offer an analysis of Bernard Williams's concept of AGENT-REGRET that shows it to be a distinct element of our moral psychology.[1] I will focus specifically on cases of Agent-Regret where the agent, through voluntary action, has caused harms that he[2] could not have reasonably foreseen.[3] I will then present R. Jay Wallace's recent attempt to show the contrary, i.e., that AGENT-REGRET is not a significant concept for *descriptive* moral psychology. Finally, I will show that Wallace's arguments are insufficient to defeat my account of Agent-Regret.

II. Unforeseeable Harms and Agent Regret: The Paradigm Case

Williams presents a paradigm case.

> The lorry driver who, through no fault of his own, runs over a child, will feel differently from any spectator…. Doubtless, and rightly, people will try, in comforting him, to move the driver from this state of feeling, and move him indeed from where he is to something more like the place of a spectator, but it is important that this is seen as something that should need to be done, and indeed some doubt would be felt about a driver who too blandly or too readily moved to that position. We feel sorry for the lorry

> driver, but that sentiment co-exists with, indeed presupposes, that there is something special about his relation to this happening, something which cannot merely be eliminated by the consideration that it was not his fault. ("Moral Luck" 28)[4]

Williams writes further.

> [T]he discussion of agent-regret about … involuntary [actions, i.e., involuntary through ignorance,][5] … helps us to get away from the a dichotomy which is often relied on in these matters, expressed in such terms as *regret* and *remorse*, where 'regret' is identified as the regret of the spectator, while 'remorse' is what we have called 'agent-regret,' but under the restriction that it applies only to the voluntary. The fact that we have agent-regret about the involuntary, and would not readily recognize a life without it …, shows already that there is something wrong with the dichotomy: such regret is neither mere spectator's regret, nor (by this definition) remorse. ("Moral Luck" 30)

Having presented these evocative passages, we now proceed to examine Agent-Regret, in an analytic manner, as a kind of emotion.

III. Agent Regret as an Emotion

A. Introduction

To begin, we can formally divide types of regret into three subcategories, with Regret-in-General being the overarching category within which they fall. These three subcategories are: (1) Agent-Regret, (2) Victim-Regret, and (3) Third-Party-Regret. For the moment, we can treat these as marking out merely formal distinctions. Substantive distinctions will emerge as we proceed.

In the passage last quoted, Williams compares Agent-Regret both to Regret-in-General and to remorse. Thus, it is an "emotion of retrospective assessment" (Wallace, sec. 2.2). In this section (i.e. §III), I will clarify the nature of that emotion. In so doing, I will use a widely shared framework which holds that prototypical episodes of emotion have a series of components (Scarantino and Sousa 6–7).[6] For purposes of characterizing Agent-Regret, we will focus on the following components: (1) phenomenological, (2) evaluative, (3) behavioral, and (4) expressive (Scarantino and Sousa 6–7). However, as we shall see, the nature of Agent-Regret as an emotion of *retrospective* assessment, to some extent, entails the merger of the behavioral and expressive components.

B. The Phenomenological Component of Agent Regret

It seems that Agent-Regret would have a phenomenological character in that we

typically speak of *feelings* of regret and *feelings* of remorse. Indeed, Williams speaks of the "state of feeling" that the lorry driver is in. It seems clear that it is an unpleasant phenomenological quality, perhaps something like that suggested by the umbrella term "distress," broadly construed. This naturally raises the question as to whether there is a single type of phenomenological state that is characteristic[7] of Agent-Regret and sufficient to distinguish it from other emotions of retrospective assessment. In "Moral Luck," Williams seems to intimate that there is.[8] Yet, in later writings, Williams indicates that a wide variety of *emotions* may be instantiated in, or manifested through, Agent-Regret, such that there is neither a single emotion that is always present, nor a single emotion that is (at least sometimes) present in Agent-Regret but nowhere else. It may be, then, that there is no single *phenomenological emotion-component* that is so much as characteristic or prototypical of Agent-Regret. Thus, in *Shame and Necessity*, Williams argues that, at least in some cases, "Agent-Regret … can be psychologically a manifestation of guilt" (93). He also holds that it may be manifested by shame (*Shame and Necessity*, ch. III, IV).[9] Further, in discussing *Oedipus Tyrannus*, Williams recounts the "terror" Oedipus experiences upon realizing what he has done (69), and how, shortly before his death, he "reflects on the *horrors* of many years before" (60; emphasis added). In other relevantly similar cases, Williams speaks of *humiliation* (72-73) and *mortification* (92). Given this array of emotional manifestations, I doubt that there is any single type of phenomenological experience common to all, or even great majority, of cases of Agent-Regret.[10]

C. The Evaluative Component of Agent Regret[11]

On Williams account, *the constitutive thoughts of regret* are "something like" the following combination: (a) "how much better if it had been otherwise"; (b) "some conception of how things might have been otherwise"; and (c) "consciousness of how things would then have been better" ("Moral Luck" 27). While perhaps sufficient for Regret-in-General, this characterization is insufficient for Agent-Regret. In Agent-Regret one prototypically thinks that "one might have acted otherwise, and the focus of the regret is on that possibility, the thought being formed in part by *first-personal* conceptions of how one might have acted otherwise" (27; emphasis added).[12] Thus, what we will call the "Constitutive Thoughts of Agent-Regret," which are present in prototypical cases, must be something like the following.[13]

CTAR$_1$: (a) The thought that, given the nature or consequences of my action, how much better if I had done otherwise;
(b) some conception of how I could have done otherwise such that the nature or consequences of my action would have been different; and
(c) consciousness of how things would then have been better

had I done so.

In addition, Williams clearly thinks that Agent Regret involves taking oneself to be, in some sense, accountable for one's actions and their consequences. In the case of unforeseeable harm, at least, this accountability gets its initial grounding in the fact that one's action caused the harm. "Those who have been hurt need a response; simply what has happened to them might give them a right to seek it, and where can they look more appropriately than to you, the cause?" (*Shame and Necessity* 70).

> Everywhere, human beings act, and their actions cause things to happen, and sometimes they intend those things, and sometimes they do not; everywhere, what is brought about is sometimes to be regretted or deplored, by the agent or by others who suffer from it or by both; and when that is so, there may be a demand for some response from that agent, a demand made by himself, by others or by both. (*Shame and Necessity* 55)

Thus, "in the story of one's life there is an authority exercised by what one has done, and not merely by what one has intentionally done" (69). "[T]he fact that you acted unintentionally does not, in itself, dissociate that action from yourself" (54). Moreover, "the significance of someone's life and its relations to society may be such that someone needs to recognize and express his responsibility for actions when no one else would have the right to make a claim for damages or be in a position to do so" (74).

None of this goes so far as to say that we are always and everywhere accountable for the consequences of our actions. Further, nothing is this account implies that we are *culpable*, in the full moral sense, for causing unforeseeable harms. However, aside from the discussion of what I call "the Reparative Response" in what follows, saying what it *does* involve would require a more a more extended analysis of Williams' critique of contemporary morality than can be given here.

