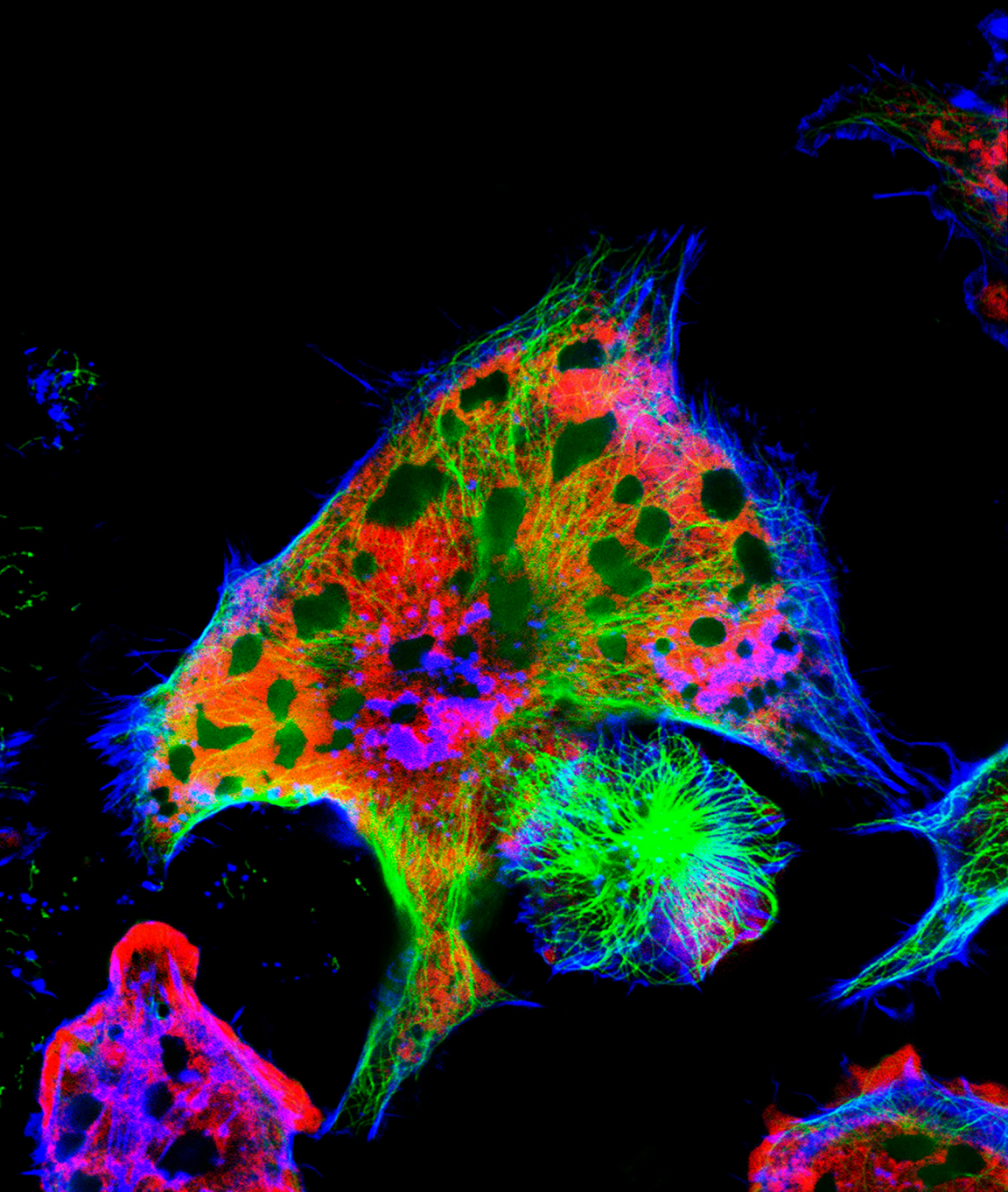
compos mentis

Undergraduate Journal of Cognition and Neuroethics
Volume 10, Issue 1



compos mentis

Student Editor
Devon Frazee

Production Editor
Zea Miller

Publication Details
Volume 10, Issue 1 was digitally published in July of 2022 from Flint, Michigan, under ISSN: 2330-0264.

The *Compos Mentis: Undergraduate Journal of Cognition and Neuroethics* is produced by the Center for Cognition and Neuroethics. For more on CCN or this journal, please visit cognethic.org.

Center for Cognition and Neuroethics
University of Michigan-Flint
Philosophy Department
544 French Hall
303 East Kearsley Street
Flint, MI 48502-1950

Table of Contents

Table of Contents (cont.)

compos mentis

Moral Responsibility in the Age of Free Will Skepticism: A Defence of Frankfurtian-Compatibilism

Owen Crocker
University of Victoria

ABSTRACT
Free will skepticism is radical in its core claim that free will is illusory. Criminal law, however, appears to presuppose that persons are free and hence, morally responsible for their actions. So, if free will skepticism is true, our current practices that hold people to account for their wrongs appears unjustified–even immoral. This paper will challenge the free will skeptic's core claim that free will does not exist and defend current practices of moral responsibility by offering (and defending) a Frankfurtian-compatibilist approach to the topics of free will and determinism.

KEYWORDS
Free Will, Moral Responsibility, Crime, Free Will Skepticism, Criminal Law, Compatibilism, Determinism, Libertarianism, Legal Personhood, Autonomy, Rationality

INTRODUCTION

Crime, free will, and moral responsibility are socially and philosophically loaded terms, which often convey different meanings to people. The criminal justice system utilizes and refers to many of these terms to justify its punishment for those found guilty of criminal activity.

The practice of holding citizens morally responsible for their behaviour is a common social phenomenon, that most do not seriously question. However, according to free will skeptics (those who argue that free will does not exist), social practices that hold persons morally responsible are unjustified and in need of reform. The goal of this paper is to demonstrate that the legal practice of holding people morally responsible can be justified using a Frankfurtian-compatibilist approach.

The paper will be divided into three parts. The first part will outline three presuppositions of the criminal law. The second part of the paper will provide the reader a short summary on free will skepticism. The third part of the paper will present an alternative approach to free will skepticism. That approach being, Frankfurtian-compatibilism. In this section, Frankfurtian-compatibilism will be described as well as defended against two possible objections.

PART ONE: CRIMINAL LAW'S PRESUPPOSITIONS: LEGAL PERSONHOOD, AUTONOMY AND RATIONALITY

In this portion of the paper, three presuppositions of the criminal justice system[1] will be discussed: legal personhood, autonomy, and rationality. It seems to me that all three of these concepts play an integral role in the criminal justice system's justification in apportioning moral responsibility.

Regarding legal personhood, Naffine writes, "…the law of persons is not a discrete field of study in the common law world, such as torts, or contract or criminal law, but is a *pervasive underlying concept throughout the different branches of law*" (2009, 15) [Emphasis Added]. Criminal law, like other legal disciplines, all

1. It should be noted that the author is writing from a Canadian perspective. In this regard, references to case law or criminal practices are referring to Canadian law and practices. Many of the themes and elements will be recognizable in other criminal justice systems, but there will likely be some distinguishing factors between various countries methods of justice.

necessarily utilize what is termed, legal personhood. Without legal personhood, there would be no persons to convict of crimes, or for that matter, persons who could facilitate the court processes of convicting and acquitting those accused of crimes. The criminal law's person is comprised of many factors in assessing one's level of moral responsibility for their behaviour. For example, a 5-year-old, in the eyes of the law, is not the proper subject of punishment for breaking and entering, whereas a 27-year-old is. In addition to age, the criminal law also presupposes that the legal person is of sound mind, demonstrating the contrary would excuse a person by means of the *not criminally responsible due to mental disorder* defence. However, if such a defence is not raised or not considered, it is assumed that the accused is "normal" regarding their cognitive abilities and decision-making processes. The law of persons, as Naffine notes, is an *underlying* concept of law (2009, 1). Legal personhood can be seen as the base conceptual unit (Naffine, 2009, 1), which grounds the other faculties and powers required by the court, including autonomy and rationality.

Autonomy, meaning self-rule, is another underlying assumption in how persons are viewed and judged by the court. The ability to self-govern and control one's course of action in life is an incredibly powerful and entrenched thought in liberal societies (Nedelsky, 2011/2012, 121). As Ferguson writes, the criminal law presupposes "that human beings are rational and *autonomous*. We have the capacity to reason right from wrong, and the capacity to *choose right or wrong*. These assumptions may be incorrect, but they are, and are likely to remain, the theoretical basis of our criminal law" (1989, 140). When considering autonomy, the court will take into consideration the circumstances and factors surrounding a person's actions to determine if they were acting in a truly autonomous manner.

As Ferguson noted above, the criminal law assumes that persons are *rational* in that we possess the ability to *reason* right from wrong (1989, 140). Those deemed to lack rationality, such as children and those who suffer from mental disorders, are generally excused from being held liable, or are at least held to a lesser degree of responsibility. In the philosophical literature, the ability to reason right from wrong is often referred to as *reason-responsiveness*. Duff describes reason-responsiveness as,

> The capacities on which responsibility depends are best understood as a matter of reason-responsiveness: a responsible agent is one who is capable of recognising and responding to

> the reasons that bear on his situation. A responsible agent is 'responsable'[2] to reasons: which means not that he is responsible only when and insofar as he is actually responsive to reasons (since we can be responsible for our very failures to respond to relevant reasons), but that he is responsible insofar as he is capable of responding appropriately to relevant reasons. (2007;2009, 39)

The ability to reason right from wrong and to weigh the relevant factors before one chooses[3] is a morally salient element that the courts will require in determining one's culpability.

At this point, it should be clear that the criminal justice system views *certain* persons as the appropriate subjects to be held morally responsible. As noted in this section, not all persons are appropriate candidates, as the elements of legal personhood, autonomy and rationality must be present. When these features are shown to be absent, the court will either not hold the person morally responsible or lower the level of their responsibility in the matter. However, all these distinctions and practices that are made in the criminal justice system take place because there is the general belief that persons are morally responsible for their actions (or that persons are at least in principle capable of being held morally responsible for their actions). However, not all persons share this belief. In the following section, free skepticism will be introduced.

PART TWO: FREE WILL SKEPTICISM

Free will skepticism, according to one of its lead proponents, Gregg Caruso, can be described as a viewpoint that "takes seriously the possibility that human beings are never morally responsible in the basic desert sense" (2021, 8). In other words, persons lack the necessary control to make themselves the proper subjects to be truly deserving of blame or praise for their actions. As will be argued below, the legal concepts of autonomy and rationality appear to make two assumptions.

2. This term is used by Duff to mean "able to respond" to reasons, hence the different spelling.

3. Or the lack thereof of demonstrating sufficient rationality which the Court would consider negligence, wilful blindness, recklessness etc. For further discussion, see Sarch, A. (2019). *Criminally ignorant: Why the law pretends we know what we don't. Oxford University Press* at page 29.

First, that the agent in question could have acted differently than they did. Second, that agents are the source of their actions in a way that allows them to be morally responsible. However, if causal determinism is true, according to many free will skeptics,[4] we have strong reasons to doubt both assumptions.

Philosophical viewpoints don't arise out of thin air. Rather, they are the culmination of many philosophical, social, and scientific theories. In the case of free will skepticism, there are essentially "two routes" that may lead someone to accept the free will skeptic's conclusion. The first of these routes relies on scientific findings. Specifically, the research conducted by Benjamin Libet was very influential to the free will skeptic's cause. In the 1980's, Libet ostensibly demonstrated the absence of free will. As Caruso and Pereboom write,

> The pioneering work…by Benjamin Libet and his colleagues… investigated the timing of brain processes and compared them to the timing of conscious will in relation to self-initiated voluntary acts and found that the conscious intention to move…came 200 milliseconds before the motor act, but 350–400 milliseconds after the readiness potential (RP)—a ramplike buildup of electrical activity that occurs in the brain and precedes actual movement. (2018, 195)

Stated differently, our conscious will to move our body in a specific way or speak a specific phrase was happening *after* the electrical activity began to perform that specific action. Libet-type experiments are intended to demonstrate that when a person makes a choice to act in a certain way, that choice is not a result of their personal deliberation, but rather stems from an unknown prior physical cause. If this is true, then such findings call into question the assumption that persons are the true source of their actions. However, such scientific findings are questionable for various reasons (Caruso and Flanagan, 2018, 196-197).[5] In addition to scientific

4. Many free will skeptics argue that free will does not exist regardless of whether it turns out determinism or indeterminism is true. However, for the purposes of this paper, determinism will be assumed.

5. For example, one can concede that ramp-like electric activity builds up before a conscious decision is made but does not need to agree that such electric activity determines how one acts. It is possible that during or after the electric buildup, one makes the decision regarding how they will act. In this sense, the electric activity is simply a precursor to being able to act, not a determinant in how one acts.

reasoning, metaphysical speculations have also played an important role in the reasoning of free will skeptics.

The second route that may lead one to accept free will skepticism is premised on the idea of determinism. Such philosophical speculations can be traced all the way back to the Stoics in the 3rd century BCE (McKenna and Pereboom, 2014, 263). Determinism is the idea that every event that occurs can be fully explained by previous causes, whether those causes be social or biological. As Dennett writes, "If determinism is true, then our every deed and decision is the inexorable outcome, it seems, of the sum of physical forces acting at the moment, which in turn is the inexorable outcome of the forces acting an instant before, and so on, to the beginning of time" (2015, 5). Essentially, according to determinists, our choices and decisions are not ultimately *rooted* in our agentic powers, but rather are necessitated by prior causal forces that precede our every action and thought. If this deterministic picture is correct, it is difficult to see how a person could have acted differently than how they did in any prior circumstance. If everything is determined beforehand, how can persons be held responsible if they couldn't have acted otherwise than how they did? In comparison to the scientific findings of Libet which focused on physical processes in the brain, deterministic theories tend to focus on the metaphysical implications of a materialistic universe on free will and moral responsibility. Therefore, assuming our actions are determined, the criminal law's practice of holding people morally responsible for their actions becomes increasingly questionable.

PART THREE

Frankfurtian-Compatibilism

In contrast to free will skepticism, a viewpoint which holds the incompatibility of determinism and free will (as well as moral responsibility), compatibilists hold that determinism and free will are *compatible* with one another. There are many variations of compatibilism–all offering different accounts of what constitutes free will and moral responsibility. As stated at the beginning of this paper, the Frankfurtian-compatibilist account will be offered as an approach that can justify the current legal practices of holding persons morally responsible for the crimes they commit.

Frankfurtian-compatibilism is premised on what are called, first and second-order desires. First-order desires are those desires which have objects as their goal. For example, eating a delicious meal or getting a good night sleep. On the other hand, second-order desires have other desires as their goal. In other words, they are "desires about desires" (McKenna and Coates, 2021). For instance, one might think to herself, that it would be great to have the motivation to study for a big test. In this sense, the subject is desiring the desire to study. The ability to be able to reason between our first and second-order desires as well as supplant previous desires with new ones is the foundation of Frankfurt's conception of free will. Compared to other animals, humans are unique in their ability to possess second-order desires which give people reasons not to carry out first-order desires. Frankfurt writes, "the statement that a person enjoys freedom of the will means (also roughly) that he is free to want what he wants to want. More precisely, it means that he is free to will what he wants to will, or to have the will he wants" (1971, 16). It is an important aspect of Frankfurt's theory that persons possess the power to change or have control other their desires. It should be noted that Frankfurt acknowledges that people's desires and how they choose to act considering those desires are "far more complicated" than the first and second-order paradigm (Frankfurt, 1971, 16). There is potentially no limit to how many higher order desires one might have, which can quickly complicate one's reasoning processes. However, Frankfurt's outline of free will coupled with the control of adding or replacing desires seems to be a sufficient grounding for how moral responsibility is attributed to people. Frankfurtian-compatibilism will be further strengthened by responding to two common objections. The first of those being, the principle of alternate possibilities.

The Principle of Alternate Possibilities

The first argument that will be considered is titled, *the principle of alternate possibilities*. When considering questions of free will and moral responsibility, the openness of the future and an agents' ability to choose their actions is of great importance. As stated in section two, determinism is the idea that "our every deed and decision is the inexorable outcome…of the sum of physical forces". So, if one views an accused's decision to commit a crime as the inexorable outcome of physical forces, then intuitions of their guilt and responsibility quickly disappear. This reasoning can be formalized into the following principle:

> PAP: A person is morally responsible for what she has done only if she could have done otherwise (Kane, 2005, 283).

As Haji notes, those who hold to PAP, "insist that alternative possibilities are required for the active control an agent must exercise in performing an action for the action to be free. They claim that determinism undermines free action or responsibility because it undermines active control by eliminating alternatives" (2012, 190). However, PAP along with its implications, is not without its dissenters.

Possibilities versus Reasons

In response to the principle of alternate possibilities, Frankfurt published a paper titled, *Alternate Possibilities and Moral Responsibility* (1969), in which he attempts to demonstrate why the principle of alternate possibilities is mistaken. In the paper, Frankfurt utilizes an *intuition pump*,[6] which is intended to make us question whether the ability to do otherwise is necessary for persons to be morally responsible. Frankfurt writes the following:

> Suppose someone - Black, let us say - wants Jones to perform a certain action. So, he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something other than what he wants him to do. What steps will Black take, if he believes he must take steps, in order to ensure that Jones decides and acts as he wishes? Let Black give Jones a potion, or put him under hypnosis, and in some such way as these generate in Jones an irresistible inner compulsion to perform the act Black wants performed and to avoid others.
>
> Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to

6. For more on intuition pumps, see Dennett, D. C. (2013). *Intuition Pumps and Other Tools for Thinking*. W.W. Norton & Co.

> take steps to ensure that he do it. It would be quite unreasonable to excuse Jones for his action, or to withhold the praise to which it would normally entitle him, on the basis of the fact that he could not have done otherwise. Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it. Jones has no alternative but to do what Black wants him to do. If he does it on his own, however, his moral responsibility for doing it is not affected by the fact that Black was lurking in the background with sinister intent, since this intent never comes into play. (1969, 835-836).

As one can see in Frankfurt's example, Jones appears to be acting in a *voluntary* manner. His decision to act is in line with his own desires and without outside interference. In this sense, Jones' act is free and is intuitively deserving of praise and blame. There appears to be little to no reason to conclude that Jones' behaviour is involuntary, even though he was unable to do otherwise. The Jones example is supposed to demonstrate that when determining if one should be held responsible for their actions, the salient moral criteria is not whether they could've acted otherwise, but instead what their *basis* and *reasons* were for acting in the manner they did. Such moral reasoning is made clear in Mckenna's following principle:

> L-Reply: A persons' moral responsibility concerns what she does do and her basis for doing it, not what else she could have done (McKenna and Pereboom, 2014, 104).

The L-Reply, like Frankfurtian-compatibilism, looks to the reasons (or basis) for why one acts as they do. It seems to me that this approach is preferable to PAP, as PAP will unjustifiably rule out many cases where one freely choose to act in a certain way. For example, if Jones' act was criminal in nature, it appears to satisfy all the relevant criminal law considerations. He acted voluntarily[7] (actus reus), intentionally (mens rea) and had no apparent defense to explain his behaviour. Therefore, Jones appears to have made a free choice, that he should be held responsible for. In light of these reasons, if the courts were to look at the *reasons*

7. Voluntarily in this sentence means physical voluntariness, not moral voluntariness.

and *basis* for one's actions, they would find Jones morally responsible. In contrast, however, if the courts applied PAP, Jones would not be responsible, since he couldn't have done otherwise. Intuitions may differ at this point, but it seems far from clear that Jones is not at least partially responsible for his decision.

It is important to note, however, that Frankfurt is not claiming that "the ability to do otherwise" is always irrelevant to one's moral blameworthiness. Rather, he is claiming that the ability to do otherwise is not a *necessary* feature of moral responsibility. Frankfurtian-compatibilism can account for situations where one is forced to act in a certain way due to extenuating circumstances.[8] On the other hand, the principle of alternate possibilities provides little use in helping the court assess the responsibility of the accused. The next thought experiment is intended to continue to pump our intuitions in the direction that alternate possibilities are unnecessary to deem one responsible for their actions.

Derivative Responsibility and Future Actions

Imagine that Alex had an incredibly strong pre-disposition and first-order desire to harm opposing fans when watching sports. Alex knows that if she went to a sports bar, it would be very likely that she would run into a fan of the opposing team. If such events took place, Alex would not be able to control herself from throwing a punch or two and defending her team's reputation. In other words, Alex could not have done otherwise than how she acted. She had no alternative than to punch the opposing team's fan. If PAP is true, then it appears she is not morally responsible for her actions, as she could not have done otherwise. However, if one applies Frankfurtian-compatibilism, Alex would be held morally responsible because a first-order desire to punch an opposing fan is contrary to moral standards. In addition, it is also true that Alex could've prevented the situation by avoiding situations where she would harm others.

8. One can imagine a multitude of examples where persons are being coerced to act against their will. In these circumstances, by applying the L-Reply, one can look to a person's first and second order desires to see if such desires are truly deserving of praise and blame. If a person has a desire to not harm someone, but is being coerced into harming someone, then their act of harming someone can be understood as a higher order desire to avoid harm themselves. In this sense, harming another by possessing a desire to avoid harm yourself is likely not an immoral desire to possess. In this sense, one is looking at the reasons and basis for why one is acting, not simply whether they could have done otherwise.

Now imagine that Alex acknowledges her inclinations and first-order desires and therefore, decides to stay away from sporting events where it is likely she will encounter opposing fans. Instead of attending sports events, she stays at home. In addition, Alex decides to go see a counsellor to eliminate her first-order desire of harming opposing fans. After months of counselling, she comes to the realization that other fans are people just like her and shouldn't be harmed. Alex now has a strong disposition and first-order desire to not harm others. So, she goes to the sports bar and enjoys a game and even a bit of banter with opposing fans. At one point, she felt her past inclination to harm opposing fans, but was reminded by her counselling sessions, and couldn't harm them. In this sense, she *can't* harm them, due to her new insight and respect for others. In these circumstances, PAP would arrive at the result that Alex is not responsible for not harming the other fans, as she couldn't have done otherwise. However, the L-reply coupled with Frankfurtian-compatibilism, can praise Alex, as it acknowledges her reasons for not harming the fans and taking the appropriate steps to mitigate her threat.[9] This thought experiment is intended to target our intuitions and ideas about what it means to be responsible from the point of view of derivative responsibility. Derivative responsibility refers to "cases where the agent is said to be blameworthy for what she did (or failed to do) by virtue of being blameworthy for the causal conditions that led to it" (Mcenna and Widerker, 2018, 14). As we saw in both scenarios featuring Alex, the conditions that led to her harming others and the conditions that led to her attending counselling and not harming others were ostensibly within her control. Therefore, appealing to our intuitions as well as the principles at play in the criminal law, Alex appears to be (derivatively) morally responsible, a feature of responsibility that seems to be worth preserving.

The determinist objection that voluntary actions require the ability to do otherwise remains unconvincing. The morally salient features of decision making can be present even if one cannot act in a way other than how one did. Frankfurtian-compatibilism (as well as the L-reply) looks to the agent's basis and reasons for acting in a certain way by examining one's hierarchy of desires. The next objection to Frankfurtian-compatibilism will challenge whether persons can be the appropriate source of their actions when the assumption of determinism is taken into account.

9. This example is based on the examples presented in Dennett, D. C. (2015). *Elbow Room: The Varieties of Free Will Worth Wanting* (New ed.). MIT Press at page 147.

Manipulation Arguments

Manipulation arguments intend to demonstrate that persons do not possess the appropriate type of control over their actions to be held morally responsible. Put differently, persons cannot be considered the proper *source* of their actions. As Caruso writes, "The basic idea behind this argument is that if an agent…is causally determined to act in a particular way…then the agent is *intuitively* not morally responsible for that action, even if they satisfy all the prominent compatibilist conditions[10] on moral responsibility" (Caruso and Dennett, 2021, 53). For those persuaded by manipulation arguments, the conditions listed above in the criminal justice system (i.e., voluntariness, rationality, autonomy, lack of external forces etc.) are insufficient to demonstrate that any person could properly be considered a candidate deserving of praise or blame.

Manipulation arguments commonly invoke imaginary malevolent agents that attempt to control a subject. After such a scenario is provided, an analogy is made between the malevolent agent and the natural causal processes that are implied by determinism. Manipulation arguments are similar to the thought experiments utilized in the previous section, as they are intended to pump and move our intuition to conclude that persons should not be held morally responsible for their actions. Pereboom[11] offers the following manipulation argument:

> Scenario 1: A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms to his second-order desires. In addition, his

10. These conditions would include that her mental states are causally efficacious, that she approves of her own behaviour, that she is responsible to reasons, and that she satisfies all relevant senses of control over her actions.

11. The original manipulation argument present's four cases. However, for our purposes here, my responses apply equally as well to the two-scenario argument, as to the four-scenario argument.

> process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire – the neuroscientists do not induce a desire of this sort (2014, 79).

As one can see in the first scenario, Plum is manipulated by the neuroscientists, who produced in him mental states that lead to his decision to kill White. In scenario one, the reader is supposed to conclude that Plum is not responsible for his actions, despite Plum's apparent satisfaction of acting on first and second-order desires. In addition, Plum is also reasons-responsive, acting within his own character, under no direct external constraint, and he could've done otherwise (as the desire was not irresistible). Despite these considerations, however, it still seems that Plum is not responsible. If one considers Plum to not be morally responsible in scenario one, then according to Pereboom, Caruso and others, one should also deem Plum to not be morally responsible in scenario two:

> Scenario 2: Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Cases 1). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first and second-order desires. The neural realization of Plum's reasoning process and decision is exactly as it is in Cases 1; he has the general ability to grasp, apply, and regulate his actions by moral reasons, and it is not because of an irresistible desire that he decides to kill (Pereboom, 2014, 78).

The second scenario is almost exactly like scenario one, however, the egotistic desires that Plum possesses are "natural" compared to scenario one, where they were implanted by the neuroscientists. If one applies these two scenarios to Frankfurt's conception of free will, it seems difficult to understand how Plum can be responsible for acting in line with his desires if such desires were not a result of him choosing those desires. According to Frankfurt, freedom consists in being "free to will what [one] wants to will" (1971, 16). However, in scenario one and two, Plum's will had either been manufactured within him or he was born with it. In both cases, it seems counterintuitive to claim that freedom consists of acting in accordance with one's will when we are simply given a will, without much say in the matter. Therefore, if Plum is not responsible in scenario one, Plum also appears to not be responsible in scenario two. By concluding that Plum is not responsible in scenario two, one is supposed to also conclude that persons are never morally responsible, as the conditions in scenario two are the conditions of the actual world we inhabit.

In response to manipulation arguments, there are two general approaches. The first set of approaches are known as hard-line replies. Hard-line replies focus on the first scenario and argue that despite the direct manipulation, Plum would still be morally responsible for his actions. The second set of approaches are known as soft-line replies. Soft-line replies attempt to articulate a principled distinction between the first and second scenario. If a principled distinction can be demonstrated, the move from scenario one to scenario two becomes suspect, thereby, allowing one to retain Plum's responsibility in scenario two. In what follows, I will present two responses, a hard-line reply as well as a soft-line reply.

A Hard-Line Reply

Regarding hard-line responses, Caruso writes, "The first problem with the hard-line approach is that it conflicts too deeply with our intuitions about source hood" (2021, 78). It is hard to dispute Caruso's claim that attributing responsibility to Plum in scenario one does not conflict with our intuitions. Plum is clearly manipulated and is unaware of such manipulation, which appears to negate his responsibility either partially or fully. If such a case were to be presented before a court, Plum would likely be acquitted of the murder. However, it is interesting to consider whether or not Plum would be morally responsible if he *knew* that he had been previously manipulated. Does *knowledge* of the manipulation change Plum's

responsibility? If one knows that they possess a first and second-order desire that may lead to murder, are they not able to have another second-order desire (or even a third-order desire) that would counter-act the harmful desires? It seems reasonable, or at least possible, that in learning that someone has manipulated you to be more egotistical, one can take appropriate steps to either eliminate that desire by acquiring a higher desire or by mitigating the problematic desire.

In a similar vein to the thought experiment where Alex possesses an inclination to hurt opposing fans, when one is aware of a potential loss of control, there is a responsibility to reasonably prevent such a circumstance from occurring. In Frankfurtian terms, if one is aware of a problematic first-order desire, one should take reasonable steps to either eliminate that desire or replace it with a good desire. This sort of responsibility (i.e. taking responsibility to prevent future actions), as mentioned above is called *derivative responsibility.* There are differences between Plum's and Alex's scenarios, however, the general principle holds. That principle being, that persons are responsible for taking reasonable precautions and steps to avoid potential harm to others. Dennett calls this sort of planning, "meta level control thinking" (2015, 73) and Fischer calls it "guidance control" (2012, 178-205). So, if Plum was not directly responsible for killing White because he was manipulated, he may still be derivatively responsible for killing White, as he failed to take the appropriate steps to mitigate the manipulation.

The court recognizes this sort of control and responsibility over our future actions. For example, in contrast to the subjective mens rea (ex. intention), there is also an objective standard of mens rea. When the court applies an objective standard of mens rea, it assesses whether a reasonable person in the scenario that the accused found themselves in, would have acted different? For example, in cases where one causes the death of another while driving, the standard applied is that of a *marked departure.* The court will take into consideration all the surrounding circumstances to determine if the accused acted in a way that constitutes a "marked departure" from what a reasonable person in the same circumstances would have done.[12] Momentary lapses of judgment or accidents do not constitute a marked departure of the reasonable person.[13] For example, an unexplained swerve into traffic or the inability to control a car in snow both constitute momentary lapses of judgment that do not incur criminal responsibility.

12. *R v Hundal,* [1993] 1 SCR 867.

13. *R v Roy,* 2012 SCC 26 and *R v Beatty,* 2008 SCC 5.

In this regard, there are existing legal tools and concepts that can be applied to cases of manipulation which would help the courts determine if a *reasonable* person in Plum's situation would have taken steps to prevent the murder of White. If Plum knew that he was manipulated and had the time to take steps to prevent the murder but didn't, such factors lean towards Plum being held partially responsible. However, if Plum had no idea that such manipulation took place, it seems that he would not be held morally responsible.

In response to Caruso's claim that manipulation arguments demonstrate that the subject is intuitively not responsible, Dennett writes, "I *am* responsible for my *abstention* whether or not the manipulation I have ordered is required. (Maybe you disagree, but I think you have to admit that it is far from obvious in my example that I am *not* responsible because I have been manipulated, which is the "obvious" intuition the other examples are supposed to pump)" (Caruso and Dennett, 2021, 59). In Dennett's mind, once Plum becomes aware of the manipulation, such manipulation ceases to be able to control Plum's ability to resist the egotistic desires, as Plum now can take steps to prevent his killing of White. Therefore, it seems that hard-line replies seriously question the moral intuition that is supposed to arise in us, once we consider the subject possessing knowledge about the manipulation of their own mental states and desires.

A Soft-Line Reply

The soft-line response that will be presented will further dissect the manipulation argument's presuppositions regarding the level of control needed by an agent in order to be held morally responsible. If we grant that Plum is not morally responsible for White's death in scenario one, does that require us to also conclude that Plum is not responsible in scenario two? There seems to be a few reasons to think otherwise.

The principled difference between the two scenarios is that the manipulation in scenario one is performed by conscious agents, compared to agentless physical causes in scenario two. It seems to be far more intuitive to conclude that there is a difference between agent manipulation and agentless manipulation, then to conclude that these two sorts of manipulation are equivalent. One should even consider whether manipulation is the correct term to refer to the natural causal influences that people experience. Caruso writes, "Softline replies are therefore unconvincing because, at best, they can only show that a *particular manipulation*

example has failed to capture all the relevant compatibilist conditions for moral responsibility, not that manipulation arguments fail "tout court" (2021, 84). Stated differently, Caruso appears to claim that one can simply manufacture a new manipulation argument that would demonstrate how one lacks moral responsibility. This might be true, but the burden of proof remains on the one convinced of the manipulation argument to think of such an example. It seems incredibly difficult to bridge the divide between agents being able to control agents, and the physical universe being able to control agents.

For example, as mentioned in the previous section, the court will not hold someone criminally responsible for a "momentary lapse of judgment" when driving their car. Persons can control their car, by speeding up and slowing down as well as a variety of other controls. However, when it is snowy, as was the case in *R v Roy*,[14] our control of our vehicles is drastically reduced because of other causal factors. Other times, people have more control over their vehicles. For example, think of a Nascar driver who can control their vehicle far better than the average person. It seems intuitive that within driving a car, there are degrees of respective freedom that can and often are influenced by causal influences.

In the same way, we as persons are influenced by a variety of causal factors including, environment, parents, genes, predispositions etc. People also face different restrictions on their physical and mental capacities. However, despite all these limitations on our control of ourselves, it does not seem reasonable to conclude that we are not morally responsible for our actions.[15] Given our understanding about control and the variations of control, we can clearly see the difference between an agent manipulating our car (via a remote control) thereby, negating our control, compared to physical processes that simply *limit* our control. Therefore, the analogy between scenario one and scenario two seems to fail due to their being a principled difference between the two scenarios. Namely, the world and the physical causes within it are not an agent which can exert the same sort of conscious and intentional manipulation on persons as a neurosurgeon can. In this sense, persons seem to possess the appropriate amount of control as well as remain an appropriate source, to be considered morally responsible for their actions. The analogy that is supposed to be drawn between scenario one and

14. *R v Roy,* 2012 SCC 26.

15. This portion relies heavily on Dennett's account of control in his *Elbow Room,* chapter 3: Control and Self-Control.

scenario two is weak and thereby, should not lead one to think that persons are incapable of being morally responsible.

CONCLUSION

As mentioned in section one, criminal law presupposes legal personhood, autonomy and rationality. These three elements ground the criminal law's practice of apportioning responsibility to person's accused of committing crimes. However, according to free will skeptics, there are good reasons to deny that persons should be held morally responsible for their actions. In response to such a radical claim, Frankfurtian-compatibilism was presented as an alternative approach intended to demonstrate that the practice of apportioning responsibility is philosophically defensible. Two primary objections (i.e. the principle of alternate possibilities and manipulation arguments) that are commonly raised against compatibilism were shown to be unsuccessful or at the very least, inconclusive. As Morse writes, "Compatibilism is the only metaphysical position that is consistent with both the criminal law's robust conception of responsibility and with the contemporary scientific worldview" (2013, 123). Therefore, it seems to me that the criminal justice system's practice of holding persons responsible for their crimes can be justified on a Frankfurtian-compatibilist basis and is not be threatened by the arguments put forth by free will skeptics.

REFERENCES

Caruso, Gregg, and Owen Flanagan. 2018. *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience.* New York: Oxford University Press.

Caruso, Gregg. 2021. *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice.* New York: Cambridge University Press.

Caruso, Gregg, and Daniel Dennett. 2021. *Just Deserts: Debating Free Will*. New York: Polity.

Dennett, Daniel. 2015. *Elbow Room: The Varieties of Free Will Worth Wanting* (New ed.). Cambridge: MIT Press.

Ferguson, Gerry. 1989. "Critique of Proposals to Reform the Insanity Defence". *Queen's Law Journal* 14 (1): 135-152.

Fischer, John Martin. 2012. *Deep Control: Essays on Free Will and Value*. New York: Oxford University Press.

Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66 (23).

Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1), 5-20.

Haji, I. 2012. "Reason, Responsibility, and Free Will: Reply to My Critics." *The Journal of Ethics* 16 (2): 175-209.

Kane, R. 2005. *The Oxford Handbook of Free Will (*1 ed.). New York: Oxford University Press.

McKenna, Michael and Justin Coates. 2021. "Compatibilism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.

McKenna, Michael, and Derk Pereboom. 2014. *Free Will: A Contemporary Introduction*. New York: Routledge.

McKenna, Michael, and David Widerker. 2018. *Moral responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (First ed.). New York: Taylor and Francis.

Morse, Stephen and Thomas Nadelhoffer. 2013. *The Future of Punishment*. New York: Oxford University Press.

Naffine, Ngaire. 2009. *Law's Meaning of Life: Philosophy, Religion, Darwin, and the Legal Person*. 1st ed. Portland, Or; Oxford: Hart Pub.

Nedelsky, Jennifer. 2011;2012;. *Law's Relations: A Relational Theory of Self, Autonomy, and Law*. Oxford; New York: Oxford University Press.

Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.

compos mentis

The Tragic Bind of the Political: An Analysis of Hannah Arendt and James Baldwin's Disagreement

Nansi Dingle
Earlham College

ABSTRACT
Hannah Arendt and James Baldwin fundamentally, and famously, disagreed on the idea of allowing love to be a guiding principle in the political sphere. Baldwin believed that to "end the racial nightmare, and achieve our country," we must come together politically "like lovers." Contrastingly, Arendt believed that political equality and action are only possible if we set boundaries between the political and social spheres, keeping out the polluting effects of any discriminatory prejudices that govern social relations – and, consequently, keeping out love. This essay will explore the metaphysical infection in Baldwin's notion of love but concentrate on what is of value in his thinking for an examination of Arendt's conception of the political sphere. The discussion of love opens into a response to the wider question of how the political sphere might be constructed with a sufficient amount of stability and yet at the same time remain open to change. For Arendt, the ever-present possibility of change is predicated on the "human condition of natality" which serves as an undercurrent to our potential for political action and freedom. Baldwin, similarly concerned with freedom for political change, uses a Christian notion of salvation to ground his faith that politics might expand to include even the most disenfranchised members of the polis.

KEYWORDS
Arendt, Baldwin, Kant, Love, Politics, Salvation, Metaphysics, Natality

> Anyone who cannot form a community with others, or who does not need to because he is self-sufficient, is no part of a city-state — he is either a beast or a god. Hence, though an impulse towards this sort of community exists by nature in everyone, whoever first establishes one was responsible for the greatest of goods. For as a human being is the best of animals when perfected, so when separated from LAW and JUSTICE he is worst of all. —Aristotle, *Politics I*, Chapter 2.

> One age cannot bind itself, and thus conspire, to place a succeeding one in a condition whereby it would be impossible for the later age to expand its knowledge (particularly where it is so very important), to rid itself of errors, and generally to increase enlightenment. That would be a crime against human nature, whose essential destiny lies precisely in such progress; subsequent generations are thus completely justified in dismissing such agreements as unauthorized and criminal. —Immanuel Kant, "An Answer to the Question 'What Is Enlightenment?'"

Suppose we take Aristotle's assertion that humans are 'the most perfected' when guided by laws and participating in a political system; we must next ask: On what foundations should politics be built? Philosophically, questions about how politics should be conducted are questions about how it is we should live together; on what basis should we build our relationships with those we *have* to interact with because of living in a shared place, shared country? How can we have an organizational structure that allows for collaboration rather than the use of violence, force, or subjugation? Political theorist, Hannah Arendt, constructed a model of politics that we are still contending with intellectually today, despite how unattainable it seems with the modern 'rise of the social sphere' and intrusion of economics into all areas of life. Arendt believed in the separation of the political realm from that of the social and the private in order to preserve the privacy of family relations and individual choices and protect the political realm as a space of speech and action in which hierarchy and force have no dominion. For Arendt, political action is only possible if we set boundaries between the political and social spheres, keeping out the polluting effects of the discriminatory prejudices

that govern social relations. James Baldwin alternatively proposed that we should ground political relations on love. This, as will be shown, is because Baldwin operated with a Christian notion of salvation. In this essay, I will explore how Baldwin's faith in a politics of love raises important questions for Arendt's notion of the political but how, ultimately, love and politics are incompatible due to the necessity of distinction within political relationships. This distinction, according to Arendt, accommodates human plurality and maintains the freedom on which political action is predicated. Baldwin's contribution to Arendt's project has to do with what I see as their shared interest in the status of natality and how the political must accommodate the ever-present possibility of newness. Written into our democratic principles must be an acceptance of progress and thus, tragically, stable political foundations must endure the instability of remaining ever open to change.

James Baldwin and Hannah Arendt's differing opinions on the modus operandi of the political realm were made clear in a letter exchange in 1962. Arendt wrote an open response to Baldwin's *Letter from a Region of My Mind*, published earlier that year, stating: "What frightened me in your essay was the gospel of love which you begin to preach at the end. In politics, love is a stranger, and when it intrudes upon it nothing is being achieved except hypocrisy" (Arendt 2006). Arendt is concerned by the intrusion of love in politics because her political theory seeks to establish the most stable foundations for the political and she saw the "passion" of love to be antithetical to this project (Arendt 2018, 242). Love, as a force, brings us closer by closing the gap between people. In Arendt's words, love "destroys the in-between which relates us to and separates us from others" (Arendt 2018, 242). This "in-between" or "worldliness" is essential for Arendt's conception of the political because the maintenance of the in-between is what preserves the human condition of plurality and allows for relationships of equality. Scholar Sean Butorac wrote of Arendt's theories, "The trouble with neighbourly love is that it absolves humans of the characteristics that make distinction possible, while the intrusion of intimate love into the public realm prevents us from maintaining equality" (Butorac 2018, 711). Politics should be defined by the separation and collectivization of equal individuals and so Arendt sought a political realm based on equal exchange and an economy of transaction. She would, therefore, keep love out because equality relies on plurality as opposed to the inevitable union or oneness created by love. In *The Human Condition,* Arendt wrote, "Love, by

its very nature, is unworldly, and it is for this reason rather than its rarity that it is not only apolitical but antipolitical, perhaps the most powerful of all antipolitical forces" (Arendt 2018, 242). Love is antipolitical because it does away with difference; it brings us together too closely for freedom. "Hatred and love belong together, and they are both destructive; you can afford them only in the private and, as a people, only so long as you are not free" (Arendt 2006). Love annuls the possibility of equal exchange and, to use Butorac's words, "eviscerates the very possibility of politics" (Butorac 2018, 711).

Contrastingly, James Baldwin believed in the utter conflation of love and political relationships. In his biography of Baldwin, James Campbell wrote, "Many writers have defended the notion that love can be brought to bear upon the solution of political problems, but no one believed it more passionately than James Baldwin" (Campbell, 2021). Baldwin proposed that we conduct our political relationships through the prism of love and that we behave politically "like lovers". His concluding remarks in *Letter from a Region of My Mind*, which Arendt so disagreed with, included the following proclamation: "If we—and now I mean the relatively conscious whites and the relatively conscious blacks, who must, like lovers, insist on, or create, the consciousness of the others—do not falter in our duty now, we may be able, handful that we are, to end the racial nightmare, and achieve our country, and change the history of the world" (Baldwin 1962, 30). The strength of Baldwin's commitment to the idea of love in politics is so great that he believed it to be the only thing that could deliver America from the "racial nightmare" that it was and is living through. Baldwin's notion of love is undoubtedly deeply influenced by his exposure to the Christian gospel of love and yet he developed faith in love's power, ironically, by first distancing himself from the Church. *Letter from a Region of My Mind,* begins with a story of Baldwin running from adolescent fears of crime and destitution and into the perceived salvation of the Church. At a relatively young age he became a minister like his father and dedicated himself to writing sermons and preaching. However, he grew sick of the hypocrisy of Church leaders who lined their pockets with funds taken from parishioners on the brink of destitution. The young Baldwin found the purported Christian love to be corrupted by greed and cruelty and reduced to a mere manipulation technique. He consequently declared that anyone wishing to be "a truly moral human being" should "first divorce himself from all the prohibitions, crimes, and hypocrisies of the Christian church" (Baldwin 1962, 11).

The "love" Baldwin found in the Church was insincere however Baldwin seems only to have become more fervent about the power of true, sincere love. He believed love to be our saving grace and that if we wish to "change the history of the world" then people must approach each other politically with love.

It seems that Arendt and Baldwin agree that love has an inherently boundary-destroying nature, but they disagree as to the political merit of this quality. For Arendt, the closeness created by love threatens plurality which she saw as the condition that "both relates and distinguishes us" (Butorac 2018, 711) in a way necessary for political action. Consequently, "For Arendt, who regards plurality as the condition of action, and therefore the precondition of freedom" (Butorac 2018, 711), love cannot be given power in any political sphere seeking to preserve freedom. Whereas, for Baldwin, the closeness created by love, makes space for a radical inclusivity that assuages his fear of the potentially violent effects of any exclusionary boundaries, even those that are there to preserve equality within the political realm. Baldwin's commitment to the idea of the necessity of love in politics seems to be because he was operating under a Christian notion of salvation. Later, this essay will further examine the relationships of Baldwin's elusive "love" to the Christian metaphysic but for now - some background. A belief in the power of salvation grounded Baldwin's interest in how love could (re) integrate those who are outcasts of the political sphere; those who might appear to be shut out by Arendt's political boundaries. In an interview given in 1973 for The Black Scholar, in answer to a question about what message he would like to leave for those "engaged in active struggle", Baldwin said the following: "I think the revolution begins first of all in the most private chamber of somebody's heart, in your consciousness" (Baldwin 1973, 42). He then goes on to explain how inclusion in politics should, through the medium of love, be made open to all. He wrote, "[The vanguard of a revolution which is now global begins with the black family.] My brother in jail, my sister on the street and my uncle the junkie, but it's my brother and my sister and my uncle. So it's not a question of denying them, it's a question of saving them" (Baldwin 1973, 42). We see from this quote that to love is to make familial and that the closeness formed is what saves people and brings them into politics, i.e., makes them visible in the political sphere and so worthy of political consideration. The "vanguard of a revolution" (the leaders of his political vision) is "my brother and my sister and my uncle", in a literal and figurative sense, regardless of their circumstances. For Baldwin, unlike Arendt, the

political must transcend transactional relationships and his way of accomplishing this is by making love, as something that is a gift with no need for reciprocity, the governing principle. Baldwin's belief in 'saving people' by loving them without expecting anything in return, transcends transactional relationships.

As we have seen, Arendt's concern that politics operate within an economy of exchange and so be bounded by transactional, as opposed to loving, relationships is contrary to Baldwin's desire to both transcend transactional relationships and overcome any boundaries that lead to exclusion. Baldwin's belief in salvation and the indiscriminate nature of his (Christian) love suggests that he would take issue with Arendt's argument that to come together as political equals, the political realm must be preserved by the use of delimiting boundaries. For Arendt, preserving these boundaries is what allows for freedom; however, Baldwin's beliefs compel us to ask, what of those kept out by such boundaries, what of the outcasts? Does Arendt's political sphere exclude the prisoner and the person suffering from addiction because they do not have a chance to enter the realm of speech and action and put their rationality to work? Under Baldwin's conception of the political, the prisoner and the person suffering from addiction can be a part of the political because they can be loved. But is this conception of the political merely an eroded and weakened version of Arendt's aspirations? Doesn't a lack of limits degrade freedom and, as Arendt argued happens when love is involved, make equal relationships, and thus true collaboration and political action, an impossibility? To respond to these questions and assess the viability of political love as a means to transcend the transactional and establish a radically inclusive political economy, we must first scrutinize the philosophical ground of Baldwin's faith in love because, although Baldwin distanced himself from the Church, his notion of salvation nevertheless seems to be predicated on metaphysical grounds. Baldwin's idea of the saving gift of love seems to be that of the unconditional love of God explicated in John 3:16 in the Bible. This love is wholly outside of an economy of exchange: "For God so loved the world that he gave his one and only Son" - the *gift* of His son, of course, has no need of, and in fact does not allow for reciprocity. Rather, this biblical love is for the sake of salvation: "For God did not send his Son into the world to condemn the world, but to save the world through him" (John 3:16). For Arendt, what renders one a political entity is their participation in an economy of equal exchange. However, the love of God that so inspired Baldwin is one in which God's unconditional giving annuls

the possibility of reciprocity and equality; what equality could there possibly be between an omnipotent, omnipresent, and omniscient God and His perpetually imperfect creations? Therefore, to seriously entertain Baldwin's integration of the gift-giving economy of love into the political, we must forgo equality - something most would find utterly unacceptable and that would undermine Arendt's entire project (hence her alarm at the suggestion).