D. The Behavioral/Expressive Component of Agent Regret

The behavioral component of an emotion concerns not only overt behavior but also its motivation. In Agent-Regret, both of these features manifest themselves in a way that is not characteristic of many other emotions in that it is an emotion of *retrospective* assessment. It is a response to what *has already been done*. Moreover, that which one has done, where it results in Agent-Regret, typically cannot be undone. However, verbal and behavioral manifestations of the *desire* to take some remedial *action* are characteristic *expressions* of Agent-Regret. Thus, in significant measure, the behavioral and expressive components of Agent-Regret merge.

To begin, Regret-in-General "necessarily involves a wish that things had been otherwise" (Williams, "Moral Luck" 31). For Agent-Regret this wish concerns, at least in part, one's own actions, either in terms of their inherent nature or their consequences.[14] Paradigmatically, it will take the form of wishing that one had not acted as one did (31).[15] Given that one has *already* taken the harmful action, Agent-Regret will be manifested through apologizing and expressing sorrow and regret, as well as through expressing and enacting desires to make reparations to the harmed person or to their loved ones. One intends that these reparations have a significance other than mere compensation (28-29). Thus, the knowledge that the victim was receiving compensatory insurance payments would *not* alleviate the agent of the relevant desires (28-29). Where one cannot make suitable reparations, "the desire to make reparations remains, with the painful consciousness that nothing can be done about it; some other action, perhaps less directed to the victims, may come to express this" (29). This holds true even if no one has the right to demand recompense from the agent.[16] We will refer to this entire syndrome of thoughts, feelings, motivations, expressions, and actions that includes and is connected with the desire to make reparations as the "Reparative Response." The latter evidences and expresses the fact that, and something of the manner in which, one takes oneself to be *accountable* for such actions.[17]

IV. WALLACE'S ATTACK ON AGENT REGRET

A. Introduction

In his recent book, *The View from Here*, R. Jay Wallace has argued that there is no distinct phenomenon of Agent-Regret.[18] We will consider each of his arguments in turn.

B. Claim: Agent-Regret Has No Paradigmatic Qualitatively-Distinct Phenomenological Component

Wallace first attacks the concept of AGENT-REGRET by arguing that it has no characteristic phenomenological component that distinguishes it from Third-Party-Regret (34–36). As we have seen, it is difficult to characterize this component of Agent-Regret, and Williams acknowledges that Agent-Regret can be manifested in terms of a number of emotions,[19] so I will not contest this point.

C. Claim: Agent-Regret's Phenomenological Component Has No Special Intensity

Next Wallace contends that Agent-Regret cannot be distinguished from Third-Party-Regret in terms of the intensity of the feeling involved. The parent of the child who was run over by the lorry driver, for instance, might experience regret with a deeper phenomenological intensity than does the lorry driver (34–36). This is true enough. However, it misses an important point. Holding everything else

constant, the intensity of the regret experienced by the agent of harm is likely to be more intense than would be that of a similarly situated third-party. Thus, if the child's parent were also the lorry driver, and not a third-party to the tragedy, we would expect the parent's regret to be all the more intense. Thus, everything else being equal, the agential relationship seems to heighten the intensity of the regret experienced.

To be fair, Wallace ultimately acknowledges that being the lorry driver is a factor likely to heighten regret. However, he construes agency as just another exacerbating factor, no different in kind from being involved by being an eye-witness or a loved one. Unfortunately, Wallace presents these claims as mere question-begging assertions. This is precisely the question at issue, and not an assumption to be made. Moreover, Wallace ignores the difference in the specific types of thoughts, memories, imagined alternative scenarios, and various types of reminders and environmental stimuli—largely centered around the fact that the one was, in fact, the agent of harm—that differentially trigger, exacerbate, and alleviate the phenomenological intensity of Agent-Regret as opposed to Third-Party-Regret. Thus, at the appropriate level of granularity, there is a difference in the intensity of the phenomenological component of Agent-Regret and Third-Party-Regret.[20]

D. Claim: Agent-Regret Has No Significant Evaluative Component

Wallace goes on to argue that Agent-Regret cannot be carved out as a suitable category in terms of the thoughts that it characteristically involves. He does acknowledge that the lorry driver would have an indexical thought partially constituting his regret which would, in principle, be unavailable to third parties. However, he then asserts that "it isn't obvious that emotions of retrospective assessment that involve such indexical thoughts really constitute an interesting *natural kind* of psychological phenomenon" (36; emphasis added). This is an extraordinary claim to make without more. First, it is not clear that we should take seriously this concern about natural kinds given that it is doubtful that the emotions as such form a natural kind.[21] As to whether Agent-Regret is an "interesting" kind, Williams' extensive writings about contemporary and Ancient Greek moral psychology shows it to be very interesting indeed (*Shame and Necessity*; "Moral Luck"; *Ethics and the Limits of Philosophy*). Of course, I cannot substantiate this claim in detail within this paper, so I invite you to read Williams' work and judge for yourselves.

Moreover, I think that Wallace grossly mischaracterizes the relevant thoughts. He asks, rhetorically, "Is it just the indexical thought that it was my agency that gave rise to the unfortunate event?" This, of course, is an absurd caricature. In the throes of regret, the lorry driver is not likely to think "It was my agency that gave rise to this unfortunate event of the child being run over." First, this mischaracterizes the indexical. The driver is going to think "I did …," not "It was my agency that gave rise to …." Moreover, we need to specify the "unfortunate

event" in its particularity and with the vivacity that it is likely have for the driver. "I ran over the child. I have killed the child." are thoughts whose poignancy is not captured by Wallace's characterization. Beyond that, Wallace simply misses the complexity of the evaluative component of Agent-Regret as set forth herein.[22]

E. Claim: The Joint Operation of the Evaluative and Phenomenological Components Fails to Establish Agent-Regret as a Significant Category

Wallace then suggests—I think rightly—that Williams' wants to distinguish Agent-Regret from Third-Party Regret in terms of the phenomenological and evaluative components operating in combination. He contends that Williams grounds his account of Agent-Regret in the fact that "[o]ne's history as an agent is a web in which anything that is a product of the will is surrounded and held up and partly formed by things that are not" (Williams, "Moral Luck" 29; qtd. in Wallace 37). Wallace acknowledges this, but contends that a more adequate view would fully recognize that "[o]ur web-like history as persons … extends in two directions, involving nonvoluntary causal influences on us as well as nonvoluntary causal effects that are brought about through our agency" (37). Wallace then claims that "there ought to be a form of regret that is adequate to both of the ways in which we are linked to the larger network of events within which we operate" (38).

However, to the extent that this last claim involves a normative-rational ought, rather than a descriptive-predictive ought, it does nothing to make Wallace's descriptive case. Construed as an argument for a descriptive claim, this seems little more than question-begging assertions. Moreover, Wallace offers an unconvincing account of our experience. The distinction between what merely happens, (which can produce Third-Party-Regret) and what we do (which can produce Agent-Regret) is perhaps the most fundamental categorial distinction moral psychology. We even make a major distinction between harms we have *done* to someone and harms we have *allowed* to happen to someone, and the more attenuated that sense of *allowing* the more the phenomenon is assimilated to the purely passive stance.

Of course, Wallace cannot really consistently maintain that agency plays no special role in regret, for he acknowledges that remorse is appropriate for some *intentional* actions whereby we inflict harm. He then presses the conventional distinction between culpability and remorse for intentional harm and non-culpability and the inappropriateness of remorse for harm that was unforeseeable.

Once again, Wallace makes a question-begging move. No one denies that there is an important distinction to be made here. But that does not entail that there isn't also an important distinction between what I cause to happen (even unforeseeably), including harm, and what merely happens irrespective of what I do or don't do, including harm. Consider cases of self-defense where the aggressor is not morally culpable. Perhaps he is suffering from involuntary intoxication or the unexpected and sudden onset of psychosis, or even seizures such that the movements of his

body are mere automatisms. Nonetheless, if such a person is about to harm another, we generally take it that the would-be victim, or a bystander, may use appropriate force, including lethal force, to protect the potential victim. By contrast, consider a case where an aggressor could only be prevented from committing homicide by a "kill shot" aimed at his "human shield," in which case he could be made to surrender. We would find it much more morally troubling to use lethal force against the passive, morally-innocent human shield than we would against the active, morally-innocent aggressor. We hold the innocent aggressor liable in a way that we don't the passive shield, and this liability is *not* purely compensatory. We don't, for instance, require that force not be used against the innocent aggressor and that the victim or his heirs bring a law suit for monetary damages. As a matter of descriptive moral psychology, agency matters.

F. Claim: The Joint Operation of the Evaluative and Behavioral/Expressive Components Fails to Establish Agent-Regret as a Significant Category

Finally, Wallace acknowledges that Williams lays great stress on the Reparative Response in differentiating Agent-Regret from Regret-in-General. Here Wallace's challenge is primarily to the normative appropriateness of Reparative Response *as such* in cases of unforeseeable harms. However, from the descriptive standpoint, with which we are concerned in this paper, it does seem to be a substantial part of our moral psychology. Moreover, as Williams writes, we would think poorly of the lorry driver were he, from the outset, to regard what has happened solely as *an unfortunate event*, rather than also *something he has done*, something that leads him to engage the Reparative Response.

V. CONCLUSION

I have argued that, contrary to Wallace's contentions, Agent-Regret is a distinct, significant feature of our moral psychology. This distinctness comes largely through its evaluative component, which originates with its characteristic indexical thoughts, which thoughts play a large role in triggering, and modulating the intensity of, the phenomenal quality of emotional distress involved. Together these elements express themselves through the Reparative Response.

NOTES

1. Technical or formal terms are capitalized, even if they are not capitalized in the original text. When a term designates a concept, it will appear in SMALL CAPS. I will follow Williams' convention of hyphenating "Agent-Regret."

2. In order to more conveniently work with the quoted texts, I have used the masculine pronoun.

3. Although agents causing foreseeable, though not purposely inflicted, harms can suffer Agent-Regret (i.e., Agent-Regret that is not remorse), in this paper we will consider only unforeseeable harms.

4. Williams focuses on cases where the agent is clearly the proximate cause of a harm that is both obvious and grievous. (Williams uses the phrase "immediate cause," rather than "proximate cause") (*Shame and Necessity* 54). Further, he focuses on harm done by human agents to persons, though there may be other sorts of harm that are relevant. Specifically, he focuses on agents who bring harm to persons other than themselves. Finally, Williams focuses on the accountability of the agent, rather than on the victim. In some kinds of cases, one might take it that the victim is at least partially at fault for putting himself in harm's way, so to speak.

5. In Williams usage, the actions of the lorry driver fall within the ambit of the involuntary in that they are "involuntary through ignorance" ("Moral Luck" 28). We will follow the clearer and more common usage and say that the lorry driver's action was voluntary, but that the relevant consequences were neither intended, foreseen, nor reasonably foreseeable.

6. There are differing views as to which of the commonly recognized components are essential to emotion.

7. I take it that a "characteristic" or prototypical phenomenological component of Agent-Regret is sufficient for our purposes. Hence, we needn't maintain that there is a phenomenological quality that is either unique to Agent-Regret nor one that is a necessary component of any instance of Agent-Regret.

8. For instance, he writes that "sentiments of agent-regret are different from regret in general, such as might be *felt* by a spectator" ("Moral Luck" 28; emphasis added), and that Agent-Regret is something that "a person can *feel* only toward his own past actions" ("Moral Luck" 27; emphasis added). Unfortunately, Williams does not directly characterize—*in general terms*—the sort of phenomenological state that is present in Agent-Regret, other than by attempting to locate the concept in relation to the more familiar concepts of REGRET-IN-GENERAL and REMORSE. Otherwise, he uses evocative examples to try to make the relevant feeling or feelings clearer.

9. In the broader context of his analysis, he also contends that "[o]ne way in which we can be helped [is] by the [ancient] Greek conception that brings (something like) guilt under the wider conception of (something more than) shame" (*Shame and Necessity* 92).

10. While I do not take this argument to provide conclusive proof of its conclusion, it can do no injustice to Wallace in that it amounts to conceding the point to one of his criticisms of Williams.

11. This component has sometimes been conceptualized as a judgment or as some other form of cognitive evaluation (Scarantino and Sousa, secs. 2, 5, 6). This analysis falls within that broad tradition, without committing to a precise specification of the nature of this cognitive component. This cognitive evaluation has been held to be either a cause of the relevant emotion or constitutive of it. Based on Williams work, I take it that the cognitive elements described in this section are at least partially constitutive of prototypical cases of Agent-Regret.

12. This thought is *only* "prototypical" in that it is not necessary that the thought be that it would have been better, all things considered, if one had acted otherwise. One can have Agent Regret even when one has, all things considered, done what one ought. The

action need only have had some regrettable aspect or regrettable consequences. "Regret necessarily involves a wish that things had been otherwise, for instance that one had not had to act as one did. But it does not necessarily involve the wish, all things taken together, that one had acted otherwise" (Williams, "Moral Luck" 31). For instance, one may face a moral dilemma where one does that which is best, but that still involves doing something wrong (Williams, "Moral Luck" 31).

13. As to why this applies *only* to "prototypical" cases, see note 12. The "Constitutive Thoughts of Agent-Regret" for non-prototypical cases would be as follows.

> **CTAR$_2$**: (a) The thought that, given the nature or consequences of my action, how much better if taking the action I ought to and did undertake did not have the regrettable aspects or consequences that it did;
> (b) some conception of how the situation could have been different; and
> (c) consciousness of how things would then have been better had the situation been different.