It may be that Baldwin's language of love is too compromised by metaphysical Christian ideals to be seriously considered as a foundation for a political organizational structure that must withstand affronts without deference to metaphysics. However, the intention behind his appeal to love, his desire to overcome the transactional and erode the delimiting boundaries of politics so that everyone might be included, is where I suggest the value of his disagreement with Arendt can be found. Baldwin's interest in transcending the transactional and allowing for spontaneous gifts (albeit the gift of love) reveals an acknowledgment of the necessity of newness in politics, something Arendt also valued because of her belief that the political realm makes space for free action and, consequently, for change. Arendt described action as "the only activity that goes on directly between men without the intermediary of thing or matter" (Arendt 2018, 7). It was the closeness implied by this direct relation of 'man to man' without the need for an exchange of items that I think Baldwin was interested in with his valorization of love as a political medium that escapes transaction. However, for Arendt, action requires separation because it "corresponds to the human condition of plurality, to the fact that men, not Man, live on the earth and inhabit the world" (Arendt 2018, 7). In Baldwin's desire to do away with distinction he reveals himself to be concerned with Man, not men. For the sake of freedom, Arendt could never allow such an ideal of oneness into her idea of politics because she believed maintaining the human condition of plurality (paying attention to relations of men, not Man) to be essential for maintaining free political discourse. Despite these differences, Baldwin seems to have grasped the importance of open possibility, the possibility of action and newness, that is so central to Arendt's philosophy. Arendt believed that there is always the potential for newness and change because of the "human condition of natality" (Arendt 2018, 9). Linking natality and political action, Arendt wrote, "the new beginning inherent in birth can make itself felt in the world only because the newcomer possesses the capacity of beginning something anew, that is, of acting" (Arendt 2018, 9). The democratic process must be a space

of possibility in which political action is made possible by an awareness and accommodation of the fact of natality. In the potential for novelty encapsulated by every political act, the political sphere affirms itself as open-ended. The Arendtian political sphere must foster action and "since action is the political activity par excellence, natality, and not morality, may be the central category of political, as distinguished from metaphysical, thought" (Arendt 2018, 9). Baldwin's concern with morality implicates him in metaphysical thought and would exclude his ideas from Arendt's conception of the political. However, his emphasis on love as that which allows us to welcome *everyone* into the political, i.e., to act based on *open* possibility, reveals a tension in Arendt's thinking and perhaps in the very foundational principles of democracy.

Even Kant, heralded as the great categorical thinker, emphasized the importance of an openness to change in the political, necessitated by the inherent impossibility of static conditions. As we have explored, it seems to be that Baldwin's appeal to metaphysical foundations and an unchanging love is hankering after this impossible stasis. Nonetheless, Kant posited that any social agreement which seeks complete stability is in fact "unauthorized and criminal" and, therefore, "subsequent generations are thus completely justified in dismissing such agreements" (Kant 1784). He writes, "Such a contract, whose intention is to preclude forever all further enlightenment of the human race, is absolutely null and void" (Kant 1784). The potential precarity of Arendt's categorical thinking (and only categories founded outside of metaphysics allow for such an admission of precarity) is what accounts for the human condition of natality and thus makes space for freedom and consequently for action. So, there is a double bind in which the separation of Arendt's categories is necessary for preserving the distinctions of plurality within the political realm which is in turn necessary for establishing the freedom to act politically, and yet the very possibility of action and freedom necessitates that there must be an openness and an admission that any (democratic) system must include the potential to be overridden. As Kant highlighted: no society can immutably *bind* itself to any form of "guardianship over each of its members" (Kant 1784). Political action, only made possible by the democratic idea of equality and freedom, is of a tragic nature. Kant shows us that the members of a polis can impose no principle of governance on themselves save a "provisional order for a specific, short time" (Kant 1784). "The criterion of everything that can be agreed upon as law by a people" is that there must be the

possibility of change. To allow for action, law (and so politics) must account for the human condition of natality and, consequently, categories must be in some way open – an apparent paradox.

The tragedy is that we must make promises in order to "form a community with others" (Aristotle 2013), and yet, as Kant demonstrated, no promise can be absolutely binding. Aristotle claimed that "a human being is the best of animals when perfected" (Aristotle 2013), that is to say when a part of the polis and participating in law and justice. Nevertheless, the possibility of action, of change, necessitated by living in the temporal world without metaphysical stability, undermines the supposed certainty of any promise that attempts to predict and stabilize the future. No contract can be immutable because the temporal world is in a constant state of flux. Similarly to Kant, Arendt was concerned with the problem of promise-making in her political philosophy. For Arendt, "the price we pay for plurality" is "the impossibility of remaining unique masters of what we do" (Arendt 2018, 244). In other words, we cannot project mastery into an inherently uncertain future; we cannot guarantee that anything remains and continues to remain. Promises made in the political sphere are intended to "at least partially" expel the unpredictability of the "basic unreliability of men who never can guarantee today who they will be tomorrow, and out of the impossibility of foretelling the consequences of an act within a community of equals where everybody has the same capacity to act" (Arendt 2018, 244). The necessity of a freedom that allows each of us to act politically (referring here to the Arendtian notion of action) cannot be coupled with any indubitable promise regarding the future, even, and this is perhaps tragic, the promise that this state of freedom continues to be preserved. The tragic bind of the political is something that democratic theorists must contend with as they question the legitimacy of a democratic public's right to decide to consensually end democracy. Arendt wrote, "Man's inability to rely upon himself, or have complete faith in himself (which is the same thing) is the price human beings pay for freedom" (Arendt 2018, 244). This inability to rely on himself and consequent inability to make any binding promise is a result of the same freedom that the (Arendtian) political sphere is designed to preserve. Despite the inability to do so irrevocably, political organization is established as a means of, in some way, (attempting) to bind men to law and justice because to accomplish anything political, we need some semblance of stability despite the impossibility of attaining it absolutely. Our need to work for

this impossible stability is why Aristotle claimed that "whoever first establishes [the city-state] was responsible for the greatest of goods" (Aristotle, 2013).

We began by exploring Arendt and Baldwin's disagreement over the merit or necessity of allowing love to govern political relationships. Arendt believed the inclusion of love to be unviable because love facilitates a closeness between people that eliminates the very possibility of politics because, for Arendt, distinction and separation are needed in the political sphere in order for us to establish equality and freedom. In opposition to Arendt, Baldwin seems to believe we cannot operate under the Arendtian separation of political and private spheres because political action begins in the most intimate and private of places, in the heart and in the family. However, his argument for love as an instrument of politics is weakened by its reliance on a metaphysical foundation. Despite the untenability of a metaphysically grounded political theory, Baldwin's philosophy of love raises important questions for those of us unpacking Arendt's thinking, specifically in relation to questions of inclusivity. Moreover, Baldwin's desire to establish love as a gift that should be given to all seems to reveal an interest in the need for openness in the political sphere, an interest Arendt explored through the lens of the human condition of natality and the consequential, ever-present, possibility of newness. Kant, in his essay "An Answer to the Question 'What Is Enlightenment?'", was similarly concerned with the need to be open to change, and, despite his renowned categorization way of thinking, he declared that it would be a crime to seek to establish a principle that might remain unchangeable for future generations. No principle can be wholly stable and yet, responding to the potentially tragic nature of democracy, Arendt lays out principles of political governance that allow for the freedom that undermines absolute stability but is necessary for democratic politics to account for natality and be open to progress.

REFERENCES

Aristotle and Carnes Lord. 2013. *Politics I*, "Chapter 2." Chicago: The University of Chicago Press.

Arendt, Hannah, Margaret Canovan, and Danielle Allen. 2018. *The Human Condition*. Chicago: The University of Chicago Press.

Arendt, Hannah. 2006. "The Meaning of Love in Politics. A Letter by Hannah Arendt to James Baldwin." HannahArendt.net. *Journal for Political Thinking*. https://www.hannaharendt.net/index.php/han/article/view/95/156.

Baldwin, James. 1962. "Letter from a Region of My Mind." *The New Yorker*.

Baldwin, James. 1973. "The Black Scholar Interviews: James Baldwin." *The Black Scholar* 5 (4): 33-42

Butorac, Sean Kim. 2018. "Hannah Arendt, James Baldwin, and the Politics of Love." *Political Research Quarterly* 71 (3): 710–21. https://doi.org/10.1177/1065912918760730.

Campbell, James. 2021. *Talking at the Gates: A Life of James Baldwin*. Berkeley: University of California Press. https://doi.org/10.1525/9780520381698

Kant, Immanuel. 1784. "An Answer to the Question 'What Is Enlightenment?'"

compos mentis

Origins and Manifestations of Death Anxiety

Amanda Galloway
Indiana University-Purdue University

ABSTRACT
In this paper I will be making the claim that death anxiety is a feature of human nature. After making a case for why it can be defined this way, I will present evidence from an evolutionary perspective as well as behavioral manifestations that can be explained by this anxiety. Such manifestations can be described as distraction, denial, and reassurance seeking. Acceptance will then be given due diligence as a special consideration. These ideas are compatible with Terror Management Theory, which has spawned a significant amount of related research in the fields of sociology and psychology. Although a relatively new theory, its related ideas have been discussed to some extent by ancient philosophers onward, and continue to be synthesized into a more cohesive understanding that can be applied to benefit the human condition on both an individual and societal level.

KEYWORDS
Death Anxiety, Human Nature, Memento Mori, Terror Management Theory, Thanatophobia, Coping Mechanisms

PART I: ORIGINS

As far back as we can study, themes of immortality have permeated ancient culture in an effort to manage the despair that accompanies death. Even the earliest piece of literature known to man, the Epic of Gilgamesh, reveals the ominous manner in which death was viewed by early civilization. Gilgamesh's myth is believed to have later given rise to the creation story found in the biblical account of Genesis, which also centers around immortality. This religious narrative portrays a man and woman living in a state of innocence, forbidden to eat fruit from the Tree of Knowledge. They eat from it anyhow and, interestingly, it is the advent of their newfound knowledge that brings about death and its corresponding anxiety. I like to think of these characters as a sort of prototype of the human species akin to our hominid ancestors. Their innocence can be equated with ignorance and underdeveloped capability for rationality. As greater intelligence awoke within them (as did our predecessors), so too did an accompanying despair regarding the finiteness of life.

In this paper, I will be explaining how death anxiety is a characterization of the human condition because of what happens when abstract thought and emotional complexity is coupled with the survival instinct. Notice how Arthur Schopenhauer attributes our fear of death to the nature of our consciousness, "Here there primarily lies before us the undeniable fact that, *according to the natural consciousness* [emphasis added], man not only fears death for his own person more than anything else, but also weeps violently over the death of those that belong to him…" (2016, 648-649). We can look at the aforementioned claim in terms of the following structure: Humans dread things they don't want to have happen. Humans don't want death to happen, therefore, humans dread death. These above premises are supported in the following paragraphs by making connections between semantic processes (which enable abstract thought and complex emotions), and the anticipatory nature of dread that ensues, with the survival instinct being a root drive. This dread is an experience that can be referred to as death anxiety, which terms we will use interchangeably.

Dread signifies that which we know will happen in the future, but that we do not wish to occur. Taking out the emotional component (i.e. the feeling of aversion), we are left with anticipation. Anticipation is a necessary (but not sufficient) condition for dread. Death anxiety cannot be experienced unless one

can anticipate the future, which requires the ability to think in the abstract. Our ancient ancestors became equipped with this abstract understanding through semantic processing. They could now reflect on the past and comprehend the future. Semantic processing is the function that assigns meaning to data, both in the present as well as when accessing our memory bank (Binder and Desai 2011), without which, it is just information with no interpretive value. Words for example, would just be arbitrary grunts and babblings. Through semantic processing, assigned sounds become a *symbol* for the object in question. Words are not tangible objects; they are *representations* of things. Furthermore, words are not the only ways in which we find representation or interpretive value. Through abstract thinking processes, our species began to perceive other creatures with whom we identify as symbolizations of our own demise.[1] By witnessing others like us die (whether through tragedy or old age), semantic memory functions allow humans to connect the dots of past occurrences and assign that same inevitable event to ourselves and loved ones in the future to come. And while other animals are equipped with "fight or flight"[2] responses to evade death in the moment, they generally don't think about it otherwise. Humans, on the other hand, do think about death otherwise because they are able to ponder the past, present, and future thanks to their more complex, abstract thought. In short, unlike any other animal we observe, human faculties have developed in such a way that we are aware of our own mortality. As Homo Sapiens came to understand the concept of future occurrences as well as cause and effect, they became more aware of mortality, and with it, existential anticipation.

You're probably wondering *why* being aware of mortality necessitates a feeling of dread just because we can anticipate. We have emotions to thank in conjunction with abstract thought. We'll get to that component after I address a potential objection. One might point out that all animals avoid death, and *they* aren't filled with existential dread. It's true that all living creatures attempt to avoid death, but because other animals lack the ability to understand abstract thought to the degree we do, they can't comprehend future identity (Duval et al. 2012) or their orientation in a future world, which means they only seem to worry about

1. See Bassok (1997) to learn more about how semantic processing plays a role in assigning meaning to objects in addition to its role in language.

2. See Understanding the Stress Response (2011) for more information on the "Fight or Flight" Response.

death in the near moment of danger. Even then, death avoidance is more of an act of instinct, rather than a true understanding of what it means to die. This instinct hasn't been consciously manifest to the degree that it has in humans. *Our* attempts and hopes to avoid death go far beyond instinct. With the ability to understand abstract concepts, hominids started becoming aware of our avoidance of danger and death even when it wasn't directly imminent.

Let's now add the feeling of aversion back in with the intellectual component of anticipation to understand how emotion ties in with dread. The feeling of aversion is the other necessary (but not sufficient) condition that defines the experience. Greater emotional complexity was an added result of Semantic Processing. As covered already, Semantic Processing's conduction of abstract thought makes possible our ability to use language. There is mounting evidence that language plays a key role in the development of human emotion (Lindquist et al. 2015). The more words that become available to us, the more complex are the ideas that can be explored, expressed, and understood. This in turn leads us to create more words as more concepts arise, with an even richer construction of our experience and even our emotions. With the advent of language through abstract thought, an entire world opened up to humans that was previously unavailable. A world of morality, a world of identity. A more intricate sense of love and loss. And with loss, a new perception of death.

Our human ability to understand our mortality as a result of reason, coupled with an increased depth of emotion as semantic processing emerged, naturally produces dismay at the realization we will die. Emotions coexist with our thoughts. They are either pleasant or unpleasant reactions to thoughts depending on what those thoughts represent. When a thought represents a condition that is perceived as beneficial to us, a subjectively pleasant emotion is the resulting effect. On the contrary, a thought representing a state we view as harmful to us results in an unpleasant emotional experience. Of course, this is a simplistic cause and effect account of emotions as thought reactions, but it is important to recognize how emotions can be influenced by our perceptions. Humans have an instinctual desire to live, an added intellectual understanding of life and death, and a natural emotional response to the fact this desire will be thwarted. The conditions of anticipation and the feeling of aversion in regard to death are both in place, inevitably resulting in dread by our definition. So long as it remains true that science has not advanced a way for our psyche to become indestructible, and

that we are cognizant of reality, we will continue to feel the same dreadful anxiety about death permeating our existence.

Being relegated to a life of despair needn't be a completely pessimistic realization though. Awareness of mortality seems to enhance the survival instinct by allowing it to manifest on a conscious level. Being upset about our demise rather than welcoming it with gladness can be useful. Imagine how we might respond to spotting a hungry lion or being held at gunpoint if we weren't at all concerned about our continued existence. To facilitate the survival instinct, it helps to be averse to death. In other words, in order to live, by default we have to try not to die. While increased brain function has made us more aware of death, the feeling of dread may be what motivates us further to do something about it. Is death anxiety itself a continued adaptation to further ensure that we will do all we can to survive? Think of dread as the ultimate adaptation, giving us even more advantage to be proactive in survival. On the other hand (and more likely), our fear of annihilation may just be a spandrel of sorts, a sort of side effect that arose naturally out of our increasing intellect. Still, scientists point out that spandrels which persist in the gene pool are rarely without some useful function in their own right (Gould 1997). Aversion to death seems to facilitate the preservation of our species and DNA.

Just as not having enough fear is maladaptive, so too is excess anxiety. It could be that moderate levels (akin to Aristotle's golden mean[3]) may actually be beneficial to our survival so long as it doesn't become too extreme. We can think of it in terms of diminishing returns which eventually lead to negative returns. Suppose there is a function of benefit commensurate with expenditure of anxiety. This anxiety may cause us to go above and beyond instinct to protect our lives so that we as a species (and consequently, our DNA[4]) survive. This is undeniably beneficial. At a certain point, there can conceivably be a peak where just the right amount of anxiety yields the most benefit for our survival. If the above scenario isn't plausible to you, at the very least we can agree anxiety is able to be "tolerated" by evolutionary processes so long as it doesn't interfere with the natural selection process.

3. See Kraut (2018) to learn more about Aristotelian Ethics

4. Dawkins (2016) popularized the Gene-centered view of evolution with his 1976 book that revolutionized the study of evolution. It asserts that genes are the basic unit which drive the survival instinct in an organism.

We must contend with the controversy about what human nature consists of. There are those who deny that what I have described above is representative of the human condition on the grounds that it is not universally observed. I propose that if something (be it trait, feature, behavior, etc.) is a tendency within a species, that something can be acknowledged as a part of its nature. The lack of apparent anxiety in every single individual shouldn't be a problem. Philosophers have long pointed out that man's rationality is what sets us apart from other species, and that reason is a distinct component of human nature. Not every human is rational or adequately uses reason though, so by that logic we would have to reject that man's rationality is our nature. Suffice it to say that what we mean is that the human species in general has the *capability*[5] of rational thought. Not every human is rational, but we as a species tend to have the ability for rational thought. Furthermore, not every trait is expressed within a species until environmental factors come into play. Science has empirically supported that behavioral expressions derive from a combination of nature and nurture. It is therefore plausible that all humans are innately predisposed to having death anxiety, which may predictably only show expression in specific environments. This may explain why some people claim that they are not as bothered by their mortality.

A noteworthy mention of environmental factors would include those that lead to feelings of significance (sometimes referred to as self-esteem, self-worth, and even a sense of fulfillment). Abeyta (2014) shows a direct correlation between feelings of significance and death anxiety. The greater the feelings of *in*significance, the greater the feelings of trepidation about our extinction. Those who have less death anxiety are observed as having higher feelings of significance. Remarkably, significance only affects the intensity of death anxiety and not the other way around.[6] The important takeaway is that we are defining human nature as a tendency within our species under the condition of being mortal, and that other environmental conditions may also affect the way in which this tendency is expressed.

5. K. Rand (personal communication, November 6, 2020), researcher and psychology professor, confirmed my examples of ways in which man has tendencies that are not found in every human by adding that man is only capable of reason, which supports my claim that tendencies like rationality and certain fears are sufficient to define human nature.

6. See Solomon et al (2015) for a worthwhile illumination on how feelings of significance affect death anxiety.

There are also two final elephants in the room that will be briefly addressed and those are the issues of suicide and murder. Does the fact that so many people voluntarily take their own life or that of another debunk death anxiety as an underlying malady? No. It is a mistake to automatically assume that one who kills is fearless of death. It is plausible that a person can be afraid to die while also finding the circumstances of life to be unbearable, and ultimately commit suicide. Many examples can be given in which people do things in spite of fear. Skydiving, ripping off a band aid, public speaking, you get the idea. Therefore, killing oneself is not a justifiable reason to dismiss death anxiety. It is also plausible one may kill another while desiring to preserve their own life. This often occurs in relatively large quantities during war, for example.

After considering the above explanations and responses to possible counter arguments, it would not be unreasonable to conclude that humans developed death anxiety as a result of evolved brain function. Regardless of whether or not death anxiety itself is an adaptation that was selected for over time or merely an uncomfortable spandrel, our species had to find a way to deal with this uneasiness lest we be weighed with too much despair, which becomes detrimental to survival outcomes. As it may be, when anxiety levels rise past the peak productivity level, benefits begin to taper off until finally we yield negative returns, rendering excess anxiety a hindrance to survival success. In order for anxiety to not reach beyond the bounds of what can be tolerated, humans have managed to develop coping mechanisms[7] in order to keep our death anxiety in check. We distract ourselves through entertainment. We engage in self-deception and denial. We seek reassurance through promises of eternal life beyond the grave as presented by religion and even science. In other words, we have developed ways to balance our terror in order to maintain equilibrium. This is why we don't observe a constant, salient state of death anxiety in individuals, especially when assessing on a superficial level. Deeper inspection (and introspection) would reveal dread running in the background if it wasn't already in the foreground.

We each have something in common with Adam and Eve. As children dwelling in a similar state of ignorance, there comes a time in human development when the knowledge of death likewise dawns upon us. In our own lives, we feel the same

7. In *The Worm at the Core*, Solomon et al (2015) describe many ways in which humans unconsciously behave in order to keep fears of death at bay. These authors are known for having a large body of studies that provisionally support their theory known as Terror Management Theory.

sorrow as that of our legendary biblical heroes when enlightened on the matter. It would be highly abnormal for this news to be met with gladness. Such a mental state could even be considered pathological. Even neutrality would be unusual and is likely simply due to the aforementioned coping mechanisms dulling reality. The very essence of our thoughts, emotions and instincts necessarily dictates that when presented with the particulars of our forthcoming annihilation, we react with dismay and dread once fully understood. It would be appropriate to once again quote Schopenhauer with the following, " If now the will, by means of knowledge, beholds death as the end of the phenomenon with which it has identified itself, and to which, therefore, it sees itself limited, its whole nature struggles against it with all its might," (2016, 651) which is why he also wrote, "the greatest fear is the fear of death," (649). Any who claim that they have never encountered such anxiety after having gained a genuine comprehension of what their annihilation means, deserves congratulations for figuring out how to defy all nature and perfectly conquer the will to live.

PART II: MANIFESTATIONS

Memento Mori. It means "remember death" in Latin and is a phrase not meant to be morbid, but to inspire. It is a precept intended to motivate one to cherish every drop of existence that remains available to us in creating a purposeful life. Marcus Aurelius, who would rather have been a philosopher, was a Roman emperor who adopted the Stoic lifestyle that embraces the above ideology. Aurelius wasn't immune to suffering his fair share of tragedy at death's hands. Later reputed to have been of a more sensitive nature and having lost nine of his own children, Aurelius had to have struggled to come to terms with life's transience. In his *Meditations*, he quotes Epictetus as saying, "When a man kisses his child, he should whisper to himself, 'Tomorrow perchance you will die'" (1945, 124). It is clear from Epictetus' direct writings that he meant this as an admonition to not become too attached to that which won't last, perhaps as an attempt to minimize future grief, which might have appealed to Aurelius as he sought ways to cope with death as anyone would. Today, many would interpret Epictetus' advice as a reminder to not take life for granted. Both interpretations provide a suggestion for handling the anxiety regarding our finiteness. The renowned philosopher Seneca (whom Aurelius would have studied) wrote in a letter to his

correspondent, Lucilius, "You are afraid to die. But come now: is this life of yours anything but death?" (Seneca and Barker 1932). It isn't new now and it wasn't new then that fear of death was an obstacle to be reckoned with for every being able to contemplate it.

We've established that the fear of death is a natural tendency for humans. If you are one of those lucky few who denies you experience any degree of death anxiety, you have simply developed a coping mechanism that allows you to manage life without constant worry. These coping mechanisms don't refute the theory; their presence actually further substantiates it. While the following tactics may be performed on a conscious level, many of us implement them for reasons outside of our conscious awareness, at least one of these reasons being death anxiety. These physical manifestations of our attempts to cope with mortality come in three main categories: Distraction, Denial, and Reassurance.

The world as a whole never sleeps. All around us can be found a myriad of Earth's inhabitants engaged in seemingly frivolous activity. We keep ourselves distracted to block out the unpleasant thoughts that plague us when our minds are forced to face our thoughts. It was Blaise Pascal who wrote, "Men, unable to remedy death, sorrow, and ignorance, determine, in order to make themselves happy, not to think on these things. Notwithstanding these miseries, man wishes to be happy….For this he must needs make himself immortal; but unable to effect this, he sets himself to avoid the thought of death" (1901, 38). If left to our own musings long enough, our inner narrative would eventually circle around to haunt us with regrets and secret yearnings that are only frustrated by our finitude. We couldn't escape the salient truth that our very death is what makes our longings so urgently loud in our head. Escape we do in the form of social media, alcohol, TV shows, shopping, YouTube, anything that keeps us from having to face our own thoughts alone. What better way to ignore our despair than to entertain ourselves with lighthearted fluff that makes us laugh and forget?

I know what you're going to say next. You're going to point out that not all entertainment is rainbows and butterflies; horror movies and violence also plaster our screens. Does this not show a large sample of people who defy this theory by their apparent delight in laughing death in the face? On the contrary. Perhaps author Stephen King explains it best when he wrote in his 1981 non-fiction book Danse Macabre, "We take refuge in make-believe terrors so the real ones don't overwhelm us, freezing us in place and making it impossible for us to function in

our day-to-day lives. We go into the darkness of a movie theater *hoping* to dream badly, because the world of our normal lives looks ever so much better when the bad dream ends" (2010). King admits to these death anxieties in his book and affirms the macabre is intended to help us escape from reality by displacing our own fears onto artificial characters who live in a fantasy world we've convinced ourselves is unlike our own.

Rather than sit with the greater discomfort of facing our own horrible ending, we easily distract ourselves whether it be watching horror films, reality TV, or Rom Coms. Unfortunately, distraction only lasts as long as the two-minute blurb or two-hour flick can run. We need more than a momentary series of distractions in order for the pain numbing effects to be sustainable. When distraction fails, we have denial.

You may wonder how denying our mortality has any credibility when we are intellectually aware that we will die. People even willingly engage in viewing violence for entertainment purposes, so how can anyone be in denial? What we fail to understand is that even when witnessing death around us, many people tend to have the idea that such an event will never happen to *them*. You would think it would be unmistakably obvious to each of us, but Irvin D. Yalom quotes one of his adult patients who reported her surprise at this revelation, "I suppose the strongest feelings came from realizing it would be ME who will die, not some other entity like Old-Lady-Me or Terminally-Ill-and-Ready-to-Die-Me" (2008, 13). In Leo Tolstoy's short story, *The Death of Ivan Ilych,* Ivan almost seems to think he is too special to die:

> The example of the syllogism that he had learned in Kiseveter's logic-Caius is a man, men are mortal, therefore Caius is mortal-had seemed to him all his life correct only as regards Caius, but not at all as regards himself. In that case it was a question of Caius, a man, an abstract man, and it was perfectly true, but he was not Caius, and he was not an abstract man; he had always been a creature quite, quite different from all others; he had been little Vanya with a mamma and papa, and Mitya and Volodya, with playthings and a coachman and a nurse...Had Caius been in love like that? Could Caius preside over the sittings of the court?

> And Caius certainly was mortal, and it was right for him to die; but for me, little Vanya, Ivan Ilych, with all my feelings and ideas- for me it's a different matter. And it cannot be that I ought to die. (2004, 121-122).

Death seems to befall other people, but not ourselves. We are too unique and exceptional of persons in comparison with others. How can this be?

It doesn't seem possible to fool ourselves into believing something in direct contradiction with what we know. Enter the world of self-deception. Setting aside for a moment the *motive* for doing this, let's define what it entails. Denial means we have refused to accept information or conclusions presented to us. Take into account the effects of confirmation bias; only accepting evidence that supports our existing or desired narrative, and discounting (rejecting) evidence that is contrary to our belief system. This seems to be a fairly common occurrence but can only be sustained to a certain extent. When there is overwhelming information that renders us unable to easily deny such facts, we have to resort to self-deception in order for denial to maintain its effectiveness. Self-deception concerns lying to oneself, and in order to lie, one has to knowingly misrepresent true information. In order to successfully deceive another, that person has to actually believe the liar. But how do we successfully deceive ourselves when the liar and the believer are of the same brain? How can our mind both know the truth and not know the truth simultaneously?

George Orwell brilliantly demonstrates this in *1984*, a novel about a dystopian society whose party member citizens are forced to believe everything that their corrupt Government feeds them, even when the beliefs are in direct contradiction to what they were forced to believe prior to. Our hero Winston Smith's job is to assist the government with literally rewriting history so there is no evidence of what was purported beforehand. For years you may be at war with Eurasia, only to experience yet again the gaslighting disturbance of finding out from the same authorities that we've never been at war with Eurasia, nor did anyone ever say that. We've been at war with Eastasia. Everyone around Winston seems to accept this as truth despite knowing yesterday that *Eurasia* is the enemy (2000). The reason Orwell's book is so compelling is not because of its fictional component, but because of its reflection of actual phenomena, like that of doublethink: to believe one thing while also maintaining a contradictory belief. One amusing example is when I hear a man claiming he hates when women wear makeup while

complaining when his significant other doesn't wear any. Another example is the woman who wants to be treated as an equal partner by having her husband do the same chores as her (i.e. washing dishes), yet she insists she can't mow the lawn like him. What about the tenet to tolerate free speech so that we may hear other ideas, but won't tolerate speech of ideas we simply don't like? The list goes on in which our levels of hypocrisy amount to an indisputable state of self-deception.

It doesn't take too much convincing at this point to understand why our brains would give us conflicting information about unpleasant realities in the world around us. We see to some degree how we can persuade ourselves to believe something that we know on another level doesn't make sense. I'll indulge a bit further though to drive this home. Most are familiar with the "fight or flight" response. When we engage in flight, we are avoiding the stimulus that poses a threat. When the realization of our mortality dawns on us, we may respond to the threat to our existence by using avoidance strategies. Drawing this together with our ability to deceive ourselves, we can see how denying the reality of our fate can be successful, even while knowing otherwise. I don't think it warrants explaining further *why* someone would want to avoid painful truths.

When denial isn't enough, we can find reassurance in religion,[8] science, and social constructs. Is it any wonder that religion in some form has been around for longer than recorded human history? One of the most appealing aspects of religion is the comforting doctrine of immortality that is found in most faiths. Virtually every religious philosophy addresses the issue of death, making attempts to provide solace for those asking life's pressing question, "what happens when we die?" Whether an afterlife involves Heaven, resurrection, non-material souls, reincarnation, or any combination of these, religion provides the promise of continued existence so desperately desired. We needn't assume that faith-based beliefs are the only themes that make promises of eternal life. Scientists do exist who are currently involved in work on Substrate Independent Minds, a visionary form of technology that hopes to be able to essentially upload human consciousness onto an inorganic platform that resists biological death. Carbon Copies is one such group that compares our brains to software data that can be

8. Dennett (2006) gives a fascinating lecture on how religion has evolved in the context of the selection process. This can be integrated with some of the earlier commentary in this paper regarding the need for humans to find ways to adapt to their death anxiety, with religion being just one of those ways in which coping mechanisms can be regarded as part of the evolutionary process.

uploaded and accessed even on separate "devices". If science can figure out how to replicate the processing of our minds and functionally preserve all of the data, they predict we can experience consciousness on a substrate, independent of our physical body, which means we won't have to perish with our flesh. We've also heard of cryogenics. It is not a myth that establishments like Alcor or the Cryonics Institute exist that preserve the deceased in hopes that a more advanced technological future can bring about revival. For those who find the above to be mere pseudoscience and religion too mythical, we can find reassurance of mortality in social constructs. Particularly fame and legacy.

If we know that we can't literally live forever, we can try to achieve the closest thing to it. It's fairly common that people dream about making a big break or accomplishing an amazing feat, garnering masses of adoring fans who will remember them long after their earthly advent. Kings have erected great edifices and statues of their likeness to make their presence known in a lasting way. Musicians and writers hope their art will be appreciated throughout generations to come. Other breakthrough prodigies fiercely protect their right to credit for their work not just for financial benefit, but for the sake of perpetual recognition. Sigmund Freud once reportedly fainted at the very thought of his work not being carried on after he passed away.[9] We know that Aurelius himself must have struggled with this due to the amount of times he reflected in his Meditations to avoid relying on fame for comfort. We sense a distinct impression that we would feel at greater peace if we knew we were to be remembered rather than slipping into an eternal obscurity.

Few will accomplish star status though, which leaves us legacy. Living on through the memory of those who live after us. The popular holiday Dia de los Muertos is celebrated by millions who honor their ancestors by displaying their photos on an offering table called an "ofrenda'. Coco is a heartwarming film that depicts a character who is terrified that when he is no longer remembered, he will

9. Sigmund Freud is known to have been obsessed with fame and to have struggled with death anxiety. He had taken special interest in Carl Jung as an apprentice, with whom he had hoped would carry on his ideas. The incidents in which Freud fainted pertain to this topic in which he interpreted some comments from Jung as a threat to Freud's legacy. One particular comment that caused Freud to pass out was Jung's allusion to Freud's Oedipal theory, which asserts that sons are jealous of their fathers and want to usurp them. In a letter to Freud, Jung once wrote, "Let me be to you as the son is to the father." It was a reference to this phrase that Freud took as an encrypted message that Jung meant to supplant him and render Freud's work irrelevant. See Silverstein, (2014) and Razinsky (2016).

disappear (Unkrich 2017). Many would relate to its pinnacle message in the form of a song entitled, "Remember Me" (Anderson-Lopez and Lopez 2017). Speaking of ancestors (of which one day we will be), descendants also play a role in perpetuating our known existence. Family crests and diaries for example, represented values and ideals that members identified with and passed down through their posterity. Today, photographs are cherished mementos that seek to preserve moments in time that would otherwise be forgotten or never even known about. If we can't be alive forever, perhaps there is peace of mind knowing that someone at least knew who we were. If we can't achieve that through fame, our posterity is more likely to retain a semblance of us in their mind. From religion to scientific pursuits to renowned fame to leaving behind a legacy, these strategies remain an active role in pursuit of rectifying our grievances about annihilation. Whether or not we will ever be able to accept this truth, to thrive with full authenticity despite our knowledge of what's to surely come, depends on whether we are able to rise above our innate tendencies. Which leads us to a final consideration. Acceptance.

Acceptance is a category of its own because of its unique quality. It doesn't come naturally like the aforementioned techniques and is contingent on factors involving a great deal of effort. Acceptance is more elusive than the more common coping mechanisms because it usually comes only after further suffering. Acceptance necessitates having to face death in all its stark reality, being fully awake to all its terrible glory without the crutches of distraction, denial, or reassurance to lull us back to sleep. While many are forced to face it when given the diagnosis of a terminal illness, witnessing violence, or undergoing a major paradigm shift, it is unlikely that anyone would want to voluntarily endure suffering when there are other anxiety reducing strategies that are working well enough. Still, for some people the aforementioned strategies don't work for various reasons. Regardless of whether or not your security blankets were ripped from under you, or you just want to embark on a life of authentic truth, you're probably wondering why you'd have to face facts with such raw exposure in order to reach acceptance.

What I explain next about obsessive compulsive disorder (or OCD for short) and phobias may seem to be off topic here, but you'll understand the relevance soon enough.[10] OCD is characterized by intolerance of uncertainty (Morein-Zamir

10. For a more comprehensive understanding of OCD and exposure therapy, see Phillipson, (n.d.). It is worth keeping in mind the parallels I've drawn between death anxiety and the content within the article while reading.

et al 2020) that manifests in obsessive and/or compulsive behaviors that are meant to reduce the effects of distress in the patient (American Psychiatric Association 2013). They often engage in rituals that are external acts to both distract from and reassure them regarding their source of anxiety. Their rituals are not merely an attempt to avoid the source of the anxiety, but to also avoid the discomforting thoughts themselves. O'Connor and Audet attest that the more one engages in these reassurance seeking behaviors, "the more they are removed from reality and the senses" (2019, 40). The root fear of OCD is often not what it appears to be on the surface. For example, a person who obsessively washes their hands isn't just afraid of germs. They're afraid of what the germs do and what it might mean if they got infected. In fact, their compulsions are reactions to the unceasing ruminations as their imagination runs wild considering all of the possible things that aren't likely to happen, but could. If this sounds unrelatable to you, think back to the old nursery rhyme about stepping on cracks and breaking backs. Most children rationally know that stepping on a crack won't actually lead to a fractured spine, but how many of you felt the need to avoid the crack anyway? You know, just in case... Hence, the handwashing or knocking on wood. Just in case. Or saying a quick prayer or wearing your favorite underwear on game days… Just in case.

Phobias are different from OCD in that sufferers are afraid of a specific known object (Fritscher 2020) and unlike OCD, they do not engage in rituals to ward off thoughts of the fear stimulus and only experience distress when faced with the threat. Spiders, snakes, and heights, for example, are some of the most common phobias known to man (Horn 2015). Although not characterized by obsessive properties, when directly threatened with or given the possibility of exposure, people will go to extreme lengths to avoid the stimulus that evokes fear. It is similar in this way to OCD in that avoidance tactics are used as a temporary compulsion to protect themselves from that which they are averse to. In any case, these two conditions are two sides of the same coin and the method of Exposure Therapy for treatment is generally the same.[11] Are you with me still? I promise I'll get to what this has to do with acceptance and why increased discomfort precedes it.

11. See "What's the Difference" (2018). Exposure with Response Prevention (ERP) is a type of Exposure Therapy (ET) that is used specifically to treat OCD. It incorporates the additional measure of prohibiting OCD sufferers from engaging in reassurance seeking behaviors before, during, or after exposure to the fear inducing stimulus. Both phobia and OCD patients engage in exposure therapy and must avoid any behaviors designed to placate the patient's exposure. In

Keeping in mind what we covered in self-deception about the ability to compartmentalize, we can understand why OCD patients recognize their rituals are irrational while still believing their rituals will solve the problem. Sometimes those lucky rituals coincidentally correlate with successful relief, which reinforces our belief that we need them. This is precisely why children will cry when they lose their favorite teddy bear and have to sleep in the dark without it. The only way to no longer depend on the ritual, is to live without it and learn that the ritual itself does nothing (or negligibly little) to prevent harm. Exposure therapy is one of the most (if not the most) highly successful modes of treatment for patients suffering from OCD and phobias (Pittenger et al., 2005, p. 36). This type of therapy exposes patients to their feared stimulus at increasing intervals of intensity. A patient deathly afraid of spiders might begin by looking at photographs of spiders and then over time have to sit across the room with a real spider. With each consecutive session the spider may be brought closer until the patient is made to tolerate actually coming into physical contact and even holding the spider. It is crucial they not be allowed to use any techniques to soothe themselves at this time, or the therapy will not be successful. After becoming habituated to the fear invoking stimulus, they learn to endure it without the ability to seek reassurance by engaging in their ritual or avoidance behavior. That means they have to sit with the discomfort of not being able to wash their hands, or check the door locks five more times or go ballistic when they see a spider lurking in a corner. What first is met with terror becomes tolerance, which becomes indifference, which sometimes even becomes acceptance. Likewise, we'd have to learn to not switch on the TV or recite a dozen Hail Mary's to escape the inevitable. We'd have to learn that while we must take reasonable precautions to not contract deadly germs or prematurely face our death by walking in front of a bus, the rest is out of our control. No amount of lock checking or research about cryogenics will prevent our obliteration. We cannot run away from the funeral casket the way we can from a spider. There is no doubt about it. We will die.

Once we have finally faced facts, we can begin to move towards acceptance. Studies like Dr. Kevin Rand's show that accepting death can lead to a greater sense of fulfillment (2016). When you are hyper aware that a certain moment will be your last, there arises a sense of urgency to live a more enriching life. When we gain acceptance, we stop fighting a losing battle, and when we stop fighting,

this paper we will refer to all types of ERP as ET.

we are freed up to spend that energy on pursuits that matter more to us. Pursuits that have more likelihood for success. Being aware of death means we are more likely to refrain from procrastinating and take steps to create lives that we can be proud of, which enables acceptance. When we feel we've made the best of what we had, it is much easier to accept the end. Think of it like this. Imagine you're just a child back when you only had a few spare dollars from doing menial chores and you managed to get into your favorite ice cream shop right before closing. Satisfied the last of your allowance was well worth the splurge, you walk out with your treat, adorned in all its sugary splendor. And then, to your utter dismay, the delicious dessert you so craved unexpectedly slips from your hand and falls splat on the dirty ground to meet its doom. It's too late to get another one and you've wasted your money. You wouldn't have minded the ice cream being gone nearly as much if you'd at least have gotten a chance to enjoy some of it first!

Life can be viewed similarly. Maybe death can be accepted; but only if we lived fulfilling lives that we had a chance to savor, would we even be somewhat ready to part ways. And in order to live such a life, it helps to feel a sense of urgency to make the most of it. Otherwise, it's business as usual and before we know it, we've run out of time. The question that remains is whether or not distraction, denial, reassurance, and acceptance are actually effective in relieving our anxiety and to what extent. That is a difficult question to answer and would warrant a lengthier discussion that won't be addressed here. What we can deduce is that these strategies likely wouldn't retain such a strong presence in our lives if they weren't fulfilling some compelling need. Sure, society and technology have evolved to make some specifics look different, but the root principles have always been the same. The mortal environment in which we live coupled with our biological make-up leads to the expression of life's greatest terror: our finiteness and disconnection in an absurd universe culminating with our death. Remembering death may motivate us to waste less time. Reflecting on our lives with less regret means we are more likely to be able to accept our end. In keeping with the idea of memento mori, we can enjoy our chance to savor life to its fullest while it lasts. Marcus Aurelius leaves us with a profound edict, "...let us pause and ask ourselves if death is a dreadful thing because it deprives us of *this*" (1945, 110). Evaluate what it is you are doing and ask yourself not if it's worth avoiding death for, but if it's worth living for.

REFERENCES

Abeyta, Andrew A., Jacob Juhl, and Clay Routledge. 2014. "Exploring the Effects of Self-Esteem and Mortality Salience on Proximal and Distally Measured Death Anxiety: A Further Test of the Dual Process Model of Terror Management." *Motivation and Emotion* 38 (4): 523–28. https://doi.org/10.1007/s11031-014-9400-y.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5*. Arlington, VA: American Psychiatric Association.

Aurelius, Walter Marcus, and Irwin Lucian. 1945. "Marcus Aurelius and His Times the Transition from Paganism to Christianity, Comprising Marcus Aurelius: Meditations; Lucian: Hermotimus, Icaromenippus; Justin Martyr: Dialogue with Trypho, First Apology."

Bassok, Miriam and Douglas L Medin (E.d.). 1997. Essay. In *Psychology of Learning and Motivation: Advances in Research and Theory* 37, 37:1–39. Academic Press.

Binder, Jeffrey R. and Rutvik H. Desai. 2011. "The Neurobiology of Semantic Memory." *Trends in Cognitive Sciences* 15 (11): 527–36. https://doi.org/10.1016/j.tics.2011.10.001.

Dawkins, Richard. 2016. *The Selfish Gene: 40th Anniversary Edition*. New York, NY: Oxford University Press.

Dennett, Daniel C. March 12, 2006. "Breaking the Spell: Religion as a Natural Phenomenon." *Nature of Knowledge Series*. Lecture.

Duval, Céline, Béatrice Desgranges, Vincent de La Sayette, Serge Belliard, Francis Eustache, and Pascale Piolino. 2012. "What Happens to Personal Identity When Semantic Knowledge Degrades? A Study of the Self and Autobiographical Memory in Semantic Dementia." *Neuropsychologia* 50 (2): 254–65. https://doi.org/10.1016/j.neuropsychologia.2011.11.019.

Fritscher, Lisa. March 9, 2020. "How Phobias Differ from Other Mental Disorders." Verywell Mind. Verywell Mind. https://www.verywellmind.com/phobias-or-anxiety-disorder-2671925.

Gould, Stephen Jay. 1997. "The Exaptive Excellence of Spandrels as a Term and Prototype." *Proceedings of the National Academy of Sciences* 94 (20): 10750–55. https://doi.org/10.1073/pnas.94.20.10750.

Horn, Allyson. May 1, 2015. "The Ten Most Common Phobias." ABC News. ABC News, https://www.abc.net.au/news/2015-05-01/ten-of-the-most-common-phobias/6439210.

King, Stephen. 2010. "What's Scary [Foreword]." Essay. In *Danse Macabre*. New York: Gallery.

Kraut, Richard. June 15, 2018. "Aristotle's Ethics." In *The Stanford Encyclopedia of Philosophy*. Stanford University. https://plato.stanford.edu/entries/aristotle-ethics/.

Lindquist, Kristen A., Jennifer K. MacCormack, and Holly Shablack. 2015. "The Role of Language in Emotion: Predictions from Psychological Constructionism." *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00444.

Morein-Zamir, Sharon, Sonia Shapher, Julia Gasull-Camos, Naomi A. Fineberg, and Trevor W. Robbins. 2020. "Avoid Jumping to Conclusions under Uncertainty in Obsessive Compulsive Disorder." *PLOS ONE* 15 (1). https://doi.org/10.1371/journal.pone.0225970.

O'Connor, Kieron, and Jean-Sébastien Audet. 2019. "OCD is Not a Phobia: An Alternative Conceptualization of OCD." *Clinical Neuropsychiatry* 16 (1): 39.

Orwell, George. 2000. *1984*. London: Penguin Classics

Pascal, Blaise. 1901. *The Thoughts of Blaise Pascal: Translated from the Text of M. Auguste Molinier*. Edited by Auguste Molinier. Translated by Kegan Paul. London: G. Bell and Sons.

Phillipson, Steven. (n.d.) "In Both the Treatment of OCD and in Living a Disciplined Life, There Is No Word More Important than 'Choice.'" Ocdonline. Accessed June 1, 2022. https://www.ocdonline.com/choice.

Pittenger, Christopher, Ben Kelmendi, Michael Bloch, John H. Krystal, and Vladimir Coric. 2005. "Clinical Treatment of Obsessive Compulsive Disorder." Psychiatry (Edgmont). U.S. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/21120095/.

Rand, Kevin L., Daniella A. Banno, Amanda M. Shea, and Larry D. Cripe. Febuary 11, 2016. "Life and Treatment Goals of Patients with Advanced, Incurable Cancer." *Supportive Care in Cancer*. https://doi.org/10.1007/s00520-016-3113-6.

Razinsky, Liran. 2016. "Publish and Perish: Freud's Claim to Literary Fame." *Journal of Modern Literature* 39 (4): 1–18. https://doi.org/10.2979/jmodelite.39.4.02.

Seneca, Lucius Annaeus. 1932. Letter 77 On Taking One's Own Life. In *Seneca's Letters to Lucilius*. Translated by Edward Phillips Barker. Oxford: Clarendon Press.

Silverstein, Barry. 2014. "Freud's Lost Metapsychological Papers." In *Freud: Appraisals and Reappraisals*, edited by Paul Stepansky, Vol. 1, 160-163. New York, NY: Routledge.

Schopenhauer, Arthur. 2016. Supplements to the Fourth Book Chapter 41 in *The World as Will and Idea,* 648-678. Translated by R. B. Haldane M.A. & J. Kemp M.A.,British Columbia, Vancouver: Kshetra Books.

Solomon, Sheldon, Jeff Greenberg, and Tom Pyszczynski. 2016. "Self-esteem: The Foundation of Fortitude." In *The Worm at the Core: On the Role of Death in Life*. New York: Penguin Random House.

Tolstoy, Leo. 2004. *The Death of Ivan Ilych.* Edited by David Goldfarb. New York: Barnes and Noble Books.

Understanding the Stress Response." 2011. Harvard Health Publishing. https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response.

Unkrich, Lee, dir. 2017. *Coco.* Burbank, CA: Walt Disney Pictures and Pixar Animation Studios, DVD.

"What's the Difference Between Cognitive-Behavioral Therapy, Exposure Therapy, and Exposure with Response Prevention?" 2018. OCD and Anxiety Center of Cleveland. https://ocdandanxietycenterofcleveland.com/anxietyblog/2018/9/24/whats-the-difference-between-cognitive-behavioral-therapy-and-exposure-with-response-prevention-and-exposure-therapy

Yalom, Irvin D. 2011. *Staring at the Sun.* London: Piatkus.

compos mentis

Reconsidering Psychological Egoism

Amanda Galloway
Indiana University-Purdue University

ABSTRACT
Empirical evidence of altruistic behavior, such as studied in evolutionary biology and psychology, is often cited as a primary reason for dismissing Psychological Egoism as a reputable theory for human behavior. Although Psychological Egoism ought to be regarded as false, the manner in which this conclusion has traditionally been arrived at is faulty. An alternative argument is presented that is not only preferable to traditional ones, but sufficient to lay the theory to rest. This paper seeks to show that without this alternative argument, flaws in traditional reasoning render PE as being worthy for reconsideration.

KEYWORDS
Altruism, Selfishness, Motive, Empathy, Egoistic

> It seems philosophical arguments against psychological egoism aren't quite as powerful as we might expect given the widespread rejection of the theory among philosophers. So the theory is arguably more difficult to refute than many have tended to suppose. (May n.d., Conclusion section, par. 1)

Despite the general consensus reflected in this quotation, I would like to explain why Psychological Egoism ought not to be dismissed lightly even though it makes more sense to conclude it's false. I argue that although many (and arguably all) human actions are egoistic, they are not necessarily so, and therein lies the flaw with Psychological Egoism (PE). I will begin by first defining the nature of this theory after which I will assess various arguments for and against it. Arguments against PE will be drawn from the perspective of one who advocates altruism in humans as a sufficient reason for dismissing PE.