14. When specified generically, the content of this wish must be characterized in this disjunctive fashion. See notes 12 and 13 and accompanying text for an explanation.

15. This is *only* the *prototypical* content of the wish. See notes 12 and 13 and accompanying text for an explanation .

16. "[T]here is an aspect to responsibility, which comes out if we start on the question not from the response that the public or the state or the neighbors or the damaged parties demand of the agent, but from what the agent demands of himself" (*Shame and Necessity* 68). "[T]he significance of someone's life and its relations to society may be such that someone needs to recognize and express his responsibility for actions when no one else would have the right to make a claim for damages or be in a position to do so" (*Shame and Necessity* 74).

17. The agent's judgement of accountability was discussed in §III.C.

18. Wallace also claims that the phenomena typical of Agent-Regret, to the extent that they bespeak moral accountability, are normatively unwarranted. I will not engage with this claim in this paper. This paper analyzes Agent-Regret only in *descriptive* terms.

19. This is discussed in §III.B.

20. Wallace also expresses his argument in a manner that obscures the significance of agency as such. This is brought to light in §§IV.E, IV.E, and IV.F.

21. EMOTION is probably a family-resemblance concept ("Emotion" 259).

22. This is discussed in §III.C.

WORKS CITED

"Emotion." *The Cambridge Dictionary of Philosophy*, 2nd ed., Cambridge University Press, 1999, p. 259.

Scarantino, Andrea, and Ronald de Sousa. "Emotion." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, 25 Sept. 2018, plato.stanford.edu/archives/win2018/entries/emotion.

Wallace, R. Jay. *The View from Here: On Affirmation, Attachment, and the Limits of Regret*. Oxford, Oxford University Press, 2013.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, Mass., Harvard University Press, 1985.

———. "Moral Luck." *Moral Luck: Philosophical Papers 1973–1980*, Cambridge, Cambridge University Press, 1981, pp. 20–39.

———. *Shame and Necessity*. Berkeley and Los Angeles, University of California Press, 1993.

Are Ambitious Evolutionary Debunking Arguments Self-Refuting?

Caner Turan
American University of Beirut

Evolutionary debunking arguments (henceforth, EDAs) present an epistemic challenge for evaluative/moral realism. The idea is that, since (i) evolutionary forces shape extensively the content of our evaluative/moral beliefs, and (ii) evolutionary mechanisms aim at reproductive success rather than attitude-independent evaluative/moral truths, evaluative/moral realism is an epistemically untenable metanormative (or metaethical) view. A possible response to the challenge is to assert that ambitious EDAs are self-refuting: since they take *all* of our evaluative/moral beliefs to be epistemically suspect, they cannot provide independent reason to believe that our evaluative/moral beliefs are debunked by the argument. My aim in this paper is to show how to plausibly respond to such an objection. To do that, I first lay out the epistemic challenge EDAs pose for evaluative/moral realism and briefly explain Sharon Street's and Richard Joyce's arguments (Section 1). I then discuss Katia Vavova's objection that ambitious EDAs are self-refuting (Section 2). I argue, contra Vavova, that the level of ambition of an EDA does *not* affect the strength of its *epistemic* premise because EDAs are essentially *inductive* arguments. Rather, the level of ambition of an EDA affects the strength of its *empirical* premise, which is the Achilles heel of any ambitious EDA.

1. Evolutionary Challenge for Realism

1.1. The Structure of an Evolutionary Debunking Argument

Evolutionary debunking arguments claim to undermine the justification of our evaluative beliefs by placing a special focus on the evolutionary origins of them.

Some of such arguments are more ambitious than the others as they try to undermine the justification of *all* evaluative beliefs,[1] while some of them are targeted at moral beliefs only,[2] and some at a certain subset of moral beliefs.[3] All EDAs, however, take a common form. They all claim that knowledge of a certain subset of beliefs is improbable, since (i) such beliefs are shaped exclusively by the mechanisms of natural selection and (ii) evolutionary processes aim at reproductive success and thus are insensitive to attitude-independent evaluative truths, if there are any. The former is the *empirical* premise, and the latter is the *insensitivity* premise. EDAs also have an *epistemic* premise, namely that if non-naturalist evaluative (or moral) realism,[4] the empirical premise, and the insensitivity premise are true, then we cannot justify the beliefs in question. These three premises constitute the blueprint of any EDA:

1. *Empirical premise.* Evolutionary mechanisms have a pervasive influence on the content of our evaluative/moral beliefs.
2. *Insensitivity premise.* Evolutionary mechanisms aim at reproductive success and not attitude-independent evaluative/moral truths.
3. *Epistemic premise.* If there are attitude-independent evaluative/moral truths, evolutionary mechanisms have a pervasive influence on the content of our evaluative/moral beliefs, and evolutionary mechanisms aim at reproductive success and not attitude-independent evaluative/moral truths, then we lack an independent reason to think that our evaluative/moral beliefs track the truth, i.e., we lack justification for our evaluative/moral beliefs.
4. *Skeptical conclusion.* We lack knowledge of attitude-independent evaluative/moral truths, if they exist at all.

The epistemic premise is the core of any debunking argument. There are many ways of forming beliefs. Think, for example, of people who rest their beliefs about an outcome of a football game, or about whether it is going to rain the following day on the behavior of animal oracles. Although their beliefs might turn out to be true, they are only incidentally true since animal behavior has nothing to do with the states of affairs in a football game or with the state of the atmosphere. Hence, we have a good reason to suppose that people who form their beliefs through a process that is not good at tracking the truth – as in the case of animal oracles – are not justified in their beliefs. Such processes are called "off-track." (Kahane 106) Hearing, on the other hand, is most of the time an epistemically reliable process, which means that it is good at tracking the truth. People whose beliefs are informed by their hearing mechanism are probably correct in their beliefs about what they hear, provided that they don't have an impaired hearing mechanism and no environmental factor is distorting their beliefs. For example, the fact that I hear Beethoven's Eighth Symphony in the radio at the moment is a good justification

for my belief that Beethoven's Eighth Symphony is now playing in the radio, provided that there is nothing unusual about my hearing mechanism and nothing in my surroundings affects its proper function. EDAs, therefore, are based on the crucial distinction between processes that track the truth and off-track processes.

1.2. Street and Joyce

According to Sharon Street's empirical premise, evolutionary processes have an enormous influence on the content of our evaluative beliefs. Although evolutionary forces do not *directly* determine the content of our evaluative beliefs, they select for "basic behavioral and motivational tendencies" (Street 113), which in turn play an important role in shaping the content of these beliefs. Street then presents her Darwinian Dilemma: either there is a relation between evolutionary influences on our evaluative beliefs and independent moral truths or there is not. If there is no causal connection between evolutionary mechanisms and moral truths, then evolutionary forces probably have a distorting influence on our evaluative beliefs, which are off-track. It would be an enormous and inexplicable coincidence to claim that our evolutionarily shaped (thus distorted) evaluative beliefs are true. Assuming a relation between selective pressures and evaluative truths is not a good strategy for the realist either, since the realist must appeal to a *tracking account* to explain the nature of that relation. Tracking account loses the scientific battle against what Street calls the *adaptive link account* because of its metaphysical commitments and its failure to explain why true evaluative beliefs promote survival (Street 126–30).