By altruistic behavior, I mean behavior motivated by the sincere desire to benefit or avoid harm to another. By egoistic behavior, I mean behavior motivated by the desire to benefit (whether it be a hedonic state or not) or avoid harm to ourselves. When I specify pure altruism, I mean altruistic behavior with no trace of egoistic motive. That is, altruism without any regard or concern for one's own benefit or well-being. The theory of PE asserts that all human action is egoistic, and so pure altruism is impossible. Consider this example: a harried businesswoman takes an opportunity to relax with a book on her porch swing, when a neighbor's child comes crying. The woman, although sympathetic to the child, wants to continue with the little time she has to read her book. She comforts the child and then gets back to reading. David Hume maintains that empathy is often the underlying motive for behavior cases like this (Hume 2010).[1] But the Psychological Egoist says otherwise: according to PE, the underlying motive for helping the child

1. Hume appeals to empathy in order to show how pure altruism is possible. He presents the example of a mother who sacrifices her health to care for a sickly child who will inevitably die. Hume writes in Appendix II of *An Enquiry Concerning the Principles of Morals*, "Tenderness to their offspring, in all sensible beings, is commonly able alone to counter-balance the strongest motives of self-love, and has no manner of dependence on that affection," (2003, par. 9). What is in the mother's self-interest here? The child will never return the favor of care in the mother's advancing age and is bringing more distress on the mother. Of course, the egoist could point out a motive like avoiding the public shame or guilt of abandoning her child. But this could easily be refuted by the altruist as they point out that guilt cannot be felt in the first place if there were no genuine empathy for the child preceding the guilt.

is self-interest. The businesswoman helps the child at least in part (and perhaps wholly) so that he will leave her alone and she can continue her leisure time.

I don't think anyone would argue that self-interested motivations aren't accountable for some human behavior. But what the egoist needs to prove is that there can never be a case where self-interest isn't present and causally efficacious. If an opponent to Psychological Egoism could come up with even one instance of pure altruism (one instance of altruism where egoistic motives are, even if present, not causally efficacious), then we can lay PE to rest as a philosophical theory. Much of the confusion lies in what is possible versus what we actually witness. Even if it were the case that every instance of altruism has been accompanied by self-interested motives, this does not mean it has to be this way. An illusion of PE is created when the egoist can argue out a (perhaps hidden) self-interested motive in every circumstance. Even the popular example used by the altruist of the soldier sacrificing his life to jump on a grenade can be construed as self-interested (Shaver 2019, section 1). Perhaps the soldier couldn't bear survivor's guilt or wanted to be remembered as a hero. The list of potential self-interested motives goes on and could be rather convincing. Inferring PE based on this would, however, be a mistake. Even if all behavior is, as a matter of fact, egoistic, this is insufficient for Psychological Egoism.

I want to concede that it is possible to observe, or at least argue for, egoistic motives in all cases. I am arguing that, despite this concession, PE is still not true on logical grounds even if it cannot be dismissed on empirical grounds. To see what I mean, consider recent researchers' attempts to disprove PE. In the *Stanford Encyclopedia of Philosophy* article on Psychological Egoism, Shaver comments on a series of experiments performed by Daniel Batson and colleagues in which they tried to show the implausibility of PE. In one of these, Batson set up a controlled experiment in which he attempted to manipulate empathy in order to see if there was a change in helping behavior towards a fictional student, Katie Banks, who was enduring a serious hardship (Batson 2019). In order to elicit and control for empathy, Batson had participants watch a video of a painful diathermy treatment before being presented with a request for help from Katie Banks. Batson divided participants into two groups as follows: The first was a watch-him/pain group where participants were supposed to focus on techniques the demonstrator used during the treatment that would contribute to the effectiveness of the message eliciting certain reactions in the audience. This would help ensure the focus was

diverted away from empathic triggers, inducing a lower amount of empathy for comparison. The second group was designated as imagine-him/pain, where the focus was on imagining the other person in pain. What Batson discovered was that a greater number of people from the higher empathy imagine-him/pain group were willing to help Katie.

The problem with the study is that empathy may not have been the reason participants helped. The egoist can think of other reasons to account for the helping behavior that are self-interested. Perhaps the participants feared judgment from a religious deity if they failed to serve their fellow beings. They even could have acted on an honor basis to contribute to societal behaviors that would in turn benefit them. Batson should have compared the imagine-him/pain group with an imagine-self/pain group to see if there was any difference between those two in helping behavior. This would have better determined whether actual empathy was at play. Imagining pain in another person should induce greater empathy in the participant in comparison to imagining pain in the self, which is oriented towards self-interest. In order to make a case against PE, helping behavior would need to be greater in the imagine-him/pain group. If the results were the same between the two groups, we would have no way of knowing whether the helping behavior was influenced by empathy any more than self-interest. However, if the results showed greater helping behavior in the imagine-him/pain condition over the imagine-self/pain condition, that would offer more support for refuting PE.

Thus, proving pure altruism exists runs into the same setbacks that egoism presents. We simply can't prove pure altruism is possible for humans even if we can imagine a world where it is. Every act of seemingly pure altruistic behavior can be countered by the egoist as having some possible self-interested motive as cause of the action. Once again, we would have to rely on empirical data to infer our best guess and even then, it might not be a sound conclusion. It would always be possible to find data to the contrary, discovering we were wrong all along in how we interpreted the original data. In the end, it seems that the theory of PE cannot be dismissed on empirical grounds. New evidence can always emerge that undermines current data, and our interpretation of it is subject to error due to insufficient understanding. Overlooking pertinent information shows how our analyses of a given situation can be underdetermined. Batson interprets his experimental results as support for altruism when it can conversely be used to support egoism depending on how the data is interpreted. For example, since

self-interest was removed from the watch-him/pain group, we could interpret this as higher egoism in the imagine-him/pain group versus the watch-him/pain group, which actually supports PE. The results can be interpreted in either direction, leaving us with no reliable answer.

In fact, there is room for the egoist to make a stronger case. He could make an argument where empathy itself is considered to be egoistic. What if the empathy one feels for doing something for the sake of another is a form of selfishness in and of itself that motivates the action? What if empathy itself is the pain or desire, stemming from one agent's identification with another, we try to remove or attain? In *A Buddhist Bible*, we find a quote from Eastern Pali Scriptures, "And what is the Suffering of not getting what one desires? ...But this cannot be got by mere desiring; and not to get what one desires is suffering" (Goddard 1966, 24). This surely applies not only to what we want for ourselves, but when what we want is directed towards the object of our desire. When our altruistic desires for another are thwarted, it is the same as suffering, as the Eastern text indicates. On the positive side, empathy itself is the pleasure experienced when witnessing the wellbeing of another. For example, I may feel your discomfort by proxy if I witness you trying to scratch an itch you can't reach. I scratch it to make myself feel better since your discomfort is my discomfort. Furthermore, your relief is my relief and I arguably acted in self-interest to scratch your back precisely because of empathy. It's worth asking whether I can count as a benefit that very feeling of satisfaction I get when I do something for the sake of another. The altruist would say no; self-interested motives are sometimes merely a by-product of the action instead of a cause of the action and are not wrapped up inseparably with empathy. But the burden of proof lies on the altruist to defend this--a burden I believe is met below.

The strength of PE is that in the background, before any desire for an object is experienced, seems to lie a desire for personal benefit. The first cause for an action is actually an "ultimate" desire, followed by an object of desire which causes pleasure (May 2021). Let's take the example of desiring water to quench your thirst. According to PE, we only want water because of an ultimate desire which includes self-benefit: avoiding pain of thirst, deriving pleasure from quenching our thirst, and to survive. Water is just a means to an end of bringing about the ultimate desire for self-benefit. And so, in this way, the ultimate desire includes the seeking of pleasure/avoidance of pain in and of itself. If we didn't have a desire

to avoid the pain of thirst and feel the pleasure of that thirst being quenched, we would not want the water to begin with.

However, there is a way for the altruist to refute this point, and, I argue, this refutation uncovers a fatal flaw in Psychological Egoism. Joseph Butler presents such a refutation, which is basically two arguments in one. His idea is that you cannot receive pleasure by seeking pleasure directly because the desire one feels for any given object is directed towards another rather than inward, which yields two different concepts, one that is essentially object-directed and another that is essentially self-directed and, thus, egoistic: "There is then a distinction between the cool principle of self-love, or general desire of our own happiness, as one part of our nature, and one principle of action; and the particular affections towards particular external objects, as another part of our nature, and another principle of action" (Butler 1999, 47). In order to receive self-benefit, we must first seek something other than self-benefit. I must actually desire that "something other" for its own sake in order to experience the benefit of ensuing pleasure. If I did not desire that "something other", then I would not feel the benefit of pleasure I expect to derive from it.

In this case, the object of desire appears to be a tool that is used merely for the purpose of deriving pleasure. Perhaps this is similar to what John Stuart Mill meant when he wrote, "The utilitarian doctrine is, that happiness is desirable, and the only thing desirable, as an end; all other things being only desirable as means to that end," (Mill 1971, para. 2). While this may be true operationally, it is still the case that desiring the object is and must be a separate condition from desiring pleasure itself. This is important because the egoist makes a mistake in equating the object with said pleasure. Because the object of desire is a separate condition, it functions as a sort of cause and effect rather than it being a unified companion with self-benefit. I want the object which causes me to have pleasure. Desiring to have pleasure as a first cause does not necessarily entail that particular object of desire. Even if it did, I must actually want and obtain that object for itself in order to receive the pleasure. Consequences, or recognizing benefits we might reap, ought not to be confused with initial motivation. Pleasure may be derived from the action, but there is a difference between pleasure resulting from an action and doing the action for pleasure itself. Genuine concern for another person may be a prior reason to do any given act while any self-interested effect is independent of

this act. Having our conscience eased doesn't appear before empathy is felt (i.e., by an agent toward another being) and that is the key there.

Supposing we accept that an egoistic motive must always be present with altruism, it is in accordance with Butler's view that self-interested benefits in some cases are only a by-product of doing something good for another. For example, if I can't sleep because I'm worried about someone, and I get out of bed to go help them, I may be merely fooled into thinking I only went to help them so that I could sleep at night. In reality, since the fact that my lack of sleep was preceded by feeling bad for the person, my egoistic feeling was only involved as a corollary. It is even only a corollary with an ultimate desire and is an independent event which does not arise from that. Even if always observed to be present along with altruistic acts, egoistic motives may only be there coincidentally. The altruistic feeling precedes the egoistic feeling, which means the root cause couldn't be self-interest. So even if the egoistic feeling always follows, if altruism is present, egoism does not have to be the cause. If the altruistic feeling is the sole cause, it doesn't matter whether or not egoism is also present and if that is the case, PE is false.

The point is that we don't need to prove the existence of pure altruism in order to dismiss PE. We only need to prove its possibility, and for that it suffices to show the actuality of impure altruism. This is because so long as impure altruism exists, pure altruism is possible. That is, so long as there is impure altruism, it is *possible* for it to be an initial cause as already explained, which makes any accompanying egoism causally irrelevant, and that suffices to show that PE is false. In the end, the answer remains somewhat inconclusive but comes down to a Descartes-style analysis as the basis of belief (Descartes 2008).[2]

In the same way Descartes shows we have no way to prove we even have bodies (see Meditation I), so is there no way to prove pure altruism exists. Taking a trip down the rabbit hole of radical skepticism takes an unreasonable turn which ends up leading to the conclusion that there's no reason to think that even impure altruism exists based on the premise that it can't be proven. However, Descartes does conclude in Meditation II that he can know that he exists as a thinking thing.

2. In *Meditations*, Rene Descartes justifies his own existence as a thinking entity by acknowledging his own ability to think, even if only apparent to himself. In fact, it is imperative to recognize that this can only be apparent to oneself. Descartes artfully demonstrates the epitome of extreme skepticism in Meditation I by arguing we have no reason to believe that anything, even our bodies, exists.

In a similar way, we can conclude by our own intimate acquaintance with and phenomenal experience of our thoughts and emotions that we have altruistic motives. This is how PE ought to be countered. Consistent with Descartes' demonstration, this can only be detected from the vantage point of one's own mind. There is no way to prove altruism exists in others nor a way to prove to others that altruistic motives occur in ourselves.[3] In fact, it is not even possible to prove to ourselves that our own actions are purely altruistic. But as I have been arguing, this is unnecessary. Since individuals get the sense that they themselves have altruistic motives, they may conclude that pure altruism is possible, and that is all that is needed to assert that PE is false. So long as altruism is experienced by any human, it is *possible* that this alone can be a motive for helping behavior even if egoistic behavior is always coincidentally present. Of course, some individuals might deny that they ever perform even impurely altruistic actions, and they might argue that anybody who thinks otherwise is in the grips of a delusion or a Cartesian madman's dream. But if the arguments in this paper works, it is reasonable for the rest of us to conclude that psychological egoism is false even if it can't be dismissed absolutely.

REFERENCES

Batson, Daniel C. 2019. An Unsettling Surprise. In A scientific search for altruism: Do we only care about ourselves? Oxford University Press.

Butler, Joseph. 1999. "Sermon 4 and 5- Upon the Love of Our Neighbor." Essay. In *Five Sermons Preached at the Rolls Chapel and a Dissertation upon the Nature of Virtue*, edited by Stephen Darwall, 46–57. Indianapolis, IN: Hackett Publishing.

3. The conventional approach to oppose PE is typically an appeal to the existence of altruism based on empirical argument. But as I have tried to show above, this is a fallacious method: it may lead to a true conclusion, but it does so via an invalid argument. Nonetheless, denying that pure altruism exists becomes extreme in the same way that skepticism can lead us to believe that other humans besides ourselves may not be sentient beings. Whether or not there is a way to prove that others have phenomenal experience is highly controversial, but it would be absurd to draw the conclusion that everyone besides ourselves is merely a robotic creature with a convincing external portrayal of thoughts and emotions characteristic of human consciousness. Since we know that we ourselves have phenomenal experience, we trust others do in the same manner.

Descartes, René. 2008. *Meditations on First Philosophy*. Oxford University Press.

Goddard, Dwight. 1966. *A Buddhist Bible: Selections from Pali Sources*. Boston: Beacon Press.

Hume, David. 2010. *A Treatise of Human Nature: An Enquiry Concerning the Principles of Morals*. Vol. 2. *Project Gutenberg*.

May, Joshua. n.d. "Psychological Egoism." In *The Internet Encyclopedia of Philosophy*. Accessed September 2, 2021. https://iep.utm.edu/psychological-egoism/#H3.

Mill, John Stuart. 1971. *Utilitarianism*.

Shaver, Robert. n.d. "Egoism: Psychological Egoism." In *The Stanford Encyclopedia of Philosophy*. Stanford University, January 15, 2019. https://plato.stanford.edu/entries/egoism/.

compos mentis

Shifting the Problem of Akrasia

Nathaniel Kastl
Texas A&M University-Corpus Christi

ABSTRACT
The Problem of Akrasia is a longstanding paradox that arises when decision making is considered to be a chiefly rational process and a person is expected to make the best possible choice for themselves (as rational people can sometimes be observed to knowingly choose the worse of two actions). This paradox presents a serious problem for most models of decision making. Recently, Donald Davidson has offered several promising solutions. In this paper, Davidsonian thought on the issue will be thoroughly investigated and ultimately found unsatisfactory. Instead, a new 'hedonistic' model of decision making will be proposed based on conclusions drawn from Davidson and Aristotle's work on the issue. In this new model, the Problem of Akrasia does not emerge.

KEYWORDS
Akrasia, Donald Davidson, Aristotle, Decision Making

A particularly handsome friend of mine had been having a eudaimonius spring semester when one of his exes reached out to see if he would come over. Now, this friend wasn't really one to be indecisive or easily confused, so it wasn't long before he came to a simple conclusion: She was bad news. After calmly thinking it over, he decided that he should text her back and tell her that he was not interested. A few minutes later, he found himself on the way to her house to do the exact opposite. Naturally, the next few days were a whirlwind that left him exactly as humiliated, emotionally battered, and angst-ful as he had predicted.

So why would this very-distinct-from-me person have made such a terrible decision for himself when he had already correctly predicted the outcome and determined that it was undesirable? This question already assumes much, but it can safely be said that my handsome friend was experiencing firsthand what philosophers have referred to as the Problem of Akrasia.

Though not the coinage of the term or the earliest commentary on the issue, a sort of de-facto early codification of the Problem of Akrasia can be found in Aristotle's *Nicomachean Ethics* (Book 7). Within, Aristotle describes and attempts to solve the observed paradox present in human behavior in which a person appears to rationally select the best course of action for themselves, then fails to act on it. Importantly, Aristotle goes on to make the following exceptions: A person cannot be said to be exhibiting akrasia if their consideration failed to include all relevant information, or if their consideration was logically flawed, or if their considerations were overwhelmed by their passions, or if the person was in any other way indisposed to proper rationality. In these cases, Aristotle is quite comfortable with the a priori explanations for their behavior. For the cases that remain, however, he offers two striking propositions:

1. Decision making is chiefly rational (outside of the aforementioned exceptions)

2. The Socratic Assumption: *or* No person knowingly and willingly makes the worse of two choices for themselves

Both of these propositions seem quite compelling, but observational (and anecdotal) evidence would seem to suggest that they contradict each other as together they paint a picture of reality that is simply not the case.

To solve the Problem of Akrasia, Aristotle suggests that the first premise is flawed. Instead of being the center of decision making, he suggests that rationality might be but one of several competing factors, and that this puzzle in particular might be accounted for via another factor for decision making: a desire for pleasure. This conclusion is somewhat distasteful to Aristotle as in the wider context of *Nicomachean Ethics*, it indicates that having great faculties of rationality does not actually guarantee that a person will not commit wrongdoing. To Aristotle, the only real defense that remains is to attempt to quell one's own desires through habit. If I had asked him to explain the actions of my friend, Aristotle probably would have told me that his rationality had been beaten that night by his lazily disciplined set of habits- and then changed the subject.

A MORE CONTEMPORARY APPROACH

Of course, the work of Aristotle is beyond steeped in what could generously be called the eccentricities of his time and lacks the more thorough structure and extensive collaborative processing of our time. In the modern era, one name is seminal with work on the Problem of Akrasia: Donald Davidson. In 1969, Davidson fundamentally altered and modernized thought about the Problem of Akrasia. Not only is his work a practical well of information and argumentation that evolved dynamically across his career, but it also serves as a hub of discussion for the philosophical community at large.

Davidson's account of the Problem of Akrasia is refreshingly distinct from the Aristotelean thought that came before it. In particular, he insists that the Problem of Akrasia is poorly served by being viewed through an ethical lens, as supposedly akratic situations can be amoral just as often as they can be within the scope of ethical consideration, like in the case of a person who acts inconsequently against their own intention, or even have morally superior outcomes, like in the case of a person who believes an evil act would be the best course of action for themselves but refrains akratically. Additionally, Davidson resists the Aristotelean position that the problem might be solved by attempting to manipulate the premises that created it.

Davidson lays out his account of the problem most clearly in his 1969 paper *How Is Weakness of the Will Possible?*. Davidson presents three premises:

1. If a person believes one course of action to be better than a second course of action, they *want* to take the first

2. If a person wants to take one course of action more than another, then they will take that course of action if they can only take one

3. Sometimes, a person will not take the course of action that they believe to be better.

Like Aristotle's, these premises are all compelling but seem to be contradictory. Unlike in the case of Aristotle, however, Davidson does not attempt to solve the problem at this level. Instead, he re-imagines what (even in his argument) is considered to be an akratic action. In this new understanding, the person acting contrary to their judgement is in fact not ignoring the parts of consideration involved in the first two premises (the selection of the best action and the transition from an opinion of a best action to a desire to take that action), but are rather performing those considerations hastily. In other words, a person who acts akratically has indeed selected a course of action that they believe to be better, has wanted to perform it, and then has performed it- only they did so without fully considering the wider context relevant. This phenomenon only seems to be a paradox because that same person can also perform the same consideration to a greater, more complete degree at a different time (before or after the fact) and as a result deem the action taken to have been against their better judgement.

Notably, this solution bears a clear resemblance to one of the Aristotelean exceptions to true instances of akrasia (the event of a consideration that lacks all relevant information). As a result, it can probably be said that Aristotle would not consider this to be a solution to the Problem of Akrasia at all, but rather a claim that what appears to be akratic actions are in fact just a type of action that he has already dismissed as unremarkable. That's not to say, however, that Davidson is incorrect or even that this observation is an objection at all. In fact, the explanation of akratic actions as simply being actions of another less interesting type is a very common theme in contemporary thought on the issue, both for Davidson in this as well as other explanations and for his peers.

Moreover, there is actually an important distinction between Davidson's account and Aristotle's exception: Davidson's account includes the secondary 'complete' consideration that causes the person taking the action to notice the issue in the first place. If this consideration was not involved in the scenario, it would indeed be the case that Davidson simply describes a person who fails to make a decision well given their capacities as accounted for by the exception. While it could certainly be argued that this, in reality, is what akratic actions are, it is not the case that this is what we *mean* when we say that we are experiencing akrasia because in this case the person performing the action, limited by the lesser amount of knowledge considered, doesn't actually know that they have done something seemingly akratic (which ostensibly is very much the point). Instead, when the wider secondary consideration is performed that does actually contain all relevant information, the person becomes aware of what might be called their better judgement and as such is then able to become aware of the fact that they have acted against it. Thus, Davidson's account is distinct in that it is representative of what is meant by the term akrasia where Aristotle's exception is not.

For a better objection to Davidson's position on the Problem of Akrasia, then, it might be more productive to refer to those he makes of himself. In his later paper, *Intending*, Davidson points out a flaw in the more granular mechanics of his explanation. Using the example of a person who 'intends' to build a squirrel house for no particular reason and never does, he introduces the concept of a 'pure intention', or an intention that cannot be connected to a consideration based on desires held at the time or a real action. This metaphysical gap puts Davidson's model in a precarious position as an unclassified sort of intention poses exactly the same problem that akrasia did in the first place: namely, that it is inconsistent with the model of decision-making being presented. In attempted solution, Davidson argues that 'pure intention' may be a sort of sibling to 'ordinary intention', which would actually be able to be identified with the consideration present in his first premise. In this way, Davidson pins 'pure intention' in place in relation to his model. Regardless, 'pure intention' lies outside of Davidson' model, connected as it may be, and this difficultly underscores a clear lack of completeness.

DAVIDSON'S REFORMULATION

A later explanation of akrasia by Davidson may be more satisfactory. In *Paradoxes of Irrationality*, Davidson makes a much more radical attempt to explain the Problem of Akrasia that even arguably violates his rule about manipulating the original premises. In this reformulation, Davidson argues that what is often called irrationality is, in all actuality, simply normal rationality that is not completely understood. He comes to this point by first identifying the concept of a 'rational core' of an action as being the reason that is given when that action is rationalized. Because Davidson believes that all actions taken by a person can be rationalized in some way, he further argues that, though some actions may be irrational in certain other senses of the word, no actions are irrational in the sense that they lack a 'rational core'. In a sense, he argues that even if there is not a good reason, a person taking an action always has at least some reason for having done so. Thus, akratic actions based on irrationality in this sense cannot be said to exist if this model is accepted. Davidson doesn't stop here, of course.

If this is the case, he argues, then the real paradox lies within the case of a person who fairly weighs two separate actions, finds one to be superior to the other, then takes the other- and moreover, the real inconsistency is introduced due to the fact that he still maintains that a person ought to act on their own best judgement (the Socratic Assumption).

Naturally, Davidson only introduces a more nefarious problem in order to later introduce a more nefarious solution- or two, rather. First, Davidson points to a surprisingly original idea: the difficulty verging on impossibility of properly communicating one's drives and the impact on how this might affect our perceptions of others acting akratically. To home in on this point, Davidson introduces two new principles- the assumption of intersubjectivity and the assumption of interpersonal interpretation. The former, intersubjectivity, posits that a person cannot truly understand another person's beliefs, desires, or intentions unless the two individuals share (or perceive themselves as sharing) a vast amount of positions that they would categorize as common knowledge. The latter is very similar. With the assumption of interpersonal interpretation, Davidson presses further into the issue and argues that one must actively believe that the person they consider is wise and that what they hold to be true actually is true, else they will often tend to consider them to be irrational. In fact, Davidson argues, as the

magnitude of the deviation from our beliefs to another person's increases, so too does the difficulty of understanding them in rational terms, rational though they may be. As a result, much of what is considered to be akratic in the first place might simply be explained by a failure in what Davidson speculates to be a profound reliance on empathy and sympathy for others different than ourselves. In other words, since the act of rational consideration is so sensitive to its starting conditions (i.e., a person's values, beliefs, etc.), a failure to properly recognize these starting conditions as valid in others can cause a person to fail to properly recognize a fully rational consideration, ultimately manifesting in the phenomena of perceiving others as acting akratically.

Davidson's second solution to his more complex formulation of the problem (and his main thesis for *Paradoxes of Irrationality*) is that many actions judged to be irrational may have been the result of a cause that was mental in nature but was not a reason. From here, Davidson expands to a partitioned theory of the mind. The main suggestion of this theory is that the mind consists of multiple parts which can sometimes act against each other. While one of these departments consists of the reason and consideration described in his previous formulations, the other designates the opposing course of action. In this theory, though only the first part matches what is meant by the more typical verbiage of consideration, both parts exhibit their own versions of reasoning and justification based on available resources, so nothing truly akratic or irrational is actually occurring. Instead, the course of action is made based on the seizure of control of one of the two competing parts after full and fair considerations have been made with differences in the action selected only resulting from differences in the source of the consideration, not the quality of the consideration itself.

In a sense, this partitioned theory of the mind is only meaningfully different from intersubjectivity and interpersonal interpretation in that the two entities are contained by a single 'person' (though this terminology gets more tenuous here). Mechanically, it works the same: The Problem of Akrasia is sourced to the dissonance caused when a controlling rational consideration is ignored due to a failure to recognize the desires and drives behind it, whether that key consideration be performed by another person or by a portioned section of the same mind. Together, the two explanations approach complete coverage of the issue by explaining issues of akrasia both within the self and in others.

PROBLEMS FOR DAVIDSON'S LATER SOLUTIONS

Of course, these explanations come with their own sets of issues. Intersubjectivity and interpersonal interpretation share the same problem as Davidson's first solution to the Problem of Akrasia- that they simply identify perceived akratic actions as a more mundane, known sort- but to a much more critical degree. Even if the scope of explanation for this solution is limited to akratic action in others, it cannot be said to fully achieve that goal. By saying akratic actions in others are merely our misunderstanding, one neglects the fact that the very interest in the Problem of Akrasia in the first place is in the confounding personal experience of it. A person who perceives themselves as having behaved akratically surely cannot be said to be doing so as a result of not being similar or sympathetic enough to themselves, and shelving momentarily the Problem of Other Minds, it would stand to reason that since the Problem of Akrasia is a widely reported phenomenon that seems to just be a part of being a person, this intersubjectivity and interpersonal interpretation resistant explanation cannot apply fully to what surely is the very same experience of akratic action in others from their own perspectives. Though some actions perceived as akratic in others may very well be simply a result of failure to understand their motives, this explanation rhetorically fails to demonstrate that all actions perceived as akratic in others are as such.

The partitioned theory of the mind does fill in this blind spot, but it has its own difficulties- mainly that to suggest such a thing would carry enormous ontological weight. Really, other than the fact that it seems to work as a solution to the Problem of Akrasia, there is little reason to think that the mind works in this way at all. It could just as easily be the case that Davidson's sealed-away consideration is instead just forgotten, or that an unknown part of the mind vies instead to select the lesser of two choices; explanatory power is an insufficient justification that becomes more egregious the more complex and counterintuitive a theory becomes.

So why, then, did my friend make such an awkward mistake in his love life? Were his faculties of reason overcome by a non-fit-of-passion desire? Was he just hasty and limited in his consideration? Did a secret part of his mind make the decision for him? Ultimately, each of these explanations are convincing only to a limited extent. Wherever Aristotle left his model of decision making too vague,

Davidson seems to have made the opposite mistake, constantly bloating into over specificity and leaving behind gaps due to a too-limited scope. Perhaps even more damning is the fact that Davidson's insistence that the solution to the Problem of Akrasia lies outside of the premises that devise it limits him entirely to argumentation that tries to recontextualize the phenomena after the fact.

Frankly, his reason for doing this is mystifying. The classic approach to addressing a paradox, which the Problem of Akrasia certainly is, is by carefully examining the structure of the argument and rejecting at least one premise or part of the reasoning. For this particular paradox, pointing to the reasoning as the issue can be more or less eliminated from consideration since the reasoning is so brief and simple, so the only clear approach that remains is to reject a premise. To his credit, Aristotle does address the Problem of Akrasia in this way when he concedes that decision making might not be a purely rational process. And, to *his* credit, the formulation of the Problem of Akrasia that appears in Davidson's *How Is Weakness of the Will Possible?* attempts to account for this adjustment with the addition of a second premise that converts a selection by reasoning to desire.

But why stop there? Surely Aristotle and Davidson's concessions both reveal that desire is a necessary component for modeling decision making. Conversely, reason does not seem to have earned its place in the model yet. In fact, as Haidt argues in *The Emotional Dog and Its Rational Tail* (2001), psychological evidence would seem to suggest that reasoning in the sober sort of way considered by Aristotle and Davidson can actually *follow* the decision-making process rather than preceding it. Haidt's evidence further indicates that what one might take as being sober consideration is actually justification and is in no way involved with real decision making but is believed to be to reinforce the illusion of self-rationality, similarly to how optical illusions and fabricated memories work by preying on the parts of our brain that already lie to us and justify our behavior. In other words, the insistence that decision making include rational consideration could very well be a mistake of correlation for causation with only the Problem of Akrasia to serve as a counterexample that occupies those cases where the correlation fails in a way that causation never would be able to. The fact alone that the Problem of Akrasia even persists indicates that a rational formulation of decision making is a poor fit and leaves desire alone in being a known component to the process.

A HEDONISTIC MODEL

So, lacking a rational consideration-based model, what would a desire based or 'hedonistic' model of decision-making look like? Simply put, this model would posit that in any given moment, a person acts on their greatest desire. This model is not a new one, nor are the most basic objections against it new. One may argue, for example, that under this model a person would never work out, get vaccinated, or otherwise subject themselves for discomfort for a future pay off since only the greatest desire in the present in considered. The response to this objection, of course, is that a desire held at a particular time can still apply to the future or involve planning. A person being vaccinated does certainly have a desire not to receive an arm pain, but this desire is outweighed by a greater desire (still in that moment) to avoid the greater discomfort of becoming ill. Similarly, the objection that people would act immorally (stealing or resorting to violence) if they only acted on their immediate desires is addressed by pointing out that there are certainly desires that could outweigh any desire to do wrong, like a desire to be moral or a desire to avoid the consequences of violating norms. Only when these desires are absent or outweighed by other desires are immoral actions possible, as reflected in real life.

When the Problem of Akrasia is approached with this hedonistic model, it simply ceases to be a paradox. In familiar Davidsonian form, akratic action would be modeled thusly:

1. If a person desires to take one action more than another action, they will

2. Sometimes, a person takes an action that is different than the action that they have formulated a rational argument in favor of

Both premises are compelling, but this time, they do not seem to be in conflict. The solvency of this model goes deeper, though. Unlike Davidson's *How Is Weakness of the Will Possible?* solution, this solution in no way attempts to argue against the existence of akratic actions. While Davidson argues that akratic actions are actually a sort of failure of rationality (akratic actions are the result of hasty consideration, so not fully rational), the hedonistic model still allows for

akratic actions (in that a fully rational consideration is made and disregarded) while still being internally consistent. Even Davidson's issue with the placement of 'pure intention' is easily resolved via the consideration that a desire (say, to build a squirrel house) could just be relatively weak and often outweighed. A person could have a pure intention in that they desire to one day build a squirrel house, but on any given day they desire some other activity more. This seemingly unimportant detail is actually key evidence that the hedonistic model succeeds where Davidson fails as it doesn't suffer from the hyper-specificity that excludes certain components of decision making.

On the other hand, perhaps the hedonistic model is so intuitive that it wraps around and collides with the Problem of Akrasia in a different way. I'm sure my handsome friend would certainly not be convinced that he did not have a very strong desire to continue having a focused and flourishing semester and another desire to avoid heartbreak- surely these weren't overpowered by a desire for a relationship he was consciously suspicious of. It would seem to be the case that he believes (based on his desires, of course) that he ought to have acted in a certain way but then acted in a different way. There are two important ways to respond to this objection. First, and most uncomfortably, is the implicit suggestion contained in the hedonistic model of decision making that pushes so many away: that we might tend to delude ourselves or be ignorant about what our real (and acting) desires actually are. The meaning of this suggestion ranges all the way from the concept of purely subconscious desires that manifest in confusing ways to the idea that we might simply be less than honest with ourselves when we desire something that we wish we had not. In the case of the dieting man who is shocked to find that he has eaten a sundae, there is a simple and tactless response: he probably just wanted to eat that ice cream more than he wanted to lose weight. Though this may seem profoundly different from the idea of purely subconscious desires, they really amount to the same thing, which is to say they are both reasonable enough explanations that cannot really be engaged with or debated because they are entirely reliant on discussion about subjective experience.

Relying on this sort of argument is unsportsmanlike in conversation and bad form in philosophy, but that's not to say that it's incorrect. The limitations of subjectivity do stop (productive) conversations in the case of a disagreement, but they don't define the truth quality of statements made in the technical unknown.

Luckily, a much more functional argument can be made to explain the disconnect between how a person believes their desires are arranged and the desires that they seem to act on. In the case of a desire that is clear and present, like a sundae or a short drive to meet an old flame, there is little doubt that acting accordingly to the desire will result in said desire being realized. However, in cases of delayed gratification, like good nutrition or one's long-term mental health, the same cannot be said. As a result, it may well be the case that these two different sorts of desires are treated differently within the hedonistic model- either by actually being of different qualitative sorts, or simply by being of different relative weights according to their immediacy. With this in mind, akratic actions are even easier to explain under the hedonistic model. To return to the case of the man with the sundae, it might be more tactful and more accurately explanatory to suggest that his known desire to eat the sundae won out because it had a more immediate outcome than his desire to be healthy. Notably, though, less immediate desires can still outweigh immediate desires given appropriate magnitude. If the sundae offered to the man came at the cost of a prison sentence starting a year after he ate the sundae, he would clearly leave it be.

Another major objection to the hedonistic model solution lies within the potential hypocrisy of the proposal. Surely it cannot be the case that Davidson's partitioned mind theory fails because it is only supported by its solvency of the Problem of Akrasia while the hedonistic model does not suffer the same critique. In all actuality, this is a real problem for the hedonistic model. In defense, one might argue at the very least that the hedonistic model has the benefit of being vastly more intuitive than the partitioned mind, which sounds frivolous but is poignant when considering the comparative burdens of proof. Additionally, the evidence offered by Haidt would seem to suggest that the rational basis on which the partitioned mind theory relies is flawed in such a way that the hedonistic model is not, nearly to the point of implication. The only real defense to this criticism, though, is Occam's razor. As problematic as it is to posit theories that exist only as a response to problems and not as an extension of observation of reality, it is an essential part of how philosophy and science both progress. Occam's Razor is an important tool for controlling this sort of theory that states that when possible, the simpler of two equally effective explanations is more likely to be reliable. Under Occam's Razor, the preference from theories like the partitioned mind to the hedonistic model is undeniable.

What about our intuition that we do have rationality, though? Do we really never act according to our rational considerations and just justify backwards? Fortunately, this question does actually have a reasonably satisfying answer baked into the hedonistic model itself: We desire rationality. If a person desires to be rational, as many do, the weight of a desire in favor of choices supported by one's reason increases. This caveat is hardly enough to brighten the outlook of such a grim model, but it does offer at least a ray of hope for consideration. In a way, just opposite to how the Aristotelean model characterized decision making as being mainly rational with components of desire factoring in, the hedonistic model is mainly based on desire, but still contains a component of rational consideration.

Is this model enough to explain what happened to my friend all those weeks ago? It's bizarrely satisfying to give up a bit of control and allow oneself to believe that actions might not be the result of thoughtful consideration, most of the time or ever. I'm sure my friend still isn't completely comfortable with the idea that what he thought he wanted wasn't really what he wanted, or that just that the timing of events was enough for his desires to manifest in such a way that brought him to harm, but a solution to the Problem of Akrasia was never going to bring comfort. For all my friend can know, that very same evening may well play out again, even despite his best intentions and loftiest thoughts about himself. One almost wonders if the reason that the Problem of Akrasia has remained unsolved for so many millennia is because the real comfort is in putting a name to the unknowability and lack of control that we find even within ourselves.

REFERENCES

Aristotle. c.335. *Nicomachean Ethics*. Translated by Christopher Rowe. Oxford: Oxford University Press.

Davidson, Donald. 1969. "How Is Weakness of the Will Possible?" *Moral Concepts.*

Davidson, Donald. 1978. "Intending." *Philosophy of History and Action* 11: 41-60.

Davidson, Donald. 1982. "Paradoxes of Irrationality." Oxford: Oxford University Press.

Gosling, Justin. 2002. *The Weakness of the Will*. Abingdon: Routledge.

Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (4): 814.

Hurley, Paul. 1992. "How Weakness of the Will Is Possible." *Mind* 101 (401): 85-88.

Mele, Alfred R. 1983. "Akrasia, Reasons, and Causes." *Philosophical Studies* 44 (3): 345-368.

Saarinen, Risto. 1993. "John Buridan and Donald Davidson on "Akrasia"." *Synthese* 96: 133-153.

Tremblay, Michael. 2020. "Akrasia in Epictetus: A Comparison with Aristotle." *Apeiron* 53 (4): 397-417.

compos mentis

Doxasticism about Moral Obligation

Kyle Kirby
Kalamazoo College

ABSTRACT
In this paper, I develop and argue for the view that an agent's moral obligations depend on their non-moral beliefs, a view which I call *doxasticism*, and I situate doxasticism within the current debate on whether an agent's moral obligations have any dependence on their epistemic or doxastic state. Three views have emerged in the contemporary literature. *Objectivism* is the view that an agent's moral obligations depend neither on their evidence nor their beliefs. *Prospectivism* is the view that an agent's moral obligations depend on their evidence. *Subjectivism* is the view that an agent's moral obligations depend on what they believe is morally best. I begin by giving a brief overview of the chief objections against each of these views. Then, I construct doxasticism from two principles: that ought implies can, and that moral obligations must be able to guide belief. In doing so, I introduce a novel modal concept of psychological possibility to describe the possibility of forming intentions to act. Lastly, I respond to two objections to doxasticism: first, that doxasticism is in conflict with robust moral realism, and second, that doxasticism unduly posits moral obligations in cases of non-veridical beliefs. I conclude that neither objection is a serious worry for the doxasticist.

KEYWORDS
Moral Obligation, Belief, Objectivism, Prospectivism, Subjectivism, Ought Implies Can

INTRODUCTION

Do an agent's moral obligations depend on their beliefs? Three central views have emerged in the contemporary literature on this question. *Objectivism* about moral obligations is the view that an agent's moral obligations depend on what is best independently of any epistemic or doxastic state of the agent. *Prospectivism* is the view that an agent's moral obligations depend on what the agent's evidence entails is best. *Subjectivism* is the view that an agent's moral obligations depend on their beliefs about what is best. In this paper, I will develop a fourth view which I call *doxasticism*, which is the view that an agent's moral obligations depend on their non-moral beliefs.

I take doxasticism to be filling a gap between the existing views. According to both subjectivism and doxasticism, an agent's moral obligations depend on some subset of their beliefs. The two views differ in what subset of their beliefs they take to be relevant; under subjectivism, the agent's moral obligations are determined by their moral opinions, while under doxasticism, the agent's moral obligations are dependent upon their non-moral beliefs.[1] Doxasticism is similar to both objectivism and prospectivism insofar as all of these views are non-substantive. Standing alone, none of these are complete moral theories, as each requires some additional principle about what is of moral value or about right-making in order to determine what an agent's moral obligations are. This is in contrast with subjectivism, which is a complete theory of moral obligation on its own; an agent is morally obligated to perform an action if and only if the agent believes that action to be morally best.

On one hand, objectivism and prospectivism both deny that an agent's moral obligations depend on their beliefs. Because of this feature, both of these theories face difficulties in cases of ignorance, when the agent's beliefs do not match the facts or the evidence of the case. On the other hand, subjectivism faces its own perennial problems of reducing morality to mere opinion and conflicting with robust versions of moral realism. If subjectivism is the only other theory on the market that can adequately handle cases of ignorance, then things look bleak for our theories of moral obligation. I argue that if doxasticism can capture the same

1. Non-moral beliefs are beliefs without moral content, which are distinct from *im*moral beliefs, or beliefs that are morally wrong to have.

insights as subjectivism in cases of ignorance while also not facing the numerous objections that subjectivism does, then we should be doxasticists.

In sections 1, 2, and 3, I briefly summarize the chief objections against objectivism, prospectivism, and subjectivism, respectively. In section 4, I construct doxasticism from two plausible and popular moral principles: that ought-implies-can, and that an agent's moral obligations must be able to provide guidance for the agent's actions. In sections 5 and 6, I respond to potential objections against doxasticism.

1: AGAINST OBJECTIVISM

Objectivism is the view that an agent's moral obligations depend neither on their epistemic nor their doxastic state (Spelman 2017, 8). Another way to state this is that under objectivism, moral obligations only depend on the hard facts, where the "hard facts" are understood to be the facts unrelated to beliefs and evidence. The paradigm case against objectivism is the following example from Frank Jackson.

Jackson's Drug Example

> Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B, and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will not completely cure it. One of drugs B and C will completely cure the skin condition; the other though will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which the killer drug. What should Jill do? (Jackson 1991, 462-463; Spelman, 9).

Any plausible objectivist account will admit of some moral principle in this case along the lines of "Jill is morally obligated to do what is best for John's well-being." Let drug C be the drug that would completely cure John. Under objectivism, Jill would be morally obligated to prescribe drug C and to not prescribe drug A. However, this seems wrong when we consider that it is generally wrong for

medical practitioners to act recklessly with regards to their patients' well-beings. By prescribing drug C, Jill is (from her point of view) making a 50/50 bet on John's life. Given that she can likely relieve John's condition with drug A without taking a massive risk on John's life, we should rather say that Jill is morally obligated to prescribe drug A instead of C.

The objectivist may respond by claiming that Jill is indeed morally obligated to prescribe drug C and not drug A, but that prescribing drug A would be an act of blameless wrongdoing. If this is the case, the objectivist needs some explanation for why intuitively, prescribing drug A not only seems blameless, but positively right. Furthermore, other cases exist in which it would not make sense to posit the objectivist obligation in the first place. Suppose that I am writing this paper while sitting in my living room, and mere feet away from me, there is a drowning child outside of my house. However, my curtains are drawn, and I cannot see the child. According to the objectivist, I would be morally obligated to walk outside my house and save the drowning child.[2] However, it seems much more plausible to say that I am not obligated to save the child that I have no awareness of, instead of saying that I am morally obligated to save the child but am engaging in blameless wrongdoing by continuing to write this paper. To maintain this picture of moral obligation would be to posit massive amounts of unknowable obligations and blameless wrongdoing in everyday life.

One further response the objectivist might make would be to say that these cases are different in one significant way: in Jill's case, she *can* prescribe drug C, while in my case, I *cannot* save the drowning child. Then, by appealing to ought-implies-can, the objectivist could explain why in my case, we need not posit that I had an obligation to save the child. Ultimately, this response fails because it betrays the objectivist's commitment to the principle that moral obligations only depend on the hard facts. Setting considerations about causal determinism aside, it is nomologically possible for me to save the child; no law of nature precludes me walking outside my house and picking up the child. Claiming that I could not save the child requires appealing to some sense of possibility that requires that I have awareness of the child, but the facts of my awareness are not included in the hard facts, and the objectivist cannot appeal to any such sense of possibility. These cases capture our intuitions that, as Jonathan Spelman puts it, "objectivism

2. I also take it that any plausible objectivist theory will generally admit of a principle such as "We are morally obligated to save drowning children instead of writing papers."

is plausible in cases where agents *know*, or at least *can know*, what is best," and implausible otherwise (Spelman, 75). This is not good enough; in searching for a general theory of obligation, we need a theory that is plausible even in cases when we aren't sufficiently informed, and so, our search continues.

2: AGAINST PROSPECTIVISM

Prospectivism is the view that an agent's moral obligations depend on their evidence. This time, Holly Smith provides yet another medical scenario against prospectivism.

Smith's Drug Example

> Harry is a physician who has to decide on the correct treatment for his patient, Renée. Careful consideration of the literature has led Harry to believe that his doing nothing (act E) is a significantly better bet than either of his alternatives, namely, prescribing drug F or drug G. Harry's senior colleague, however, knows that Harry has made a mistake. While Harry's evidence does suggest that act E is a significantly better bet than one of his alternatives (i.e., his prescribing drug F or drug G), Harry's evidence does not suggest that act E is a significantly better bet than his other alternative. In fact, Harry's evidence suggests that his other alternative is a *slightly* better bet than act E. Harry's senior colleague tells Harry this, and Harry justifiably believes her, but before he can ask her which of his alternatives is the slightly better bet, she is called away to deal with an emergency. (Smith 2010, 5; Spelman, 76).

Similarly to objectivism, I take it as a desideratum that any plausible prospectivist theory admit of a principle such as "Harry is morally obligated to do what his evidence suggests is the best bet for Renée's well-being." In this case, given that Harry's options are limited to act E, drug F, and drug G, Harry knows something about what his evidence suggests is the best bet: not-E. Either drug F or drug G is a better bet according to the evidence, but Harry does not know which one. Once again, according to prospectivism, our agent must make a coin flip. However,

being so reckless with a patient's behavior is wrong, and thus we should reject prospectivism.

Prospectivism also faces a much graver, Gettier-like problem in defining what exactly an agent's evidence is. Consider two potential answers given by Michael Zimmerman: an agent's evidence is either the evidence the agent could avail themselves of, or the evidence the agent does avail themselves of (Zimmerman 2014, 73). If an agent's evidence is the former, then there are cases like Harry's in which the agent could avail themselves of the evidence (for instance, if Harry had researched more thoroughly ahead of time) but does not, in which case the evidence can play no role in the agent's decision making. If an agent's evidence is the latter, then one could simply fail to avail themselves of any evidence for anything, thereby trivially absolving themselves of any moral obligations. In either case, how an agent's evidence is defined is too broad or too strict to be satisfactory for a general theory of obligation.

3: AGAINST SUBJECTIVISM

Subjectivism is the view that an agent's moral obligations depend on what they believe is morally best. Spelman uses the cases from Jackson and Smith to motivate the case against objectivism and prospectivism. Because subjectivism is supposedly the only other theory on the market, this implicitly motivates subjectivism (Spelman 2017). The costs of moral subjectivism are generally believed to be great, but much of Spelman's paper is dedicated to showing that the costs are not as unpalatable as we may have been led to believe. He considers four arguments from Zimmerman against subjectivism and responds to them with varying degrees of success. I will be addressing the two responses I find to be least successful: his response to Zimmerman's objection that subjectivism implies that Hitler did nothing wrong, and his response to Zimmerman's objection that subjectivism implies that it is trivial to become morally infallible.

The first response is to the objection that "the Subjective View implies that, on the assumption that he was doing what he believed to be best, Hitler did no wrong. But it is grotesque to think that such a perverse belief could render mass murder morally permissible" (Zimmerman 2008. 14). Spelman's response to this case is to bite the bullet; if Hitler believed that his commanding genocide was morally best, then Hitler did not violate a moral obligation by commanding genocide. However,

Spelman claims that there are other factors that mitigate the "grotesque" nature of this claim. One consideration is that under these assumptions, while we cannot claim that Hitler did something wrong in commanding genocide, it is possible that Hitler did something wrong in forming his belief that commanding genocide was best (Spelman, 82). For this to be the case, subjectivism requires that Hitler have certain beliefs (such as the belief that we should be careful in how we form our beliefs) which in turn would give Hitler an obligation for the responsible uptake of beliefs, and we could rightly criticize Hitler for violating that obligation. Spelman does not see this as a problem; he suggests that there is already widespread agreement that we should be careful when forming our beliefs. Still, there is some small possibility that Hitler's beliefs were arranged in such a way as never to put him in violation of a moral obligation, but Spelman believes that this would require so many layers of mitigation and unlikely scenarios that the conclusion that Hitler did nothing wrong would not be so unacceptable (Spelman, 83).

Spelman and I have opposing intuitions about whether these many layers of mitigation successfully render the conclusion acceptably not-grotesque. However, Spelman does make a point that I agree with, which is that, in extremely rare cases, when assumptions grind against many of our normal intuitions, our intuitions about the conclusions of such cases are less reliable (Spelman, 84). As it stands, I have two serious problems with Spelman's response to this objection. The first is that one need not have intuitions about particular rare cases to hold the belief that "What Hitler did in commanding genocide was wrong, period, and it is impossible that what he did wasn't wrong." When considering the plausibility of that belief against belief in subjectivism, from which belief it would be better to start moral theorizing seems an open question. Philosophers who favor a top-down approach to moral theorizing, seeking to recreate our first-order intuitions like "Genocide is categorically wrong," would see this as providing reason to reject subjectivism, rather than reason to revise our intuitions in rare cases. The second problem that I have concerns an implication of subjectivism to the Hitler case that Spelman curiously omitted. Spelman accurately notes the negative statement that in these special cases, we cannot claim that Hitler did anything morally wrong. What Spelman omits is that in the cases where Hitler believes that commanding genocide is morally best, Hitler does something positively morally right in commanding genocide. The possibility of this conclusion seems undeniably grotesque compared to the possibility of the negative statement.

The second response is to the objection that subjectivism entails that it is a trivial capability of agents to be morally infallible. It seems that we make moral mistakes all the time, so subjectivism must be mistaken. Spelman's response in defense of subjectivism is that "At times, our moral beliefs are dispositional rather than occurrent" (Spelman, 86). In these cases, given that we do not immediately know what our moral beliefs are, it does not follow that we would be able to trivially fulfill them. Furthermore, many of our intuitions about our moral mistakes stem not from the fact that we are violating our moral obligations, but instead from the fact that we are aiming to know and achieve what would be morally best, independently of our beliefs and obligations. As such, moral deliberation, solicitations of moral advice, and our intuitions that we make moral mistakes can be explained away. This defense seems to bite the bullet a little too strongly; while Spelman has provided several mitigating reasons to blunt the conclusion, it still seems quite easy to become morally infallible regarding the fulfillment of one's moral obligations.

There is also a problem with the fact that our moral beliefs are not always so clear-cut; if an agent's moral beliefs in a situation could be only dispositional, then there is a problem regarding whether those beliefs are operative in the agent's decision making. If they are, then that requires some story of how those beliefs impact the decision-making process without becoming occurrent. If they are not, then subjectivism loses an advantage that it has over objectivism and prospectivism because an agent's moral obligations are once again dependent on something that plays no direct role on their decision making. Furthermore, subjectivism might have undesirable commitments in the philosophy of belief: if one is a belief-eliminativist and a credence-realist, it is unclear how subjectivism can account for this. Credence, as a fine-grained attitude, either cannot account for the fact that we either have moral obligations or we do not, or must account for this fact with line-drawing vagueness at some credial threshold (Jackson 2020, 1). Given these outstanding objections, subjectivism also seems to be an unsatisfactory general theory of obligation.

4: CONSTRUCTING DOXASTICISM

Spelman's position is akin to David Lewis's position in *On the Plurality of Worlds*. Both saw problems with an assortment of views in the field and adopted

a theory with great costs to solve those problems. If there is a difference here, however, it is that unlike some of Lewis's opponents, we *can* have "paradise on the cheap:" a theory that solves our problems with objectivism and prospectivism without committing to the great costs of subjectivism (Lewis 1986, 136). Enter, doxasticism: the view that an agent's moral obligations depend on their non-moral beliefs. Like objectivism and prospectivism, doxasticism is a non-substantive moral theory: in a given case, it requires a substantive moral principle to determine exactly what an agent's moral obligations are. In Jackson's example with Jill, a plausible doxasticist principle would be "Jill is morally obligated to do what the most efficacious treatment for John's condition is according to her beliefs." Unlike subjectivism, which would require Jill to have a moral opinion in order for her to have a moral obligation to prescribe drug A, doxasticism can accommodate our intuitions in this example with only Jill's belief that drug A is the most efficacious drug.

The two Drug Examples that I used as motivation against objectivism and prospectivism share a common quality: each hinge upon the agent in question lacking crucial information, which leads us to the conclusion that their moral obligation cannot stem from that information, whether it be information about the hard facts of the case or information about what their evidence entails. I believe that the intuitions which led us to those conclusions can be explained by deeper principles of normative ethics. As such, my aim is not to further develop a top-down account of doxasticism by presenting more cases and finding a theory that fits our intuitions. Rather, my goal is to develop a bottom-up account of doxasticism by constructing the mild claim that moral obligations depend on non-moral beliefs from two plausible and popular principles in normative ethics: that ought-implies-can, and that moral obligations must be able to guide belief. Some consequences follow from this method. It may be that the principles I use are false and thus my construction fails, or that I could make a stronger claim by appealing to more or different principles. I do not believe it necessary to further address these possibilities here.

Ought-implies-can is not one single principle, but rather a family of principles, not all of which even relate to morality (King 2019, 8).[3] While there are many senses of "ought," the relevant one for the doxasticist is the sense of "ought" that pertains to an agent having a moral obligation. As for the other term in the

3. See Motiz (2012) and Wedgewood (2013) for non-moral conceptions of ought-implies-can.

implication, "can" is generally understood to be a modal notion; in some sense, a subject S can do action A if it is possible for S to do A. Typically, this form of possibility has been characterized as nomological possibility; in Kant's formulation, the principle is stated as "The action to which the "ought" applies must indeed be possible under natural conditions," where possibility under natural conditions is definitionally equivalent to nomological possibility, or possibility given the laws of nature (Kant 2007, A548/B576). Nomological possibility is not a quality restricted to actions; while the objects of moral obligations are actions, in statements like "It is possible that it will rain tomorrow," nomological possibility can be applied to non-action states of affairs or propositions.

Nomological possibility, as far as kinds of possibility go, is quite narrow, especially when contrasted with logical or metaphysical possibility. However, when we discuss actions, it seems that there is an even narrower form of possibility to which we can appeal. Recall the previous example of the drowning child outside of my house that I am unaware of. Earlier, I alluded to some sense of possibility according to which it is not possible for me to save the drowning child. While that sense of possibility is not available for the objectivist, it is available for the doxasticist. This form of possibility is what I will call psychological possibility.

If nomological possibility is the possibility of a state of affairs given the laws of nature, we can characterize psychological possibility as the possibility of actions given the laws of *human* nature; or in this case, the laws of psychology. This is a mere characterization though; I do not intend to claim that there is such a thing as human nature. Rather, I wish to appeal to certain facts concerning the psychology and philosophy of action that are relevant to our discussion of moral obligations. First, we must understand the things we are morally obligated to do. We are morally obligated to do actions, but what exactly are those? One direction in which we can search for answers lies in the philosophy of language within speech act theory. In rebuffing a hypothetical person skeptical of the claim that linguistic communication involves acts, John Searle makes the following case:

> ...when he takes a noise or a mark on paper to be an instance of linguistic communication, as a message, one of the things that is involved in his so taking that noise or mark is that he should regard it as having been produced by a being with certain intentions. He cannot just regard it as a natural phenomenon, like a stone, a waterfall, or a tree. In order to regard it as an instance of linguistic

> communication one must suppose that its production is what I am calling a speech act. (Searle 2013, 222)

In this passage, Searle makes a critical observation: what makes an action an action is not simply a physical property, but the logical presupposition of some intention behind the action. There is, admittedly, a wider sense of actions in which unintentional acts could be called actions. We could say that sleepwalking is an action, or that our involuntary heart beating is an action of the heart, or that we can anthropomorphize non-living things as taking actions, like a volcano erupting. However, this wider sense of action is not relevant for a discussion of moral obligation; these unintentional or non-living actions are simply not the kinds of things to which we ascribe moral qualities. As far as we are presently concerned, the actions we are interested in are intentional actions.

Once we grant that the objects of moral obligations are intentional actions, it follows that these actions require particular beliefs. In the earlier drowning child case, it was not psychologically possible for me to save the child because I did not have the requisite kind of beliefs to form the intention to act to save the child; namely, I was completely unaware of the child's existence at all. Therefore, I had no moral obligation to save the child- because my having an obligation to save the child would have entailed that it was psychologically possible to save the child, which would have required that I have the requisite beliefs such that I could have formed the intention to save the child. As such, there is an *awareness* condition on an agent's beliefs for their having a moral obligation: the agent must have certain descriptive beliefs such that it is possible for them to form the intention to perform the action of the obligation. We may formalize psychological possibility as the following:

> *Psychological Possibility*
>
> An action is psychologically possible if and only if it is nomologically possible for the agent to form the intention to perform the action.

Given this definition, the principle of "ought implies can" is the condition that for an agent to be morally obligated to perform an action, it must be psychologically possible for the agent to perform that action.

This condition alone is insufficient for explaining our intuitions in the Drug Examples. Given what I have stipulated, it is still psychologically possible for Jill to prescribe drug C and for Harry to prescribe drug G; it might appear bizarre or irrational to an observer, but there is no lack of awareness of the possibility of prescribing either of those drugs. Rather, what Jill and Harry lack are the beliefs about which drugs would be most effective (or in Harry's case, what the evidence entails about each of the drugs). The next principle to which I will appeal to explain our intuitions in the Drug Examples is that moral obligations must be able to guide action. How I will understand this principle is that for moral obligations to be able to guide action, when an agent has a moral obligation, some aspect of the obligation must provide a basis that is available to the agent and that upon reflection provides reasons to act. What could this basis be? With a modification to the first Drug Example, we can rule out the fact of the obligation itself serving as that basis.

Forgetful Jill

> While Jill is deliberating over which drug to prescribe John, she suddenly remembers a past case like John's, and recalls that she justifiably concluded that the correct drug was drug C. However, upon this realization, Jill consults the available evidence once again, and still cannot find any evidence that would inform her of which drug between B and C cured or killed; she seemingly cannot find the evidence she remembers having previously, nor does she remember what the evidence exactly is. What should Jill do?

In this example, past-Jill is a stand-in for the objectivist case, as Jill is still unaware of which drug is the completely curing drug. Even if Jill is right, and she accurately remembers prescribing drug C previously, to prescribe drug C in this case when she would have no evidential basis for prescribing C over B would still be a great risk to John's life. Without knowing why she would be obligated to prescribe drug C, above and beyond knowing the fact that she could be obligated to prescribe drug C, Jill's presumed obligation to prescribe drug C fails to provide her with adequate guidance, and therefore must not be her actual obligation at all. Furthermore, if the fact of the obligation must be available as a reflective basis

whenever an agent has a moral obligation, then it seemingly becomes impossible not to be aware of one's moral obligations. However, we frequently do not know what our moral obligations are, so this must be incorrect.

My proposal here is that moral obligations must be able to guide action in the sense that the agent has some beliefs in alignment with the explanation of the moral obligation; in other words, the agent must have some ability to know why they have that obligation. Daniel Fogal and Olle Risberg provide a recipe for explanations of particular moral facts:

> *Particular descriptive explanans*: particular natural fact(s) (e.g. *a* is a lie).
>
> *General moral explanans*: general explanatory moral principle (e.g. lying is wrong).
>
> ---
>
> *Particular moral fact*: particular moral fact (e.g. *a* is wrong). (Fogal and Risberg 2020, 175)

This tripartite model of moral explanations accounts for the supervenience of moral properties on natural properties and applies to talk of moral obligations, as the property of having a moral obligation is a moral property. It works through the application of general moral principles to particular natural facts, and the two function together to explain a particular moral fact. The primary goal of moral explanations is to explain *why* particular moral facts obtain, as opposed to merely stating that they do. If there is no necessary link between an agent having a moral obligation and an agent having some awareness of why they would have such an obligation, then all fulfillment of moral obligation would essentially be a matter of luck. In the above case, the explanation for why an agent would be morally obligated to save the child follows the same recipe:

> *Particular descriptive explanans*: a child is drowning.
>
> *General moral explanans*: agents have moral obligations not to let children drown.

Particular moral fact: I have a moral obligation not to let the child drown.

However, the fact that I do not know that the child is drowning is a matter of luck. If I am indeed still obligated to save the drowning child, there is no quality or quantity of moral reasoning that will allow me to fulfill my obligation, as I lack the beliefs that would allow the relevant moral principle (that we ought not let children drown) to fit my descriptive understanding of the situation. Thus, we arrive at the undesirable conclusion: if access to the moral explanation is not necessary for an agent to have a moral obligation, then it seems that there is no function for moral reasoning or justification doing the right thing.

So far, I have shown that at least some awareness of the moral explanation for a particular moral obligation is necessary for an agent to have that obligation. In each of our examples so far, the awareness that the agent lacks is the descriptive awareness, rather than the general moral principle. Jill presumably possesses the awareness that what would make her action right is acting for John's well-being, and I certainly know that saving drowning children is right. The reason our examples are all instances of descriptive ignorance rather than moral ignorance is because the requisite awareness of the moral explanation required for having a moral obligation is the descriptive rather than the moral explanans. To illustrate, consider the following case in which the agent has the appropriate descriptive beliefs but lacks the moral beliefs.

Egoist Earl

> Earl lives in a world where the correct moral right-maker is the principle of utility. After reading "Famine, Affluence, and Morality" by Peter Singer in Earl's moral philosophy class, Earl learns of the descriptive facts regarding human suffering around the world. However, Earl's current moral beliefs dispose him towards an egoist view of right-making, and as such, Earl does not donate to any effective altruism causes. (Singer 1972)

What are we to make of Earl's moral obligations? Given that we have stipulated that subjectivism is not the correct right-maker in Earl's world, it seems that any observer in Earl's world would be warranted in blaming Earl for violating his moral

obligation. Earl's moral obligation only required him to have certain descriptive beliefs for the general right-making principle of utility to apply. The case could be changed such that Earl never did learn of the descriptive facts surrounding suffering in the world, but note that in that case, we would not say that Earl has a moral obligation to donate to an effective altruism cause, just as we did not say that I had a moral obligation to save the drowning child of which I was unaware.

Given the scope of our arguments from psychological possibility, we may claim that descriptive beliefs can affect what moral obligations we have, but we cannot make the parallel claim for moral beliefs. This results from the fact that descriptive beliefs carry modal force that bears on psychological possibility, while moral beliefs do not. Remember that psychological possibility requires the nomological possibility of the formation of an intention for an action. What intentions it is nomologically possible for us to form is a matter of our descriptive beliefs rather than our moral beliefs, and as such, it is the descriptive component of the moral explanation that an agent must be aware of rather than the moral component. On the one hand, if I believe that there is a wall in front of me, and I have auxiliary beliefs about the physical world such that I do not believe that I can walk through physical objects, then it seems nomologically impossible for my beliefs cause the intention to walk through the wall. On the other hand, if I believe that a course of action is merely wrong rather than physically impossible, it is much more conceivable that it is nomologically possible to intend that action anyways, regardless of the fact that I likely won't. Moral beliefs simply lack the modal force on psychological possibility that descriptive beliefs possess; perhaps if there was an instance of some anomalous person who was causally determined to always form intentions in accordance with what they believed to be morally right, this would need revision, but no such counterexample seems to exist. Furthermore, such an example would make moral beliefs relevant qua their arbitrary specification in some anomalous causal principle, rather than qua being moral beliefs.

In addition to our earlier awareness condition on an agent's beliefs for their having a moral obligation, we now have an additional guidance condition. For an agent to have a particular moral obligation, the agent must have beliefs that would fit the descriptive explanans that partially explains why they have that obligation. This is the dependence that I claim doxasticism, the view that moral obligations depend on our non-moral beliefs, captures. What substantive principle is true, if any, is a much larger question for which the bulk of normative ethics is responsible.

However, the agent need not possess the beliefs about the moral principle in order to have a particular moral obligation that arises from it. Instead, the agent only must possess the descriptive beliefs so that the principle can apply and explain the agent's moral obligation. This view is highly advantageous as it captures the informational dependence of obligations that subjectivism accounts for, while not making the further claim that morality is a matter of an agent's moral opinion. Restricting the dependence of moral obligations to non-moral beliefs side-steps the kind of relativization that makes subjectivism unappealing while preserving our original intuitions that led us to reject objectivism and prospectivism. In this way, doxasticism is paradise on the cheap.

5: OBJECTION FROM ROBUST REALISM

Classical theories of moral subjectivism are often excluded from the label of "moral realism," despite the fact that under subjectivism, there are still moral statements that are truth-apt and indeed, true. The belief-dependence of moral facts under many subjectivist theories may allow them to be considered "realist," but only in a weak sense. This weak sense of realism is contrasted with what David Enoch calls "robust realism," which is his label for metaethical non-naturalist views that see moral truth as "perfectly objective, universal, absolute" (Enoch 2011, 1). Given the great amount of progress recently made in advancing metaethical non-naturalism, it would be a serious concern if the belief-dependence of moral facts under doxasticism was incompatible with robust moral realism.

Luckily, Enoch provides a potential solution for the compatibility of doxasticism and robust realism. For many moral realists, only a certain kind of mind-dependence is problematic (Hanson 2018, 47). Enoch provides a criterion of normative relevance for determining which kinds of mind-dependence are acceptable: a normative truth is acceptably mind-dependent if it is only mind-dependent to the extent that the mind-dependence is normatively relevant (Enoch, 3-4). Furthermore, the normative truths must themselves not be constitutively mind-dependent, but rather must be mind-dependent according to deeper normative truths that themselves are not mind-dependent. Doxasticism fits this criterion: it specifies non-moral beliefs as normatively relevant, and the truth of doxasticism itself would not be mind-dependent if robust moral realism is true.[4]

4. There may be views of "sophisticated subjectivism" that can be realist in Enoch's strong, robust

The particularities of Enoch's theory aside, this resolves a tension between belief-dependent theories of obligation and the desire for a realist theory that is also *objective*. Typically, if an object of theorizing is belief-dependent, we are inclined to categorize it as subjective, but the criterion of normative relevance states that if the belief-dependence itself is just a specification of another objective principle (like doxasticism), then the belief-dependence is of no problem for an objective and realist theory.

6: OBJECTION FROM VERIDICALITY

The form of dependence doxasticism claims to hold between an agent's moral obligations and their non-moral beliefs is most easily explained by the supervenience relation generally taken to hold between the moral and natural facts.

Doxasticist Supervenience

> For every property of having a moral obligation *M*, if an agent is *M*, then that agent has the property of having a set of relevant non-moral beliefs *N* such that necessarily, every agent that is *N* is *M*.

One challenge that arises from this principle is what I will call the Veridicality Challenge. The Veridicality Challenge stems from the fact that under the principle of doxastic supervenience, the hard facts of the world are entirely trimmed out of the supervenience base. As such, none of the agent's beliefs need be veridical, but there seem to be cases where obligations that stem from non-veridical beliefs will result in morally undesirable outcomes. To illustrate, consider the case of hallucinatory harm.

Hallucinatory Harm

> Clare and Dale are in two different worlds and share many natural properties. In fact, they share every natural property concerning

sense. For example, Spelman's theory could be one such theory, if the truth of subjectivism itself is not mind-dependent.

> the content of their relevant non-moral beliefs. However, Clare and Dale differ in one crucial respect. While most of Clare's non-moral beliefs are veridical, Dale is massively hallucinating; while he believes that he is in the exact same physical situation in his world that Clare believes she is in her world, Dale's differs greatly. Both believe that they see a drowning child, and go to save them, making identical physical motions. Clare successfully saves the child by pulling them forcefully by the arm up and out of the pond, while Dale, making the same motions, harms an innocent bystander relaxing on the ground.

According to the Veridicality Challenge, it seems obvious that Dale and Clare do not share the same moral obligations, and because Dale and Clare share every property concerning the content of their non-moral beliefs, doxastic supervenience must be wrong. I will present and respond to two ways that this denial can be justified. The first I will call the Denial from Moral Competence, and the second I will call the Denial from Impossibility.

According to the Denial from Moral Incompetence, Clare and Dale cannot share the supervenient obligations because Dale is not a morally competent agent. For a person to have a moral obligation, there must be a deeper constitutive sense in which they are obligation-apt, and Dale does not meet whatever constitutive criterion this is. This is further supported by the fact that Dale is not morally responsible and blameworthy for the harm that he causes.

We can begin with an investigation of the relationship between moral responsibility and being obligation-apt. Gary Watson distinguishes between two "faces" of moral responsibility: it has one face of attributability, and one face of accountability (Watson 1996). On the attributability face of responsibility, we merely attribute actions to moral agents insofar as their action is the product of their moral agency. On the accountability face, we adopt a valenced attitude and hold agents responsible as blameworthy or praiseworthy.[5] It is true that Dale is not morally responsible for the harm he causes in the sense that we would not hold him accountable or blameworthy. Dale is also not morally responsible for the harm in the attributive sense; we would not attribute Dale's action to his moral qualities. While Dale is delusional, his delusions are merely sensory delusions; Dale's moral

5. See McKenna (2012, 16-17) for a description of neutrally valenced attributive responsibility.

faculties are as sharp as anyone else's. In this way, Dale differs from paradigm cases of moral incompetence, such as Susan Wolf's case of JoJo, a boy whose moral education is corrupted by his dictator parent (Wolf 1987, 53-54). JoJo is a paradigm case of the exculpation of moral responsibility from moral incompetence precisely because his moral faculties are compromised. No such parallel can be drawn from Dale. While Dale, in both senses of moral responsibility, is not responsible for the harm he causes, Dale still is morally responsible for *something*. Dale still tried to save an apparently drowning child, and he is responsible for that attempt; in fact, he may even be praiseworthy to some small extent. We can further specify which sense of responsibility his attempt arises from: Dale's attempt to save a drowning child is the product of his moral agency, and it is this constitutive dimension of moral responsibility that is required for obligation-aptness, so the Denial from Moral Incompetence fails.

According to the Denial from Impossibility, Clare and Dale cannot share the supervenient obligation because Dale cannot possess it, for the reason that there is no drowning child for Dale to be obligated towards. This is no problem for doxasticism; while Clare and Dale do not have obligations identical in referent, they do have obligations identical in content, which is all that is required to have indistinguishable supervenient moral properties. Clare and Dale's obligations are not to the drowning child directly, but rather to the content of their non-moral beliefs that identify an object of obligation. In Clare's case of veridical belief, the object in the obligation can then be further identified with the drowning child, and in Dale's case, the obligation can then be further identified with his hallucination of a child, but there is no reason that these further identifications must be contained within their initial moral properties of having a particular moral obligation.

The Denial from Impossibility encounters a different problem as it commits us to skepticism about our moral obligations. If whether we have a moral obligation depends on there being a world outside of ourselves such that our obligations are to the beings in that world, then our knowledge of our obligations would require us to know the world to which we are obligated. However, this requires defeating traditional skeptical hypotheses such as the possibility that we are brains in vats, dreaming, being deceived by an evil demon, or in this case, merely severely hallucinating. Accepting doxastic supervenience thereby comes with an advantage when compared to views that face this problem; if moral obligations

supervene merely on non-moral beliefs, then even if we are, say, brains in vats, we can still have moral obligations, and we can still come to know what they are.

CONCLUSION

What have I argued in this paper? The most important point is that the existing debate over what moral obligations depend on is seriously flawed, because it is missing the view that I take to be right: doxasticism, which occupies the sweet spot of maintaining belief-dependence while not encountering the substantial objections to subjectivism, a quality that each other view in the field lacks. In some sense, doxasticism is simply a lengthy development of the idea that you can only ever act on the information that you have and cannot be obligated to act on the information that you do not.

REFERENCES

Enoch, David. 2011. *Taking Morality Seriously: A Defense of Robust Realism.* Oxford: Oxford University Press.

Fogal, Daniel, and Olle Risberg. 2020. "The Metaphysics of Moral Explanations." In *Oxford Studies in Metaethics: Volume 15*, edited by Russ Shafer-Landau, 170-194. Oxford: Oxford University Press.

Hanson, Louise. 2018. "Moral Realism, Aesthetic Realism, and the Asymmetry Claim." *Ethics* 129 (1): 39-69.

Jackson, Elizabeth. 2020. "The relationship between belief and credence." *Philosophy Compass* 15 (6): 1-13.

Jackson, Frank. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101, no. 3: 461-482.

Kant, Immanuel. 2007. *Critique of Pure Reason*. London: Palgrave Macmillan.

King, Alex. 2019. *What We Ought and What We Can*. New York: Routledge.

Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Wiley-Blackwell.

McKenna, Michael. 2012. *Conversation and Responsibility*. New York: Oxford University Press.

Motiz, Mizrahi. 2012. "Does 'Ought' Imply 'Can' from an Epistemic Point of View?" *Philosophia* 40 (4): 829-840.

Searle, John. 2013. "What is a Speech Act?" In *Philosophy in America*, edited by Max Black, 221-239. New York: Routledge.

Singer, Peter. 1972. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1 (3): 229-243.

Smith, Holly. 2010. "The 'Prospective View' of Obligation." *Journal of Ethics and Social Philosophy* 1: 1-9.

Spelman, Jonathan. 2017. "Moral Obligation, Evidence, and Belief." PhD diss., University of Colorado Boulder.

Watson, Gary. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24 (2): 227-248.

Wedgewood, Ralph. 2013. "Rational 'Ought' Implies 'Can.'" *Philosophical Issues* 23: 70-92.

Wolf, Susan. 1987. "Sanity and the Metaphysics of Responsibility" In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand Schoeman, 46-62. Cambridge: Cambridge University Press.

Zimmerman, Michael. 2008. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press.

Zimmerman, Michael. 2014. *Ignorance and Moral Obligation*. Oxford: Oxford University Press.

compos mentis

Undue Burden: How White Feminism Misdirected the Politics of Reproductive Healthcare and Failed the People Who Need It Most

Carlyn Kirk
University of South Carolina

ABSTRACT
White women have historically dominated the abortion movement. Both contemporary "pro-choice" and "pro-life" partisan sides have neglected the concerns of minority women in the conversation. In this paper I examine this fact and seek to show that minority women are fundamentally crucial to any conversation we should be having regarding abortion. Because much of our contemporary understanding of abortion has been derived from the concerns of white women, we have excluded the needs and desires of minority women nearly altogether. To compound this problem, we have already seen the detrimental effects of stricter legislation that hinders access to reproductive healthcare services; minority women bear an unequal burden from strict legislation, they are simply more likely to face adverse consequences from such restrictions than white women. In this paper I use feminist epistemology to illustrate how crucial it is to incorporate the voices of women at the margins. In the wake of the impending Supreme Court decision on Dobbs v. Jackson Women's Health Organization (2021), this paper seeks to analyze just how little has been done to address the party most privy to the topic- minority women.

KEYWORDS
Standpoint Theory, Intersectionality Theory, Abortion, Minority Women, Feminism, Epistemic Injustice

There is no time more pressing to consider the way people are discussing reproductive healthcare, and the magnitude of the effects of abortion legislation, than today- this paper comes at a timely moment in United States history, awaiting the impending Supreme Court decision on Dobbs v. Jackson Women's Health Organization (2021), there likely will be a change in the way the Supreme Court treats abortion services. Numerous states have begun to write laws directly countering current precedent on abortion services and are attempting to halt access to abortion long before viability, even as early as four weeks; politicians have aggressively sought to pass anti-sexual and reproductive healthcare policies under the guise of a "pro-life" agenda. Likewise, the change in policies have been coupled with a change in judicial leadership. With new judiciaries come new judicial preferences, many of which have publicly opposed access to abortion services, signaling a new age not just for access to reproductive healthcare, but for women's rights.

Today, abortion has melded into the atmosphere of politics; conversations remain centered around the mainstream, partisan "pro-choice" or "pro-life" sides. Abortion, which was at one time considered a conversation that remained between the discretion of a woman and her healthcare provider, has transformed into conversations of ethics, privacy, autonomy, and judicial overreach. Nearly 30 years ago, the Supreme Court of the United States issued the landmark decision on Roe v. Wade (1973), which fundamentally transformed our conception of abortion; subtle but quite profound differences in the way abortion was discussed began to take hold in the wake of Roe. At the same time, the pro-choice movement, consisting primarily of white women, began to establish their prominence as the head of reproductive rights activism. White liberal women have now become the face of the battle for reproductive rights, ushering in a landslide of Women's Marches, and iconic cultural symbols such as the "pink pussy hat" and period art; and while these women have effectively cast light on misogynist healthcare policies, the movement has unintentionally pushed the narrative of the cisgender white woman to prominence. This narrative does little to shine light on the systemic discrimination that minority women face in reproductive healthcare, misrepresenting or altogether neglecting the needs or experiences of such women; yet it is the forgotten women who face the biggest consequences from legislation hindering access to abortion services.

As it stands today, most conversations regarding abortion fail to adequately address these people- the legislators, judiciaries, pro-choice, and pro-life groups have neglected the women most affected, most in-need, and most worthy of attention in this conversation. It is imperative, in the wake of current socio-political activity, that this conversation shifts to incorporate minority groups; an intersectional approach helps us to understand the interlocking effects that identity, oppression, and privileges have, and in turn, why these are foundational to *any* conversation about abortion. An analysis of judicial decisions and abortion conversations dating back to the 1930s showcases how white feminism has fundamentally shaped the way the United States frames abortion (Green 2016). Given the likelihood that the Supreme Court will soon overturn Roe vs. Wade (1973), it is imperative we begin a public discussion about abortion that utilizes an intersectional approach. This discussion should highlight the needs of minority women who have been historically left out of these conversations, and who are most likely to be affected by changes to reproductive healthcare policies.

In this paper, I begin this public intersectional discussion of abortion by framing the reproductive healthcare discussion in the United States as a narrative told mostly from the standpoint of white women. Social and political problems that are framed almost exclusively from the perspective of white women can be epistemically problematic, especially when it results in overly burdensome legislation. Firstly, the problems are likely to be framed in such a way as to exclude issues not important or pressing to white women, specifically excluding problems that may predominately affect Black and Brown women. Secondly, the solutions to these problems are then likely to reflect the problems in which they are responding to, further neglecting Black and Brown women's concerns from the discussion. Although it may not be surprising that the abortion narrative is told primarily from the standpoint of white women, there is a sad irony to it. While white women remain in the forefront of the conversation, it is Black and Brown women who will ultimately bear the unequal burden of restrictions to abortion rights. A discussion of intersectionality theory and Black women's standpoint theory emphasizes the importance of theorizing from and with women at the margins. I then provide evidence from recent states where abortion rights are being overturned or eroded—it is evident in these states that Black and Brown women do bear a disproportionate burden from these restrictions. When we combine a dominate narrative told from the standpoint of white women with the way in which abortion

restrictions disproportionately affect Black and Brown women, it is evident that we are likely to miss much of what is at stake. With this suspicion in mind, I conclude this thesis with a discussion of contemporary news regarding reproductive justice that shows us that significant change is still necessary to these conversations.

I. HISTORY OF ABORTION ADVOCACY

Pro-abortion advocacy most notably began with doctors in the 1930s, forty years before we see the landmark case of Roe v. Wade (1973) take center stage. At the time, pro-abortion and birth-control advocacy was stained by a "dark racial component" (Green 2016) due to its ties to the growing eugenics movement; abortion and birth control were seen as means to controlling Black and Brown populations' ability to reproduce, and many proponents of abortion sought to use abortion for this end. Many advocates valued both birth control and abortion for their ability to perpetuate their own racist agenda. The early 20th century saw a rise in support for the use of birth control and abortion, but it was far less about women's rights and more about perpetuating the eugenics movement in the United States. This movement specifically sought to limit the reproductive abilities of those seen unfit; those who were "unfit" included Black, poor, or incapacitated minority women (Green 2016). Pioneers of the reproductive healthcare movement were often proponents of the eugenicist movement; Margaret Sanger, the founded of Planned Parenthood, was herself a well-known advocate of the eugenics movement. While abortion and birth control have now become a beacon of hope for many women (of all races, incomes, and other distinguishing factors) who are in need of reproductive healthcare services, there is no denying the harmful impacts it once had on minority women.

Fast-forward thirty years, and abortion began to be legalized in several states, beginning with California. This said, the mainstream pro-life and pro-choice movements, with which we characterize our contemporary abortion politics, had yet to be formed. Much of the politics of birth control and abortion at the time had been molded by the past desires of the eugenics movement. Even as the pro-abortion side moved starkly away from eugenics as a motive and more toward protecting women's health and autonomy, it remained deeply stained by the racist underpinnings of the movement's origins. Furthermore, the pro-abortion side still left minority women out of the conversation; the concerns of the pro-abortion side

did little to address the same concerns of Black and Brown women. There were many debates at the time that connected questions of abortion and birth control to questions about gender and sexuality (Green 2016), neither of which were dominant concerns for Black and Brown communities at the time. These minority women were far more concerned with the effects of abortion and birth control and its role in controlling their fertility- what they truly desired was the ability to determine *for themselves* when to have children and how many to have, as well as ensuring that they had adequate resources to raise them. In other words, they were less concerned with abstract questions, but rather sought concrete rights that ensured bodily autonomy as well as the ability to legislate for themselves matters of reproduction without hinderance. Minority women sought something that ran far deeper than simply the right to abortion, or to "own their sexuality" so-to-speak, they wanted justice.

At the same time, abortion services and birth control were deeply opposed by the Catholic Church. The Catholic Church believed that life begins at conception and sought to enforce bans on what they considered to be deeply immoral activity- birth control and abortion were seen to be dangerous and wrong activities. The Catholic Church opposed abortion, *not* because they simply valued fetal life over a woman's right to choose, but rather because they believed they were protecting women as well as supporting them and their families. In fact, the Church also sought to expand pre-natal healthcare policies to provide more insurance for women, aid for poor women, and better adoption practices in order to support the women who were more likely to be targeted by abortion and birth control advocates (Green 2016). Despite this, the Catholic Church also alienated many potential anti-abortion advocates because they did little to acknowledge or consider the needs and desires of minority women either (Green 2016). On both sides of the debate, the concerns of minority women were not discussed, even despite their seemingly dominant needs in the conversation. While reproductive healthcare advocacy had deeply racist roots for its participatory nature in the United States' eugenics movement, the predominantly white, Catholic anti-abortion movement's dismissal of Black and Brown women's concerns alienated these women; the Church did not address their needs despite the likely alignment of their goals at the time.

This history of abortion in the United States and the way we discuss it was fundamentally changed by the Supreme Court's (7-2) landmark decision on Roe

v. Wade (1973). In this decision, the Supreme Court ruled the due process clause in the 14th amendment guarantees one the fundamental right to privacy. It is in this "right to privacy" that lies the woman's right to choose whether or not to receive an abortion. Nonetheless, the right to privacy is limited by the state's legitimate interests in protecting the health of pregnant women and the so-called "potentiality of human life." The relative weight of the woman's right to privacy in contrast to the state's interest in protecting the potentiality of life varies over the course of the pregnancy. The law reflects favorability towards privacy rights early in the pregnancy, but transitions to favor the interest of the state as the pregnancy carries on. In other words, the court ruled that the state may not regulate abortion in the first trimester, in the second trimester the state may regulate it in a manner that reflects "reasonable" restrictions to protect maternal health, and in the third trimester the state has the authority to fully restrict abortion due to the "viability" of the fetus, although exceptions can be made in dire circumstances (Roe v. Wade, 1973).

This is a transformative period for abortion rights in the United States. Prior to this decision, privacy was not a part of the abortion discussion. Indeed, publications discussing abortion dating back to six years prior to Roe did not once mention privacy in relation to the right of abortion (Vecera 2014). The first noted publication that links the two concepts was released just one day after the court released their decision on Roe (Vecera 2014). This indicates a monumental shift in the framing of abortion because the conversation shifts from the health and safety of a medical procedure and/or concerns of eugenics and becomes focused on a constitutional right to privacy. As our social conception of abortion became radically changed by the release of the judicial decision for Roe v. Wade (1973), so did the way that mainstream liberal reproductive-rights activists framed the issue.

In the wake of Roe v. Wade (1973) the mainstream pro-choice and pro-life movements that we know today began to take hold. The transformation in the way we discussed abortion was monumental, these advocacy movements were centered more around discussions of sexuality, gender, and constitutional rights, rather than concerned with the health and safety of women. The pro-abortion and anti-abortion sides managed to solidify into the partisan "pro-choice" and "pro-life" sides we see take prominence today. On both sides of the debate, white women seemed to take center stage and as a result, their concerns did as well. As time passes, the white feminist narrative has taken over the pro-choice

movement. Today there has been an influx of media portrayals of the pro-choice side, the year 2020 saw an unprecedented number of documentaries detailing this- 12 in total (Mattei 2021). In each documentary filmmakers and advocates used white protagonists as the "stand-in for a 'universal' experience that renders legible the basics of abortion restriction to the widest possible audience, which is odd considering the cultural specificity of white women's woes" (Mattei 2021). Likewise, other contemporary means of advocating for reproductive healthcare fail to fully push forth concerns of minority women, failing to approach the issue with an intersectional approach. Efforts, such as the Women's March, have made a push to persuade people on the importance of diversity and intersectionality, but still seem to resoundingly resonate with white women; 75% of participants at Women's Marches on January 26, 2017 self-reported themselves as women and 70% reported themselves as white (Heaney 2019).

Current discussions about abortion do not adequately—if at all—discuss the pervasive effects of legislating minority women's bodies and how they will be directly affected to a much greater degree than that of non-minority women. The narrative that pushes for reproductive *rights* rather than *justice* has been told time and time again, and media reflects this: coverage of the pro-choice movement specifically targets white women with its messaging (as it has historically done so). The white woman in the abortion debate has been decidedly pushing the struggles of constitutional rights to privacy and abortion access without giving thought to the inequitable weight that legislation has on minority women- it has pushed the struggles of the young cis white woman seeking to terminate their pregnancy to the forefront of our attention. This narrative, while certainly important, has been overtold; "a single story carries so much undue weight that it has been contextualized by its overrepresentation" (Mattei 2021).

In the 90's, women of color, and in particular black women, began pushing for reproductive *justice* rather than reproductive *rights*. This difference in language speaks to a desire for right to have or not to have children, as well as to dismantling the systems that enable their systemic oppression. The pro-choice movement, in contrast, tends to speak to a desire for reproductive rights, focusing more on the constitutional right to privacy more than the fundamental difference in equality that exists beyond mere access to abortion. Reproductive justice speaks to a desire for true justice and equality, a type of justice that spans across gender, race, disability, and class inequalities. Reproductive justice, in this sense, would account for the

way in which legislation primarily does or does not affect certain minority groups-often having disproportionate burdens on various minority women. Today, the movement for reproductive justice still exists, but it is far overshadowed by the so-called "pro-choice" side. White feminism has pushed specific activist groups to the forefront of our attention, such as Planned Parenthood and NARAL Pro Choice America. Yet, it is important to consider that minority women have found the messaging of the pro-choice movement largely meaningless to them (Price 2011). The rhetoric simply does not include the concept of reproductive justice, rather it remains largely concerned with broad questions about privacy or constitutional rights. And while these may be pressing concerns to which we still need answers, they simply do not speak to the same concerns of minority women—many of whom are more concerned with their right to choose for themselves whether or not to have children, bodily autonomy, and the way that restrictive abortion legislation perpetuates systemic oppression.

The pre- and post-roe discussion foreshadowed the era of political and social discourse we see today; abortion has become a mainstream partisan issue, concerned with the constitutionality of a woman's right to choose, rather than with the direct and pervasive effects that legislation has on women's health, most notably in minority groups. The discourse has effectively stalled—our opinions have become so divided, so engrained, that we cannot seem to change our opinions. A primary factor for this delay is due to the way we choose to discuss reproductive healthcare: what we argue about, the terms we choose to employ, and the very people who are discussing it. By choosing to amplify only certain voices in the abortion conversation, we have allowed ourselves to miss much of what is at stake. It is likely that by seeking to re-contextualize the abortion conversation, i.e., incorporate standpoint and intersectional theories, and amplify the voices and concerns of the affected minority communities, we may be able to understand abortion in a new light. We may, in fact, be able to make better legislative decisions in regard to abortion when we understand the pervasive and direct effects that any legislation has on these minority communities. While the needs of white women are surely important to the conversation, they can blind us to the needs of others. We shouldn't focus on the constitutional right to privacy, what we *should* be concerned with is how narrowly we've considered this problem.

II. STANDPOINT THEORY

Feminist standpoint epistemology is crucial to reframing the abortion debate by showcasing *why* it is so important to contextualize, understand, talk to, and hear from minority women. Standpoint theory is a feminist epistemology that departs from the assumption that a knowledge can be amassed from a universal point of view. Rather standpoint theory emphasizes that knowledge is always produced from a particular vantage point- it is inherently perspectival. The idea is that *any* analysis is "only properly understood in the social contexts in which they arise, and in terms of the biases and prejudices those contexts generate" (Bowell n.d.). Standpoint theory effectively moves beyond any standard, established way of understanding the world through the lens of the privileged, rather, it seeks to understand the power relations that fundamentally shape the world in which we live in. To suggest that there can be any one true account of the world, or people's experiences in it, contrasts deeply with the reality of our day to day lives; our lives, our decisions, our rights are all shaped by our identity statuses, ways in which we fit in to and interact with our world, and how we are all ultimately acted upon due to such statuses.

It should be fundamental to any social analysis, that one must seek to understand the power relations that shape the struggle between the marginalized and the non-marginalized. Thus, in order to understand these dynamics, one must actually hear from and speak to the marginalized community; it is only those who have been marginalized who can truly understand the influences that shape this. In this sense, a woman has a far better understanding of the way being a woman can affect one's role in society because they must live in a "man's world." To the same end, Black and Brown people are imperative to understanding the interlocking effects that allow for systemic oppression and racism in the white man's world.

Feminist standpoint theory tells us that firstly "knowledge is socially situated," that marginalized groups are better able to understand the workings of a system that oppresses them, and ultimately, that *any* research concerned with power relations should begin with these minority groups (Bowell n.d.). The economically privileged, white activist has only a thin understanding of the workings of such a system- race, gender, and class work together to create a uniquely important perspective on the debate. This fact is absolutely crucial to framing the abortion discussion- white feminism has narrowly focused our discussions of reproductive

healthcare but failed to consider the very women who are *most* affected. A history of abortion discourse in America has shown how both sides of the debate alienated minority women, and how white feminism has taken prominence in the conversation. It is imperative to reframe this debate to include minority women, hear from these women, and consider these women in order to truly understand the *extent* to which legislation hindering access to reproductive healthcare can truly affect these groups.

Experience is tied to knowledge, and marginalized communities have richer, more complex experiences in the world because of the way they must interact with the dominating class. Women must live in a sort of "man's-world" reflected in the division of labor, the pay gap, social stigmas, and other such ways that gender oppression and forced compliance is reflected in our society. It is women who are deeply aware of these workings for they are the ones who must live their lives in accordance with the rules of the "man's-world" in which they were born into. The consequence of this is that the dominant class (men), only have a limited view of the world for they cannot inherently understand what it is like to live as a woman for they are simply *not oppressed* by fact of their gender. To extend this further, minority women also have a deeper understanding of the world that the white women simply cannot see- it is historically marginalized minority communities who are forced to comply with systemic discrimination in our society. In this way, white men or women simply cannot create an accurate framework of our world for they have an idealized vision of the world- they do not interact with or face the consequences of systemic discrimination. Minority women, in the context of the abortion debate, have been marginalized by abortion legislation and often left out of the conversation. It is deeply problematic that we have neglected minority women in the abortion conversation because it is these women who have deeper, more complex understandings of abortion as a complex issue. They are better able to contextualize the ways in which abortion access deeply affects the lives of women (especially of minority women) for they themselves bear the brunt of the effects of strict legislation.

In order to incorporate feminist standpoint theory into our abortion conversation, we need to shift the conversation away from the white feminist narrative, and empower, speak to, hear from, and understand the needs and desires of minority women. As a society, must ask ourselves which advocacy groups must we observe? Whose voices should we amplify? The standard pro-choice

side gives significant weight to advocacy groups such as Planned Parenthood and NARAL Pro Choice America- we have all heard from these groups, we understand their message, and have amplified their voices. We need to shift the conversation to include advocacy groups that include representation of minority women and their interests, namely groups such as the National Black Women's Reproductive Justice Agenda. Organizations such as this work in tandem with feminist standpoint theory for they allow minority women to speak on abortion *in their own voice.*

III. EMPIRICAL RESEARCH

Research shows that stricter legislation that hinders access to abortion disproportionately affects minority women. An intersectional analysis of the medical, economic, and legal impacts quickly shows just how burdensome this legislation is for minority women. Due to various socioeconomic factors as well as unequal access to health care, minority women are more likely to seek an abortion than white women: "African-American women are five times more likely than white women to obtain abortions; Latinas are twice as likely as white women to do so" (Cohen 2008). Furthermore, survey research shows that African American women are more likely to underreport abortions compared to White and Latina women, which further showcases the importance of race in predicting abortion behavior (Price 2011). The higher prevalence of African American and Latina women seeking abortions is likely due to a wide range of ethnic disparities that exist in access to health care and as a result in their health outcomes (Cohen 2008). These disparities are also a likely cause of unintended pregnancy rates across minority communities, as minority women are more likely to have difficulty finding "high-quality contraceptive services and in using their chosen method of birth control consistently and over long periods of time" (Cohen 2008). When we put these sexual health concerns into the context of other disproportionate health outcomes that minority women face, e.g. diabetes and heart disease to cervical cancer and sexually transmitted infections (Cohen 2008), it becomes clear that these disparities are a symptom of a much larger epidemic, namely the systemic discrimination that persists in our society.

Data also shows us that minority women are more populated in states that are likely to overturn access to abortion. It should come as no surprise that the states most likely to diminish or altogether outlaw access to abortion should

Roe v. Wade (1973) be overturned, are those that typically fall as "right-leaning" states, namely Texas, Mississippi, South Carolina, etc.… Conversely, states that are less likely to outlaw any access to abortion tend to swing left politically, namely California, New York, Washington, Oregon, and other such states. The states with the highest demographic populations of Black and Brown people as of 2021 as percent of the total state population are Louisiana, Mississippi, Alabama, Georgia, South Carolina, Maryland, and Delaware (U.S. Census Bureau 2022). These states are typically conservative in their voting patterns and thus more likely to adopt restrictive abortion laws, with the exception of Maryland and Delaware who at times vote more moderately. States such as Texas, which has already passed legislation limiting access to abortion via the Texas Heartbeat Act (2021), have large minority populations when compared to many left-leaning states. There is a direct link, due to basic population demographics, between restrictive access to abortion in these states and its effects on these minority communities.

Conservative states also typically limit proper education programs about reproduction. This is problematic because without proper understanding of "safe sex" and how to properly use birth control, unwanted pregnancies are far more likely to occur (Pettus and Willingham 2022). For example, Mississippi law dictates that sex education in public schools must emphasize abstinence in order to mitigate unwanted pregnancies and sexually transmitted diseases (Pettus and Willingham 2022), and Mississippi isn't the only state relying on this method. Other states that stress abstinence-only sexual education fall in line with many of the typically conservative states who house higher minority populations, namely Alabama, Georgia, Louisiana, South Carolina, and Delaware (University of Southern California 2017). Extensive research has shown us that abstinence-only education is largely ineffective- "they do not delay sexual initiation or reduce sexual risk behaviors" (Columbia University 2017). There is a significant correlation between the geographic location of minority communities and limited access to evidence-based sexual-education or abortion services. This data coincides, ultimately, with a higher rate of abortions or unplanned pregnancies and sexually transmitted diseases amongst minority women (Cohen 2008). Alabama does not require sex education in schools, and any sex-education provided must emphasize abstinence (Rice et al. 2018). Alabama also has one of the highest rates in the nation of unplanned pregnancies, gonorrhea, and chlamydia (Rice et al. 2018). Simply put,

people are more likely to have an unplanned pregnancy because the legislation that governs their area does not provide them proper reproductive education.