Joyce's EDA aims exclusively at our moral beliefs. His empirical premise is that we have a complete and "confirmed non-moral genealogy" (Joyce 190). Joyce's insensitivity premise claims that the best explanation of our basic moral judgments is that "they are expressions of underlying 'design features' of human psychology" (140). Just as we cannot justify our beliefs about the battle of Waterloo if they are caused by an imaginary belief pill that is insensitive to the facts about Waterloo, so we cannot justify our moral beliefs if they are generated by biological processes that are insensitive to proposed moral facts or truths. And Joyce's conclusion is that "we have no reason to believe in moral facts" (210).

The important difference between Joyce's and Street's respective EDAs is that while Joyce thinks his EDA supports moral skepticism, Street thinks her EDA ultimately supports the truth of anti-realism. Joyce adopts a realist semantics *only* about moral propositions – that is, only moral propositions refer to categorical reasons. Street, on the other hand, believes that *all* evaluative discourse should be viewed in realist terms. Since realism itself is an evaluative claim, we "must be to adjust our metaethical view so as to become antirealists" (Street 141). Thus, Street's EDA implies not only moral skepticism but also complete evaluative skepticism.

2. Ambition and Strength

It is important to note that if the debunking argument places the burden of proof on the realist, the argument then collapses into a pervasive skepticism about *all* of our beliefs.[5] The debunker's aim is to undermine a limited set of beliefs using scientific evidence. But if the debunker's argument asks the realist to provide an independent reason to think that their beliefs are not mistaken, then the empirical premise becomes superfluous and the argument's target extends to our entire body of belief.[6] We may never provide good (independent) reasons for the truth of *any* of our beliefs due to *some* possible distorting factor, but this general epistemic worry has never been the debunker's concern. Rather, the debunker is concerned with the rationality of our evaluative or moral beliefs. To this end, she gives scientific evidence that gives us reason to think that our evaluative or moral beliefs are mistaken due to the distorting influence of evolutionary forces on such beliefs. Thus, the burden of proof must be on the debunker: she needs to give us evidence of error (a good reason that we are probably mistaken in our evaluative or moral beliefs) that follows from an empirical claim (evolutionary evidence), rather than asking the realist to give independent reason that our beliefs are not mistaken. Only then she can selectively claim that we cannot rationally maintain our evaluative or moral beliefs.

When formulating the empirical premise, the debunker determines the *target* of her argument: empirical evidence could show that *all* of our *evaluative* beliefs have been shaped by evolutionary processes, or it could show that *all* of our *moral* beliefs are determined by such processes. Street's ambitious EDA chooses the former target, while Joyce's less ambitious EDA chooses the latter. On Vavova's account, Street's argument is less likely to succeed than Joyce's argument due to the epistemic principle that she calls "the Inverse Rule of Debunking." However, she thinks that both accounts are ambitious enough to fail.

2.1. The Inverse Rule of Debunking

Vavova's claim is that the more ambitious a debunking argument becomes the less prospect of success it has. This is because the debunker tries to give us good reasons to think that we are mistaken about a certain body of beliefs, and what makes a reason good is its independence from the set of beliefs that are called into doubt. For example, if the aim of your argument is to undermine all perceptual beliefs, it would become illegitimate to base any of your premises on the truth of your beliefs that are formed through your senses. Street's EDA calls all of our evaluative beliefs into doubt; thus, the independent ground that reveals our mistake in our evaluative beliefs cannot involve any evaluative claim. Joyce's EDA can employ our nonmoral evaluative beliefs as an independent ground for the evidence of error, since its target encompasses all of our moral beliefs. Vavova calls this epistemic principle "The Inverse Rule of Debunking," according to which "[t]he

potential strength of a debunking argument is inversely proportional to its ambition" (Vavova 98).

Although I don't see any compelling reason to reject the principle,[7] I argue that ambition of an EDA does not really affect the strength of its *epistemic* premise. Rather, the relation between ambition and strength of an EDA becomes relevant only with respect to the scope of evolutionary influence on the content of our beliefs, since EDAs are essentially inductive arguments. The right strategy to debunk the debunker is thus to attack the *empirical* premise of her argument, which determines inductive strength of any EDA.

2.2. Self-Refutation Argument

Since the debunker's aim is not to deductively prove that our evaluative or moral beliefs are *necessarily* wrong but merely to show that such beliefs are *probably* wrong given our evolutionary and epistemic conditions, EDAs are *inductive* arguments. An inductive argument infers from a limited number of observations to a general, probabilistic conclusion. For instance, when we reach the conclusion that the sun will probably rise tomorrow from the set of observations that sun has risen regularly so far, we give an inductive argument. Similarly, an EDA reaches the conclusion that our evaluative or moral beliefs are *probably* wrong from an evolutionary explanation of such beliefs. When we are given an EDA, we realize that there is a discrepancy between what we (or realists, to be more specific) take evaluative/moral judgments to be and how evolutionary psychology describes them.[8] We also make a distinction between objective and subjective judgments.[9] From these observations we infer that our evaluative/moral beliefs are probably mistaken, just as we infer that the sun will probably rise tomorrow. Making such inferences is just one of the things our minds naturally do.

When we infer the probabilistic conclusion of an EDA from the scientific evidence it provides, we make an assumption about epistemic reasons, namely that scientific evidence has the power to undermine our intuitions. But what makes us believe that scientific evidence is epistemically more reliable than our intuitions? It is perfectly possible that having this evaluative claim is also an adaptation. Wouldn't this then threaten the kind of EDA that calls *all* of our evaluative beliefs into doubt? This is Vavova's objection to Street's EDA (87–9). Vavova argues that Street's EDA targets both practical and epistemic reasons, both of which have been shaped by natural selection. The idea is that if we cannot trust any of our evaluative beliefs, then we cannot trust our beliefs about whether our evaluative beliefs are debunked by the argument. "[T]o evaluate we must rely on the evaluative" (Vavova 89); however, if the argument aims to undermine *all* evaluative judgments, then we lack the resources to determine whether the targeted beliefs are debunked. Hence, Vavova concludes, Street's EDA is self-refuting.

I don't think Vavova's strategy delivers a knockout blow to Street's argument. Recall that an EDA is an inductive argument and that inductive arguments allow

for their conclusion to be false due to their probabilistic nature. Street does not claim that our evaluative beliefs are *necessarily* wrong but that they are *probably* wrong given our evolutionary and epistemic conditions. The conclusion of Street's EDA allows the possibility that some of our evaluative beliefs turn out to be true, and it follows that these true evaluative beliefs could include some of our beliefs about epistemic reasons, science, mathematics, and so on.