Minority women are more likely to be adversely affected by other sociologic factors, such as wealth and class structure disparities, which in turn can further limit their ability to seek an abortion should they live in an area where they cannot receive it legally. Wealth can often overcome legal hurdles, either though the procurement of safe, illegal abortions or by allowing women to travel to areas where abortions are legal. "All are equal before the law…" (University Declaration of Human Rights) but the law does not consider the fact that there are "long-standing and substantial wealth disparities between families in different racial and ethnic groups" (Bhutta, et al. 2020). As of 2019,

> White families have the highest level of both median and mean family wealth: $188,200 and $983,400, respectively. Black and Hispanic families have considerably less wealth than White families. Black families' median and mean wealth is less than 15 percent that of White families, at $24,100 and $142,500, respectively. Hispanic families' median and mean wealth is $36,100 and $165,500, respectively. (Bhutta, et al. 2020)

This data shows us that historically subjugated women tend to have considerably less capital than white women- this is crucial for understanding how they are more likely to be forced into unwanted pregnancies. Given the pending Supreme Court decision on Dobbs v. Jackson Women's Health Organization (2021), which may likely overturn Roe v. Wade (1973) and allow states to legislate abortion themselves, this will directly affect those low-income women who cannot afford to seek an abortion elsewhere should they need to (Millhiser 2021). It is the women who live in affluent communities and have significantly higher median and mean incomes, that are disproportionately white women, who are less likely to be burdened by such restrictions. Unlike their Black and Brown counterparts, these women have the means to travel to states with more lenient abortion legislation and thus are less likely to be forced into unwanted pregnancies or burdened by state-led restrictive abortion regulations.

IV. CASE ANALYSIS

Today, we see a new rise in state-sponsored restrictive abortion legislation, which has been written directly in contrast to federal law and the precedent set forth by Roe v. Wade (1973) (as well as Planned Parenthood of Southeastern Pennsylvania v. Casey (1992) which reaffirms the Roe v. Wade decision). Currently, the Supreme Court has heard arguments for the case of Dobbs v. Jackson Women's Health Organization (2021) and is projected to release their decision in the coming months, approximately June or July 2022. Moreover, the Supreme Court will be ruling on the constitutionality of Mississippi's "Gestational Age Act" which was enacted in the state in 2018. This act dictates that all abortions (with very few exceptions) are prohibited after 15 weeks (Dobbs v. Jackson Women's Health Organization, 2021). Currently, this law is in contrast to the current federal precedent which only allows for states to ban abortion when the fetus is deemed viable, at the beginning of the third trimester, with "reasonable restrictions" to protect maternal health in the second trimester (Roe v. Wade, 1973). Fifteen weeks is the second week of a woman's second trimester. Should the court rule that Mississippi's law banning abortion after 15 weeks is constitutional, it would effectively overturn the standing precedent of Roe v. Wade (1973) and of Planned Parenthood of Southeastern Pennsylvania v. Casey (1992)—this would mark a monumental shift in the United States for it would allow states to legislate abortion prior to viability according to their own desires. Should this happen, "24 states and three territories could quickly take action to prohibit abortion, according to [an] analysis of state abortion laws… already, 12 states have 'trigger bans' in place, designed to ban abortion immediately if Roe falls" (Center for Reproductive Rights). This shift would dictate a new future for a woman's right to bodily autonomy, for privacy rights, for health care, and for women's rights in its totality.

What is alarming is not simply the Court's pending ruling on Dobbs v. Jackson Women's Health Organization (2021), but rather their treatment of the Texas Heartbeat Act (2021). The Texas Heartbeat Act (2021) is one of the most aggressive anti-abortion laws that has taken effect in the United States since the establishment of Roe. It places a ban on all abortions after a fetal heartbeat is present, this occurs at 6 weeks gestational age (Millhiser 2021). This Act has no exceptions for cases of rape or incest. It is unique in other ways as well as it

can only be enforced by citizens with a $10,000 bounty, not by law enforcement. Because of the doctrine of sovereign immunity, which states that a sovereign or state cannot commit a legal wrong and is therefore immune from prosecution, one can only sue the state official that is enforcing the law that one is challenging. In the case of the Texas Heartbeat Act (2021), there is *no state official* that is enforcing this law- only private individuals. This legislation was specifically written in this manner so as to keep people from being able to challenge Texas on this law as there will only be private lawsuits against individual abortion providers, or those who "aid and abet" abortion (although it is not specifically defined what that means) (Millhiser 2021). Despite the emergency injunction filed to block this bill, the Supreme Court upheld it in a 5-4 decision, despite the way it attempts to circumvent how courts typically function. Chief Justice Roberts wrote in his dissenting opinion that "the nature of the federal right infringed does not matter; it is the role of the Supreme Court in our constitutional system that is at stake"; in other words, the law is unprecedented in the manner that it attempts to insulate the states from responsibility (Millhiser 2021). The Supreme Court's decision to uphold a law directly countering the very essence of how courts are supposed to function seems to signal a new age for abortion rights. Some believe this law is only tolerated by the Supreme Court because of the essence of the law, and because the new conservative majority on the court (6-3) largely opposes abortion rights.

V. CONTEMPORARY NEWS ABOUT REPRODUCTIVE JUSTICE

Today, discussions of reproductive healthcare have only just begun to recognize how fundamental minority women are to the conversation and push to incorporate them. Historically, the conversation has been dominated by the concerns of white women despite the disproportionate burden minority women bear. Contemporary media outlets that target young women with its messaging, such as Teen Vogue, speak to a new wave of feminism that seeks to incorporate the people at the margins. A recent article they released titled "Transgender People Tell Their Abortion Stories in *Trans Bodies, Trans Choices*" speaks to a similar desire for reproductive justice, one that incorporates a larger and more inclusive community in abortion discourse. However, articles such as these are far from the majority; most news about abortion says little about minority women or about

the concept of reproductive justice at all. The subject of reproductive justice in news is tailored to a niche, "woke" audience—people who are likely already more aware of the way that legislation disproportionately affects minority groups and often perpetuates systemic discrimination. Inclusive social media groups such as @feminist on Instagram, as well as groups such as "SisterSong Women of Color Reproductive Justice Collective" and "In Our Own Voice: National Black Women's Reproductive Justice Agenda" attempt to close the information gap, enlighten people about reproductive justice, and amplify minority voices- but these voices are not yet mainstream voices.

The vast majority of news generated about reproductive healthcare, however, is simply parroting back mainstream, divisive "pro-choice" and "pro-life" viewpoints. It only takes a cursory glance at any major news source to see that the concept of "reproductive justice" has yet to reach prominence. There is little being reported about the pervasive effects of strict reproductive healthcare legislation on minority groups, about what minority groups think and feel about such problems, nor about their intended goal of true reproductive justice. Consider the following: most (52%) Americans prefer getting their news from a digital platform (Shearer 2021), of these people 26% of them get news from a news website (as opposed to social media or a search engine) (Shearer 2021). The most popular news websites are CNN, MSN, New York Times, and Fox News (Majid). A quick search on any of these platforms for terms such as "abortion," "reproductive healthcare," or "reproductive justice" elicits hundreds of news articles—but one would be hard-pressed to find one that actually discusses the concept of reproductive justice, or the impact on minority women. Rather, articles are focused on the already highly publicized "pro-choice" and "pro-life" sides, adding little depth to a conversation about its impacts on minority groups. Simply put, there is a gaping hole in coverage about what matters to minority women and their fight for reproductive justice.

Significant change is still necessary to reproductive healthcare conversations; we have only just begun to skim the surface of an intersectional approach to abortion rights. Niche media outlets have only just begun to speak to the gravity of abortion discourse and legislation on minority women. Widespread change is yet to come. Much of the conversation is still dominated by the needs and concerns of white women. White feminism has pushed dominant groups such as NARAL Pro-Choice America or Planned Parenthood to the forefront of the

conversation, and these groups garner significant media attention and detract from the issues of minority women. The message and goals of organizations such as these are already well-understood, yet the goals of minority women and their desire for reproductive justice are not yet widely discussed. Conversations that spotlight minority women's concerns to the conversation remain elusive in government, in mass media outlets, and throughout much of our contemporary abortion discourse.

VI. CONCLUSION

Today we are on the cusp of a monumental Supreme Court decision, one that could overturn the long-standing precedent of Roe v. Wade (1973) and may likely have vast repercussions for all women here in the United States. But more to the point: this burden will be shouldered primarily by minority women. The gravity and impact a decision of this sort will have on these women cannot be understated- should the court rule against our current precedent, the results will be swift and harsh, women and families everywhere (but especially minority women) will reap the consequences. Historically, we have pushed these women out of the conversation. As I've discussed in this paper, abortion discourse has often failed to incorporate the concerns of minority women and is primarily centered around the needs of white women. White feminism has fundamentally transformed our conception of abortion as a social issue, paving the way for contemporary, partisan sides known as "pro-choice" and "pro-life" yet neither of these sides has truly accounted for minority women. It is up to each of us to make the decision to incorporate these women, we cannot wait until it is already too late; now is the time to shift the conversation, rather than wait to see the detrimental effects strict legislation will have on these minority communities. More precisely, we must consider how exactly we have failed minority women in the past and how we must urgently make changes in order to keep from failing them time and time again; these women deserve better, they deserve a place in a conversation that so radically impacts them, a conversation they have been pushed out of time and time again.

REFERENCES

"Abstinence-Only Education Is a Failure." 2017. *Columbia University Mailman School of Public Health*. August 22. https://www.publichealth.columbia.edu/public-health-now/news/abstinence-only-education-failure.

"America's Sex Education: How We Are Failing Our Students." 2017. *University of Southern California*. September 18. https://nursing.v.edu/blog/americas-sex-education/.

Bowell, T. n.d. "Feminist Standpoint Theory." Internet Encyclopedia of Philosophy. Accessed April 1, 2022. https://iep.utm.edu/fem-stan/.

Bhutta, Neil, Andrew C. Chang, Lisa J. Dettling, and Joanne W. Hsu with assistance from Julia Hewitt. 2020. "Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances." *The Fed - Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances*. September 28. https://www.federalreserve.gov/econres/notes/feds-notes/disparities-in-wealth-by-race-and-ethnicity-in-the-2019-survey-of-consumer-finances-20200928.htm.

Cohen, Susan A. 2008. "Abortion and Women of Color: The Bigger Picture." *Guttmacher Policy Review* 11, no. 3 (Summer): 2-12. https://www.guttmacher.org/sites/default/files/article_files/gpr110302.pdf.

"Dobbs v. Jackson Women's Health Organization." n.d. Oyez. Accessed April 15, 2022. https://www.oyez.org/cases/2021/19-1392.

Goodwin, Michele. 2020. "The Racist History of Abortion and Midwifery Bans: News & Commentary." *American Civil Liberties Union*. July 1. https://www.aclu.org/news/racial-justice/the-racist-history-of-abortion-and-midwifery-bans.

Green, Emma. 2016. "The Progressive Roots of the Pro-Life Movement." The Atlantic. https://www.theatlantic.com/politics/archive/2016/02/daniel-williams-defenders-unborn/435369/

Heaney, Michael T. 2019. "Analysis | Is the Women's March Focused on White Women – or Does It Promote Intersectional Activism?" *The Washington Post*. WP Company. July 8. https://www.washingtonpost.com/politics/2019/07/08/

is-womens-march-focused-white-women-or-does-it-promote-intersectional-activism/.

Macaya, Melissa. 2021. "Texas Abortion Law Supreme Court Ruling Dec. 10, 2021." *CNN*. Cable News Network. December 11. https://www.cnn.com/politics/live-news/supreme-court-texas-abortion-law/index.html.

Mattei, Shanti Escalante-De. 2021. "Is the Abortion Rights Movement Becoming a White Woman's Story" *Bitch Media*. October 5. https://www.bitchmedia.org/article/abortion-becoming-white-woman-issue.

McNamara, Brittney. 2022. "Transgender People Tell Their Abortion Stories in 'Trans Bodies, Trans Choices.'" *Teen Vogue*. Teen Vogue. March 14. https://www.teenvogue.com/story/transgender-people-tell-their-abortion-stories-in-trans-bodies-trans-choices.

Millhiser, Ian. 2021. "SCOTUS & The Future of Roe V. Wade." Interview by Terry Gross. *Fresh Air,* NPR, September 7, 2021. https://www.npr.org/2021/09/09/1035550744/scotus-the-future-of-roe-v-wade

Parenthood, Planned. 2022. "Supreme Court Refuses to Block Texas Abortion Ban, Greenlights Bounty-Hunting Scheme." *Planned Parenthood*. Accessed May 15. https://www.plannedparenthood.org/about-us/newsroom/press-releases/supreme-court-refuses-to-block-texas-abortion-ban-greenlights-bounty-hunting-scheme.

Pettus, Emily Wagster, and Leah Williamham. 2022. *ABC News*. ABC News Network. February 1. https://abcnews.go.com/US/wireStory/minority-women-affected-abortion-bannedlimited-82599673.

Price, Kamala. 2011. "It's Not Just About Abortion: Incorporating Intersectionality in Research About Women of Color and Reproduction." *Women's Health Issues* 21, no. 3 (May-June 2011): S55-S57. https://doi.org/10.1016/j.whi.2011.02.003

"Reproductive Justice." 2022. *In Our Own Voice*. Accessed April 15. https://blackrj.org/our-issues/reproductive-justice/.

Rice, Whitney S, Bulent Turan, Kari White, and Janet M Turan. 2018. "Norms and Stigma around Unintended Pregnancy in Alabama: Associations with Recent Contraceptive Use and Dual Method Use among Young Women."

Women & Health. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6320308/#:~:text=In%20Alabama%2%20in%20particular%2C%20an,Prevention%20(CDC)%202014).

"Roe v. Wade." Oyez. n.d. Accessed February 16, 2022. https://www.oyez.org/cases/1971/70-18.

Shearer, Elisa. 2021. "More than Eight-in-Ten Americans Get News from Digital Devices." *Pew Research Center*. Pew Research Center. January 12. https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/.

Swift, Jaimee. 2021. "We Cannot Leave Black Women, Trans People, and Gender Expansive People behind: Why We Need Reproductive Justice." *Black Women Radicals*. Black Women Radicals. September 14. https://www.blackwomenradicals.com/blog-feed/we-cannot-leave-black-women-trans-people-and-gender-expansive-people-behind-why-we-need-reproductive-justice.

"U.S. Census Bureau Quickfacts: United States." 2022. Accessed May 15. https://www.census.gov/quickfacts/US.

Vecera, Vincent. 2014. "The Supreme Court and the Social Conception of Abortion: The Supreme Court and Abortion." Law & society review 48, no. 2: 345–375.

"What's at Stake." 2022. *Center for Reproductive Rights*. March 31. https://reproductiverights.org/case/scotus-mississippi-abortion-ban/roevwade-at-risk/.

compos mentis

The Meaning That Matters

Trinity Lopez
Lawrence University

ABSTRACT
This paper looks at Kripke's interpretation of *Philosophical Investigations*. First, we begin with a picture of the world. An individual uses their awareness of *physical* and *non-physical* things to discover *meaning* in the world. Once an individual is aware of *meaning*, they understand an *accordance*. Individuals talk about their *meaning* of *accordances* and, through their use of language, come to *agree*. Then, this paper investigates an example, introduced by Wittgenstein, that captures his *paradox*. The section after that uses Kripke's interpretation of the *paradox* to make it more concise. The *paradox*, as written by both Wittgenstein and Kripke, claims that there is no fact about an individual determining whether their current *meaning accords* with their past *meaning*. Kripke offers the solution of *agreement*; an individual remembers what they have previously learned as their community's *agreement* to check whether their current *meaning accords* with their past *meaning*. We end our observation of Kripke's interpretation with his private language argument. Kripke's private language argument claims that an individual cannot *mean* anything other than what their community *agrees* with and, thus, that the concept resulting from the phrase "private language" is a semantic contradiction. I argue that the private language argument does nothing more than claim that the phrase "private language" cannot exist; the resulting concept can exist but requires different phrasing to capture its *meaning*. A "private behavior" or "private frame of reference" does not require *agreement* to exist and, additionally, does not create a semantic contradiction.

KEYWORDS
Wittgenstein, Kripke, Thing, Accordance, Meaning, Agreement, Private Language

I will be introducing my interpretation of Kripke's *Wittgenstein on Rules and Private Language*. In that book, Saul Kripke offers his view of Wittgenstein's *Philosophical Investigations*. So, I will be offering my interpretation of Kripke's interpretation. First, we will develop a picture of the world. Then, we will use this picture to observe Wittgenstein's *paradox*, which we make concise using Kripke's interpretation. The *paradox* questions whether an individual's current *meaning* is the same as their past *meaning*. Kripke answers by saying that individuals use *agreement* to determine whether their current *meaning* is the same as their past *meaning*.

Regarding Kripke's solution of *agreement*, we will find worries suggesting that Kripke's view diverges from Wittgenstein's work. These worries show us that Kripke's take on the private language argument goes further than it should; while Wittgenstein claims that the phrase "*private* language" results in semantic contradiction, Kripke takes it further by saying that the concept of the phrase "*private* language" is impossible. But while the meaning of "*private* language" may be an impossible concept, the meaning of "*private* behavior" and "*private* frame of reference" are not. This leads us to assert that his interpretation should be revised. My thesis is that it is possible for each of us, as individuals, to have a concept of a *private* thought process. "*Private* language" just might not be the correct phrase for the concept.

I. A PICTURE OF THE WORLD

Let's begin with a picture of the world. (It could also be called a picture of everything.)

A. Things

As it is ordinarily used, the word "physical"[1] refers to things that can be pointed to. For example, suppose you are on a stroll, wandering along a river. You see a tree. You can point to the wavering waters along the river; they are "physical". Or you can point to the tree, a bush, those flowers, a duck, some geese, fish, and so on; they are all "physical" things too. In the next paragraph, I suggest that *physical* things will be clarified when we add another condition to the phrase "can be pointed to".

1. Quotation marks signify that the words within them are either spoken or written language.

A *physical* thing can be pointed to but cannot be contained within a mind's experiences of non-perceptual mental states. Returning to the stroll example, the river, tree, animals, and many other things can be pointed to. But these things are not, themselves, contained within your mind's (non-perceptual) mental states! For example, the river is not within your mind; it does not flow into your thoughts, causing you to see it. Rather, the river is something external to you that causes your sight of a river. The same is true of the tree, a bush, those flowers, a duck, some geese, and fish; you must interact with these things in some way (for example, by changing your eye's or ear's view) in order to sense them. These things exist outside of the mind and, thus, are *physical*. This notion of *physical* things – being those things that are not contained within mental states – suggests that there may be other things that are, in fact, contained in mental states. Thus, we are led to add *non-physical* things to our picture.

A *non-physical* thing is the opposite of a *physical* thing; it cannot be pointed to but can be contained within the experiences of mental states. If you imagine the stroll again, there are many *non-physical* things involved in your experience of wandering down a river. For example, you become aware of a tree because you become aware of *non-physical* parts of the concept of a 'tree'.[2] A 'tree' has some necessary parts: 'a trunk', 'branches', 'leaves', 'bark', 'canopy', 'veins', and so on. Your 'tree' concept has been organized with respect to a specific structure: 'the trunk has little trunks emerging from it, there are leaves on the little trunks, the leaves have a pattern to their veins, the canopy looks a certain way, …'. From this specific order, you have your *non-physical* concept of 'tree'. All of the conceptual information I have just written are composed of *non-physical* things; they cannot be pointed to, but they are the things contained within experiences of mental states. You cannot point to 'tree' but you can think of a 'tree' or contemplate what a 'tree' is.

B. Relations

In our stroll example, we have implied that there are relationships between *physical* and *non-physical* things. To see this, consider how you become aware of

2. In contrast with quotation marks, apostrophes denote a concept. Concepts are one of the many groups of *non-physical* things. For example, the written phrase "the concept of a tree" is communicated by the following form: 'tree'. If I write the word tree without use of apostrophes, I intend for you to imagine a picture of a tree. If I write the word 'tree' with apostrophes, I intend for you to imagine your concept of a tree.

a tree. An assortment of *physical* parts – such as a trunk, branches, leaves, bark, a canopy, and so on – fits with a specific *non-physical* concept – such as 'tree'. *Physical* parts cause your awareness of their corresponding *non-physical* concepts because they relate in a specific way.

Why does 'tree' relate to a trunk, branches, leaves, barks, a canopy, and the tree's other parts? I suggest that things relate to one another because they have a *meaningful*, structured relationship. Relations are *non-physical*; you can sometimes point to their corresponding things, but doing so does not capture the structure of relation between those things. Instead, the structure of relation between those things is a *non-physical* thing within a mind.

We will call this *non-physical* relation – the one between things and their effects upon a mind – *meaning*. *Meaning* will be very useful to us; it is the focus of Wittgenstein's *Philosophical Investigations* and, thus, the focus of Kripke's interpretation as well. Often, Kripke and Wittgenstein synonymize "meaning" with "use". Wittgenstein writes: "[…] the meaning of a word is its use in the language." (Wittgenstein 2001, §43). This does not need much explanation; it explicitly identifies *meaning* as "use". Kripke does not do so explicitly but, rather, embeds the claim – that of synonymizing "meaning" and "use" – within an example of addition. Kripke writes: "…Jones now means addition by '+' if he presently intends to use the '+' sign in one way…" (Kripke 2002, 77). In other words, an individual *means* something by using it in a certain way.

In my words, *meaning* is a structure of relation between things, where the related things are either *physical*, *non-physical*, or a combination of these types. *Meaning* is a structural relation. For reasons that will soon become clear, I will notate a relation as "*R*". When two things – 'a' and 'b' – relate, I notate this as "a*R*b".[3] I will emphasize that relations are *non-physical* by writing them in apostrophes. When an individual speaks or writes, they show that they are aware of a specific structure – and, thus, *meaning* – of *R*. Individuals put *meaning* into an empty *R*.

A good linguistic test to see whether "*meaning*" and "structure of relation" are the same things is to consider the following questions. On one hand, "how does that relate?" is a question about the structural relation between things that have been said in the past. While "how" begins a question about structure ('*R*'), "does", "that", and "relate" add the concept 'relation' to the structure of the

3. Where 'a' and 'b' may be replaced by anything.

question and identify the content (the things) under consideration. Thus, with our relation notation, this is a question that reflects the structure: ''"that"*R*b'. On the other hand, "what does that *mean*?" is also a question of structural relation. "What" asks for content ('a' and 'b'). "*Mean*" expands the question (from ''a' and 'b'') to ask about relation and structure ('a*R*b'). So, both questions – one about relation and the other about *meaning* – investigate the structure of a relation ('a*R*b').

Now that we have tested whether "*meaning*" and "structure of relation" are the same thing, we can make our discussion a bit more specific by investigating two subgroups of *meaning*. Paul Grice distinguishes between two of the most prominent forms of *meaning*.[4] To show that our notion of *meaning* – being a structural relation – is sufficient, we will now apply our picture of the world to Grice's two forms of *meaning*. First, take, for example, the sentence "those spots *mean* measles". In this sentence, the word "*mean*" establishes that its speaker is aware of two distinct things: *physical* spots and the *non-physical* concept of 'measles'. In our relation notation, we have: 'those spots *R* 'measles''. This sentence can be restated by choosing a word other than "*mean*" for the relation between its things, for example, by using the sentence "those spots correspond with 'measles'". This restatement – by substituting the word "*mean*" with "correspond with" – signifies that the relation between the relevant things is structured in a specific way; when we say things correspond, it is because they have a similar, related, structure. Thus, the sentence "those spots *mean* measles" reflects a specific, corresponding, relationship between these things: spots and 'measles'.

Second, consider the sentence "The recent budget *means* that we shall have a hard year". Similar to the spots and measles sentence, this sentence identifies two distinct things: 'the recent budget' and the relevant group's 'hard next year'.[5] (In our relation notation, we have: ''recent budget' *R* 'hard next year''). The word "*means*" identifies a specific relationship between 'the recent budget' and a 'hard next year' and, thus, the sentence may be paraphrased as: "because of the recent budget, the next year will be hard".

4. For this paper, it is not essential to know what Grice's two forms of *meaning* are. Rather, I leave implied that we investigate both forms.

5. Both of these things have a combination of *physical* and *non-physical* things but let us leave this debate behind because, for our present purpose, it is not necessary.

Grice's first sentence can replace "*mean*" with the phrase "'a' corresponds with 'b'". But, in contrast, Grice' second sentence can replace "*means*" with the structure "because 'a', 'b' will be". Both sentences set up a relation between things – in the form 'a*R*b' – but the structure of these relations are different.[6]

The main point is that *meaning* is a very broad and abstract notion of relation. *Meaning* acts as if it is an empty relationship, into which the thinker of that *meaning* incorporates the structure that reflects the way they believe two or more things relate.

Let us introduce yet another word – very closely related to *meaning* – often written by Kripke and Wittgenstein: *accordance*. Once an individual has discovered a *meaning* (once the empty structural relation has been filled), they have become aware of an *accordance*. When things relate to one another, they have a certain intended structure. These intended structures – *meanings* – are what humans hope to discover (in the broadest sense of 'discover') so that they can feel like they know an *accordance*. In other words, an *accordance* is a *meaning* that an individual has become aware of.

The individual's becoming aware of an *accordance* does not entail that the *meaning* has vanished; rather, once *meaning* is discovered, it exists alongside the newfound *accordance*. For example, suppose you see a vibrant bush while on your stroll. It has berries of a magnificent blue, something you have not been aware of before (it has a structure of bush that you do not recognize). 'Blue' is a part of this bush and it is structured on the bush in a specific way: blue is only on the berries and not on the stems or leaves, it is a shimmering blue, it effects the experience that individuals have of the entire bush, and so on. You ask yourself "what could these blue berries *mean*?". The blue may *mean* that they are poisonous or perhaps luscious and delicious. Maybe it *means* something else – or nothing. Regardless, there is a structured relationship between this blue and this berry bush; it seems like the blue part of the bush *means* something. As of now, the to-be-discovered *meaning* is not an *accordance* because you are not aware of what the relation between the bush and the blue is.

If you eat the berries, you will become aware of the structural relation between the bush and the blue; you will become aware of a *meaningful accordance*. You could get sick, have a delightful meal, or have any other possible experience. Let us suppose you get sick. In this situation, you become aware of the *accordance*

6. Much like the various interpretations of the accessibility relation in modal logic.

''I eat these blue berries' *means* (*R*) 'I get sick''. After you have become aware of a *meaningful* relation (*R*) between two or more things (a, b, c, d, e, …), you have become aware of an *accordance*.

There are many different types of *meaning* that reflect different structural relations. *Meaning* can be a one-sided relationship. For example, the sentence "the recent budget means that we shall have a hard year" *means* that a recent budget will cause a bad year for a company. But it does not follow from the same sentence that a bad year will cause a non-desirable budget; the 'bad year' could refer to bad social (non-professional) interaction in the workplace, which does not entail – but may contribute to – a bad budget.

Meaning can also be a double-sided, *symmetrical relation*. A *symmetrical relation* (which I may abbreviate as *Rs*) occurs when one thing relates to a second thing in the very same way that the second thing relates to the first thing ('a*Rs*b' iff 'a*R*b' & 'b*R*a'). For example, if you see a tree, you have become aware that the 'tree' relates to the *physical* trunk, branches, leaves, barks, canopy, and so on. Further, you may become aware that the *physical* trunk, branches, leaves, barks, canopy, and other parts relate to the concept 'tree'. In other words, you may become aware of a *symmetrical relation* (''tree' *Rs* trunk, branches, leaves, barks, canopy, …') between the *non-physical* 'tree' and the *physical* parts of the tree (''tree' *R* trunk, branches, leaves, barks, canopy, …' & 'trunk, branches, leaves, barks, canopy, … *R* 'tree'').

I'm sure that there are many other types of *meaning*. I encourage you to discover them.

So far, we have distinguished between *physical* and *non-physical* things, introduced *meaning* and *accordance*, and established two types of relations between things (*R* and *Rs*). This is not a complete picture; it is confined to a single individual's awareness of the world. In order to be able to understand each other's awareness, individuals must *agree*. Thus, to complete our picture, we must investigate *agreement*.

C. The Relation of *Agreement*

Let's call an *agreement* a *symmetrical relation* between two or more individuals' awareness of *accordances*. If an individual's thought is seen as one thought bubble, an *agreement* can be seen as the merging of two thought bubbles. So, an *agreement* is an embedded relation because it connects two different *meanings*,

which are relations themselves (but not necessarily a *symmetrical relation*). The following two paragraphs explain this by outlining our method of *agreement*.

First, for an *agreement* to happen, an individual must become aware of the *symmetrical relations* that a second individual is aware of. If I say "the tree is next to the river" and you listen to me, then you will become aware that I am aware of the relation 'the tree is next to the river' ("'tree' *Rs* 'river'"); you become aware of what I *mean* by my words; you become aware that I have put 'next to' into my empty *meaning* of '*Rs*'. You have gained knowledge about my mental states; you know, given the language we speak, some things I am aware of, such as that river, that tree, 'river', 'tree', 'next to', and so on.

The second criteria of *agreement* is for a second individual to become aware of the *symmetrical relations* that the first individual is aware of. After you are aware that I think 'the tree is next to the river', suppose you want to confirm that you think so too. In that case, you would say something like "I think so too". Once you say it, I become aware that you are aware of the same *symmetrical relation* that I am aware of – namely, 'the tree is next to the river'. Thus, I have gained some knowledge about your mental states; it seems like we have the same *meaning*; it seems like we have both put the same *meaning* – 'next to' – into the structural relation '*Rs*'. Since we are each aware that each other are aware of the same *meaning* of a *symmetrical relation*, we *agree*. In as few words as possible, *agreement* occurs when two or more individuals understand the same *meaning*.

A quick note: we do not have to be aware of the same exact *meaning* in order for us to *agree*. For example, my *meaning* of 'next to' may be confined to 'close enough' while your *meaning* of 'next to' may be 'within a 1-foot distance'. Even though we may have different *meaningful* concepts of 'next to', we can still *agree* about whether or not the 'tree' is 'next to' the 'river'– given that our *meanings* resemble each other enough.

We began investigating what 'things' are; there are two main kinds of things: *physical* and *non-physical*. Then, we saw that our experiences of *physical* things cause *non-physical* things to exist. Then, we saw how they relate, giving rise to *meaningful accordances*. *Meaning* is an empty structural relation, into which a mind put things and, once an individual has become aware of a *meaning*, they develop an *accordance*. There are many types of *meaningful* relations; we investigated one-sided relations and *symmetrical relations* for two examples. Finally, we attempted to understand another form of *meaningful* relation: *agreement*. Once

two or more individuals have become aware that each other are aware of the same *meaning*, they *agree*.

Now that we seem to have a complete picture of the world, we will turn to Kripke's interpretation of *Philosophical Investigations*. In doing so, we will identify a *paradox* and offer a solution to it.

II. KRIPKE'S INTERPRETATION

In this section, we first look to Wittgenstein for an example of the *paradox*. Then, we clarify the notion of the *paradox* using Kripke's interpretation. With the *paradox* out of the way, we will look at Kripke's solution and his take on the private language argument.

A. Wittgenstein's Paradox

Philosophical Investigations gives a great example of the *paradox*:

> […] Now we get [a] pupil to continue a series (say +2) beyond 1000 – and he writes 1000, 1004, 1008, 1012.
>
> We say to him: "Look what you've done!" – He doesn't understand. We say: "You were meant to add *two*: look how you began the series!" – He answers: "Yes, isn't it right? I thought that was how I was *meant* to do it. […] (Wittgenstein 2001, §185).

What should we tell the pupil to assist understanding the rule '+2'? Not much would help this situation. The pupil is stubborn: he persists that he is following the rule because, for example, he simply skips saying every other number in the series. If we tell him "You are not following '+2' right", he may respond "Yes I am, you just don't know that I am. I leave every other number in the series implied; I know what each number should be.". He insists that he is aware of the *accordance* between the rule '+2' and the series '1000, 1002, 1004, 1006, …'; he communicates his awareness of this *accordance* by saying "1000, 1004, 1008, 1012, …" and he says "I know what I *mean*, even if you don't.".

This seems problematic and, thus, we have our *paradox*. The following two paragraphs capture the *paradox* but it will be made more concise when we develop Kripke's interpretation.

If a student could always be misinterpreting the rule we are teaching, then how would we ever know that the student is learning the correct rule? We can't say we do. We use the words that the student tells us – "1000, 1004, 1008, 1012, …" – and the student could have uttered these words *according* to any *meaningful* interpretation of them. The student is the only person aware of their own interpretation and so, it is possible that the student's *meaning* of '+2' does not *agree* with our use of that rule.

To be sure that the student uses '+2' in a way that *agrees* with their community's use of it, they should have said "1000, 1002, 1004, 1006, …". However, we are simply clueless as to what the student *means* by their 'mistaken' interpretation. It seems like no matter what a student says, they could be misinterpreting their past learning. When an individual learns a new *meaning*, it is always possible for them to become aware of an *accordance* that their teacher is not aware of.[7] This could lead to misunderstanding, which is the focus of Kripke's interpretation of the *paradox*. However, to presuppose my problem with Kripke, I claim that a mistaken interpretation of *meaning* could also lead an individual to improve their own understanding of relations.

Wittgenstein explicitly words the *paradox* as: "[…] no course of action could be determined by a rule, because any course of action can be made out to accord with the rule." (Wittgenstein 2001, §201). I soon offer the *paradox* in my words but suggest that you think about this quote for yourself before I give Kripke's interpretation and my (somewhat) concise interpretation of it. Now, we turn to Kripke's interpretation of the *paradox*.

B. Kripke's Paradox

Kripke's *paradox*, in my words, is: there is no fact that determines whether an individual's current *meaning accords* with their past *meanings*. Given our picture of *meaning* and *accordance*, my interpretation of the *paradox* claims that an individual's current awareness of a *meaningful* structure does not necessarily *accord* with the exact structure of the individual's past *meaning*.

7. This presupposes my counter to Kripke. But this is not problematic because it is consistent with his interpretation of the *paradox*.

Throughout *Wittgenstein on Rules and Private Language*, Kripke uses a rule called 'quaddition' – in contrast with 'addition' – to brew an example of the *paradox*. 'Quaddition' is a rule supplementary to 'addition'; once one of the 'added' numbers is greater than '57', 'quaddition' is followed instead of 'addition'; once 'quaddition' is followed, the sum of the two numbers is '5'. Just as a pupil in Wittgenstein's example could claim that their utterance of "1000, 1004, 1008, 1016, …" *accords* with the rule '+2', a pupil in Kripke's example could claim that their utterance of "68+57=5" *accords* with the rule 'addition'.

For example, suppose I am in a room – isolated from everyone else – and I add 68 Tonka trucks to 57 tonka trucks. I become aware of a new *accordance* and, thus, decide to follow 'quaddition'. 'Quaddition' tells me that the answer to '68+57' is '5'. How could '68+57' equal '5'? Which *accordance* could I have possibly become aware of? Based on the processes of 'addition' that others have taught me, it seems like the answer is always '125'. But, the *paradox* questions whether an individual, regardless of what processes they have been taught, can find a fact that determines whether their current *meaning* of 'addition' *accords* with their past *meaning* of it.

When searching for such a fact, no individual will find it. There is no fact – neither within their mind nor their surroundings – that determines whether an individual's current *meaning* of 'addition' *accords* with their past *meaning* of 'addition'. What determines the *meaning* of 'addition' is an interpretation instead of a fact; putting *meaning* into an empty structural relation is a method of interpretation rather than experimentation. An interpretation is not a fact itself but, rather, the way that facts are seen. For example, suppose you see a black figure in the corner of your vision and think "Did I see someone?". When you attempt to look at it, it disappears. The figure is not a fact; it is a part of the way that you were interpreting the facts around you, and, upon closer observation, you discover that there are no facts – no things around you – that *accord* with the figure. If the black figure had stayed within your vision, you may have had a new interpretation – influenced by a new awareness of *accordances* – that led you towards the claim "spirits are real". Applying this to 'quaddition', it is possible that, when encountering any new problem, an individual *means* to follow 'quaddition' because they claim to have discovered a new *accordance* between the problem '68+57' and the answer '5'.

Put more generally, an individual can always claim that they have discovered a new structural relation of *meaning* because individuals discover *meaning* through

interpretation instead of through facts. When a student believes, for example, that 'quaddition' should be followed instead of 'addition', they think they have become aware of a *meaningful* relation (thus, an *accordance*) between 'adding numbers greater than 57' and '5'. The student has found a new interpretation of 'addition' which they believe should replace our currently *agreed* upon interpretation of 'addition'. However, Kripke's solution – as discussed in the next section – prohibits any individual from discovering their own *meaning* from a newfound *accordance*.

Kripke writes "The relation of meaning and intention to future action is *normative*, not *descriptive*." (Kripke 2002, 37). In other words, the rules that we teach students – like 'addition' – are not learned through the student's direct experience of the world. Instead, rules are taught by learning the tips and tricks that other people have already figured out; rules embed past peoples' awareness of *accordances* into language. Once a community has *agreed* that they are aware of the same *accordance*, they create a rule – using language – that will be passed down to future generations; the community believes that the rule – such as 'addition' – will benefit the collective knowledge of their kind and, so, they pass it on. Kripke's point in the quote above is that individuals do not learn rules from their individual experience of *accordances*. Instead, an individual must interpret *physical* language in order to relate their (intellectual) ancestors' past awareness of *accordances* into their current awareness of *accordances*.

Learning language is learning past *meaning* and interpreting it into the learner's current, individual, *meaning*. There is a worrisome gap in this process made prevalent by the *paradox*; since we communicate language *physically* and we must indirectly interpret *non-physical meaning* from *physical* language, it is always possible that an utterance does not *accord* with the audience's interpretation of that utterance's *meaning*. As Wittgenstein puts it, "Following a rule is analogous to obeying an order. We are trained to do so; we react to an order in a particular way. But what if one person reacts in one way and another in another to the order and the training? Which one is right?" (Wittgenstein 2001, §206). I'd say: neither is right; all language can be misinterpreted but this does not entail that the individual lacks *meaning*. Kripke would disagree; he does not take it this far – or takes it too far – in saying: "[Wittgenstein] has shown *all* language, *all* concept formation, to be impossible, indeed unintelligible." (Kripke 2002, 62). While Kripke submits to the *paradox*, Wittgenstein only offers it as a question and

a mistaken one at that. After his explicit statement of the *paradox*, Wittgenstein claims this:

> It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contended us at least for a moment, until we thought of yet another standing behind it. What this shews is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases. (Wittgenstein 2001, §201).

This begins to show us that Kripke's interpretation diverges from Wittgenstein's work; perhaps there is a *non-physical* behavior that is not effected by the *paradox*. Maybe we follow rules without relying on facts by training our minds using our ancestors' language. While interpretations seem to vanish and alter, I feel – quite literally – something that does not.

C. Kripke's Solution

Kripke's solution to the *paradox* is *agreement*. For example, before I was taught the rule 'addition' by my teacher, I was uncertain whether the way I currently add things *accords* with the way I have added things in the past. Kripke would say that prior to learning a rule, an individual can never know whether their current *meaning accords* with their past *meaning*. Instead, the importance of *agreement* surfaces when I am able to show that I 'add' correctly – by solving novel problems and giving my answers to them – and when my teacher is able to tell that I follow 'addition' similar to the way that they do. My teacher saw that our answers were the same for a large number of problems and, thus, they inferred that we were both aware of the same *accordances* – between, for example, the problem '38+23' and the answer '61'. In other words, the teacher became aware of a *symmetrical relation* between our awareness and, so, uttered something like "right" to affirm that we were aware of the same *meaning* for the same *symmetrical relation*. Because two people – I and the teacher – were aware that each other were aware of the same *symmetrical relation*, we understood that we *agreed* about the *meaning* of 'addition'.[8]

8. This sentence is past tense, an important tense for understanding the *paradox*. This becomes

This appears to solve the *paradox* because it gives the individual a frame of reference for checking whether their current *meaning accords* with their past *meaning*. For example, if I, a good time after learning 'addition', come across a novel problem such as '3,482+9,382', then I will be able to remember some of the times that my teachers and I had *agreed* about 'addition'. If I feel like I might be wrong, I will recall the experiences under which I learned 'addition' so that I can know whether my current *meaning accords* with my past *meaning*. The main idea is that, by recalling past experiences of *agreement*, an individual is able to have a reliable frame of reference to check whether their current *meaning agrees* with both their individual past *meaning* and their community's past *meaning*. (But why do we need a reliable frame?)

This frame of *agreement* has a twofold purpose. First, the frame of *agreement* allows an individual to know whether their *meaning* makes sense. Second, it allows an individual to know whether their *meaning* is useful. Kipke says:

> All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives. (Kripke 2002, 77-78).

However, I feel that these two purposes alone do not give assertions their *meaning*. Circumstances of *meaning* do not get to the core of *meaning*. Rather, *meaning* is a mindful, *non-physical*, frame of reference which is caused by the circumstances relating to a *meaning*. In other words, *meaning* is caused by circumstances, but a *meaning* is not only determined by its relevant circumstances; while we may *physically agree*, we may fail to *non-physically* understand each other because of the gap in our learning process (discussed in II.B. of this paper). With respect to the second purpose (utility), why must *meaning* have "a role in our lives"? Often, individuals find *meaning* that is useful, but only to them. For example, I have created the saying "Go with the flow. Take it slow. Let it go.". This saying's *meaning* is only for me; I'm sure that, to many others, it *means* something vastly different than what I mean by it. I use it to reach a state of mind which I have recently discovered. Without more explanation of this state of mind, you (most likely) do not know what I *mean* by the phrase.

---

clearer in the next paragraph.

Thus, I believe individuals have a *private* frame of reference – independent from a community's *agreement* – that the individual uses to understand things. I find that Wittgenstein would agree (not italicized).

D. Kripke's Private Language Argument

Kripke uses '*agreement*' as an essential premise within his interpretation of the private language argument. To make this interpretation clear, we must first investigate *privacy*.

Something is *private* when only one individual is aware of it; when an individual is aware of an *accordance*, and no other is aware that the individual is aware of that *accordance*, then the *accordance* is *private*. In other words, something is *private* when it has not yet been *agreed* upon.

Kripke uses '*agreement*' for his private language argument. The argument goes as follows. A language must have *agreement* in order to be a "language". Something "*private*" cannot have '*agreement*' because it has not yet been *agreed* upon. Thus, if we put "*private*" and "language" into one phrase that *accords* with a concept, the resulting concept is a contradiction; the phrase "*private* language" is a semantic contradiction; the *meaning* of "*private* language" cannot exist without incurring problems.

Kripke takes the private language argument a step further by claiming that the concept of a "*private* language" – regardless of its *according* phrase – does not exist. However, this is a step that Wittgenstein does not take. I believe Wittgenstein would say that the phrases "*private* behavior" and "*private* frame of reference" do not entail a semantic contradiction and, thus, its *according* concept could exist. If I am right, we must revise Kripke's interpretation to sufficiently reflect Wittgenstein's work. Briefly forget that we are interpreting others and ask yourself: don't we often have *private* thoughts that we choose not to say?

III. A PROBLEM WITH KRIPKE'S SOLUTION

Kripke's private language argument claims that an individual's *meaning* cannot exist unless other people *agree* with that *meaning*. But the private language argument does not entail that an individual's *meaning* does not exist; rather, the point of it is as simple as saying: the phrase "private language" is a semantic contradiction. The private language argument makes a claim about "language",

not about whether there exists a *private meaning*, *private* thoughts, a "*private* behavior", a "*private* frame of reference", or any other phrase sufficient for the concept we are looking for.

Although "*private* language" is a phrase of semantic contradiction, there could be processes – not "language" or 'language' – that are *private*. The *paradox* seems as if it is deeply problematic; if it is plausible, then individuals are always uncertain about what they *mean*. If I, for example, do not know whether my *meaning* of 'addition' has changed throughout my use of it, then it seems like I do not know whether 'addition' – and even my mind – is reliable. But the most important phrases in the past two sentences are "seems as if" and "seems like";[9] what seems to be is not always what is. Without these phrases, the *paradox* prevents an individual from having a *private*, *meaningful* behavior. With them, the *paradox* does no such thing.

Kripke's interpretation claims that *meaning* is never *private*; it seems like he believes that he does not *mean* anything *privately*. But without *private meaning*, the individual has no system of reference that may be used to check whether their community's language *accords* with that individual's *private* processes, whether or not those processes are "language" or 'language'. Even though the *paradox* is problematic when interpreting *meaning* as determined by *agreement*, it does not raise a problem when interpreting *meaning* as a system of reference; the individual's system of reference is not necessarily determined at all. My mind – my system of reference – often does things I do not approve of, but I trust it anyway.

I think Wittgenstein would reject Kripke's interpretation. Continuing a passage I introduced earlier, Wittgenstein writes:

> […] Suppose you came as an explorer into an unknown country with a language quite strange to you. In what circumstances would you say that the people there gave orders, understood them, obeyed them, rebelled against them, and so on?
>
> The common behaviour of mankind is the system of reference by means of which we interpret an unknown language. (Wittgenstein 2001; §206).

9. The word "seem" is important for the literature concerned with Kripke and Wittgenstein, but only if you are interested. If so, I suggest looking at the words "seems", "seems like", "as if", and similar words after you have finished reading this paper.

Wittgenstein seems to say that individuals interpret language through *private* behavior or a *private* system of reference. It seems "common" to all of mankind, but the process itself is *private*. Think of *private* behavior as an individual mind's processes, such as carrying, counting, and putting together numbers to solve an 'addition' problem. These processes do not always follow language; they are learned through interpretation of language but, with enough practice, become automatic processes.[10] When an individual has become comfortable with 'adding', they do not, within their minds, tell themselves "Okay, '8+7=15', so I have to carry the one and put the five here. Now, '1+6+5=12', so I have to put twelve next to the five. …". Rather, they do the problem quickly, the speed of which (roughly) depends on the extent of their practice. For example, do the problem "92+34" and think about the "*private* behavior" that you follow.

Let us translate the problem with Kripke's solution of *agreement* into our picture of the world. I, an individual, do not need *agreement* to *mean* something. For example, if I point to someone and say the sentence "that guy's such an Austin", you do not know what I *mean* by 'Austin'. There are countless interpretations of 'Austin' that could make the sentence true or false because only the people involved with the inside joke would know what the sentence *means* and, thus, extract the correct interpretation. There may be people that have *agreed* on what the joke *means*, but it is also possible for me to have a *private* joke and chuckle to myself (as I just did, but don't try to figure out what it was). Thus, although Kripke's solution may be a solution to the question of whether an individual's *meaning accords* with their community's *meaning*, it is not a solution to the question of whether an individual's *meaning accords* with itself. *Agreement* connects two thought bubbles, but it does not connect one thought bubble to itself. Going beyond the *private* joke example, individuals often put language about *accordances* into the world and, frequently, their *meaning* fails to be correctly interpreted.

To make it even clearer, let us look at the problem from another angle. The problem with Kripke's solution is that there is not a second individual capable of completing a symmetrical relation of awareness regarding my current awareness of an *accordance*. I can be aware of an *accordance* when there is no one aware that I am aware of that *accordance*. I could dis*agree* with everyone (as may be

10. Kripke calls these automatic processes "brute inclinations"; see page 15.

the situation when an individual seems to have Schizophrenia, be high on LSD, or believe in the spirit world).

I *mean* this: the thoughts I have are not fully determined by what I have interpreted as my community's *meaning*. Although *agreement* helps me in countless ways, I do not want to be committed to the claim that my *private* processes have no *meaning*. *Agreement* points to the *meaning* that I should follow, but I decide which *meaning* I actually follow. *Agreement* suggests to me what I should put in my empty structural relation, but I decide what actually goes in there. Wittgenstein says: "The line intimates to me which way I am to go" is only a paraphrase of: it is my *last* arbiter for the way I am to go." (Wittgenstein 2001, §230). Focus on that one word: "*last*". To emphasize my problem with Kripke's solution, I leave it to you to interpret your own *meaning* for this quote. (Although, I have already, implicitly, given you my interpretation of it.)

Why does all this matter? I suggest this: do not worry about the "genius" of other people. Instead, when you learn something new, trust your *private* thoughts. Then, see whether your *private* thoughts *accord* with what you are learning. In doing so, you will save yourself from much work, time, and, most importantly, from dangerous thoughts that may lead you to a mindset of insecure stupidity. You're not 'dumb', I know with certainty. Or, rather, should I put it like: "We are all 'dumb', endlessly figuring things out."?

REFERENCES

Kripke, Saul. 2002. *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.

Wittgenstein, Ludwig. 2001. *Philosophical Investigations*. Translated by G.E.M. Anscombe. Malden: Blackwell Publishing Ltd.

compos mentis

Passing as Fuckable: Objectification, Sexualisation and the Conditions of Autonomy

Clara Moreton
University of Essex

ABSTRACT
In this paper, I attempt to pull apart the conceived dichotomy between sexual objectification and autonomy. In much of the literature, sexual objectification is often conceived as in a purely antagonistic relationship with autonomy. I argue that, though sexual objectification does restrict women's autonomy, but in a particular way that to be sexually objectifiable becomes the conditions under which any autonomy is possible for women. Sexual objectification is the relation on which gender is predicated, and thus autonomy within gender is only provided when these relations are adhered to. As an exemplary case of this, I look at 'passing' for trans women, to see how becoming sexually objectifiable is once a restriction of autonomous self-expression, while also being a condition for the access to material resources necessary for autonomy. I therefore conclude that it is not enough to demand that women individually refuse participation in sexual objectification. For the relations of sexual objectification to be properly dismantled, there must be a recognition of these other structures of oppression that make refusal impossible, and active resistance against these structures.

KEYWORDS
Feminism, Transgender, Sex, Objectification, Autonomy

Sexual objectification does not merely restrict the autonomy of women; rather it is the horizon on which women's autonomy is conceived. In this essay, I argue that Catherine McKinnon and Andrea Dworkin were right to say that hierarchy is inherent to gender as a social structure. In this way, women's autonomy becomes restricted to the realm of sexual objectification in her self-conception. However, I propose that McKinnon and Dworkin are wrong to pose sexual objectification and autonomy as in tension with one-another. Instead, sexual objectification is taken as the conditions for autonomy at all. In my view, "passing" (as it is used in trans spaces to mean being perceived as a non-trans) for trans women is an exemplary case of this, wherein sexual objectification serves as the conditions under which autonomy can be achieved. Hence, I argue that recognition of a "conditioned autonomy" under hierarchical gender is vital for resisting the unfreedom of sexual objectification that pervades gender.

A standard definition of sexual objectification is that it is an imposition of (patriarchal) social meanings that constrict my own capacity for self-presentation. Sexual objectification limits women's capacity for self-expression by delimiting what is sexually objectifiable as the only possibility (Jütten 2016, 35). The options available to women are only those that are valued to the extents to which men might be willing to have sex with them. Hence, this desire for self-worth is not, as Timo Jütten writes, an "autonomously chosen conception of the good" (Jütten, 25), but a reproduction of an imposed inequality that makes real autonomy (a "conception of the good") impossible.

In this way, sexual objectification becomes the way in which women are expected to relate to themselves and, in particular, their bodies. Simone de Beauvoir writes that this leads to a "doubling" of the self; in accordance with the demands to be an attractive object for others, she must take herself as an object too and become Other to herself; she thereby exists "*outside herself*." (Beauvoir 2009, 349). Imposed sexual objectification centres the existence of women on men, and as such are forced to take up this stance of the objectifying gaze. As John Berger writes, "The surveyor of women in herself is male: the surveyed female." (Berger 1972, 47). The language of policing and "surveying" evokes a sense of authority for this masculine surveyor. If self-worth is determined by the extent to which one is sexually attractive to men, this masculine gaze becomes the way of measuring this self-worth internally - through the eyes of the objectifier. Dworkin writes that these standards prescribe a woman's "mobility, spontaneity, posture, gait, and the

uses to which she can put her body" (Dworkin 1974, 201). For Dworkin, the result of this imposition is the limitation of, not just the capacity to visually represent myself to the world, but to fully realise what I might be capable of. Since self-worth for women is often predicated on objectification, the numerous ways of flourishing and developing myself (for myself) become concealed. Thus, sexual objectification restricts autonomy from the inside; it becomes an ideological mechanism by which women are limited to their relationship to men.

For McKinnon and Dworkin, the relations of sexual objectification run deeper; gender itself is the social structure organised by relations of sexual objectification. Sexual objectification is the reduction of a person to the status of a mere thing; but sexual objectification is more than an *act* inflicted on women - it is the *situation* of women.[1] Hence McKinnon's metaphor: "Women live in sexual objectification the way fish live in water." (McKinnon 1991, 149). For McKinnon, the very meaning of "woman" is necessarily determined by the relations of sexual objectification. Just as a fish cannot live without water, the meaning of "woman" would not exist without these relations. Correspondingly, Dworkin's account of "man" represents the inverse: objectification makes a man "feel *his* own power and presence." (Dworkin 1989, 104). She writes that objectification is a necessary condition if "he is to be a person." - though perhaps it would have been more accurate to write "If a person is to be *he*". (Dworkin 1989, 106). Therefore, the meaning of "Man" is oriented around the objectification of women, just as being objectifiable is the meaning of "Woman". Gender thus is, at its core, the hierarchical relations of sexual objectification; individual instances reproduce this hierarchy that is already imposed on them.

There is something akin to Heideggerian phenomenology here. For Heidegger, meaning of a thing is primarily understood in terms of its practical relationship with myself, and the contexts of this engagement. In his standard example, a hammer is a hammer by virtue of the ways that I use it (i.e., through hammer*ing*) and through the contexts in which it is used (i.e., a workshop). Further, it is always interpreted as *for-the-sake-of* some possibility (i.e., making a chair). This set of practical relations constitutes its ontological makeup. Conversely, this relation to "ready-to-hand" things is constitutive for who *I* am too, insofar as its use discloses

1. It is important to note that this is a *pejorative* definition of Gender; McKinnon and Dworkin are, of course, not arguing that women are reducible to men's objectification of them. It is, in their view, the ideological assumption that underlies gender norms as a whole, though.

to me the possibilities for my future (Heidegger 2010, 55-88). There is, then, this reciprocity (not to be confused with equality) wherein the use of a thing constitutes it's meaning, whilst simultaneously disclosing the meaning of who I am (and might be). Gender has a similar structure for Dworkin and McKinnon.[2] The meaning of "woman" is generated in its 'use' – in the practice of sexual objectification. This, in turn, gives meaning to the idea of "man". It might be said that, just as the totality of these relations constitutes the World for Heidegger, this relation between man and woman as giving one-another meaning constitutes the World of gender. This conception of gender as inherently hierarchical is important because it demonstrates that the lack of autonomy in objectification is not a contingent component of the meaning of womanhood, that can be easily discarded. Rather, it seems difficult to see that autonomy and womanhood be reconciled at all.

The practice of passing, in the context of being a trans woman, exemplifies a complexity in this view. On the one hand, passing seems to limit autonomy; in becoming perceived as a woman, thus as available for objectification. On the other hand, it seemingly grants autonomy, since, through being recognised as a cis woman, there is some freedom from transmisogyny. Talia Mae Bettcher, like McKinnon and Dworkin, argues that gender (in patriarchal settings) is organised along the lines of relations of sexual objectification. For Bettcher, gender norms (gendered signifiers such as clothes, body language, etc.) function to communicate a naive notion of "biological sex" that is understood as synonymous with genital status.[3] She writes that genitals constitute the "deep, concealed reality of sex" in the ideological consciousness of hierarchical gender (Bettcher 2007, 55). This structure is therefore set up to serve a "sexually manipulative heterosexuality" because it specifically exists to communicate to men who is available to be

2. It is worth noting that Heidegger strongly opposed the strict opposition of "subject" and "object." I use "object" here only to be consistent with the language of McKinnon and Dworkin. Indeed, the relation between the objectifier and the objectified is not a neat opposition either. If what it means to be a woman is to be objectified by men, then conversely, what it means to be a man is to objectify women. Thus, it is not only that there is an "I" that imposes itself on some Other, but that this very act of objectification seems to, in return, produce a broader meaning and structure - a gendered World. There is certainly more to be said about the phenomenological implications of this reading, but this task is far outside of my scope.

3. This conception of biological sex, Bettcher notes, is not based on any coherent - let alone scientific - notion of biology, but rather on a naturalised notion of the sex, whereby the genitals one is born with is the sex that one "ought" to have. Bettcher elaborates on this idea in her article *Trapped in the Wrong Theory* (2014).

sexualised. And, again echoing Dworkin and McKinnon, she argues that this communicative relation between gender and "real" sex exists specifically because "... a man needs to know a person has a vagina for the same reason a man needs to know about sexual willingness without actually having to ask." (Bettcher 2007, 57). Bettcher's development of McKinnon and Dworkin is significant because we see that what it means to "pass" as a woman is to become someone who can communicate to others that they are "biologically female." Indeed, this is consistent with the common-sense use of the term in trans spaces; to pass is to be taken as cisgender. If McKinnon and Dworkin are right, then to be taken as a woman is to be taken as someone whose existence is centred on men's sexual satisfaction. And this is visible in the fact that it is common for cis men to believe that they play a vital role in constructing a trans woman's womanhood; one writer explicitly states this, saying that he realises the "important role [he] can play in reinforcing their sense of femininity." (Nicholson 2020, 269). Above, it was noted that women often adopt this masculine, Othering gaze toward themselves to judge their own self-worth. In his own words, we see the origins of this gaze laid bare; a man believes that the identity of a trans woman is constituted, at least in part (which is enough) through his own attraction to her. Thus, her capacity for self-determination is overwritten, instead taken as predicated on the sexual enjoyment of men.

Parallel to this, there is also a clear sense in which passing not only provides autonomy but acts as a necessary condition for autonomy. Bettcher argues that trans women, under hierarchical gender, are always understood as either "evil deceivers'' or "make-believers". The (perceived) deception of trans women lies in the fact that she has falsely communicated to others that she is someone (in the eyes of a patriarchal society) for whom sexual objectification is acceptable (a woman). Thus, trans women are taken as "deceivers" in this way, as never quite achieving this ethereal, innate status of the Woman. Consequently, recognition of trans women *as* women is often attainable only through passing. As Bettcher writes, "either pass as the opposite sex or be read as openly fraudulent" (Bettcher 2014, 403). For many trans women, particularly for trans women of colour or poor trans women, the latter is not an option, and certainly not a route to autonomy. For example, a recent report showed that 1 in 3 employers would not hire a trans person. (Crossland Employment Solicitors, 2022). Another study indicated that 87% of people would not consider dating a trans person. (Blair and Hoskin 2019,

2074-95) Of course these studies (or those participating) assume that they always *know* who is cis or trans. But many trans people live full-time being perceived as cis *because* passing can be the difference between getting a job and not. Trans people may not even reveal that they are trans to romantic partners, out of fear that they would be taken to be "deceivers" or even rapists. These cases make it clear that passing (and hence becoming objectifiable) is a condition for basic survival, let alone the autonomy of self-expression. Hence, passing reveals how being sexually objectifiable can simultaneously suppress one's autonomy (by restricting the possibilities for self-presentation) whilst also being a necessary condition for the material resources which make any autonomy possible. There is something of the Kantian double-bind here: "Argue as much as you will… but obey!" (Kant 1992, 8:35). For Kant, it is necessary for the flourishing of free enlightenment and progress that one can criticise and deliberate, but ultimately must obey the state (as to not disrupt the status quo *too* much!). Similarly, a woman is "free" to participate in the labour market, to have many romantic partners; but, only so long as her expression as a woman does not *truly* disrupt the relations of sexual objectification that "moral sex" is predicated on.

It might be objected that trans women can be sexually objectified whilst (and as a result of) not being perceived as cis women. If sexual objectification forms the meaning of womanhood, then the sexual objectification of trans women by men would surely be the same. However, the sexual objectification of (non-passing) trans women is quite different to the sexual objectification of cis women. Bettcher says that non-passing trans women are often interpreted as "make-believers", and that they are "represented as whores - sexually available and disposable" (Bettcher 2007, 52). Julia Serano argues that it is the (seemingly) voluntary occupation of femininity that men interpret as hyper-sexual. Femininity is already taken as sexual (made clear in the way that women wearing feminine clothing is often used as a way to suggest that she is "asking" for sex). Given this sexualisation of femininity, there is a common corresponding belief that someone becoming a woman must only be doing so for the sexual satisfaction of men (Serano 2014, 254). Therefore, we might say that disposability then comes as a result of the "make-believer" trans woman signifying the sexuality inherent to femininity whilst transgressing the ideological ontology of the "natural" woman. For Bettcher, Sex is a *normative* construct; it makes certain demands on us, first and foremost that the genitals we are born with are the ones we *ought* to have

had, and that we have a duty to communicate these moral genitals to others through gendered presentation (Bettcher 2014, 397). Thus, a man who is castrated is still a man because he "ought" to have been. The naturalisation of gender into this "moral sex" ideologically functions to further obscure the possibility for emancipation from gender. The fact that trans women can be the object of sexual attraction for straight men (evidenced by the popularity of trans women in porn) then calls this normative construct into question. Either a man has to believe that he might have sexually objectified someone who isn't a woman (which, as noted previously, is the foundation of Man as a gender), or agree that gender is a social construct, thereby de-naturalising and casting doubt upon his right to sexually objectify women. Thus, trans women are constructed as make-believers, and usually, simultaneously, as deceivers who trick men into sex. To these men, they are women insofar as they are sexually objectifiable, and men insofar as they reveal this tension in gender. One can clearly see why passing grants some autonomy in this context. While passing certainly cannot help one escape sexual objectification, it certainly makes trans women appear less "disposable". Thus, the sexual objectification of trans women (due to their being trans) does not grant them the status of (cis) womanhood, but rather only serves to put them at further risk of violence.

In my view, many feminist writers do not take seriously enough the fact that sexual objectification, and the gendered beauty standards that are generated through it, is often the *only* way that any autonomy can be achieved. Dworkin writes that, if we are to be liberated from patriarchal standards, we must reject them. She calls on women to stop "mutilating our bodies" (Dworkin 1974, 107). Additionally, it has been argued that "association with imposed social meanings may undermine their status" as people with "equal social standing to men… regardless of the voluntariness of their own choices" (Jütten 2016, 22). In other words, while a woman might voluntarily engage with gender norms that men interpret as "asking" for objectification, this choice ultimately leads to being taken as lesser than, thereby reducing their ability to be seen as full individuals. The takeaway here must be that women should be cautious about the stereotypes they voluntarily engage with. Dworkin, again, writes that "the object is allowed to desire if she desires to be an object" (Dworkin 1989, 109). A woman might choose to become objectified, as in Nussbaum's account of objectification wherein two people consensually objectify one-another for a greater physical

intimacy (Nussbaum 1995, 249-91). But this "choice" is really only a concession to the patriarchal structure - or an inevitable result. The conclusion that follows is that, since sexual objectification restricts autonomy, sexual objectification must be rejected in order to be autonomous. Certainly, sexual objectification (and gender with it) must be resisted and the caution against a "free choice" approach to objectification it not necessarily unwarranted; but, as a result of taking autonomy and sexual objectification as a strict dichotomy, these accounts fail to see that *individual* rejection itself presupposes a kind of autonomy. Resistance to sexual objectification cannot merely be a matter of refusing to engage with men and femininity. Sexual objectification is embedded society, caught up with capitalist and racist oppression for all women. This is made clear when black women's natural hair is regarded as "unprofessional" because it does not conform to white beauty standards. As a result, black women often have to spend more time and money making their hair conventionally "professional" in this way or they risk not being hired for jobs. Individual rejection of this beauty standard is therefore a luxury that not many can afford. Hence, the relationship between sexual objectification is far more complex; sexual objectification, for many, is the *condition* for autonomy. The autonomy it provides, of course, is always conditioned by this relationship of sexual domination which conceals the broader unfreedom at work. However, one must recognise the autonomy all the same *as* autonomy, since to do otherwise is to conceal the ways that sexual objectification is mandated in a much more thorough way. It is not enough to suggest that women simply stop engaging with behaviour that men take to be objectifiable; instead, we must pay greater attention to the ways that gendered violence, capitalism and racism make this rejection almost impossible. Then, we can find ways to make this rejection possible more wholly and abolish the limits of our freedom.

To conclude, sexual objectification certainly does limit the autonomy of women, insofar as it constructs women as possible objects from the start; it attempts to form their very existence around the sexual satisfaction of men. However, the relationship between sexual objectification and autonomy is not dichotomous. For many trans women, passing is provides autonomy (by making the satisfaction of material needs possible) only on the condition that she become someone for whom sexual objectification is possible for her. Examining the ways that sexual objectification can be a *condition* for autonomy is necessary for recognising the ways in which sexual objectification is a pervasive structure. Perhaps it is accurate

to say that sexual objectification limits autonomy primarily through making it impossible to live autonomously otherwise. Recognising sexual objectification as pervasive is necessary to see what is required to make the 'otherwise' possible. It is only through resisting these other kinds of oppression concurrently that people can be emancipated a gender system that provides autonomy to women only insofar as its practice does nothing to reveal what they are fundamentally denied: the freedom for something else.

REFERENCE LIST

Berger, John. 1972. *Ways of Seeing*. New York: Penguin.

Bettcher, Talia Mae. 2014. Trapped in the Wrong Theory: Rethinking Trans Oppression and Resistance. *Signs: Journal of Women in Culture and Society*. 39:2, 383-406

Bettcher, Talia Mae. 2007. Evil Deceivers and Make Believers: On Transphobic Violence and the Politics of Illusion. *Hypatia*. 22:3. 43-65

Blair, Karen L., and Rhea Ashley Hoskin. 2019. "Transgender Exclusion from the World of Dating: Patterns of Acceptance and Rejection of Hypothetical Trans Dating Partners as a Function of Sexual and Gender Identity." *Journal of Social and Personal Relationships* 36, no. 7 (July): 2074–95. https://doi.org/10.1177/0265407518779139.

Crossland Employment Solicitors. 2022. "Transgender Discrimination in UK Workplaces" Accessed 15th January 2022. https://www.crosslandsolicitors.com/site/hr-hub/transgender-discrimination-in-UK-workplaces

De Beauvoir, Simone. 2009. *The Second Sex*. trans. Constance Borde and Sheila Malovany-Chevallier. New York: Vintage Books.

Dworkin, Andrea. 1989. *Pornography: Men Possessing Women.* New York: Plume.

Dworkin, Andrea. 1974. *Woman Hating.* New York: Plume.

Heidegger, Martin. 2010. *Being and Time*, trans. Joan Stambaugh. New York: State University New York Press.

Jütten, Timo. 2016. "Sexual Objectification." *Ethics* 127: 27-49.

Kant, Immanuel. 1992. 'An Answer to the Question: " What is Enlightenment? "' in *Kant, Political Writings*. Cambridge: Cambridge University Press.

McKinnon, Catherine. 1991. *Toward a Feminist Theory of the State*. Cambridge: Harvard University Press.

Nicholson, Geoffrey H. 2020. On Being a Transamorous Man. *TSQ* 1 May; 7 (2): 268–271.

Nussbaum, Martha C. 1995. "Objectification." *Philosophy & Public Affairs* 24 (4): 249–91.

Serano, Julia. 2016. *Whipping Girl: A Transsexual Woman on Sexism and the Scapegoating of Femininity*. Second Edition. Berkeley: Seal Press.

compos mentis

Moral Anger as an Expression of Agape Love: Can Moral Anger Lower One's Status?

Kacper Mykietyn
University of Toronto

ABSTRACT
Myisha Cherry argues that moral anger is an expression of agape love and, consequently, a justified response to oppression. She upholds that by expressing moral anger one brings the aggressor to the point of equality within the social hierarchy. I will contend that lowering someone's position to the point of equality is very unlikely due to the complexity of the oppressor's psychological makeup. I will, however, show that this is not necessarily a problem since bringing someone slightly below or above the point of equality might still express agape love.

KEYWORDS
Moral Anger, Agape Love, Myisha Cherry, Fighting Oppression, Social Hierarchy, Racial Injustice

1. INTRODUCTION

In this paper, I will contribute to Myisha Cherry's conversation on moral anger as agape love. In the first section, I shall explain what agape love and moral anger are, to then show how the former is an expression of the latter. In the second section, I will briefly explain one characteristic of moral anger, namely, its focus on one's status in the social hierarchy. I shall then argue that bringing someone to the position of equal status is very unlikely (if not impossible) because of the complexity of the oppressor's psychological makeup and reasons they decided to oppress others. However, I'll contend that this is not necessarily a problem since one might express moral anger as agape love while lowering one's status slightly above or below the point of equality. As long as the degree of our moral uncertainty is not too high we are allowed to express moral anger.

2. MORAL ANGER AS AN EXPRESSION OF AGAPE LOVE

First, agape love entails that we love someone for "their own sake" (Cherry 2019, 158). It is an emotion that is "universal and impartial" (Cherry 2019, 158). Put simply, agape love is directed toward all members of the moral community (that includes both the oppressed and the oppressors). To express such love, one needs to try to "understand the other from their shoes'' (Cherry 2019, 159). After all, we express agape love to others because we try to understand their perspective and not because we hope for a reward; consequently, the scope of agape love cannot exclude even enemies. However, agape love is not expressed to others only but might also manifest itself in a form of self-care (Cherry 2019, 160). Cherry claims that one needs to first love oneself in order to "extend" (Cherry 2019, 160) that love to others. Lastly, she says that agape love is also active and directed at change, that is, it aims at the restoration of the moral community and achieving the common good.

Second, moral anger is "a judgment that one has been wronged" (Cherry, 160). For instance, a black person might experience moral anger as a result of racial discrimination. Such anger recognizes that the oppressor perceives themselves as superior to the person of color. Hence, the aim of moral anger is "leveling the wrongdoer's status with one's own" (Cherry 2019, 160). In other words, moral anger attempts to help the oppressor realize that their feeling of superiority is

mistaken and that they are equal to others. Cherry distinguishes moral anger from the kind of anger that is sudden and unreasonable. The anger she has in mind is rational and communicative; it sends a message that we disagree with the way we are treated. As a result, the purpose of moral anger is not revenge, but rather transformation. Throughout such anger, the oppressed aims to transform the oppressor, not to gain revenge and make someone suffer.

Having explained those concepts, we can now better understand how moral anger expresses agape love. First, anger expresses agape love because it "communicates respect" (Cherry 2019, 165). When the oppressed get angry at the oppressor, it is because they recognize the "humanity" of the persecutors (Cherry 2019, 165), that is, they recognize the significance of the oppressor's actions. To act otherwise, would be to deny the moral agency of the oppressor. Second, moral anger expresses agape love since it has "the moral community's best interest in mind" (Cherry 2019, 166). Both anger and love aim at change and not distraction; they attempt to improve the community and bring the offender to justice. Anger, in such a situation, also has a "preventative" (Cherry 2019, 165) function since it aims to communicate to other members of society that a certain action is wrong and should not be undertaken again. Lastly, anger expresses love since it is concerned with the wrongdoer's own moral status (as I alluded to above). Those expressing moral anger understand that it is in the oppressor's own interest that they realize their mistakes and fix their moral standing.

3. CAN MORAL ANGER AS AN EXPRESSION OF AGAPE LOVE LOWER ONE'S STATUS?

One additional characteristic of moral anger, which will be crucial for my argument, can be found in Cherry's analysis of Martha Nussbaum's definition of anger. Nussbaum pointed out that when we get angry at someone, we naturally aim to push them down in the social hierarchy. By pushing someone lower, we act as someone more rational and morally superior; as a result, we increase our status in the social hierarchy. Anger, therefore, commits "the error of status focus" (Cherry 2019, 160) since it analyzes the situation in terms of one's social position. Cherry, however, believes that such an understanding of anger does not really capture the nature of the phenomenon. She agrees that moral anger involves lowering someone's position, but she does not perceive it as an issue. She claims

that such lowering is not problematic until the harmed one does not lower the aggressor below one's own status. Put differently, by expressing moral anger, one lowers the aggressor to the point of equality which is after all not objectionable. In such a process of lowering "the angry agent reminds the wrongdoer of their human failure and thus their humanity, but also their equality with others" (Cherry 2019, 161). For example, a racist person might be lowered in the social hierarchy so they are equal to people of color. Moral anger in such a situation aims to eradicate the racist person's false sense of superiority.

To analyze that response let me first, for the sake of clarity, distinguish between three reductive functions that anger might play in someone's position in the hierarchy. (Such a distinction is introduced by me, not Cherry.) Anger might either 1) reduce one's position in the hierarchy below the point of equality, 2) reduce one's position in the hierarchy to the position that is still above the point of equality (insufficient anger), or 3) reduce one's position to the point of equality.

Moral anger, as understood by Cherry, aims at (3) only. In this paper, I will argue that (3) is virtually impossible or at least very unlikely to achieve. However, I will argue that the fact that moral anger cannot achieve its aim is not as problematic as it seems since achieving (1) and (2) is also an expression of agape love.

First, I will argue that it is very unlikely that one manages to lower the status of the oppressor to the point of inequality because the individuals and harms they cause are so complex that no one knows how much lowering is needed. Hence, it is doubtful that the aim of moral anger can be achieved. To make my argument let us compare two cases:

A. In the first case, a twelve years old white boy refuses to socialize with people of color

B. In the second case, a white middle-aged man refuses to date women of color, since as a rule, he does not find them attractive.

Perhaps if we precisely analyze the two cases, we will be able to determine what kind of anger is justified in each case. Let me, therefore, add some detail. In the first case, the boy refuses to socialize with people of color; however, his beliefs are to a greater extent a result of indoctrination. Hence, the boy is, perhaps, not fully (or perhaps not all) responsible for his racism. In the second case, the man

takes a racist and sexist attitude toward women of color. However, his views on what is and what is not attractive are, at least to some degree, influenced by modern media and capitalism. Moreover, he seems to be racist only on a personal level. When it comes to the problems on a systemic level, he is more open. For example, he supports feminist movements and often votes for women of colour to increase diversity in the government and positions of power more generally.

Having established the details, one can attempt to express one's moral anger in a rational manner that aims at change and, consequently, brings both agents to the point of equality. We could, for instance, try to explain to the boy the wrongness of his actions; however, we would try not to raise our voice or at least, raise it, but only to a certain degree. Otherwise, our anger could turn out from being communicative to being hysterical or simply inappropriate. On the other hand, in the case of the middle-aged man who is more aware of the consequences of his actions, we would not have to be as careful in expressing our anger. We can speculate that using swear words would not prevent our anger from being communicative because of man's maturity and familiarity with some degree of profanity. However, yealing itself does not seem sufficient either. To express rational and communicative anger it would be useful to appeal to the feminist values the man already upholds to show how their application on the personal level should make him work on his behaviour. The purpose of such thought experiments is to show that the more we know about the person and the reasons they decided to harm another, the more rational and effective our moral anger will be.

The problem, however, is that in real life the aggressor does not provide the oppressed with the history of their upbringing nor any details about their current personal life. The harmed one often knows literally nothing about the oppressor except the fact that they harmed them. Moreover, even if the oppressed tries to improve the moral standing of the oppressor and get to know them, the oppressor will most likely refuse to cooperate. Consequently, the process of lowering one's status to equality will be even more difficult. Hence, if a random person harms me, then it is very likely that my anger will bring them either slightly below or above the point of equality, but not to the point of equality as such.

Sometimes the anger we express might lower one's position by let's say 85%, but it will not bring them to the state of equality with the oppressed. For example, the twelve-year-old (from the first case) might begin to play with non-white

children; however, he will still subconsciously prefer white kids. In such a situation, the child realizes that he is not better than others and he is lowered almost to the point of equality; however, the point itself is not yet reached. (Whether the existence of subconscious racism only brings us closer to the points of equality is debatable, but let us assume that it is close enough.) What's important is that the anger which brings the boy lower in the hierarchy still manages to express agape love. Even though the anger did not lower the boy to the place of inequality with others, it was still directed at change and improving his moral standing. Agape love was, therefore, expressed.

Similarly, moral anger expresses agape love when it lowers the twelve-year-old's status slightly below the point of equality. Being morally angry at the child in a reasonable and communicative way might make him feel like he needs to reevaluate his actions; it is possible that the way we communicate will make him feel as having a lower status than he actually should. However, if our moral anger is really communicative, then such feelings will be only temporary and the twelve-year-old will eventually enjoy a status equal to those of his peers. In a similar way, one might feel worse after taking an injection; it might hurt at first, but the long-term effects will be positive. Such an expression of anger once again manifests agape love since it cares about the moral standing of the child and understands that temporary inconvenience will eventually pass and the moral community will be improved as a whole. Thus, I have shown that getting to the point of equality at which moral anger aims might be more difficult than it seems; this, however, is not a problem since going slightly above or below such a state is still an expression of agape love.

REFERENCES

Cherry, Myisha. 2019. "Love, Anger, and Racial Injustice." In *The Routledge Handbook of Love in Philosophy*, edited by Adrienne M Martin, 13–23. New York : Routledge.

compos mentis

Surviving Libet: Why Neuroscience Did Not Disprove Free Will (Yet)

Joachim Nicolodi
King's College London

ABSTRACT
The idea that we have free will in at least some of the decisions that we make is under heavy attack by recent neuroscientific studies which largely go back to a ground-breaking article published by Benjamin Libet and colleagues in 1983. Libet discovered a specific electrical charge in the cortex, the so-called *readiness potential*, that can be measured significantly before the conscious decision to act arises. This, Libet thought, is the definite proof that it is not *us* that consciously initiates our decisions but unconscious brain activity, thus definitely disproving the existence of free will in any ordinary sense of the term. This paper claims that Libet was wrong to draw such a conclusion from his research. I will explain how his experiments were conducted and then consider two ways one might challenge his conclusion. One popular way of doing this has been unsuccessful, while another gives us good grounds to refute Libet's argument. It is plausible to say that the experiments conducted by Libet, and all Libet-style experiments that came after, capture but a very small percentage of the decisions that we make, and it is therefore premature to draw such stark conclusions from them. I will suggest a new methodological approach to solve this problem and hopefully move the whole debate in a more fruitful direction. Rather than racking our brains about Libet-style experiments and developing more sophisticated approaches to escape or accommodate their empirical observations, we should wait for research to be published that captures the complexity of decision-making more adequately.

KEYWORDS
Free Will, Determinism, Philosophy of Psychology, Neuroscience, Benjamin Libet

INTRODUCTION

Ordinarily, we think that at least some of our decisions are the product of free will. This notion is dismissed as illusional by most contemporary psychologists and neuroscientists:

> It seems we are agents. It seems we cause what we do. […] It is sobering and ultimately accurate to call this all an illusion. (Wegner 2002, 341-342)

> The phenomenological feeling of free will is very real […] but this strong feeling is an illusion, just as much as we experience the sun moving through the sky, when in fact it is we who are doing the moving. (Bargh 2008, 148-149)

> Free will *is* an illusion. Our wills are simply not of our own making. […] Your brain has already determined what you will do. (Harris 2012, 5, 9)

This scepticism largely traces back to a ground-breaking study published by Benjamin Libet and colleagues in 1983. This paper aims to show that Libet – and indeed all Libet-style experiments that came after – do not succeed in showing that free will is an illusion and our intuitions about conscious decision-making are misguided.

Here is an overview of what follows. In part one, I will present a definition of free will and then a summary of Libet's study and the philosophical argument he based on it. Free will can be understood in a plethora of different ways, so we do well in getting clear on what we mean when discussing it in this paper. In the second and third parts of the essay, I will discuss two lines of attack that can be brought up against Libet. While the first one fails, the second gives us very good reasons to doubt Libet's conclusions on the inexistence of free will: the experiments from which those conclusions are drawn simply do not capture all the decisions that we care about when debating free will. I will conclude by proposing a novel methodology that might enable neuroscience to answer the question once and for all.

I. THE TARGET AND THE CHARGE

Most people believe that we have free will in at least some of the decisions that we make (Nahmias et al. 2005). What exactly free will amounts to is a much-debated question within philosophy and trying to come up with an adequate definition would lead us beyond the scope of this paper. For our purposes, it will suffice if we stick to the definition of free will that most people associate with the term, which goes roughly like this:

> A subject S freely wills an action *φ* at time *t* if and only if it was possible, holding fixed everything up to *t*, that S choose or do otherwise than *φ* at *t*.[1]

Let us unpack and elaborate on this in the following. Say I am confronted with a decision and have to choose between option *a*, going out, and option *b*, staying home to study. I freely will action *a* if and only if I could have also chosen *b*. If I did not have that possibility, my decision would have been coerced or at least predetermined, thus leaving no room for my conscious volition to intervene in any way. But if the possibility is given, the decision is "up to me" and thus free because it was initiated by my consciousness that could have also pulled in another direction.

Now, of course it is in principle always possible to end up with a different decision, simply because a different past could have led to a different outcome. In one possible universe, I forgot that I have an assignment tomorrow, thus ending up choosing *b* over *a*. To avoid that our decisions are always free because they always could have turned out differently, we need to add the caveat of "holding fixed everything up to *t*": we only consider the actual world and the decision made therein, not the outcomes in all possible universes.

To sum up: I act according to my free will if and only if I could have chosen otherwise, that is to say, if I could have consciously decided to do otherwise than I did. Traditionally, also our idea of moral responsibility is based on this fundamental freedom, for could an agent really be blamed for his actions if they were determined by forces completely outside of his influence? I will run with this definition of free will and the connection to consciousness throughout the paper

1. I follow here roughly what Timothy O'Connor calls the "Categorical Analysis" of free will (O'Connor 2022).

since Libet himself adopts a similar one in numerous places (Haggard and Libet 2001) and uses "free will" and "conscious will" interchangeably (Libet 2005). It is also the definition that is prevalent in folk psychology, and the one that most psychologists and neuroscientists who comment on the issue associate with the term. Most of these psychologists and neuroscientists believe that Libet's study conclusively proves that this traditional notion of free will is an illusion and indeed Libet himself thinks that free will does not hold up to neuroscientific scrutiny. But how exactly does his study put the notion of free will in dire straits?

In the 1983 study conducted by Libet and colleagues, participants had to perform a simple physical task that would, according to Libet, classify as a typical free voluntary act. Whenever they felt like it, participants had to flex their wrists and, at the same time, tell when exactly the conscious decision to do so arose. For that purpose, an oscilloscope clock was placed before the participants with a spot of light revolving around the periphery of the screen so that the subjects could tell where the spot was when they were first aware of their intention to act. During this task, the neuronal activity was measured with an electroencephalogram (EEG), and the findings were indeed striking. The researchers found a specific electrical charge in the mesial motor area of the cortex ("readiness potential", RP) that begins on average 550 msec before the act but also 350-400 msec before the conscious decision (Libet et al. 1983).[2]

We can already see how this is problematic for the definition of free will given above. The *brain* determines the flexing of the wrist, and not the individual's conscious decision. It is an unconscious neuronal process that initiates our actions and not our consciousness, which is also why we cannot deliberately choose to do otherwise: whatever we do, there is always a RP outside our conscious control that determines which course of action we are going to take. Libet concludes that our traditional view of how we make decisions is fundamentally misguided, as is our way of assigning blame and praise. He grants that our volitions still have a role to play, however only in vetoing our actions while the RP is already present and never in initiating them (Libet 2005). Although Libet leaves this possibility open and wants to grant at least some sort of conscious freedom, we can still say with confidence that he radically contradicts the typical notion of free will and leaves

2. It should be noted that the term "readiness potential" was not coined by Libet but goes back to the German "Bereitschaftspotential" in a paper published by Kornhuber and Deecke in 1965. Also the experiment itself can be seen as a precursor for Libet's, which however gained much more fame.

us only with "free won't" (Haggard and Libet 2001, 48). To get a clearer picture of the debate and the approaches that different thinkers took, we can summarise Libet's argument as follows (although he never formalises it this way):

1. We have free will if and only if at least some of our decisions are made consciously.

2. Neuroscience shows that no decision is made consciously.

C. We do not have free will.

The argument is logically valid, as can be readily seen. We now have to determine if it is sound. Most of the approaches in the literature can be summarised as either denying the first or the second premise. To give an adequate picture of the debate and come up with a satisfying conclusion of whether or not Libet is successful, I will consider these approaches in turn in the following two sections.

II. ROAD ONE: DENYING THE NEED FOR CONSCIOUSNESS

The first possibility to attack Libet's argument is to deny premise one. This road has been taken by David Rosenthal (2002), Neil Levy (2005), and Daniel Dennett (1984). Although their approaches differ, they all deny that we necessarily need to make conscious decisions to have free will. They accept premise two – neuroscience showing that decisions are always the product of subconscious brain processes – but they deny that consciousness and the possibility to do otherwise are necessary for free will. Broadly speaking they all argue, therefore, in compatibilist terms: although Libet proves that there exists a local, neural determinism, this does not exclude the possibility of free will. How do they argue for this?

Rosenthal argues that whether or not our actions are free does not depend on whether they are consciously caused, but on whether they are "fitting comfortably within a conscious picture we have of ourselves and of the kinds of things we characteristically want to do" (Rosenthal 2002, 219). If we reflect on ourselves, we all have a certain picture of who we are and what we want, and actions are free if they align with this self-conception, no matter how they are caused. Rosenthal's

theory has some obvious flaws, which is why I will not spend much time on it. The most devastating is that it radically contradicts our moral intuitions if we grant, as stated in the beginning, that freedom and moral responsibility go hand in hand. If we rob a shop and knock out the cashier, it would be of no excuse whatsoever to say that it contradicts the conception we have of ourselves and is something we would have never pictured ourselves doing.

More interesting proposals are made by Dennett and Levy, which are sufficiently similar to treat them together.[3] They hold that the first premise of Libet's argument is incoherent: decisions are *necessarily* unconscious, which is why – when Libet says decision-making has to be conscious to count as free – he is making demands that free will cannot even in principle fulfil. The question of consciousness is simply irrelevant to the question of free will, which is why the first premise lacks any basis and Libet's experiments do not affect free voluntary decisions in any way.

Levy asks us to consider an ordinary example of decision-making, namely deciding if we should accept a job offer in a different city (Levy 2005, 71). If we look close enough, Levy says, consciousness does not play the role we normally think it does. We do not *decide* that this particular reason, being close to our family, say, is more important than a higher salary. We just *become aware* of these different values, but consciousness does not have any influence in assigning them. Neither can we consciously determine the outcome of our deliberation; it happens on a sub-personal level (that is, not on the level of the conscious agent), and then just appears before us. Dennett argues in the same vein, although not directly mentioning Libet:

> But those same decisions can also be seen to be strangely out of our control. We have to wait and see how we are going to decide something, and when we do decide, our decision bubbles up to consciousness from we know not where. We do not witness it being *made*; we witness its *arrival*. (Dennett 1984, emphasis by the author)

I think this line of reasoning has some initial plausibility. Indeed, there seem to be many cases in which a decision just happens to us and we do not have

3. Dennett criticises Libet also on other points, for instance in *Freedom Evolves* (2003). However, I think the point he makes here is particularly interesting.

any conscious control over it. Choosing a romantic partner might be one good example, but also choosing a particular kind of career or university course. When we are in love it is very hard to pinpoint why we chose that person over another, and – when we are asked why exactly we love that person – awkward quandaries are commonplace. The same can be said in seemingly more conscious decisions, for instance, when choosing a career as an artist or choosing to study literature. In these cases, too, it is hard – if not impossible – to track exactly how this decision was made. One can of course start talking about passions or talents, but it still does not seem to be an active decision but rather something that happens on the subconscious level.

However, I do not think Dennett and Levy succeed in showing that decisions are *necessarily* like this and cannot be initiated otherwise. There are cases in which it is quite clear that conscious reasoning was involved in the process. To give just one very striking example it is worth looking at Effective Altruism. The whole point of Effective Altruism is to come up with a value system that enables us to assign each possible action a precise numerical value (so-called "quality-adjusted life years", or QALYs), so that different actions can be compared and the option with the highest value chosen (MacAskill 2015, 61). Moral questions like these seem to be the exact opposite of decisions as described by Dennett and Levy. Every aspect of the decision is meticulously thought through and quantified to eventually lead to the best possible outcome.

Of course, Dennett and Levy could say that here, too, the value of the different options just *appears* before us, independently of our consciousness. However, this would lead to the extreme conclusion that a great part of moral philosophy, and indeed all of value theory, is utterly useless since it rests on unconscious processes we do not have any influence on. This highly counterintuitive conclusion is reason enough to refute Dennett's and Levy's claim about the nature of decision-making, which is why I will not go into their compatibilism any further.[4] It is also reason enough to abandon the first way of refuting Libet's claims. The more plausible way is to argue against the second premise, as I shall explain in the following section.

4. Levy further supports his claim by arguing that influencing our decisions is an attempt to control our control system, and therefore conceptually impossible because it would lead to an infinite regress. However, I think his account of a "control system" needs to be fleshed out in more detail to be a valid counterpoint. Like this it is rather ad hoc and unclear, as Mele remarks as well (Mele 2008, 111).

III. ROAD TWO: QUESTIONING THE SCOPE OF THE EXPERIMENT

As we have seen, trying to refute the first premise of Libet's argument and unlinking free will from consciousness altogether likely leads to some pitfalls. The more promising approach is to deny premise two: that neuroscientific experiments like Libet's prove that no decision is made consciously. One obvious way to do this is by trying to find flaws in the studies themselves and questioning the underlying methodology. However, such methodological flaws will be hard to find since his research has been backed up by many Libet-style studies in more recent years (e.g. Lau et al. 2007, Banks and Isham 2009). The results seem to be accurate (with some obvious deviations in the exact timing of the RP), which is why I will argue that premise two is too strong *even if* we accept Libet's neuroscientific findings. The most devastating way of doing this, I find, is to argue that it is too rash to generalise the findings of Libet-style studies to *all* decisions that we make.

The tasks in Libet's experiment exhibit at least three features that make such a generalisation highly problematic: (i) The tasks are completely *disinterested*, in the sense that the person has no particular reason to flex the wrist at that point and not another. The same can be said for both Lau et al. and Banks and Isham, where the subjects were told to push a button at a point of their choosing and then, again, report the time of their awareness to do so. All these tasks are comparable to Buridan's famous ass, which has to decide between two equidistant and equally big haystacks and therefore has no reason to choose one over the other. (ii) The tasks are spontaneous physical movements that do not need much conscious deliberation to begin with. (iii) The outcome in the experimental tasks is clear, while we normally do not know how the distal decisions that we make will play out.

In this sense, the disinterested, spontaneous, and proximal tasks in Libet-style experiments represent but a very small number of the decisions we make in our lives and seem to be incomparable to the decisions we normally talk about when debating free will. As we have seen, the traditional notion of free will is closely connected to moral responsibility, which is why we are normally concerned with ethical dilemmas when we talk about free will. Consider Sartre's famous example of a young man who is undecided whether he should join the resistance or stay at home and care for his frail mother. Dilemmas like these are, obviously, entirely different from the tasks in Libet's experiments since the person is interested in the

yet unknown outcome and, therefore, spends much time evaluating the different options. Generalising from Libet's experiments to these kinds of cases would be extremely bold, to say the least.

It is important to note that Libet predicts and addresses this kind of worry, however, he does so only briefly. He says that "it is common in scientific researches to be limited technically to studying a process in a simple system; and then to find that the fundamental behaviour discovered with the simple system does indeed represent a phenomenon that appears or governs in other related and more complicated systems" (Libet 2005, 559). Libet gives the example of Millikan and Fletcher measuring the charge of a single electron in one isolated system, saying that it is valid for electrons in all systems. Although Libet does not flesh out this reply in any more detail, I think he has a point. Generalising from case studies is an essential part of the scientific methodology, indeed without it, science would hardly be possible. As the economist Henry Mintzberg pointedly remarks: "If there is no generalizing beyond the data, no theory. No theory, no insight. And if no insight, why do research?" (Mintzberg 2017, 187).

However, it is key that the domain of the generalisation is sufficiently similar to the findings of the case study. And this is not the case with Libet and the Libet-style experiments conducted so far. As we have seen, there are at least three striking differences between the tasks in the research and more complex, moral decisions: (i) disinterest, (ii) lack of deliberation because of the simple task required, and (iii) proximity and knowledge of the outcome. These differences make a generalisation highly problematic since the domain is exceedingly different.

This clearly shows that the neuroscientific research made so far is insufficient to show that all our decisions are made unconsciously, simply because of the striking differences between the simple, Buridan-type tasks in the experiments and more complex, ethical decisions. And, as we have seen, to deny premise two it suffices to show that there are at least some decisions that might be made consciously. This is not to say that more complex decisions are *necessarily* conscious. It is just to say that we do not yet have the research to make any definite conclusions about it. The second premise of Libet's argument fails, and consequently his argument as a whole. Libet's experiment does not demonstrate that our intuitions about free will are misguided since conscious decision-making might still play a role on some occasions.

I will conclude with a brief suggestion about how non-Libet-style experiments could be conducted in the future to yield answers about whether more complex decisions are unconsciously initiated too. To my mind, the best way of conducting such an experiment would be with typical prisoner's dilemmas, a standard example in game theory. I would suggest something roughly like this: put two participants in separate rooms, construct a fictitious scenario that they can imagine (e.g. the classic example of two captured robbers) and then give them options to choose from plus the possible outcomes, of course not with avoiding prison years as an incentive but maybe with winning a small amount of money. If both cooperate with each other and remain silent, they both win £30. If one of them "confesses" to the police and the other does not, the first wins £60 and the latter 0. And if they both confess they both win £15.

This situation would be sufficiently simple to be studied in laboratory conditions, but it would fulfil the three criteria for more complex decisions mentioned above. We are interested in the outcome, we have to think carefully about the different options and the psychology of the other person, and we do not know how our decision will play out. If the results would show a similar neuronal activity as in typical Libet-style experiments, it would indeed be highly problematic for our intuitions about conscious decision-making. As of now, however, the research gives us neither reason for enthusiasm nor worry about free will.

CONCLUSION

Let us take stock of what we have achieved in this essay. I started by giving the definition of free will that is most commonly used within psychology and neuroscience but also the public discourse. Going into too much detail would have led beyond the scope of this paper, and also missed the target of the discussion since it is explicitly the traditional notion of free will that Libet and most neuroscientists after him challenged. I then turned to the experiments conducted by Libet and colleagues and critically assessed the argument against the traditional conception of free will that he based on them. Responses to this argument can be roughly grouped into two categories: those denying premise one and those denying premise two. I argued that philosophers pursuing the first route – like Rosenthal, Levy, and Dennett – are not able to successfully refute Libet. The more promising way of doing this, I concluded, is to deny premise two by arguing that

Libet's experiments only prove that a small percentage of our decisions are made unconsciously, not decisions *in toto*. By explaining how slightly more complex decision-making processes could be investigated in laboratory conditions, I hoped to show how neuroscience could give us a definite answer to the matter at hand.

REFERENCES

Banks, William P. and Eve A. Isham. 2009. "We Infer Rather Than Perceive the Moment We Decided to Act." *Psychological Science* 20 (1): 17-21.

Bargh, John A. 2008. "Free will is un-natural." In *Are We Free?: Psychology and Free Will*, edited by John Baer, James C. Kaufman, and Roy F. Baumeister: 128-154. Oxford: Oxford University Press.

Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: The MIT Press.

Haggard, Patrick and Benjamin W. Libet. 2001. "Conscious Intention and Brain Activity." *Journal of Consciousness Studies* 8 (11): 47-63.

Harris, Sam. 2012. *Free Will*. New York: Simon & Schuster.

Lau, Hakwan C., Robert D. Rogers, and Richard E. Passingham. 2007. "Manipulating the Experienced Onset of Intention after Action Execution." *Journal of Cognitive Neuroscience* 19 (1): 81-90.

Levy, Neil. 2005. "Libet's Impossible Demand." *Journal of Consciousness Studies* 12 (12): 67-76.

Libet, Benjamin W. 2005. "Do We Have Free Will?" *The Oxford Handbook of Free Will*, edited by Robert Kane, 551-563. Oxford: Oxford University Press.

Libet, Benjamin W., Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity Readiness-Potential." *Brain* 106 (3): 623-642.

MacAskill, William. 2015. *Doing Good Better: Effective Altruism and How You Can Make a Difference*. New York: Penguin Press.

Mele, Alfred R. 2008. "Recent Work on Free Will and Science." *American Philosophical Quarterly* 45 (2): 107-130.

Mintzberg, Henry. 2017. "Developing theory about the development theory." *Handbook of Middle Management Strategy Process Research*, edited by Steven W. Floyd and Bill Woolridge, 177-196. Cheltenham: Edward Elgar Publishing.

Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. 2005. "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility." *Philosophical Psychology* 18 (5): 561-584.

O'Connor, Timothy, and Christopher Franklin. "Free Will." In *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), edited by Edward N. Zalta.

Rosenthal, David. 2002. "The Timing of Conscious States." *Consciousness and Cognition* 11 (2): 215-220.