As a matter of fact, our beliefs about epistemic reasons are more likely to be true compared to our beliefs about concrete evaluative matters, even though the former are also beliefs. Our beliefs about epistemic reasons are *beliefs about beliefs* because they are about whether our beliefs are epistemically trustworthy. But *beliefs about beliefs* are categorically different from *beliefs about specific cases*. Notice the difference between two kinds of statements: (1) "Our evaluative beliefs are probably false unless supported by empirical evidence," and "Our scientific beliefs are probably false unless supported by empirical evidence;" (2) "My wife is beautiful," and "The earth is flat." The former kind of statements is more likely to be a product of *reasoning* rather than *biological conditioning*, and thus it is more likely to be true. This is because the reasoning that is involved in the former kind of statements forces one to distance themselves from their (possibly distorted) beliefs about specific cases and make them realize that such beliefs are epistemically vulnerable. And this process allows room for belief revision.

Moreover, there is a gap between the evolutionary origins of something and its truth: the fact that our genes partly determine our acceptance of evolutionary theory does not make evolutionary theory wrong, or the fact that our genes caused us to engage in mathematics does not render mathematical propositions wrong. Similarly, conceding that our beliefs about epistemic reasons have been shaped by selective pressures does not entail that those beliefs and the probabilistic conclusion of Street's EDA are wrong.

It is difficult to declare an EDA to be self-refuting unless its epistemic premise renders our beliefs about epistemic reasons wrong. I am pointing to the difference between the statements "All *X*s are wrong" and "All *X*s are *probably* wrong," when both statements are themselves instances of *X*. It is more difficult to call the latter self-refuting because it leaves open the possibility of itself (or any other instance of *X*) being true. Also, even self-refuting statements or arguments can give us important clues about the issue being discussed. Take, for example, Mark Twain's famous statement "All generalizations are false, including this one." It is one way to approach this generalization about generalizations (or belief about beliefs) by claiming that it is self-refuting and thus it does not establish *anything*. It is also possible to take it to show ultimately that *some* generalizations are true. If the statement is true, then all generalizations but the statement itself are false. If the statement is false, then some generalizations are true. In both cases, some generalizations are true. Likewise, it is possible to take the statement "All of our

evaluative beliefs are *probably* false, including this one" to show ultimately that *some* of our evaluative beliefs are true.

What about Joyce's EDA? According to Vavova, EDAs that target moral beliefs, such as Joyce's EDA, still target too much (90–3). Such EDAs claim that we are probably mistaken about morality because our moral beliefs have been shaped by natural selection, which is an off-track process. However, Vavova argues, to show that there is no relation between adaptive moral beliefs and true moral beliefs, we first need to know something about the contents of true moral beliefs and adaptive moral beliefs. And this requires us to make assumptions about what morality is like. Otherwise, morality could be about anything and accordingly we would have no reason to think that adaptive moral beliefs and true moral beliefs do not coincide. EDAs like Joyce's make such assumptions (e.g., moral judgments are rationally authoritative) but they at the same time call our entire body of moral beliefs into doubt. They thereby render their own moral assumption illegitimate. Thus, EDAs that declare *all* of our moral beliefs to be epistemically suspect cannot give us independent reason to think that we are mistaken about morality.

I don't think such an objection refutes Joyce's EDA. Since the epistemic premise of an EDA takes the form of a *reductio*, it is essential for any EDA to assume something about morality or about normative domain in general. Otherwise, it would not go through. If your aim is to debunk moral realism, you should first assume that moral realism is true. If your aim is to debunk our moral beliefs altogether, you should first make an assumption about basic commitments and presuppositions to morality. The epistemic premise is a conditional: *If* our moral beliefs have such-and-such features, then they are *probably* wrong, considering the extensive influence of evolutionary forces on the content of those beliefs and the insensitivity of evolutionary processes to the truth of them.[10]
Moral beliefs could, of course, have different features than it is assumed by the argument. It is possible that the correct account of morality is an *anti-realist* one. In that case, only the assumed conception of morality, that is moral realism, could be debunked. This is where Joyce seems to go wrong. He takes the conclusion of his argument to have debunked morality in general. However, he thereby dismisses anti-realist conceptions of morality that could escape the evolutionary challenge.[11] Although there seems to be nothing wrong with making a metaethical *assumption* to get the argument going, dismissing alternative conceptions of morality could possibly create a problem for the debunker, if their intention is to debunk morality as a whole.

2.3. Ambition and Inductive Strength

I have argued that the level of ambition of an EDA does not affect the strength of its *epistemic* premise because EDAs are essentially inductive arguments. Focusing on the epistemic premise and declaring more ambitious EDAs to be self-defeating do not remove the skeptical worry that we might be mistaken in our

evaluative/moral beliefs. As long as one admits that our beliefs are heavily shaped by natural selection, it is natural and plausible to think that objective morality could simply be an illusion. And this worry remains even if we think the argument is self-defeating. Does this mean there is no relationship whatsoever between an EDA's ambition and strength?

There *is* such a relationship, but the level of ambition only affects the strength of the *empirical* premise. The only way to ease the skeptical worry seems to be to show that there is *no extensive* evolutionary influence on our beliefs. Many philosophers and evolutionary biologists agree that certain capacities and tendencies relevant to evaluative thought and behavior, and *some* of the content of our evaluative beliefs *can* be explained by evolution. However, there is much less agreement among them on whether evolutionary forces have a *pervasive* influence on the content of our evaluative beliefs. The idea is that the effects of *human culture* and *moral reasoning* on the contents of our evaluative beliefs can be thought of independently from the effects of biological evolution on such beliefs.[12] I cannot go into detail and review the literature due to space limitation; however, for our purposes, it will suffice to say that the more ambitious an EDA is the more difficult it is to provide a complete evolutionary explanation of the beliefs in question.

If the debunker can show conclusively that her empirical premise is true, her argument will get very strong. However, the more sets of beliefs she claims to have determined by our biological nature, the *more difficult* it gets to provide a complete evolutionary origins story. For instance, it would be more difficult for a debunker who tries to debunk all of our evaluative beliefs to prove her empirical premise than a debunker who aims only at our moral beliefs. But provided that both debunkers succeed in their respective tasks, the former debunker's argument would get inductively stronger than the latter debunker's argument, for her empirical premise would encompass a greater number of beliefs. Thus, the level of ambition of an EDA has an effect on the strength of its *empirical* premise, and accordingly on the inductive strength of the argument.

3. Concluding Remarks

I have rejected Vavova's claim that Street's and Joyce's EDAs target too much and become self-refuting. This is because EDAs are essentially *inductive* arguments with a *probabilistic* conclusion, which allows *some* of our beliefs to be true. Furthermore, even a self-refuting debunking argument could indicate the truth of a subset of what is being debunked. The "Inverse Rule of Debunking" is true, but not for the reasons Vavova provides. In other words, the level of an EDA's ambition affects the argument's strength but not because more ambition causes an internal contradiction. Rather, the more set of beliefs an EDA calls into question the harder it becomes to provide a complete evolutionary origins story. The

empirical premise is the chink in an EDA's armor.