Wegner, Daniel M. 2002. *The Illusion of Conscious Will*. Cambridge: MIT Press.

compos mentis

The Hopeful Capacity of Octopuses

Levi Satter
Brooklyn College

ABSTRACT
This work considers whether the experience of hope is solely within the purview of persons, or if invertebrates with complex nervous systems also have the capacity to hope. If one accepts that increasingly complex conscious experiences arise from successively more complex biological communication within a body, then any cognition or emotion may be experienced by a sufficiently complex organism. Hope is experienced both cognitively and emotionally. According to Snyder's Model of Hope Theory the cognitive experience can be divided into pathways thinking and future goal orientation, whereas the emotional experience is what arises from the interaction of cognitions and environmental pressures. Octopuses are often defined by their intelligence and the flexible way they pursue future goals. In 2021, their emotional life was brought to the forefront when they were identified as having sentience, or the conscious capacity for sensory and subjective experience. Because of their flexible goal-oriented thinking and the fullness of their emotional experience they have the biological complexity which gives them the capacity for experiencing hope. Therefore, humanity's moral circle should expand to treat them as moral patients, like other vulnerable populations. It is not the automatic tendency of humans to treat other organisms, especially invertebrates, as moral patients. Using social narratives to hold octopuses in moral patienthood that is reflective of the vibrancy of their lived experience may be used to extend the moral circle of humanity like it has done for other vulnerable populations.

KEYWORDS
Animal Consciousness, Moral Psychology, Hope, Emotions, Cephalopods, Ethics

INTRODUCTION

Octopuses[1] have risen in popular culture in the last decade as researchers and naturalists documented the vibrancy of their lived experience. They are aquatic invertebrates with experiences seemingly entirely foreign to those of humans. Yet to those who work with octopuses, their experiences allowed them to be "sensitized to the other; especially wild creatures" (Ehrlich and Reed 2020). They have been shown to be sentient and documentaries about their lives have won an Oscar (Ehrlich and Reed 2020), yet the fullness of their experience has yet to be delineated. Understanding their experiential life is the work of future experiments. Yet, philosophical questions may be used to consider the directions for those efforts. As such this paper will question, does an octopus have the capacity to have the relatively complex experience of hope? Part 1 considers the terminology used in relation to this question. Part 2 considers octopus sentience as well as their capacity and exhibition of hope. Part 3 concludes with implications that may follow from recognizing that octopuses do have the capacity to hope.

If one assumes that mind, consciousness, and emotions all have their basis in neurological and cellular processes, then increasingly complex organisms can have increasingly complex but similar experiences. This is similar to the idea of homologous and analogous structures in evolutionary biology. Bird and bat wings are homologous structures because they developed from forelimb structures in a common reptile ancestor. However, birds and insects have wings that are analogous because similar structures developed along completely separate evolutionary tracts. Therefore, even evolutionary disparate organisms, i.e., humans and octopuses, may have analogous experiences if their neurological structures have sufficient overlap.

Hope is a cognitive and emotional experience that is future oriented and based on the ability of an organism to identify a future preferred outcome that is different from the present and to consider a variety of ways in which the preferred outcome can be accomplished. In human psychological studies the emotional experience of hope rises with the cognitive experiences of an individual feeling able to accomplish a goal as well as their ability to identify alternative pathways in the face of barriers (Snyder 2002). Octopuses have an exceptional cognitive

1. The plural to octopus is octopuses, not octopi. Octopus is derived from the Greek word októpus. Thus, the plural is with an "es." The ending "i" is used for words derived from Latin.

capacity which allows them to identify numerous ways to accomplish a goal. While they are not self-aware, they do have personalities with a sufficient unity of experience and temporal understanding to satisfy the cognitive aspects of hope. Recently sufficient evidence was collected to designate octopuses as being sentient which means they are capable of having complex and differently valenced feelings (Birch et al. 2021). Their emotional life must be inferred from field reports and inductions from their handlers. Research on octopuses' experiential life is ongoing and increasing due to developments in neuroscience, cognition, and technological advances. Yet globally, octopuses are not recognized as within humanity's moral circle, in fact they are often disregarded and treated as only worthy of being eaten (Gritzer 2019). This may be due to their alien nature as aquatic invertebrates. It is uncomfortable to attribute moral patienthood to other animals, yet increasingly social mammals have been understood as being worthy of human moral considerations, e.g., dogs, whales, elephants, etc. Despite octopuses being scientifically recognized as sentient and it being widely distributed in the news cycle (Baker 2021; Hunt 2021; Pandey 2021; Tran 2021), legally they are still not protected, and morally few people consider the feelings of their calamari. Laws are often derived from the moral attitudes of their citizens; thus, it makes sense that legal protections are lacking because moral considerations for octopuses are also lacking.

Hope is a central human experience. There is a vivacity to hope which allows one to endure. Many choices to end one's life occur because of a pervasive feeling of hopelessness in human beings. Because of its central nature to human experience, it may be uncomfortable to consider that other organisms may be capable of experiencing hope even in a less aware form. This is especially true for organisms which have yet to be widely considered in humanity's moral circle. Octopuses' capacity for experiencing hope indicates a complex and vibrant experiential life that requires their recognition as moral patients and challenges the understanding of any cognitive emotional experience being solely accessible by human consciousness.

PART 1: TERMS WITHIN THE EXPERIENTIAL CONTINUUM

Mind, consciousness, and emotions are not specific to human experience. They all arose progressively as organisms interacted with their environment.

Sensation is the primary aspect of experience, but an organism does not have to be conscious to experience sensation. It is what allows even single cellular organisms the ability to interact with their environment, seeking out food and avoiding things that may cause damage. Therefore, at its base, experience begins with sensation of the environment which either attracts or repels the organism (Ginsburg and Jablonka 2010; Godfrey-Smith 2016).

Mind is not something that exists outside of experience, but it is inextricably tied to the interaction of sensation of the environment with an organism's body. As an organism senses its environment, information is incorporated into the body of the organism allowing it to act more effectively within its environment. Mind describes this reactionary activity, from sensation, to incorporation, to action. Therefore, even single cellular organisms have rudimentary forms of minds. During the evolutionary process more complex minds develop as the needs of the organism's body increase. In a single cellular organism reaction is all that is necessary, but multicellular organisms need ways of communicating information which they gain from the environment to the other cells within their body. Therefore, intercellular signaling is required, meaning the mind of a multicellular organism would have to incorporate intrabody sensation as well as environmental information to allow the organism to interact effectively within its environment. This process increases to more rudimentary organisms, like worms, up to human beings, and can even be used to explain the collective intelligence of group interactions (Schermer 2022).

A useful metaphor for the mind is that of fire. Fire is the process of combustion in the visible spectrum of light. It exists as long as the chemical process of combustion occurs, but once it is finished the fire no longer exists. The same is true of mind. If sensation leads to activity within a body, then mind exists within the body.

Just because an organism has a mind does not mean that it has consciousness. Although, consciousness too exists on a spectrum. Consciousness is the awareness of a mind that it exists within an environment. Consciousness arises first through the ability of associative learning (Ginsburg and Jablonka 2010). Associative learning takes place when an organism can store information about past objects within memories which then can change their future interactions. It is more complex than mind since it requires the ability to store information about the world, whereas mind arises from the immediate sensory experiences within the world. As the associations within an organism become more complex, they can

take on a preferential quality leading to motivating the organism to seek or avoid objects based on past experience as opposed to only present experience which is the role of the mind. Thus, as species exhibit higher levels of consciousness, they can exhibit more complex preferences and temporal cognitions. These preferences and consistent ways of an organism interacting within the world are termed personality. The integration of all this learning leads to a progressive understanding of the organism as an agent within the world, and at its most complex leads to the self-aware consciousness exhibited by human beings.[2]

A useful simile to understand consciousness is that it is like an organism's discriminatory capacity for sound. The first experience of consciousness was akin to white noise (Ginsburg and Jablonka 2010; Godfrey-Smith 2016). Then as the associative capacities of organisms increased with evolution so too did their experience of consciousness. By applying the simile, human beings would be capable of experiencing the world as a complex symphony presented by the interaction of many instruments, whereas progressively less fully conscious organisms would have a steadily reduced ability to distinguish both the parts and the whole of the symphonic experience.

Emotions are mental states that occur because of neurophysiological changes within an organism. They serve the function of orienting the organism and adapting its behavior to succeed in its environment. They do this by inhibiting irrelevant behaviors and making relevant behaviors more likely. They have evolved along with cognition and are updated throughout a person's life as they interact with the world (Barrett 1998). In human beings it is the hormonal endocrine system interacting with the nervous system that produces emotional experience. As with mind and consciousness, emotions in humans are not fundamentally different from other organisms but differ in their complexity and their ability to be consciously experienced (Panksepp 2005; de Waal and Andrews 2022). In line with the functional view of emotions, an organism's emotional life increases with its progressively complex cognitive experience (Panksepp 2011). This happens because evolutionarily they both arise gradually aiding the species in successfully navigating their environment (Cosmides and Tooby 2000).

2. This view of consciousness does not attempt to give a neurobiological basis for the "hard problem" of consciousness (Chalmers 1995). It does however attempt to pinpoint the evolutionary process from which consciousness arose. How experience arises from neurobiological processes must continue to be debated within other papers.

From an evolutionary perspective, emotional states likely arose from more basic physiological sensations feelings. Examples of feelings are hunger, sexual desire, interest, fear, and joy, etc. The purpose of feelings is to orient an organism's behaviors to their environment based on homeostatic needs and environmental stressors. Because of their immediate importance they are necessarily consciously experienced. Emotions are capable of being experienced both consciously and subconsciously which leads to a more complex conscious experience. Additionally, emotions, unlike feelings, are used not only for subjective orientation, but also for social signaling of internal states. Therefore, an individual can feel anger, and the emotion of frustration when they try not to express their anger inappropriately. Additionally, this is why an organism can feel joy in relation to working towards a goal while experiencing the emotion of hope.

Two emotions in particular orient an organism towards perceived future circumstances, fear and hope. Fear is negatively valenced and is associated with perceived immediately present threats. In response to the threats the organism either chooses the actions of fight, flight, or freeze. Evolutionarily fear-based responses are most associated with behaviors which lead to the least physical pain for the organism. In modern times it is most commonly associated with preemptive avoidance behaviors (Sylvers et al. 2011). These behaviors are often reflexive; in line with characterization of fear as a basic emotion[3] (Ekman 1999; Ekman 2016). There is significant debate in the emotion literature concerning the labeling of emotional and affective states; however, the emotions which most researchers agree have a strong empirical basis are fear, anger, sadness, disgust, and happiness (Ekman 2016).

Hope is a positively valenced temporally motivating state which occurs after an individual identifies a preferred future goal and identifies ways to meet that goal. It has both a cognitive and an emotional component which are mutually reinforcing. The cognitive component is the identification of a preferred future. This differs from fear in that the preferred future outcome associated with fear is merely the absence of a currently threatening stimulus. Whereas the preferred future outcomes associated with hope can be propositional, reached through imagination[4].

3. A basic emotion essentially means that it is fundamental and ubiquitous to emotional experience.

4. What about "false hopes?" All hopes are orientations towards a preferred, and sometimes improbable outcome. The only difference between a hope that is labeled as false or one that is

Hope is associated with, but distinct from, subjective experiences such as optimism, self-esteem, self-efficacy, belief in problem solving abilities, and social support (Syder 2002; Hobfoll et al. 2003). In psychological literature, hope is most commonly measured by the Adult Hope Scale, which is a self-report measure. The scale differentiates hope into two factors, agency and pathways thinking (Snyder et al. 1991; Snyder 2002; Cheavens and Ritschel 2014). Agency is first the identification of a goal and second the subjective expectation that an individual can use pathways to accomplish that goal. Pathways thinking is the planning of how to accomplish a goal, and if necessary, the ability to consider and implement alternative routes to the goal. As the goal gets closer to being accomplished, an individual experiences positively valenced feelings and emotions like joy, happiness, or the emotional state of hope. If barriers arise in their attempts to accomplish a goal, they experience negatively valenced emotions, like anger, fear, or sadness. The cognitive trait of hope is the primary focus of hope research because it has temporal stability, as compared to the more variable emotional state of hope (Snyder 2002). Additionally, individuals who have high hope scores are more likely to navigate around barriers to their goals even when they are not experiencing the emotion of hope because of their propensity for pathways thinking (Snyder 2002; Cheavens and Ritschel 2014).

Considering the complexity of cognition and emotional richness that is required to experience hope, it has only been associated with human beings. Basic emotions and feelings have been identified in other organisms, but many complex states, like hope, have been considered to be special to personhood. This likely has to do with the agential quality of hope. Other organisms are not recognized as moral agents, although human beings do regularly assign agency

---

not is whether it is considered even remotely possible, and whether it is eventually accomplished. To an outside observer the success of the actions legitimizes the methods that were used; meaning a hope that was originally labeled "false" would be retrospectively understood as not "false" if the goal was accomplished. On the other hand, a hope that ends in the ultimate failure of the individual is retrospectively understood as a "false hope." Until the retrospective analysis of the hope happens it cannot meaningfully be labeled as a "false hope."

Delusional hopes which are not based in reality, like those of individuals experiencing episodes of psychosis, may be the best example of a "false hope." However, individuals who experience extreme delusions often have low scores on the Adult Hope Scale, indicating they experience the emotion of hope rarely. This probably has to do with being unable to actualize their goals or imagine alternative pathways towards their completion (Snyder 2002).

to animals when they seem to make a forward-looking plan and carry it out to completion. For example, a pet getting onto the counter and stealing treats.

Additionally, the emotional experience of hope appears to be vital to continuing one's life. Without it one finds the emotional state of despair, which if it continues for long enough, can lead to the wish to end one's life. "The sense of the unmanageable, of helplessness, of invasive negativity about the future is, in fact, one of the most consistent warning signs of suicide" (Jamison 1999, 94). As a morbid irony, after an individual makes it a goal to kill themselves, and decides on the pathway to accomplish that goal, they often exhibit a lifted mood and more energy for about 10 days before the attempt[5] (Snyder 2002). This vital aspect of hope to human consciousness is likely one of the reasons that it has not been viewed as something that is experienced by other organisms, even ones with intellectually or emotionally complex lives.

There is something personal about hope to our conceptions of being human. Thus, the challenge lies in accepting that human experience of the mind, consciousness, and emotion exist on a continuum with other animals, and that all "uniquely human" experiences might also be experienced by a sufficiently complex organism, including hope. Octopuses may be such an organism. They have complex intellectual lives and are capable of establishing future goals while using multiple pathways to accomplish their goals in the face of barriers. Additionally, they have been recognized as sentient which requires the ability to feel both positively and negatively valenced emotions (Birch et al. 2021). Therefore, on the surface, it appears that they have the capacity to experience the motivating state of hope and its emotional counterpart. Considering the importance attributed to hope within human consciousness it is necessary to consider this in more detail.

PART 2: CAPACITIES OF OCTOPUSES

Octopuses are highly intelligent creatures that not only learn, but also play (Montgomery 2015). Unlike humans who have a definitively centralized mind, they have a dispersed mind in which the arms seem to have significant autonomy[6].

5. Thus, supporting the idea that the emotional experience of hope relates to the cognitive experiences of agency and pathways thinking.

6. In fact, 60% of the neurons within their body are in their arms, and 40% are within their central brain.

The metaphor of a conductor with a jazz band is useful when considering the interaction of the octopus's central and distal neuronal structures (Chiel and Beer 1997; Godfrey-Smith 2016, 105). Despite the dispersed nature of its mind, memories are stored within the central brain. Learning which occurs with one arm is capable of being utilized with a different arm (Mather 2021b). Additionally, the front two arms are most commonly used for reaching tasks, and all arms are capable of being used in a visually directed fashion which requires a top-down signaling approach (Mather 2021a). Therefore, while the mind of an octopus is more dispersed than that of a human it is a singular organism that acts upon the world not as nine different minds, but as one central mind which has eight largely autonomous partners.

Therefore, a reasonable question is, do octopuses even have the cytoarchitecture[7] to experience hope? Hope is correlated with activation of the medial orbitofrontal cortex (Wang et al. 2017). The frontal cortex is the seat of executive function and personality and has been identified as the primary cortex differentiating human cognition from the cognition of other organisms (Semendeferi et al. 1997). However, the Cambridge Declaration on Consciousness, concluded that a frontal cortex is not necessary for consciousness, therefore organisms with no frontal cortex, e.g., birds and octopuses, still possessed the neural architecture necessary for complex cognitive affective experience (Low et al. 2012)[8]. Thus, it is reasonable to consider whether it is possible for organisms, like octopuses, to have the experience of hope despite not having a frontal cortex.

Cognitive Capacity- Pathways Thinking

Octopuses are capable of using their sight to locate prey within a jar and use flexible learning to open the lid by twisting (Anderson and Mather 2010). It takes them time to learn this skill, and some must be shown how to do it first by watching a handler do it. This type of manipulation is not one that would be found in the wild; therefore, octopuses must be able to incorporate novel information which is not evolutionarily relevant into mental representations to access later.

Despite being predators, octopuses are regularly prey as well. As such, octopuses are exceptionally patient in the face of danger, and will stay still while

7. Groupings of neurons which together form central nervous system structures.

8. The authors of the declaration were "prominent cognitive neuroscientists, neuropharmacologists, neurophysiologists, neuroanatomists and computational neuroscientists" (Low et al. 2012).

camouflaged for long periods, or remain in their dens when in the presence of a predator. If the predator follows them, they shoot ink leaving a blast of ink behind. One octopus used all these methods to escape a pyjama shark. When that did not work it tried a different tactic. It collected many different seashells and rocks with its tentacles from the ocean floor to make a shield ball. This shield ball was a type of compound object that the octopus employed at multiple points during its life. The pyjama shark kept attacking despite the shield ball, so the octopus tried yet another strategy. It dropped its defensive posture and ultimately ended up riding on the back of the shark (Ehrlich and Reed 2020). The octopus utilized many different pathways towards its goal of not being eaten by the shark. This use of pathways thinking in response to danger is especially adaptive considering the solitary life that octopuses lead. They do not have social groups to keep watch while they are distracted thus their flexible responsiveness to danger is high (Mather 2019b).

Octopuses exhibit pathways thinking when playing as well. This was seen in the wild when an octopus was chasing a school of fish and swatting at them. It made the school of fish move but the octopus did not seem to be trying to eat them, it was merely enjoying exploring its agency with them (Ehrlich and Reed 2020). It is easiest to see octopuses play in captivity. Their high intelligence requires that they are stimulated, otherwise they become agitated and often try to escape (Montgomery 2015). If they are effectively stimulated by being given novel objects and by being handled by their caretakers, then they show significantly greater adjustment (Montgomery 2015).

Octopuses that are given novel objects first try to bring the objects to their mouth. Then they proceed to go through four further identified stages of play with the object (Kuba et al. 2014). First, they explore the object with their arms. Then they pass it between two arms continuously or push and pull the object, or they pull the object along with them. The third and fourth stage of play has to do with the frequency of engaging with the object and the variability of interactions with it. This is an example of pathways thinking because the octopus has a goal of understanding a novel object and then finds multiple ways to engage with that object that are not solely related to its immediate use. As has been shown, pathways thinking is something that octopuses use in all aspects of their lives.

Cognitive Capacity- Distinctive Personalities and the Temporal Experience of Goals

Octopuses not only have complex intellectual and emotional lives, but also a personality attached to it. This is in line with the understanding that mind and consciousness arise from the interaction of an organism's body with its environment and internal states. It is important not to anthropomorphize organisms and attach personalities to something which is just responding reflexively (Godfrey-Smith 2016; Dennett 2019). Yet it does not appear to be that this is the case with octopuses. Personality is a set of consistent behaviors, cognitions, and emotional patterns. Considering the intellectual capacity of the octopus mind, their consistent behaviors, and their emotional expressiveness, it seems necessary to conclude that they do in fact have personalities.

In captivity octopuses use sight to distinguish handlers they prefer from handlers they do not (Montgomery 2015, 52). They make it known when they do not like handlers by spraying them with their siphon. This categorization happens quickly since they were able to discriminate between two unfamiliar handlers who were dressed the same, but either gave them food, or a slight pain instead with increasing certainty over two-weeks (Anderson et al. 2010). As well, octopuses accomplish routine tasks, like where they prefer to penetrate a clam shell with their beak, in individually different but consistent ways reflective of different thinking styles, i.e., personalities (Mather 2008). Their wide variability in personality has even led the Seattle Aquarium to develop a personality test for the octopuses in their care (Montgomery 2015, 52-53). Godfrey-Smith sums up the experimental literature well with the observation, "One message of octopus experiments is that there is a great deal of individual variability" (2016, 54).

Like humans, octopuses have both avoidance goals and desire goals. An example of an avoidance goal is learning to avoid a negative stimulus. Octopuses are able to learn where they experienced a negative stimulus, e.g., pain, and then avoid that situation in the future. This requires the event to be encoded into the octopus's long-term memory and then retrieved at a later date (Birch et al. 2021). Additionally, octopuses have been known to be upset with lights inside and outside their tank, so they shoot water with their siphon to break the lights, thus turning it off (Godfrey-Smith 2016).

Desire goals are also commonly experienced by octopuses. For example, an octopus waited for its handlers to leave for the night, then escaped from its

tank, walked three feet, and went into another tank which housed the flounders. After eating a fish, it would leave the flounder tank and traverse back into its own tank. It did this regularly until it was caught by a handler who arrived early one day (Montgomery 2015; Corpuz 2016). This required future planning, and consideration of entities which were not currently in presence, i.e., its handlers. Leaving the cage and moving into the other tank after people had left for the night required future planning. This cannot be explained by a reflexive response to being hungry because the octopus was fed regularly, and it only left its tank after everyone had left for the night. The octopus had a goal, eating the flounder, which it consistently waited to enact until after it would not be seen moving by its caretakers.

Octopuses have personalities and consider goals in a temporal fashion, but are they self-aware? One of the most common first steps to answering the question of agentic self-awareness is to see whether the organism has self-recognition. The most common test to analyze whether an animal has self-recognition is the mirror test. Octopuses do not pass the mirror test, but questions have been raised as to whether the mirror test is valid for all organisms including those whose vision is not their primary sense (Kohda et al. 2019). Even though octopus visual sensation is effective it is possible that they self-recognize using chemical sensation (Mather 2021a). This is supported by a chemical recognition mechanism that was identified as the reason octopus arms do not become attached to each other, and why octopuses do not treat their amputated limbs as food (Nesher et al. 2014). Research in this area is still ongoing, but it must be acknowledged that there is as of yet no compelling evidence for selfhood in octopuses (Birch et al. 2020). This aspect will be discussed in relation to the experience of hope in Part 3 of this paper.

Emotional Capacity- Sentience and Experiential Reports

Birch and colleagues reviewed over 300 scientific studies and found evidence for the emotional experience of octopuses among other invertebrates (2021). From this study all cephalopods were concluded to be sentient[9]. They concluded

9. The recognition of sentience requires certain countries to give the organisms more importance when considering them in future legal discussions. In 2021 the U.K. recognized the sentience of cephalopods and decapods, e.g., crabs and lobsters. The U.S. has not recognized their scientifically recognized status legally, although "boiling lobsters alive without stunning was already illegal in the U.S." (Baker 2021). Despite being recognized as sentient, within the U.K.

that there is sufficient evidence for a high confidence that cephalopods experience the feelings of pain, pleasure, hunger, thirst, warmth, joy, comfort, and excitement (Birch et al. 2021). As evidence for their emotional experience octopuses often turn a dark red color with skin folding into horns appearing when they are feeling aggressive (Godfrey-Smith 2016, 117; Montgomery 2015, 119-120). They further have been shown to signal this aggression to other conspecifics by making themselves tall and spreading their webs (Scheel et al. 2017). Additionally, octopuses at the Steinhart Aquarium in California get frustrated and act out when they are bored. So, the handlers regularly stimulate them and provide them with novel experiences and objects to keep them content and stop them from acting out (Newitz 2015). Octopuses have the ability for a wide array of feelings which are the precursors to emotional experience (Birch et al. 2021). Emotional expression is still being explored scientifically; however, experiential reports indicate a complex emotional life that influences octopuses' behavioral activity.

These emotional states can be further inferred by octopus dreaming. Octopuses like other cephalopods seem to sleep and even dream[10] (Godfrey-Smith 2016, 133-135). Octopuses rest for a long period of time with a neutral gray color to their skin, which is their sleep stage. Then after some time they begin to change color suddenly, with a pattern very similar to human sleep stages (Malinowski et al. 2021). Consistent with the idea of an emotional life, their skin changes color in similar ways as their skin does when they are awake, i.e., darker red colors with horned skin show more small agitated movements and lighter more gray colors with smooth skin show less agitated movements. The issue of knowing whether an animal is sleeping and does in fact dream is similar in difficulty to whether an animal has consciousness. However, from the experiential evidence researchers conclude it is highly likely that they do dream, but it has yet to be empirically supported beyond a doubt (Godfrey-Smith 2016, 133-135; Nature by PBS 2019; Malinowski et al. 2021).

It is impossible to experience the cognitive emotional life of another organism. Famously, Nagel showed there was no way to understand the consciousness of a bat (1974). However, this inability to experience the consciousness of another extends to other human beings as well. Therefore, all assumptions of consciousness

---

they can still be "sold to untrained handlers, transported in ice-cold water, boiled alive without stunning them and other extreme slaughter methods" (Baker 2021).

10. For an example of this, watch "Octopus Dreaming" (Nature by PBS 2019).

must be made with a cognitive leap; a belief that through collecting behavioral evidence that consciousness can be inferred with a degree of certainty. The sentience designation and behavioral evidence support the idea that octopuses have vibrant emotional lives.

PART 3: OCTOPUS HOPES

While it may be intuitively appealing, it would be a mistake to say that hope is a uniquely human emotion. As has been argued, consciousness comes about gradually, and many animals experience emotional states (Ginsburg and Jablonka 2010; de Waal and Andrews 2022). Therefore, any sufficiently complex organism can experience the motivational and emotional state of hope.

Octopuses are organisms that followed an entirely distinct evolutionary path to human beings. From a shared ancestor of a flatworm both humans and octopuses developed complex neuronal organizations and highly similar visual sensory mechanisms. While their mind is more decentralized, they still have top-down and bottom-up capabilities including, learning, memory, and emotional life. Both empirical and subjective evidence point to the fact that octopuses have distinct personalities and preferential attitudes. More research is needed to understand the extent of octopus consciousness, but they live vivid sensory lives and exhibit complex behaviors which are indicative of intelligent and emotional cognition.

Octopuses are capable of pathways thinking to achieve one's goals and are motivated to accomplish novel future goals in line with their personalities as opposed to a reflexive unconscious way of attaining goals. Since they also feel emotions, it logically follows that while engaging with a future oriented agentic task they could feel a sense of hope that they will accomplish the task. Therefore, octopuses have the capacity to experience the motivating state and emotion of hope while interacting with the world.

Hope has widely been considered to be a uniquely human experience as it is a desire for one's future self. But it ultimately arises from the cellular interactions within human bodies and brains. It may be uncomfortable to think that similar to the analogous structures of a bird and insect wing, neurological complexity leads to the capacity for similar cognitive and emotional experiences between humans and octopuses, especially one as complex and vital as hope. Since octopuses are not self-aware, they will not experience hope in the vital way that humans do, where

living without it is often the reason for suicide. However, self-awareness merely augments the experiential fullness of hope. A cognitive emotional experience can be felt without being understood by the organism experiencing them. Therefore, while octopuses' hopes for their future are likely more rudimentary in nature than humans, e.g., focused on more interesting experiences or preferred foods, they still have the capacity to have desires for their future selves and thus experience hope.

One might suggest that it is impossible to say whether an octopus experiences hope. Considering the fact that they are invertebrates and have a less centralized nervous system they may be too different to humans to experience an emotion like hope. The sentience report showing they can feel many emotions including joy and fear, as well as the ability to develop connections with their caretakers indicates that they share at least some similar experiences with humans (Birch et al. 2021; Montgomery 2015). Ultimately it is impossible for any individual to say that another person experiences hope, yet we believe that they do. This is because of a recognition of the capacity of another human being to feel the experience which we have labeled hope. Human beings are not fundamentally different than other organisms, which is why we are, for example, able to derive conclusions about the efficacy of medicines from animal testing. If our capacities are not fundamentally different from animals', except in terms of increased complexity, then any animal with sufficient capacity of experience should be able to have similar experience. What octopuses naturally lack is spoken language, longer lifespans, and an automatic propensity for social groupings. Therefore, it would not be safe to assume that they had the capacity to feel a social emotion, e.g., shame. Hope is not exclusively a social emotion; in fact, it often originates individually. Since octopuses have the capacity for hope, and capacity, as it is with humans, is the strongest evidence for the experience of hope, then it should be concluded that octopuses do experience hope.

Extending Moral Circles- Moral Patienthood of Octopuses

It may be uncomfortable for people to consider that another organism can experience hope, especially one as seemingly different from us as an octopus. In part, this is because it is easier for humans to empathize with other social mammals because they act similarly and share similar physical characteristics (Mather 2019a). This bias leads to increased attention in media and research attention

being given to mammal species who only make up .2% of total species worldwide and especially invertebrates being discounted in moral considerations (Mather 2019a). Invertebrates make up 90% of global species, but their experiential life is discounted despite it having a wide range of complexity (Horvath et al. 2013). Octopuses, and cephalopods in general, have a vibrant experiential life. Thus, it is a mistake to conclude from this bias that social mammals hold a monopoly on conscious experience and moral consideration.

This suggests that we should afford octopuses more moral rights. They cannot be seen as traditional moral agents because that would require that they understand our societally developed morals. Additionally, their lack of self-awareness does not allow them to be seen as full moral agents. But it is morally required to hold space for them as moral patients. A moral patient is one that is given rights which recognize the responsibility of other moral agents to treat them with concern for their wellbeing. The capacity for hope helps show the need for holding moral space for octopuses because it is an example of a complex motivational and emotional state as opposed to a more basic emotional experience identified when attributing sentience. At present octopuses, like all invertebrates, are not a part of humanity's moral consideration unless the individual's moral circle includes all animals, like Jainists or vegans (Anderson 2019). Octopuses share the capacity to hope with human beings, though their experience of it is likely very different. In time as humans become more comfortable attributing complex cognitive emotional experiences to other species it is likely that many other species will also be found to have the capacity to experience hope. The conditions which must be met are, the organism is capable of cognitive flexibility in achieving future goals and is capable of experiencing emotions. An organism's self-awareness likely increases the vital nature of hope that is experienced by human beings, but less aware organisms can still have the emotional experience of hope arise from its cognitive mechanisms. Octopuses should be treated as moral patients in a way that is reflective of their conscious experience.

Vulnerable populations who have reduced agency, or are unable to act as moral agents, are commonly treated as moral patients, e.g., people in vegetative cognitive states or children. Children are an especially helpful parallel when considering octopuses. They have many cognitive and emotional capacities that develop as they age. It is only through interaction with other more expert individuals, especially parents, that they are then able to put words to their

cognitive and emotional experiences. Therefore, they too experience the emotion of hope in more rudimentary forms until, as they develop, it begins to take on a vital, and often subconscious aspect in their lives. Their capacity for complex experience and human identity makes it easy to hold them as moral patients. Yet a similar capacity for experience also exists within many organisms, including octopuses.

Increased demand for octopuses, and declining levels in the wild have led to the development of octopus farms, which are on track to be fully operational within a few years (Marshall 2021). Only recently octopuses were put on the list of sentient animals according to U.K. law (Baker 2021). Globally they are given very few legal considerations like this. In fact, they are commonly enjoyed fried, i.e. calamari, boiled, or even eaten alive (Gritzer 2019). To hold them in moral patienthood that is reflective of the vibrancy of their experience requires maintaining their individual identities when they cannot. Narratives must be told which allow insight into their lives. This can be done through Oscar winning documentaries, field notes, empirical research, and philosophical scholarship (Ehrlich and Reed 2020; Montgomery 2015; Godfrey-Smith 2016; Mather 2019b; Birch et al. 2021). Social narratives lead to identity constructions that help hold an entity within one's moral circle (Nelson 2002). It is imperative that the experiential life of all organisms, including octopuses, continues to be mapped. Human moral considerations should be updated along with the empirical data and philosophical conclusions. Octopuses have a complex conscious experience which includes the capacity to hope. Octopuses should be treated as moral patients and given legal considerations which reflect the richness of their mind, consciousness, and emotions.

REFERENCES

Anderson, Ross. 2019. "A Journey Into the Animal Mind: What Science Can Tell Us About How Other Creatures Experience the World." *The Atlantic*, March, 2019. https://www.theatlantic.com/magazine/archive/2019/03/what-the-crow-knows/580726/?utm_campaign=the-atlantic&utm_medium=social&utm_source=facebook&fbclid=IwAR3FpK-BxyJFc2Oy91vLc2Cflo6rUPLxVPx0ROq9HmI-mRNWV9sw9wfitBY.

Anderson, Roland C., and Jennifer A. Mather. 2010. "It's all in the cues: Octopuses (Enteroctopus dofleini) learn to open jars." *Ferrantia* 59: 8-13.

Anderson, Roland C., Jennifer A. Mather, Mathieu Q. Monette, and Stephanie R.M. Zimsen. 2010. "Octopuses (Enteroctopus dofleini) recognize individual humans." *Journal of Applied Animal Welfare Science* 13 (3): 261-272.

Baker, Harry. 2021. "Octopuses, squids and lobsters could become 'sentient beings' in the UK." *Live Science*, November 25, 2021. https://www.livescience.com/cephalopods-and-crustaceans-recognised-as-sentient-in-uk.

Barrett, Karen Caplovitz. 1998. "A functionalist perspective to the development of emotions." In *What develops in emotional development?*, 109-133. Boston, MA: Springer.

Birch, Jonathan, Alexandra K. Schnell, and Nicola S. Clayton. 2020. "Dimensions of animal consciousness." *Trends in Cognitive Sciences* 24 (10): 789-801.

Birch, Jonathan, Charlotte Burn, Alexandra Schnell, Heather Browning, and Andrew Crump. 2021. *Review of the Evidence of Sentience in Cephalopod Molluscs and Decapod Crustaceans*. The London School of Economics and Political Science. https://www.lse.ac.uk/News/News-Assets/PDFs/2021/Sentience-in-Cephalopod-Molluscs-and-Decapod-Crustaceans-Final-Report-November-2021.pdf.

Chalmers, David J. 1995. "Facing up to the problem of consciousness." *Journal of Consciousness Studies* 2 (3): 200-219.

Cheavens, Jennifer S., and Lorie A. Ritschel. 2014. "Hope theory." *Handbook of Positive Emotions*: 396-410.

Chiel, Hillel J., and Randall D. Beer. 1997. "The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment." *Trends in Neurosciences* 20 (12): 553-557.

Corpuz, Mina. 2016. "New England aquarium has its own octopus escape story." *The Boston Globe*, April 14, 2016. https://www.bostonglobe.com/metro/2016/04/14/new-england-aquarium-has-its-own-octopus-escape-story/3ShjEIp3tdIAqbLGPLtuSO/story.html.

Cosmides, Leda, and John Tooby. 2000. "Evolutionary psychology and the emotions." *Handbook of Emotions* 2 (2): 91-115.

de Waal, Frans B.M., and Kristin Andrews. 2022. "The question of animal emotions." *Science* 375 (6587): 1351-1352.

Dennett, Daniel. 2019. "Review of Other Minds: the octopus, the sea and the deep origins of consciousness." *Biology and Philosophy* 34 (2): 1-6

Ehrlich, Pippa., and James Reed, directors. 2020. *My Octopus Teacher*. Netflix. 1 hr., 30 min.

Ekman, Paul. 1999. "Basic emotions." *Handbook of Cognition and Emotion* 98 (45-60): 16.

Ekman, Paul. 2016. "What scientists who study emotion agree about." *Perspectives on Psychological Science* 11 (1): 31-34.

Ginsburg, Simona, and Eva Jablonka. 2010. "Experiencing: A Jamesian approach." *Journal of Consciousness Studies* 17 (5-6): 102-124.

Godfrey-Smith, Peter. 2016. *Other minds: The octopus, the sea, and the deep origins of consciousness*. New York: Farrar, Straus and Giroux.

Gritzer, Daniel. 2019. "The best ways to cook octopus: We explore a variety of methods for cooking octopus, including using sous vide, a pressure cooker, and simmering it on the stovetop for perfectly cooked tentacles every time." *Serious Eats.* October 10, 2019. https://www.seriouseats.com/how-to-cook-octopus#:~:text=Octopus%20can%20be%20eaten%20raw,starts%20out%20much%20more%20tender.

Hobfoll, Stevan E., Robert J. Johnson, Nicole Ennis, and Anita P. Jackson. 2003. "Resource loss, resource gain, and emotional outcomes among inner city women." *Journal of Personality and Social Psychology* 84 (3): 632.

Horvath, Kelsey, Dario Angeletti, Giuseppe Nascetti, and Claudio Carere. 2013. "Invertebrate welfare: an overlooked issue." *Annali Dell'Istituto Superiore di Sanità* 49: 9-17.

Hunt, Katie. 2021. "Lobsters and Crabs Are Sentient Beings and Shouldn't Be Boiled Alive, UK Report Says." *CNN*, November 22, 2021. https://www.cnn.

com/2021/11/22/europe/uk-sentient-beings-crabs-octopus-and-lobsters-scn/index.html.

Jamison, Kay R. (2011). *Night falls fast: Understanding suicide*. New York: Vintage.

Kohda, Masanori, Takashi Hotta, Tomohiro Takeyama, Satoshi Awata, Hirokazu Tanaka, Jun-ya Asai, and Alex L. Jordan. 2019. "If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals?" *PLoS Biology* 17 (2): e3000021.

Kuba, Michael J., Tamar Gutnick, and Gordon M. Burghardt. 2014. "Learning from play in octopus." In *Cephalopod Cognition* edited by Darmaillacq, Anne-Sophie, Ludovic Dickel and Jennifer Mather, 57-71. Cambridge, England: Cambridge University Press.

Low, Philip, Jaak Panksepp, Diana Reiss, David Edelman, Bruno Van Swinderen, and Christof Koch. 2012. "The Cambridge declaration on consciousness." In *Francis Crick Memorial Conference, Cambridge, England*: 1-2. http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf

Malinowski, J. E., D. Scheel, and M. McCloskey. 2021. "Do animals dream?" *Consciousness and Cognition 95*: 103214.

Marshall, Claire, 2021. "The world's first octopus farm– Should it go ahead?" *BBC*, December 20, 2021. https://www.bbc.com/news/science-environment-59667645.

Mather, Jennifer A. 2008. "To boldly go where no mollusc has gone before: Personality, play, thinking, and consciousness in cephalopods." *American Malacological Bulletin* 24 (1): 51-58.

Mather, Jennifer A. 2019a. "Ethics and care: For animals, not just mammals." *Animals* 9 (12): 1018.

Mather, Jennifer A. 2019b. "What is in an octopus's mind?" *Animal Sentience* 26 (1): 1-29

Mather, Jennifer A. 2021a. "Octopus consciousness: the role of perceptual richness." *NeuroSci* 2 (3): 276-290.

Mather, Jennifer A. 2021b. "The Case for Octopus Consciousness: Unity." *NeuroSci* 2 (4): 405-415.

Montgomery, Sy. 2015. *The Soul of an Octopus: A Surprising Exploration Into the Wonder of Consciousness*. New York: Simon and Schuster.

Nagel, Thomas. 1974. "What is it like to be a bat?" *Readings in Philosophy of Psychology* 1: 159-168.

Newitz, Annalee. 2015. "Captive Octopuses Get Bored Unless You Give Them Puzzles." *Gizmodo*, October 26, 2015. https://gizmodo.com/captive-octopuses-get-bored-unless-you-give-them-puzzle-1738731062.

Nature by PBS. 2019. "Octopus Dreaming". September 23, 2019. Educational video, 1:48. https://www.youtube.com/watch?v=0vKCLJZbytU.

Nelson, Hilde Lindemann. 2002. "What child is this?" *Hastings Center Report* 32, no. 6: 29-38.

Nesher, Nir, Guy Levy, Frank W. Grasso, and Binyamin Hochner. 2014. "Self-recognition mechanism between skin and suckers prevents octopus arms from interfering with each other." *Current Biology* 24 (11): 1271-1275.

Pandey, Manish. 2021. "Octopuses Feel Pain and Need Legal Protection, say MPs." *BBC News*, June 18, 2021. https://www.bbc.com/news/newsbeat-57528249.

Panksepp, Jaak. 2005. "Affective consciousness: Core emotional feelings in animals and humans." *Consciousness and Cognition* 14 (1): 30-80.

Panksepp, Jaak. 2011. "The basic emotional circuits of mammalian brains: do animals have affective lives?" *Neuroscience and Biobehavioral Reviews* 35 (9): 1791-1804.

Scheel, David, Stephanie Chancellor, Martin Hing, Matthew Lawrence, Stefan Linquist, and Peter Godfrey-Smith. 2017. "A second site occupied by octopus tetricus at high densities, with notes on their ecology and behavior." *Marine and Freshwater Behaviour and Physiology* 50 (4): 285-291.

Semendeferi, Katerina, Hanna Damasio, Randall Frank, and Gary W. Van Hoesen. 1997. "The evolution of the frontal lobes: a volumetric analysis based on three-dimensional reconstructions of magnetic resonance scans of human and ape brains." *Journal of Human Evolution* 32 (4): 375-388.

Shermer, Michael. 2022. "Ogi Ogas– Journey of the mind: How thinking emerged from chaos." Interview by Michael Shermer. *The Michael Shermer Show*.

Skeptics Society, April 2, 2022. Audio, 1:41:51. https://www.skeptic.com/michael-shermer-show/ogi-ogas-journey-of-mind-how-thinking-emerged-from-chaos/

Snyder, Charles R., Cheri Harris, John R. Anderson, Sharon A. Holleran, Lori M. Irving, Sandra T. Sigmon, Lauren Yoshinobu, June Gibb, Charyle Langelle, and Pat Harney. 1991. "The will and the ways: development and validation of an individual-differences measure of hope." *Journal of Personality and Social Psychology* 60 (4): 570-585.

Snyder, C. Richard. 2002. "Hope theory: Rainbows in the mind." *Psychological Inquiry* 13 (4): 249-275.

Sylvers, Patrick, Scott O. Lilienfeld, and Jamie L. LaPrairie. 2011. "Differences between trait fear and trait anxiety: Implications for psychopathology." *Clinical Psychology Review* 31 (1): 122-137.

Tran, Tony. 2021. "United Kingdom Declares Octopuses, Squids Are Sentient Beings: The Country Says the Creatures Can Feel 'Pain, Pleasure, Hunger, Thirst, Warmth, Joy, Comfort and Excitement.'" *Futurism,* November 23, 2021. https://futurism.com/the-byte/united-kingdom-octopus-sentient.

Wang, Song, Xin Xu, Ming Zhou, Taolin Chen, Xun Yang, Guangxiang Chen, and Qiyong Gong. 2017. "Hope and the brain: trait hope mediates the protective role of medial orbitofrontal cortex spontaneous activity against anxiety." *Neuroimage* 157: 439-447.

compos mentis

Asger Jorn, Modified Art, and the Importance of the Relation of the Human Animal to Capitalist Society

Chelsea Schwartz
Indiana University-Purdue University-Indianapolis

ABSTRACT
In this paper I will analyze modern art and its relationship to the human experience and contrast this against how capitalism and the subsequent transition to a consumerist society has changed the nature of art. I will do so by looking at a few different views of art, starting with Leo Tolstoy and Theodor Adorno and following up with a focus on the work and values of Asger Jorn. Asger Jorn was a Danish artist who incorporated a variety of different mediums and applications of art, ranging from written publications and paintings to express his views on modern art and its relation to what he refers to as the human animal. I will showcase the way Jorn's different political and artistic affiliations, including being a founding and key member of notable avant-garde movements such as the COBRA art movement and the Situationists International, incorporated themes of community, expression, and artistic engagement and critique to challenge and bring forth change from the degradation of human life that follows consumerist culture.

KEYWORDS
Aesthetics, Art, Asger Jorn, Situationists International, COBRA, Capitalism, Consumerism, Theodor Adorno, Leo Tolstoy, Avant-Garde, Culture Industry

What unifies all art of the past and present, and thus has allowed different forms of art movements to bridge cultural lines, is that it is a uniquely human experience. In this paper I will focus on the state of the modern art experience and how the transition of art movements has not only paved the way for vandalistic arts such as the modification series by Asger Jorn but highlight the ways that this style is imperative in maintaining the status of art as a universal act of human expression. I will do this by first outlining the ways in which the transgression of art through the industrial and technological revolutions of the 19th and 20th century has left no choice for the art world other than a route of radical ideals to survive and how this is ultimately the most beneficial route for modern art to take. I will showcase Jorn's view on the status of art in contrast to both the previously held views of art and the views of art at the time of his career, in relation to both the Russian writer Leo Tolstoy and his definition of art, as well as a specific mention of the German philosopher Theodor Adorno whose views on the art industry were relevant when Jorn was an active artist. I will then discuss the relation of Jorn to the evolution of art and where it stands today. Finally, I will specifically look at the background of Jorn and show how his unique history and affiliations make him the perfect candidate to represent this revolutionary view, with a specific focus on his piece *Ainsi on s'Ensor (Out of This World—After Ensor, 1962)*, a painting from his modification series that I think best encapsulates this transition, both through its homage to another artist, James Ensor, and simultaneously an homage to the anonymous artist who unknowingly provided the base for which Jorn was able to create this painting. In this way, Jorn exemplifies the importance of the capacity of art to adapt to society while simultaneously challenging it, showcases the importance of the collective aspect of art making, and ultimately defines art as an integral key to understanding what he calls the "human animal".

The main reason it is so difficult to analyze a piece of art and understand its significance is because the very nature of art and aesthetics are so interwoven that it is far too easy to reduce art to the beauty it contains. The field of aesthetics can be thought to represent that which is beautiful but the aesthetic beauty of art is only one aspect, the likes of which is not always helpful when determining the significance of art. In order to break down the true meaning of art, one needs to start by determining how to decide what makes art beautiful. In the late 1800's, Leo Tolstoy rejected the necessity of beauty in art as he addressed the critical question of the nature of art in his aptly titled *"What is Art?"* Here

he evaluates what exactly constitutes a work as an *art*work in particular and its role as a condition of humanity. Tolstoy's definition of art puts an emphasis on the emotional experience of art. He says: "Art is a human activity, consisting in this, that one man consciously, by means of certain external signs, hands on to others feelings he has lived through, and that other people are infected by these feelings, and also experience them." (Tolstoy, 1897, 43) Tolstoy's theory of art characterizes art as something infectious, in which the artist infects the viewer with the emotion behind the art they have created. Tolstoy's view then creates a definition of art that makes art an inherently social act, one essential to human life, that allows us to communicate often complicated emotions and feelings with others across cultural borders and throughout the passing of time. Tolstoy's definition of art is not a perfect one by any means—for starters, it seems to reduce art to whether or not the artist has elicited an emotion properly, which takes away from the time and effort that artists put into the technicality of their craft. Tolstoy's definition is also too limited to be considered as a comprehensive theory of art as it does not take into account the crucial question of how to decide what is art, or more specifically in relation to his view, how to decide what an emotionally charged social interaction looks like in regards to art. Different people can and will have different reactions to different pieces of art and it seems too alienating to implement a definition of art that declares that art can only be good when it has communicated what the artist was intending to emote, instead of allowing for it to be open to different interpretations and emotional responses from different people across different times. While it may be true that art cannot be reduced to a mere transmission of human feeling, it is still important to remember that the inherent emotional labor that goes into art is a crucial aspect of what makes it unique from other technical crafts. I feel it is important to mention Tolstoy's view because it seems then that while there are flaws, it is a good base to start with in relation to Jorn as the themes of both human emotion and human nature in general were ones Jorn repeatedly emphasized in his artistic endeavors.