NOTES

1. Sharon Street's EDA in her 2006 paper is generally considered as an example of such an ambitious argument (cf. Kahane 114–7; Das 430–2; Vavova 80–1).
2. See Richard Joyce.
3. See Joshua Greene.
4. Typically, non-naturalist evaluative or moral realism is committed to three claims: (1) a claim about moral judgments being truth-apt (cf. Sayre-McCord 5), (2) a claim about moral principles being attitude-independent (cf. Shafer-Landau 15), and (3) a claim about the moral facts obtaining (ibid.). The first claim distinguishes realism from non-cognitivist accounts, according to which moral judgments do not try to describe some feature of the world but they are expressions of our emotional or other non-propositional attitudes. The second claim distinguishes realism from relativist accounts, which assert that moral truth is constructed by individual preferences or social conventions. And the third claim distinguishes realism from nihilistic accounts that reject moral truth and value altogether. When I use the term 'realism' throughout the paper I refer to *non-naturalist* realism, unless stated otherwise.
5. I agree with Vavova (82–4) on this point.
6. No empirical premise is needed for the following argument to go through: (P_1) There are infinitely many possible states of belief; (P_2) One of those states of belief is correct; (P_3) It is unreasonable to adopt a belief if we don't have good reason for our belief being correct; (P_4) We have no good reason to think that we are in the correct state. (C) It is unreasonable to believe that we are in the correct state. An EDA that places the burden of proof on the realist takes this specific form and ultimately asks the realist to justify *all* of their beliefs, not just their evaluative or moral beliefs (cf. Vavova 83–4).
7. In fact, the principle reflects the nature of moral and political disagreements very well. Think of two people (person *A* and person *B*) you disagree with on a moral matter *M*. Imagine that you agree with the person *A* on many other moral matters, whereas you disagree with the person *B* on *all* moral matters. Your disagreement with *A* has a significant debunking potential since you have a common independent ground from which you can provide evidence of error. Your disagreement with *B* has no such debunking potential. It is not even possible to evaluate your disagreement with *B* as you lack a common independent ground to move forward (cf. Vavova 99).
8. According to the realist, evaluative/moral claims refer to *categorical* (attitude-independent) reasons. However, evolutionary psychology tells us that such claims are based on evolved psychological dispositions that favor adaptive behaviors. That is, we form our evaluative/moral beliefs not as a response to attitude-independent evaluative/moral truths but simply because they promote reproductive success. This conflicts with the presupposed categorical nature of morality.
9. Any argument that moves from scientific evidence to evidence of error in our beliefs indicates the distinction between objective and subjective judgments or standpoints. For example, there is an important difference between the statements "It is cold out here"

and "It is 20 degrees Fahrenheit," even though both refer to the same state of affairs. While the former reflects a *subjective* outcome of an interaction between one's body and a state of affairs (some species could feel warm in those conditions), the latter reflects an *objective* fact. Facts about temperature and the fact that evolutionary forces shape psychological dispositions (both in humans and other primates) that favor adaptive behaviors are thus *objective*, whereas beliefs such as "It is cold out here" or "It is wrong not to help distant strangers" are *subjective* in this context.

10. To reiterate, the epistemic premise and the conclusion it leads to allow the possibility of some of our moral beliefs' being true. And these true moral beliefs could include beliefs about the nature of moral beliefs. Nothing in the argument precludes this possibility.

11. Street does not fall into this trap and acknowledges that anti-realist conceptions of morality are safe against her EDA (152–4).

12. Cf. Nagel 142; Sober 93–113; Ayala 258–60; FitzPatrick; Prinz.

WORKS CITED

Ayala, Francisco J. *Evolution, Explanation, Ethics and Aesthetics: Towards a Philosophy of Biology*. Academic Press, 2016.

Das, Ramon. "Evolutionary Debunking of Morality: Epistemological or Metaphysical?" *Philosophical Studies*, vol. 173, no. 2, 2016, pp. 417–35. *SpringerLink,* doi: 10.1007/s11098-015-0499-9.

FitzPatrick, William J. "Why There Is No Darwinian Dilemma for Ethical Realism." *Challenges to Moral and Religious Belief*, edited by Michael Bergmann and Patrick Kain, Oxford University Press, 2014, pp. 237–55.

Greene, Joshua. "The Secret Joke of Kant's Soul." *Moral Psychology, vol. 3*, edited by W. Sinnott-Armstrong, MIT Press, 2007, pp. 35–79.

Joyce, Richard. *The Evolution of Morality*. MIT Press, 2006.

Kahane, Guy. "Evolutionary Debunking Arguments." *Noûs*, vol. 45, no. 1, 2011, pp. 103–25. *Wiley Online Library*, doi: 10.1111/j.1468-0068.2010.00770.x.

Nagel, Thomas. "Ethics without Biology." *Mortal Questions*, Cambridge University Press, 1979, pp. 142–46. *Cambridge Core*, doi: 10.1017/CBO9781107341050.012.

Prinz, Jesse J. "Is Morality Innate?" *Moral Psychology, The Evolution of Morality: Adaptations and Innateness*, vol. 1, edited by Walter Sinnott-Armstrong, MIT Press, 2008, pp. 367–406.

Sayre-McCord, Geoffrey. "The Many Moral Realisms." *Southern Journal of Philosophy*, vol. 24, Supplement, 1986, pp. 1–22. *Wiley Online Library*, doi: 10.1111/j.2041-6962.1986.tb01593.x.

Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford University Press, 2003.

Sober, Elliott. *From a Biological Point of View: Essays in Evolutionary Philosophy*. Cambridge University Press, 1994.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies*, vol. 127, no. 1, 2006, pp. 109–66. *SpringerLink*, doi: 10.1007/s11098-005-1726-6.

Vavova, Katia. "Debunking Evolutionary Debunking." *Oxford Studies in Metaethics*, vol. 9, edited by Russ Shafer-Landau, Oxford University Press, 2014, pp. 76–101.