So then, how does this conception of art hold up, especially in relation to modern art? Tolstoy wrote *"What is Art?"* shortly after the Industrial Revolution had completely changed the economy of Europe and North America alike. The advancement of technology at this time brought in not only inventions such as the automobile and the airplane to allow people to move from place to place faster, but also the advancement of the camera and the progression of the field

of photography. With these modern advancements, the days where art was a necessary means of capturing an objective moment were gone—one doesn't need to paint a person to encapsulate their memory if one can simply take a photo. The photo will not only be produced more quickly than previous methods but will often also be considered a more accurate representation of what that person looks like. As technological advancements were happening at rapid speeds, art movements in the 20th century were moving more quickly as well. The time of art movements lasting over hundreds of years at a time, such as the Renaissance movement, which spanned over the 14th, 15th, and 16th centuries, was over. The art movements of the 20th century lasted for much shorter periods of time, spanning only several years at a time with more overlap between movements. The increasing presence of consumerism in modern society is one likely culprit of this shift as the transition to a focus on capital gain meant that every aspect of human life was becoming faster, more disposable, and, most notably, revolved around the monetary value it contained. In addition to the increase of technology and consumer ideals happening at this time, artists were living in a world where the wounds of two major world wars were still very fresh. Movements such as the abstract expressionist movement, the first internationally renowned art movement originating in America, dominated Western culture and was a way for post-war artists to express themselves and the trauma from the wars in new art which focused on the artist and their expression as opposed to more objective paintings of people and landscapes of the past. Artistic camps varied between those dedicated to furthering this immersion into an increasingly technological form and those that opposed the influx of technology and desired a shift towards an emphasis on the humanistic tendencies behind art. At a time where artists and theorists alike were caught between an influx of new information and unsure of the proper route to take, philosopher Theodor Adorno was a prevalent figure and a hefty critic of what he deemed the shift into a "culture industry", a term he coined to represent this capitalist influence on popular culture at the time in his book *"The Dialectic of Enlightenment,"* written alongside fellow German philosopher Max Horkheimer. The 20th century reality of capitalism was no longer operating under the same constraints outlined by Karl Marx in relation to the beginnings of Industrial Revolution and capitalism. This meant that alienation in the 20th century is no longer seen from the perspective of a worker who is beat down with savage repression but instead is reflected in the ways that the everyday

person has become subdued with an influx of illusions and consumer goods, the value of which is imbued by mass culture and marketing characteristics that they don't actually possess. This transition of society and culture from actual needs to those artificial needs based on capital is the key concept behind Adorno's culture industry critique. This standardization of capitalism determining societies needs has created a society where people are less and less able to both think for themselves and to assess life critically, and therefore have become psychologically dominated by the capitalist infiltration of everyday life. Artists might seem to have gained more freedom of expression with the evolution of modern art but this freedom is an illusion as the artist is now bound to the demands of the consumerist standards in order to create art that society deems beautiful. Adorno provides more detail on what he believes beauty to be by contrasting it with the concept of the ugly, which he outlines in his "*Aesthetic Theory.*" Adorno's view holds beauty to be a harmony which the deformity of ugliness interrupts (Adorno, 1970, 46). For Adorno, this relationship between beauty and ugliness is important in determining a cohesive definition of art and its societal role through the ways that these themes reflect the role of modern art in society. Art cannot be reduced to that which is beautiful because art needs ugliness to reflect the ugliness of society. Beauty is not enough on its own to account for the impact of society on art and actually is a reflection of people's aversion to that underlying ugly truth of the world which makes them uncomfortable. It is important to understand Adorno's theory in relation to Jorn as there are undeniable similarities between the philosophy of Adorno and Jorn's outlook on society and capitalist culture. Adorno and Jorn both believed in the power of art to reflect and criticize society but the main difference lies in the way that Adorno overall holds too far of a pessimistic outlook. Adorno's theory holds little hope for an outcome where art is able to make significant change in this commodification of practical life whereas Jorn remained playfully optimistic about art's ability to construct meaningful societal change. Art may be an ever-elusive topic to both navigate and define, but the beauty of Jorn is that he exemplified the beauty and purpose of art in a way that both revealed its importance and place in society while allowing it to still retain an air of mystery. This places art somewhere in between the realms of Tolstoy's matter-of-fact definition of art as expression while also allowing art to showcase its role in human emotional expression and societal and political critique.

It is important to continue to address and emphasize the change that was happening in this post-war society supported by a dedication to reform and a new capitalist economic basis, with an emphasis on immersing every aspect of life into the constraints of that which could be commodified, and see that art was no exception. The ability of art to overcome the influence of a culture increasingly dominated by this influence can only be produced by an equally resilient form of art, one that emphasizes a shift in the priority of art from one focused solely on the value it contains in relation to a price value by utilizing itself as a means of challenging the norms of that society. Jorn's importance is best seen in the way he sought a different route not only for art but the artist and all of humanity. He co-founded and participated in several coalitions that combined artistic and political elements in an attempt to rally against the increasing themes of banality he felt corresponded with this transition of post-war society. Groups that Jorn was affiliated with include: the Danish Helhesten group (1941-1944), COBRA (1948-1951), the International Movement for an Imaginist Bauhaus (IMIB, 1953-1956), the Situationist International (1957-1972), and the Scandinavian Institute for Comparative Vandalism (SICV, 1961-1965). Jorn's affiliation with these groups, all of which he co-founded in addition to being an active member, illustrate an emphasis on his resistance of the increasingly present influence of capitalism on art. It was through these groups that Jorn built a resistance network through like-minded artists and utilized these relationships to spark conversation on the state of modern art. While the groups differed in views, themes of community, collaboration and expression opposed to consumerism were important themes touched upon throughout his involvement in each organization. In addition to self-exploration, Jorn was a trained artist, having enrolled himself in Fernand Léger's private art school in 1936. One year later, he returned to his home of Denmark where he was enrolled in the Royal Danish Academy of Fine Arts in Copenhagen. Jorn's varying group affiliations with fellow artists in addition to his more formal training highlight the way that he was dedicated to learning and immersing himself in the art world, allowing him to absorb influence from a wide range of styles and people. I will focus on Jorn's involvement with the specific groups of COBRA and the Situationists International, as I think these were his most foundational affiliations and the ones that had the most significant impact on him as an artist and thus reflect the strongest in his artistic output, especially in relation to his views on modern arts role in capitalism.

The first of these was the COBRA group, a group of artists, mostly known as painters but included poets, photographers and anthropologists alike with mutual intentions of criticizing the nature of the state of the post-war society in which it was formed (Kurczynski, 2019, 161). In this era, there was a theme of renewed hope, an emphasis on a fresh start after the end of the horror of two major World Wars. The problem with the mainstream societal view of change then was the emphasis that this fresh start was built around the increasing capital culture that grew from the Industrial Revolution. This culture rooted in institutionalizing every aspect of human life was slowly turning traditionally creative ventures, such as art, into something that could no longer be differentiated from other commercial ventures. COBRA emphasized a restructuring of this culture into an art industry that was no longer an industry but instead a realm where artists could pursue what they deemed true creative and meaningful artistic endeavors. One way in which they attempted to achieve this goal was to replace the traditional Western influence on a solitary artist, an idea which had been increasingly motivated by this influence of capitalism making art more about the monetary value of the artist's labor as opposed to art as a social interaction amongst artists and audience alike, with a view that instead emphasized the importance of collective art making. COBRA as an art group began in an untraditional way—the artists gathered and held discussions, collaborating with each other on their thoughts and ideals just as frequently as they did their art. The group collaborated in this way over the three years in which their coalition spanned and combined their artistic abilities to instill themes of artistic collaboration and elements of experimentation into a society they saw as intent on stifling the genuine creativity of humanity. COBRA showcased their emphasis on the collective by working together on several murals and publications, highlighting the importance of art as a social interaction in contrast to the ideal of individualism. COBRA was much more than just an avant-garde art group but instead a revolution of human nature through the implementation of political theory on art for societal change. It is in this way that COBRA can be seen as an integral part of understanding Jorn and his art and how the themes implemented in this group, although short-lived as a collective, followed him throughout his life.

The second movement crucial to understanding the psyche of Jorn was the Situationists International. This group was initiated out of a dynamic between Jorn as the leading artistic figure and Guy Debord, who was the leading theorist and

wrote *"The Society of the Spectacle"*, a pivotal text in understanding the concept of and beliefs behind the group. In this book, Debord described the current state of society as having been reduced to a spectacle or what he considered a mere image of what it once was. For the Situationists International (here-on out referred to as SI), authentic relationships between beings have been replaced with relationships of ownership, the likes of which is only the appearance of actually owning anything. The commodification of everyday life has replaced authentic human relationships with relationships based on the commodities provided by capitalist infrastructure. The "Situation" aspect of the name came from the group's emphasis on a tactic deemed "Situationism", in which the members attempted to revolutionize against their environment by creating situations that forced the everyday person to be removed from their immersion into the spectacle, even if only for a moment. They used mediums such as graffiti, posters, and road signs to deliver their message, cleverly vandalizing the city, in an attempt to use these alternative forms of art and messages to politically critique commodity culture. Similar to COBRA, the group was known for working as a collective, publishing several articles both as individuals and sometimes as an anonymous group, in their journal *Internationale Situationniste.* While this group started out with a focus on art as a vehicle for political change, the dynamic grew increasingly political with less of an emphasis on art and it is for this reason that Jorn disbanded from the group in 1954 although he didn't sever his ties completely; he continued to offer monetary support to Debord and the group for years after he left (King, 1998, 6-7).

Both of these groups are important not only for the ways they provided artistic influence and sparked political controversy in society but because they are so representative and integral to the evolution of Jorn as both a person and artist. COBRA and the SI are but just two small pieces of the enigmatic puzzle that is Asger Jorn. Understanding Jorn and his beliefs is a path riddled with playful contradictions, much like the theme of playfulness and humor that he imparted in much of his artwork. Jorn balanced the initial trainings of Fernand Léger, a strict mentor of Jorn and an artist with a prominent focus on cubism and modernist art, from an academic standpoint to his eventual transition to a more spontaneous personal style much different than that of Léger. He met his COBRA co-founder Christian Dotrement at the International Congress of Revolutionary Surrealism in Brussels yet went on to later criticize surrealism in a COBRA journal (Kurcynski,

2019, 165-167). His work can be seen as influential on the post-war trend of abstract expressionism yet he was a key member of the SI which rejected this trend, stressing instead an importance on the way art can express itself through persons and their lived life through the groups concept of situations. It is evident that although he is frequently grouped with both the surrealism and expressionist movements, he opposed being associated directly with either of these styles. Jorn also criticized COBRA for a lack of political focus in their art yet left the SI when the Debord insisted the group shift away from a focus on art and the group became too theoretically politically focused. When one looks at the development of both Jorn's personal and professional life, they can see a clear resistance to being pinned down, having never settled into one specific art style nor allowing himself to be boxed into one neat ideology. It is this aspect of Jorn that is both charming and unique, and why I feel he is the perfect candidate to represent the complexities so often attributed to understanding art. Both art and Jorn in particular resist being catalogued because they are nuanced and multi-faceted much in the way that human nature innately is. People grow and evolve in the same way that art does, which is why it is at best a complex and uniquely human experience that, much like human existence itself, should be allowed to experiment and exist for people without the implication of monetary value skewing the priority of it's potential.

It is this resistance to change and emphasis on human complexity that can best be exemplified in Jorn's concept of the human animal, seen encapsulated in a work by Jorn that was discovered posthumously, titled originally in Danish as *"Mennsekedyret"*, which can be translated into English as *"The Human Animal."* Jorn's interest in animal nature and it's relation to humanity can be seen especially in the work produced during the COBRA era, which placed an emphasis on child-like art styles, using motifs of animal representation and themes of Nordic myths to show an opposition to a Humanist emphasis on the importance of Western culture as the embodiment of culture and human nature. Jorn stressed the importance of art in relation to human nature but disagreed with these commonly held beliefs of post-war Humanism that put an emphasis on Western culture as being the standard for that which was good art. COBRA, and Jorn specifically, instead put an emphasis on embracing a more primitivist approach, although not from a standpoint of fetishism that can be seen displayed frequently at this time but instead as a nod to the importance of the inclusion of all cultures, not just those

from the West, in the art world. Jorn's continual emphasis on community amongst artists and appreciation of the influence and contribution of others can best be seen in his Modifications series, in which during the years of 1959-1963 Jorn frequented flea markets and purchased paintings that embodied the Academic style art at the time, art that put an emphasis on Western 20th century trends such as impressionism, and revalued the art with his own modifications.

Jorn's concept of modifying art was a form of *détournement*, a method implemented frequently by the SI, which involved taking a pre-existing image (such as a painting in Jorn's case) and changing the meaning of the original piece by changing it (Jorn, 1959). The way Jorn executed his own personal method of détournement was through his modifying of the thrift store painting purchases he made, combining both a critique of the culture the art represented while simultaneously allowing his re-imaginings of the previous art to show a way that art can instill change in that very culture. Jorn's modification series is also a very critical example of Jorn's unique ability to tread the line between the dichotomy of individual expression combined with collective effort in the spirit of political and social reform. By taking artwork that Jorn felt exemplified the art of the bourgeois, art which represented the very transition of art towards banal, mass-produced pieces, and vandalizing it with his own personal touches, Jorn had incorporated themes of collectiveness, humor, and an aspiration for political change into a unique series. Jorn did not mean for these modifications to be seen primarily as a critique of the work of the former artist but rather used it as an ironic opposition to this increasing influence behind the popularity of that particular style of art, the influence of capitalist exploitation of art. While the overlying theme of Jorn's modification series involved the base of a repurposed piece of art, as stated previously usually depicting some representation of what was considered academic art at the time, such as paintings of portraits and landscapes that represented the kitsch art of the time, and combined with the addition of Jorn's vision, this is the extent to the physical similarities behind the paintings. However different they may appear on the surface, the modifications hold a similar underlying essence as a critique of culture but vary greatly in their physical representation of this matter. The painting from this series that I feel is the most critical to represent Jorn and his ideals expressed not only throughout his lifetime but especially throughout this series is his painting titled *Ainsi on s'Ensor (Out of this World—after Ensor), 1962*. In this painting, we can see what can best be described as a somber scene

of a hanged man which Jorn modified with touches of dark humor—most notably the inclusion of a vulgar mask onto the face of the hanged man, which was a nod to painter James Ensor, who's work was of notable influence on Jorn. It is at this point that I will focus most specifically on how this painting in particular captures the spirit of Jorn as a person and an artist as well as combining the themes of creative collectivity, artistic influence, and political and social reform in a uniquely humorous, ironic and revolutionary way to represent the ties between humanity and human nature and art that Jorn stressed throughout his life.

The first notable theme of this work is the aspect of collectiveness that it incorporates. As displayed by his involvement in several groups of artists coming together to create and inspire, Jorn believed in and stressed whole-heartedly the importance of the relationship art had on the social life through these recurring themes of the collective artist over the individual artist working in solitude. In the bottom left-hand corner of *Ainsi on s'Ensor,* one can make out the signature of what is most likely the original artist of the piece. Jorn often kept the original artists signature intact when modifying a painting, which can be seen as a representation of his commitment to using détournement not as a method of appropriating previous art forms but to pay homage to a forgotten artist while critiquing the institution of which the art represented. Through this, it can be seen that Jorn considered this a way of art to be repurposed and revalued in a way that plays on these themes of the community of art and represented a social form of art, focused on rebellion against the society of which both the original artist and Jorn were apart of. And thus, through this revaluing of the previous work, Jorn highlights this importance of the human relationships behind art, both between the artist or artists and the audience. He shows the way that art is a mode of human expression that can change and adapt, just as the society in which it was created changes and adapts. Jorn states in an essay titled *"Détourned Painting",* published by the Rive Gauche Gallery for an exhibition catalogue: "ALL WORKS of art are objects and should be treated as such, but these objects are not ends in themselves: They are tools with which to influence spectators. The artistic object, despite its seemingly object-like character, therefore presents itself as a link between two subject, the creating and provoking subject on the one hand, and the receiving subject on the other. The latter does not perceive the work of art as a pure object, but as the sign of a human presence." (Jorn, 1959) One can see in this the way that Jorn intended for his détourned paintings to represent a rebirth

of the concept of painting, straying away from the focus on the institution of art while instead highlighting a focus on the social relationship not only between artists but the way that art should be an essential component of the relationship between the artist and their life in general.

Another important theme captured in *Ainsi on s'Ensor*, a theme that is also representative of Jorn's dedication to the importance and influence of artists on one another, is the homage he pays to James Ensor. James Ensor was a Belgian artist of the late 19th and early 20th centuries, producing art associated with the expressionist style prominent at the time. In 1937, a young Asger Jorn was introduced to Ensor through the German art collector Herbert von Garvens, who had an extensive collection of Ensor's work (Andersen, 1994, 93-96). It was this point in time, a pivotal moment in Jorn's career as he was coming into his own artistic style, that the influence of Ensor on Jorn was born. Jorn appreciated the themes of dark humor and angst that Ensor incorporated in his art, the influence of which can be seen throughout much of Jorn's work. Jorn's emphasis on the collective then was not just an importance on artists physically working together but also the ways that artists can learn and adapt through each other. Through the combination of Jorn's academic training, time spent learning and working with many different styles of artists throughout his various group affiliations, and inspiration he drew from others such as Ensor, he was able to create a unique voice in his art, a voice that represented him, his ideals and the influence on his peers on both of those aspects. Jorn strived to use his artistic voice to help transition popular art from its current focus to one that put an emphasis on all artists' ability to have this same personal experience with art.

Lastly, the most important theme I feel that is integral to this piece is in the very element of détournement, specifically in his examination and incorporation of kitsch through the modification of that very style to highlight the problems within the culture it represents. *Ainsi on s'Ensor* is a reflection on aesthetics in its anti-popular aesthetic nature. On the surface one can easily argue that it is not a beautiful painting but it is this unsightly and haunting appearance that draws attention to its revolutionary components that makes it both crucial to the modern theory of aesthetics as well as its relation to society. The design Jorn implemented disrupts the norm and interrupts the viewers thought process, forcing them in to the situation at hand and thus making them evaluate what this means for not only this specific piece of art but art in general. It sparks a social commentary on what

the nature of art is and offers a new idea of art, one that strays from art that is exclusive to those of status and driven by commodity and offers instead a form of art that represents the very heart of art to Jorn and it's importance as a mode of satisfying a basic human need for expression, a need Jorn felt was at the center of all human behavior. Jorn's own spin on détournement through the implication of his modification series was yet another exemplar of his ability to expertly tread the dichotomy between individual expression with the collective effort, as he took a concept used by many in the SI and executed it with his personal style through these modified paintings as a means to illustrate an artistic response which was accessible to a society confronted with the reality of the dilapidated state of the art—a response through which he sought to inspire societal reform. In an excerpt from Jorn's essay, *"Détourned Painting",* he includes the following poem at the beginning:

> Be modern,
> collectors, museums.
> If you have old paintings,
> do not despair.
> Retain your memories
> but détourn them
> so that they correspond with your era.
> Why reject the old
> if one can modernize it
> with a few strokes of the brush?
> This casts a bit of contemporaneity
> on your old culture.
> Be up to date,
> and distinguished
> at the same time.
> Painting is over.
> You might as well finish it off.
> Détourn.
> Long live painting. (Jorn, 1959)

An introduction meant to be read as a tongue-in-cheek reflection on the current state of culture, combining a reference to the famous saying: "The king is dead;

long live the king" as a way to highlight the parallels between the notion that "painting is dead" as a mourning of what painting once was while simultaneously endorsing the transition with the line "long live painting". This ironic contrast through a historical reference highlights the adaptability of art in a fashion that is very true to character for Jorn. Jorn's reference here is a perfect example of the ways he contested the current state of art by use of direct opposition to it through his modifications and through this opposition, attempted to spark a revival of the true spirit of both painting and art as a whole in action. Jorn's claim that painting is dead is an example of his resistance to the existential and unproductive angst associated with both artists and art critics alike at the time in relation to the state of art. Jorn did not truly believe in the death of painting but instead held on to the notion that art could produce meaningful change within itself through itself.

It is undeniable that the world of aesthetics, and its relation to popular society, has always been complicated but has become increasingly so as the lines between the two have become more and more blurred. Art has long been viewed through a superficial lens, a lens that makes it hard to determine exactly how to define what art is at its core. This inclination to reduce art to its aesthetic beauty in addition to the increasing commodification of the everyday has attempted to turn art into something which can be bought and sold much like any other commodity. Asger Jorn resisted this, rallying for a reality in which art can be what it was intended to be, a mode of expression, a place where any average person can express what it means for them to be unapologetically human without the worry of whether or not the outcome will make a good financial project. Our increasingly modern society calls for increasingly modern solutions, and rather than take a defeatist stance of pure pessimism, Asger Jorn maintains his optimism in art as he roots for the average person. He spent his life exploring and creating in an attempt to help facilitate a shift to a society where art can simply be what humanity needs it to be—while simultaneously understanding that what that looks like is as equally complex of an answer as our very nature as human animals. Although Jorn might not have revolutionized the art world completely, he dedicated his lifetime to exploring art as it related to his own human condition and repeatedly attempted to bring about a world where this was the norm through his art. Asger Jorn's art, especially the piece *Ainsi on s'Ensor,* might not always be what society deems as aesthetically beautiful but that is because art does not always have to be reduced to these very rigid standards. The beauty in this piece, and all of Jorn's work, thus

stems from the meaning behind it and how it relates to humanity and our relation to what it means to be human in capitalist society. Jorn's work is refreshing and hopeful in an era where the outlook is perpetually dismal, offering little room for optimism or possibility for change. Beauty can be seen in Jorn's work through its emphasis on the power of creativity to create new social relationships through art. It showcases an accessibility of art, outside of the culture industry in which the everyday person can create art simply to express their inner human animal with no regard for the capital they may acquire.

REFERENCES

Tolstoy, Leo. 1897. *What is Art?* New York, NY: Thomas Y Crowell.

Horkheimer, Max, and Theodor W. Adorno. 1969. *Dialectic of Enlightenment*. Stanford, CA: Stanford University Press.

Shield, Peter. 1998. *Comparative Vandalism: Asger Jorn and the Artistic Attitude to Life*. Farham, UK: Ashgate Publishing.

Kurczynski, Karen. 2014. *The Art and Politics of Asger Jorn: The Avant-Garde Won't Give Up*. Warham, UK: Ashgate Publishing.

Kurczynski, Karen. 2019. "Asger Jorn and Cobra – A Many-Headed Beast". In A Cultural History of the Avant–Garde in the Nordic Countries 1925–1950. Leiden, The Netherlands: Brill. doi: https://doi.org/10.1163/9789004388291_008.

Jorn, Asger. 1959. "Détourned Painting." *Situationists International*.

Jorn, Asger, and Niels Henriksen. 2012. "The Human Animal." *October* 141: 53–58. http://www.jstor.org/stable/41684276.

Debord, Guy. 1977. *Society of the Spectacle*. Detroit: Black & Red.

compos mentis

A Defense of Sally Haslanger's Sociopolitical Account of Race

Neil Van Kanegan
Calvin University

ABSTRACT
In this paper I outline Sally Haslanger's "sociopolitical" account of race, which describes the term in a way that is ameliorative, helpful, and practical for those involved in social justice efforts and racial reconciliation. Haslanger's account isn't the only one in the field of social ontology to pursue these goals, however. Chike Jeffers provides a rival ameliorative account that emphasizes cultural impacts on race, responding and objecting to the differences that he identifies between his account and Haslanger's. Furthermore, I defend the definition that Haslanger proposes from the objections that Jeffers raises, and I proceed to argue that Jeffers' account in turn is subject to a litany of problems that make his definition of race unworkable as an alternative to Haslanger's. These problems arise in part due to the fact that Jeffers seeks to reclaim a positive notion of "white pride," and are compounded by Jeffers' failure to adequately explain how certain types of racism and injustice are excluded from both his reclamation of white pride as a positive and ameliorative term, and from his account of race as a whole.

KEYWORDS
Chike Jeffers, Culture, Ethnicity, Metaphysics, Race, Sally Haslanger, Social Justice, Social Ontology

ACKNOWLEDGMENTS
I would like to thank Dr. Kevin Timpe for his helpful comments on the second draft of this paper, Dr. Christina Van Dyke for her comments and questions during the presentation of this paper, as well as Madeline MacRAE for keeping me motivated throughout the editing process.

I. INTRODUCTION

Like Sally Haslanger's "Tracing the Sociopolitical Reality of Race," a contribution to *What is Race? Four Philosophical Views*, this paper defends a normative and functional definition of race. I move past the idea that there must be a firm consensus on the definition of race, as I believe that debates regarding this consensus permit little progress in establishing practical and functional accounts of the term (Haslanger 2019, 8-11). In this paper, I argue that Haslanger's construal of race as a sociopolitical phenomenon (within the United States) carries strength as a philosophical and functional definition of race.

Haslanger's presentation of race in this essay also stands in contrast with definitions she has previously proposed, as her goal here is strictly explanatory and utilitarian in terms of pointing out and diagnosing racism and the reasons for and results of racialization (Haslanger 2019, 24-25). Earlier definitions proposed by Haslanger were intended to define race "[as a term] whose reference is fixed by ordinary uses, but whose content is discovered empirically using social theory," and thus define race in more theoretical terms (Haslanger 2010, 169). I take on Haslanger's current view because I believe that her sociopolitical account provides the most conceptual clarity and explanatory power in terms of individual agency, cultural diversity, (including cultural outliers to racial groups) and the way in which races arise. I consider objections to Haslanger's account of race by Chike Jeffers, and I provide replies to his objections.

II. BUILDING A FUNCTIONAL DEFINITION OF RACE

Haslanger outlines her sociopolitical account of race through an explanation of "racialization" as follows:

> *Social/Political Race (SPR)*: A group G is *racialized* relative to context C iff members of G are (all and only) those
>
> (i) who are observed or imagined to have certain bodily features presumed in C to be evidence of ancestral links to a certain geographical region (or regions)--call this "color";

> (ii) whose having (or being imagined to have) these features marks them within the context of the background ideology in C as appropriately occupying certain kinds of social position that are in fact either subordinate or privileged (and so motivates and justifies their occupying such a position); and
>
> (iii) whose satisfying (i) and (ii) plays (or would play) a role in their systematic subordination or privilege in C, that is, who are *along some dimension* systematically subordinated or privileged when in C, and satisfying (i) and (ii) plays (or would play) a role in that dimension of privilege or subordination.[1] (Haslanger 2019, 25-26)

The definition that Haslanger provides has many strengths. One initial and readily apparent strength it has is its agency-based language, which describes how agents attribute characteristics to other agents. This language enables Haslanger to describe how individuals interact with and ascribe conceptions of race, as opposed to making group-based generalizations regarding epistemic access and intent, which miss the fine-grained detail of racial interactions in their low-resolution analyses (Pappas 2004, 28).

In Haslanger's view, people are *racialized* by the context they find themselves a part of, particularly by other members of that context. This process of racialization occurs due to the perception by others that certain biological characteristics place them in a group thought to have ancestral links in a certain region. It is important to note here that Haslanger also thinks of this link between person and geography, or person and ancestry as being made by an outside observer, instead of a self-reflecting or self-categorizing agent. This is also to say that in addition to groups, individual people can be racialized if they are perceived to have biological or ancestral links to a racialized group. Expanding from Haslanger's view, this categorization on an individual basis likely occurs mentally, as in a close person-to-person interaction; racialization occurs instantly as the first thing we notice about someone is often the color of their skin.[2] This mental categorization is then

1. Italics Haslanger's.

2. A person's gender expression and level of physical ability may also be among the first things that someone notices when they encounter another person. As this paper is meant to serve the purposes of intersectionality, it isn't my intent to rank order the importance of these attributes or argue that the color of one's skin takes priority in recognition over the other two attributes I

reflected on our interactions with people, including how we treat them, what our assumptions are about them, and what topics we engage (or avoid engaging) them with in conversation.

The overall coherency of this definition can be seen in the natural and fluid way that (ii) flows from (i). Because person X is racialized in context C, they are placed in either a subordinate or privileged group due to the associations of their race with said group. Interestingly, according to Haslanger, the features that ground racialization (though often imaginary) can be enough to "mark" someone as either being privileged or subordinate. In Haslanger's view, perceptions, regardless of their truth, are incredibly powerful in their ability to maintain and uphold societal structures, for better or worse. Additionally, Haslanger claims that this action of marking reinforces the idea that a racialized person belongs to a privileged or subordinate group, tying a distinct form of perception to the perpetuation of two specific sociological categories.

The justificatory nature of racialization and marking is explained in greater detail as Haslanger describes the third part of her definition of race. To Haslanger, these two actions contribute to upholding systemic subordination and privilege in context C, although she denies any causality between marking and the existence of privilege and subordination. I think that she would attribute the causal origin of these two groups to the unfortunate and systemically unjust outcomes of history. The apparent location of persons within these categories may be the result of either generations of slavery and oppression, or the profiteering that resulted from this oppression. Many other historical outcomes may causally contribute to the creation of these two categories. The maintenance of these categories on the other hand, is a result of collective affirmation of the reality of these groups, regardless of the fact that they are reductive and subject oneself to the fallacy of black-and-white thinking (i.e., someone is either privileged or subordinate, there is no in-between). In simpler terms, marking and racialization contribute to the persistence of these two groups but aren't responsible for bringing them about. The aforementioned collective and continual affirmation of the existence of these groups that allows them to persist, while certain systemic historical outcomes caused them to arise.

mentioned. Rather, I intend to bring attention to how we perceive skin color in order to explain Haslanger's account of race and racialization.

In terms of the overall scope of Haslanger's definition, her ultimate goal in proposing her new definition of race, and in exposing the inadequacy of our current understanding, she explains, is to structure an understanding of race on different terms in opposition to common perception and provide an understanding of race through the lenses of society and culture. This debunking approach is also meant to be used by those seeking to identify and root out instances of injustice in their communities, and thus provides a much more ameliorative and functional definition of race that contrasts with theoretical definitions she has posed in the past two decades.

Her most recent account provides a re-assessment of our prior beliefs and applications of those beliefs, causing us to "motivate a new relationship to our practices." Haslanger argues that the current grounding for racialization is flawed due to its reductive grouping and reliance on characteristics and attributes that may or may not exist. Thus, Haslanger debunks our current conception of race and how we racialize people by describing how we think of race, and how racialization occurs (Haslanger 2019, 30-32).

Furthermore, when looking at the content of Haslanger's definition as a whole, we can see that it isn't necessary for an individual to participate in a shared culture, language, or set of practices to be considered part of a racialized group. Seeing as Haslanger defines a race as a group of people who have been racialized to similar geographic origin, ancestry, or the apparent presence of shared physical features, various cultures, whether original (i.e., deriving from a pre-racialized status) or reactionary to racialization, can arise. Racialization is a process that occurs immediately upon seeing someone, regardless of their burqa, kente cloth, sari, or t-shirt.

Here, Haslanger makes a distinction between race and ethnicity. Ethnicities entail those cultural practices as defined by art, language, and geography, and can precede racialization. She argues that the process by which these ethnicities are placed hierarchically in context C is described by racialization. In addition to this, once the imposed racialization has ended, ethnicities and ethnic identities can continue existing. Haslanger develops her definition further on an explanatory basis by attributing greater power to the process of racialization as an explanation for race, when compared with other "cultural" explanations and definitions of race.

Haslanger explains modern racial groups by positing that they are combinations of various ethnicities that have been racialized together as a kind of ethnic amalgam (Haslanger 2019, 27). She also establishes a third category in addition to race and ethnicity to explain how people of various ethnicities have responded to being racialized:

> …there are three relevant types of groups: ethnicities, pan-ethnicities, and races. Ethnicities have distinctive cultures. Races typically consist of people from multiple cultures. Pan-ethnicities emerge when multiple groups are racialized and treated as one group, and form an identity and way of life as a result. So Hmong, Japanese, Khmer and Korean are ethnicities. They are all treated as Asian in the United States, and *Asian Americans*[3] form a pan-ethnicity. (Haslanger 2019, 28)

This third group, the pan-ethnicity, forms a different kind of category than the previous two that Haslanger has described, in terms of its specific origin.

Whereas races are ascribed and attributed by observers separate from the racialized person in question, and ethnicities are established on a basis of community, ancestry, culture, and geography, pan-ethnicities result from the reaction of a group of ethnicities to being racialized. This reaction consists in the affirmation and sense of community that arises between ethnic groups when they are racialized together. This does not mean, however, that all Asians share a cultural identity, as a Nepalese person living in Nepal may not think of themself as being Asian, because they might not have ever been racialized by anyone. What this means is that Asian-Americans as a pan-ethnicity *may* share a cultural identity that they have adapted for themselves, and that Asians outside of America obviously don't fall under the pan-ethnicity of Asian-American, nor might they think of themselves as even being Asian (Haslanger 2019, 28).

Additionally important to note is that shared culture in Haslanger's view, (in contrast with the view of Chike Jeffers) is not a defining feature of race. As mentioned before, people can be racialized based on perceived common ancestry or physical features. Participation in a common culture *can* be a characteristic of a racial group, but only ethnicities and pan-ethnicities are defined primarily in terms of a shared culture.

3. Italics Haslanger's.

Contrary to diminishing the role of culture in race however, Haslanger believes that shared cultural practices on a racial level can be responses to racialization that act as powerful coping mechanisms that can serve a vast array of purposes, often outstripping the need to comfort members of an oppressed group. This isn't to say that culture on a racial level *must* be a response to oppression, merely that this is one form that culture can take in a racial context and thus culture would still exist even in the absence of racialization and oppression. A potential example of this can be seen in the gospel songs that enslaved people in the Americas sang as they worked. This coping method used in response to oppression has since heavily contributed to the thriving contemporary genre of Black Gospel Music, and music from this genre has in turn been sampled by artists such as Kanye West and Chance the Rapper, showing that this response to oppression has culturally outstripped its original intent and is sometimes even being applied to currently pervasive issues of injustice that African-Americans are faced with.

Haslanger also cites that culture (as a whole) is dynamic and versatile, forming from communal interaction and enjoyment, personal identity, and responses to the external world, along with responses to oppression and societal problems. Clearly, we shouldn't reduce culture to a simple etiological result of racialization, although some parts of culture can be interpreted as a result of racialization as we have seen thus far.

Furthermore, Haslanger defends her construction of race by arguing that her definition is merely one of many explanations that can be outlined based on the questions that are asked of race regarding its origin, role, and influence. Although she acknowledges that other descriptive accounts could provide a somewhat accurate explanation of how we currently conceive of race, as mentioned before, Haslanger seeks to provide an ameliorative and aspirational understanding of the concept that aids activists, minorities, and anti-racists in defining social problems and fighting systemic oppression while also helping social ontologists, philosophers, and sociologists to understand where race comes from and how culture interacts with race and is integrated with it. Definitions and clarifications regarding terms such as 'marking' and groups such as 'privileged,' 'subordinate,' 'ethnicity,' and 'pan-ethnicity' help further these goals as well (Haslanger 2019, 29).

Finally, Haslanger argues that our conception of race should allow fluidity between races. In her mind, a just world is one which allows for the fluidity of

members between racial categories as she describes that she comes from a family of both mixed racial and religious backgrounds. This means that Haslanger defines one's individual participation in culture (rather, one's ethnicity, and further, one's racial self-image) as being deeply voluntary in addition to being built and engaged with on a basis of personal volition and agency.

III. JEFFERS' OBJECTIONS TO HASLANGER'S SOCIOPOLITICAL ACCOUNT OF RACE

Haslanger's definition of race does not come without its due controversy. The objections I consider come from Chike Jeffers. From the beginning of his argument in his contribution to *What is Race? Four Philosophical Views*, Jeffers makes it clear that his utmost priority in the metaphysics of race is to create explanations of racial phenomena that are useful in application to societal problems and are able to combat any efforts that "lead us astray in that enterprise" (Jeffers 2019, 176-177). Thus, the projects and methods of both philosophers are similar (as both seek to provide ameliorative accounts) despite the fact that their conclusions differ. Although Jeffers sometimes finds a reasonable degree of similarity between his cultural construal of race and Haslanger's sociopolitical view, he highlights an array of differences.

Jeffers states that he and Haslanger agree that the project of defining race should be to clarify terms regarding racial phenomena that have typically been taken to have bases in genetics and biology. Furthermore, he argues that his methodology is similar to Haslanger's in that it seeks to devise a definition of race that serves a practical purpose through re-evaluating the way race is commonly thought of while making suggestions for improving or eliminating these prior definitions. Another goal that Jeffers and Haslanger share is to "compare different metaphysical stances on the nature and reality of race by asking what significance they accord to our differences in appearance on the basis of ancestral place of origin" (Jeffers 2019, 192). An illustration of this goal can be seen in how Haslanger asserts that perceived characteristics can be powerful and hold perceived reality regardless of whether these characteristics objectively exist. In other areas however, as will be seen, Jeffers' and Haslanger's goals differ to some degree.

In Jeffers' mind, the key difference between his construal and Haslanger's is not the methodology of their inquiry, but rather the conclusions that are reached as a result of this inquiry. Both philosophers arrive at social constructivist conclusions agreeing on the point that race is fundamentally a phenomenon that arrives via social forces instead of being grounded in biology. Jeffers' conclusion however, has a cultural bent while Haslanger offers a more social or sociopolitical explanation of this phenomenon. This isn't to say however that Haslanger's definition doesn't involve culture.

While Haslanger addresses how culture arises within races and how culture interacts with race, she distributes explanatory power in racial contexts to other factors such as perception, marking, and categorization. It's merely the case that Jeffers emphasizes culture more and gives it a greater role in his explanation of race as well as in his objections and replies to Haslanger's main contribution in *What is Race? Four Philosophical Views*. Jeffers does this in order to emphasize the idea that race is an emergent expression of culture which responds to racism and racialization in an identity-shaping manner.

Jeffers also identifies some differences in terms of the overall aim of his project when compared to Haslanger's. While both Jeffers and Haslanger seek to disrupt old ways of thinking and re-evaluate our relationships to our practices, Jeffers' primary goal is both to destroy any grounds that a particular racial group might use to claim superiority, and to reorient racial practice towards revealing diversity and the fertile and fruitful dialogue that comes from discussions of race (Jeffers 2019, 192-193).

Jeffers clearly defines his cultural constructionist view early in his response to Haslanger:

> The reason my view can be identified as a kind of cultural constructionism is because it takes culture to be fundamental from a normative standpoint, for I hold that the value of cultural difference is the reason we may value race and hope to see it live on indefinitely, rather than take its destruction to be our goal, at least in the long run. (Jeffers 2019, 194)

Jeffers takes this to be a major point of difference between his construction of race and Haslanger's. For Haslanger he argues, culture arises naturally from ethnicities, but arises somewhat artificially from races due to a response to racialization. On

the other hand, he thinks that culture is inherent to race and is defined by it. He also thinks that race isn't a negative and should be valued for its ability to foster culture and community. This is Jeffers' position because he sees race as an expression of cultural variation among people. Thus, in a way culture and race are somewhat inseparable in his view.

Jeffers also sees no need to include Haslanger's categorization of the pan-ethnicity in his philosophy. Instead, he refers to ethnicities as well as races and wants to do away with the additional category. He says that his main challenge in discarding pan-ethnicities is explaining the importance of culture to race. It also seems that Jeffers believes that races as groups or categories do all the explanatory work that Haslanger's pan-ethnicities do, so for Jeffers it would be pointless to posit the existence of an additional group. Additionally, Jeffers objects to an understanding of races as units containing individuals who don't share the same culture. Jeffers argues that ethnic groups can be broadened, revealing the way we linguistically refer to unified cultures. Using the example of an "eastern" ethnicity, we can refer to Gujarati culture. Expanding from there, we can refer to Indian culture. Expanding from there, we can refer to Asian culture and so on (Jeffers 2019, 196).

Surprisingly, Jeffers points out that his construal of race as it interacts with culture could be used to justify an idea of white pride or pride in any racial identity for that matter. Jeffers argues that "the end of racism [doesn't require] the end of whiteness." Thus, Jeffers holds that culture is fundamental to a philosophical and self-evaluative conception of race and that "white people" (along with members of other races, respective to their own cultures) "should cherish white culture." Jeffers clarifies, however that the kind of "white pride" his conception of race justifies is much different than the modern-day alt-right and white supremacist version.

Instead, an ideal white pride would consist of white people being able to appreciate past and current white culture in light of collaboration with people of color in dismantling systems of oppression. This version of white pride (here using Jeffers' terminology) is divorced from any conception of white superiority. Jeffers also argues that a goal for modern social justice movements should be to redefine racial pride in terms of helping other racial groups in eliminating oppression. This said, there are no normative descriptions or evaluations in Jeffers' view of white

pride that explain how this form of pride might aid in doing this (Jeffers 2019, 198-199).

Jeffers also objects that Haslanger's concerns (about the exclusion and conformal pressure that would be incurred were race to be defined primarily in terms of culture) are unwarranted. He posits that once racism is eliminated, there will be no need for oppressed races to form resistant and reactionary cultural movements and hence, members of these races won't feel forced or pressured to participate in these reactionary movements. He makes this claim to support the argument that culture is a necessary aspect of race and should be viewed positively in its functional integration with race (Jeffers 2019, 200-201).

IV. MY REPLY TO THE OBJECTIONS OF JEFFERS

I believe that the proposed alternative understandings and objections leveled against Haslanger's definition of race as provided by Jeffers are ungrounded and flawed. In this section I argue this point seeking to defend Haslanger's definition of race, her conception of the role of culture, and the way she thinks culture interacts with race. Furthermore, I believe that Jeffers gives culture too large of an explanatory role in his conception of race as outlined in his response to Haslanger.

To begin my defense of Haslanger's goals, I think it is helpful to understand where Jeffers' own goals go awry so I can examine how the rest of Jeffers' objections fail. Primarily, Jeffers aims to strike down any definition of race that could be used to support the idea of the supremacy of a single race. I think that this goal is very ambitious and admirable and should be a key aspiration of any definitional account of race but where Jeffers is concerned, we encounter an inconsistency. This inconsistency is formed by Jeffers' later affirmation of ideas of racial pride. Jeffers argues that racial pride isn't something to be eliminated but kept and revised. While I agree that having pride in one's race, and further, pride in the achievements of one's own culture are a benefit, I think that Jeffers fails to adequately explain how this racial pride won't devolve into racial supremacy which he is clearly opposed to.

While Jeffers seems to make an attempt at anticipating and rebutting this objection, he fails to adequately do so. He tries to explain away this objection by simply stating that "the possible persistence of white cultural identity I countenance is necessarily divorced from the widespread treatment of whiteness

as supreme." While Jeffers also says that his view "ought to be rejected if it gives non-accidental support to white supremacist calls for white pride," (Jeffers 2019, 199) I believe his view *accidentally* fails to rule out white supremacist calls for white pride and thus incorrectly conceives of a way that racial identity and culture should be reconciled. This entailment of Jeffers' account should not be disqualified from discussion on a basis of the metaphysic seeing as Jeffers seeks to provide an ameliorative account of race, which is hampered and harmed by Jeffers' failure to exclude racist versions of white pride from his conception of this term.

The failure I reference is encountered when Jeffers attempts to explain what a redeemed white pride will look like. He claims that in a post-racist future, white people won't draw on ancestral/national history or perceived achievements of their race for sources of pride but will (or rather should) take pride in standing side by side with other racial groups in fighting against racism and systemic oppression. I would certainly say that Jeffers paints an optimistic and glorious picture of the future of racial pride (which he affirms for all races, not merely Whites) but how will our conception of racial pride make this radical shift?

Current conceptions of racial pride are based on histories and cultures that are rich and complex. Some white people credit themselves with being the founders of Christianity and the western world. Some Asians take pride at the sight of the Taj Mahal or the Great Wall of China. Is it certain that fighting racism could provide this same richness and complexity? Jeffers seems to have nothing to say on the matter. We would assume based on his previous position that Jeffers would answer "yes" in response to this question but based on the implications of his account, it doesn't seem likely that his definition can support this assertion. How he would justify the idea that fighting racism can enrich cultural identities in this way is a complete mystery.

Jeffers stumbles on a linguistic issue as well. He wisely believes that using the terminology of "white pride" has the potential to be harmful for current discussions of culture and race so he accordingly understands why his talk of racial pride could be met with some apprehension. However, he looks forward to the day when all racial pride can be cumulatively embraced and affirmed as he thinks that this will also entail the redemption of the terminology of racial pride. How this transformation will occur is left ambiguous as well. As a whole, Jeffers provides an account of racial and social identity that fails due to an inability to exclude certain types of racism and injustice.

As a point of defense for Haslanger, I think that Jeffers' claim that culture is "fundamental from a normative standpoint," and his claim that races should be preserved because they embody cultural differences don't represent vast gulfs between his position and Haslanger's (Jeffers 2019, 194). Certainly under Haslanger's system it can be said that culture is normative as Haslanger evaluates ethnicities, pan-ethnicities, and races in terms of their relations to culture. Additionally, Haslanger can describe a greater quantity of phenomena using her three categories (providing accounts of culture on all three levels) when compared to the explanations provided by Jeffers' categories of race and ethnicity. Haslanger would certainly agree that races embody cultural difference however, we wouldn't be in danger of losing our culture should racial categories be eliminated, in her view.

Racialization plays a key role here as well. Can we currently consider an idea such as "Hispanic culture" without racializing Hispanic/Latinx people? I don't think we can. Furthermore, Haslanger's conception of race appears more dynamic as races only appear to have corresponding cultures due to the artificial and created nature of the racial category itself. Thus, how can we say that culture is "fundamental" to race?

I believe that Jeffers' intent to eliminate Haslanger's category of the pan-ethnicity would be a mistake as well. The strength of this category is that it preserves and values the individual volition and will of a racialized agent. This account of agency helps document how members of differing ethnicities respond and adapt to being racialized, creating a narrative of a unified, pan-ethnic culture that reveals how agents interpret and reflect on their individual cultural experiences and practices in light of their place within one or more communities. The kind of linguistic expansion that Jeffers discusses appears to play a similar role only that it reveals the way in which we as racializing people think about the identity and race of others (Jeffers 2019, 196). This exploration by Jeffers doesn't have the same power of Haslanger's account of the pan-ethnicity which describes how *racialized people think about their own identity.*

Finally, I think that Haslanger's concern is warranted. If culture were to be the sole cause for race, then people would be excluded from racial groups for cultural non-participation. The argument that Jeffers puts forward stating that in a post-racist society there will be no reactionary movements, and hence no feelings of exclusion from these reactionary movements is concerning, as it fails to

explain why reactionary cultural movements will be done away with. As previously mentioned, cultural practices that react to oppression and injustice have adapted and evolved but show no sign of fading away.

Haslanger acknowledges that these reactionary movements can evolve into cultural waves centered merely around the enjoyment, entertainment, and ritual practices of the members of a racial group, existing independently of any oppression or reaction towards systemic injustice. Jeffers would be wrong to assert that exclusion couldn't occur in this kind of cultural environment. A powerful counterexample to Jeffers' assertion of the non-existence of racial exclusion in a post-racist society can be seen in music. I have heard Asian-American and African-American parents deride their children for "listening to white music." This accusation doesn't reprimand the child for being unwilling to participate in the project of reactionary cultural movements (although it may in some contexts). Rather, it makes the assertion that the child isn't satisfied with their own culture and is trying to supplant it with another. We can see from this counterexample that exclusion from racial groups can occur with or without the existence of reactionary cultural movements within races.

V. CONCLUSION

In this paper I have sought to defend Sally Haslanger's sociopolitical construction of race as argued in "Tracing the Sociopolitical Reality of Race," a contribution to *What is Race? Four Philosophical Views.* I have argued against the many and varied objections of Chike Jeffers as published in "Jeffers' Reply to Glasgow, Haslanger, and Spencer," a contribution to the aforementioned volume. I have evaluated and compared Haslanger's account using the metrics of explanatory power and conceptual clarity when applied to cultural diversity and outliers to cultural groups, individual agency, and the origin of race as a category that people are placed into. Finally, I have concluded that Sally Haslanger's definition of race is sound as a functional and philosophical explanation of the term, standing up to the objections of Jeffers.

REFERENCES

Haslanger, Sally. 2010. "Language, Politics and 'The Folk': Looking for 'The Meaning' of 'Race'." *The Monist* 93 (2): 169–87.

Haslanger, Sally. 2019. "Tracing the Sociopolitical Reality of Race." In *What Is Race? Four Philosophical Views*, 4–37. New York: Oxford University Press.

Jeffers, Chike. 2019. "Jeffers' Reply to Glasgow, Haslanger, and Spencer." In *What Is Race? Four Philosophical Views*, 176–202. New York: Oxford University Press.

Pappas, Gregory. 2004. "Distance, Abstraction, and the Role of the Philosopher in the Pragmatic Approach to Racism." In *Pragmatism and the Problem of Race*, 22-32. Bloomington: Indiana University Press.

cognethic.org