Program of the 71th Annual Meeting of the New Mexico Texas Philosophical Society

May 21-22, 2021

Friday Morning Session 1

Nick Harding, University of Southampton
"Monogamy is not Immoral: Sexual and Romantic Exclusivity can be Justified by Appealing to the Difficulty of Managing Jealousy"
Comments: **Alex Gillham**, St. Bonaventure University

Alex Gillham, St. Bonaventure University
"Deprivationism, Abortion, and The Impairment Argument"
Comments: **Nick Harding**, University of Southampton

Christopher Stratman, University of Nebraska-Lincoln
"Ectogenesis and Misogyny"
Comments: **Seth Goldwasser**, University of Pittsburgh

Friday Morning Session 2

Xuanpu Zhuang, Bowling Green State University
"Public Justification and Changes of Personal Commitments"
Comments: **Daniel Grasso**, University of Missouri, St Louis

Reuben Sass, Rice University
"Defending a Connection between Anger and Judgments of Blameworthiness"
Comments: **Carissa Phillips-Garrett**, Loyola Marymount University

Claire Kirwin, Clemson University
"Sympathy for the Devil?: The Guise of the Good Remastered"
Comments: **Leonard Kahn**, Loyola University New Orleans

Friday Morning Session 3

Eric Bayruns García, California State University, San Bernardino
"Belief Content and Rationality: Why Racist Beliefs Are Not Rational"
Comments: **Alexandra Lloyd**, University of Colorado, Boulder

Greyson Abid, University of California, Berkeley
"Towards a Two-Factor Approach to the Cross-Race Effect"
Comments: **Justin Bell**, University of Houston-Victoria

Zachary Simpson, University of Science and Arts of Oklahoma
"How the "As If" Becomes True: Fictions from Vaihinger to Appiah and Beyond"
Comments: **Robert Tierney**, University of Houston

Friday Afternoon Session 1

Walter Barta, University of Houston
"Against Diminishing Marginal Utility"
Comments: **Lamont Rodgers**, Houston Community College

Sean Donahue, University of Southern California
"Autonomy and Information Overload"
Comments: **Emil Badici**, Texas A&M, Kingsville

Emil Badici, Texas A&M, Kingsville
"Newcomb's Problem and the Prisoner's Dilemma: two problems or one?"
Comments: **Xuanpu Zhuang**, Bowling Green State University

Friday Afternoon Session 2

Robert Tierney, University of Houston
"Rehabilitating Agent-Regret"
Comments: **Paul Carron**, Baylor University

Ian Hosbach, University of North Carolina at Charlotte
"Becoming-with-Others, Becoming-Careful"
Comments: **Sarah Woolwine**, University of Central Oklahoma

Daniel Grasso, University of Missouri, St Louis
"Rational Resolve as Magnification: A Response to Holton"
Comments: **Jean-Paul Vessel**, New Mexico State University

Friday Afternoon Session 3

Jennifer Wargin, Texas A&M University
"A Proposal For Disrupting Partisan Seating In Congress"
Comments: **Christopher Stratman**, University of Nebraska-Lincoln

Eric Shoemaker, University of Toronto
"The Equal Opportunity To Be A Legislator: Why Randomly Selecting Legislators Is More Democratic Than Electing Them"

Comments: **Jennifer Wargin**, Texas A&M University

Sarah Woolwine, University of Central Oklahoma, and **Justin Bell**, University of Houston-Victoria

"John Dewey, Moral Theory, and Intellectual Disability"

Comments: **David Beisecker**, University of Nevada-Las Vegas, and **Jessica Ain**, University of Nevada-Las Vegas

President's Address

Jean-Paul Vessel, New Mexico State University

"Was Socrates the Wisest Man in Athens, and If So, in Which Respects?"

Saturday Morning Session 1

A.G. Holdier, University of Arkansas

"On the Epistemic Threats of Counterfeit Memorials"

Comments: **Pranav Ambardekar**, The Ohio State University

Caner Turan, Tulane University

"Are Ambitious Evolutionary Debunking Arguments Self-Refuting?"

Comments: **Katherine Deaven**, University of Wisconsin-Madison

Pranav Ambardekar, The Ohio State University

"Why Knowledge Might Not Entail Belief"

Comments: **Peter Hutcheson**, Texas State University

Saturday Morning Session 2

Jackson Hoerth, Temple University

"The Dual Role of Kant's Imagination and Nature as Art"

Comments: **Daniel Dal Monte**, Temple University

Carissa Phillips-Garrett, Loyola Marymount University

"Judgment and Sanction in Aristotle's Account of Blame"

Comments: **Harrison Lee**, Baylor University

Paolo Verdini, University of Alberta

"The Mirror as the function for the Self and the Other: Merleau-Ponty Between Phenomenology and Child Development"

Comments: **Zachary Simpson**, University of Sciences and Arts of Oklahoma

Saturday Morning Session 3

Jared Oliphint, Texas A&M University

"Beyond Types and Tokens: A Case for Expanding the Ontology of Words"

Comments: **Savvas Ioannou**, University of St Andrews

Paul Kelly, University of Wisconsin-Madison

"On the Plausibility of a Dispositional Analysis of Intrinsic Value"
Comments: **Reuben Sass**, Rice University

Daniel Dal Monte, Temple University
"Methodological TI: An Ambiguous Position Unable to Secure the Scope Restriction on Spatiotemporal Predicates"
Comments: **Jackson Hoerth**, Temple University

Saturday Early Afternoon Session 1

Seth Goldwasser, University of Pittsburgh
"Skills and Conceptions: How We Know-How"
Comments: **Sean Donahue**, University of Southern California

Alexandra Lloyd, University of Colorado-Boulder
"#MeToo: An Epistemic Account"
Comments: **A.G. Holdier**, The University of Arkansas

Saturday Early Afternoon Session 2

Joshua Anderson, Virginia State University
"In Defense of Character Consequentialism"
Comments: **Kenneth L. Brewer**, University of Texas at Dallas

Lamont Rodgers, Houston Community College
"Against a Consequentialist Analysis of Utopia"
Comments: **Walter Barta**, University of Houston

Saturday Early Afternoon Session 3

Daniel Flores, Houston Community College
"Does God Love You? Love, Timelessness, and Subjective Experience"
Comments: **Ian Hosbach**, University of North Carolina at Charlotte

Harrison Lee, Baylor University
"A Reductio for Divine Ideas Creationism"
Comments: **Vanessa Voss**, Lone Star College

Saturday Late Afternoon Session 1

Katherine Deaven, University of Wisconsin-Madison
"Relative Significance Controversies in Evolutionary Biology"
Comments: Daniel Flores, Houston Community College

Kenneth L. Brewer, University of Texas at Dallas
"'We Know You Better Than You Know Yourself': Aesthetic Taste and Recommender Systems"
Comments: Claire Kirwin, Clemson University

Saturday Late Afternoon Session 2

Marie Le Blevennec, Boston University
"Do Victims of Injustice Have a Fairness-Based Duty to Resist Them?"
Comments: **Joshua Anderson**, Virginia State University
Leonard Kahn, Loyola University New Orleans
"Buying Luxuries and Saving Lives"
Comments: **Caner Turan**, Tulane University

Saturday Late Afternoon Session 3

Savvas Ioannou, University of St Andrews
"Intuition of Mind-Brain Distinctness: Why Do We Have It?"
Comments: **Paul Kelly**, University of Wisconsin-Madison
Alec Oakley, University of Victoria
"Emotion and Expectation"
Comments: **Greyson Abid**, University of California-Berkeley

Business Meeting

Institutional affiliations reflect the institutions of the participants at the time of the conference.