A PRIORI

Brown University Undergraduate Journal of Philosophy Vol. VIII

A Priori :

Brown University Undergraduate Journal of Philosophy

Volume VIII

Brown University
Department of Philosophy
2025

Masthead

Letter from the Editor

Dear Readers,

On behalf of the A Priori Editorial Board, we are excited to present volume VIII of A Priori: The Brown University Undergraduate Journal of Philosophy. Since Brown undergraduate students founded A Priori in 2016, the journal has seen growth and improvement every year. This progress is due in no small part to the hard work of our amazing team. Our editors, readers, designers, and web developers have made A Priori into a journal that we are proud to be a part of and to stand behind.

Thank you to everyone who submitted papers to the current volume of A Priori. This pool of paper submissions was our most competitive year to date. We received over one hundred outstanding submissions, which made adjudicating between papers quite challenging. Given the fierce competition this year, we would like to encourage the authors who we could not publish this year to apply again next year and to continue pursuing philosophy in the ways that are meaningful for them.

To those who we are able to publish in this volume, thank you and congratulations Not only were all of your papers outstanding, but they also complement each other to produce a cohesive yet diverse journal. We are immensely grateful for the diversity and rigor of your thought and wish you the best regardless of whether you plan to pursue philosophy moving forward.

We also wish to thank the Brown University Philosophy Department and Brown's Philosophy Department Undergraduate Group for supporting this journal. Without this support, we would not be able to print and distribute physical copies of A Priori. In addition, we are grateful to everyone who has assisted A Priori, especially Katherine Scanga and Kalum Alldredge, for their administrative support, Professor David Christensen for his efforts to re-establish A Priori after a brief period of inactivity, and every student on the A Priori team for their hard work and commitment to philosophical discourse.

A Priori Editors-in-Chief
Lucy Cooper-Silvis,
David Tapper,
Katherine Warren
Providence, Rhode Island, May 2025

Table of Contents

Stoic Compatibilism: Uncovering Moral Responsibility in a Deterministic World

Zach Weis

Abstract: This paper addresses a puzzle in Stoicism. While the school emphasizes virtue, control, and responsibility, it also maintains that all events are "fated" or "determined." This apparent incongruency in Stoicism has led to an ongoing scholarly debate about the tenability of Stoic compatibilism. In this paper, I insert myself into this debate. I do so by analyzing Stoic compatibilism through the lens of Frankfurtian compatibilism, asserting that even without the possibility to act otherwise, the Stoic understanding of freedom creates rational grounds for moral responsibility. I contend that even though the Stoics believe that our actions are causally determined, their understanding of rational assent allows actions to be attributable to agents. I suggest that this was the primary concern for the Stoics and that forcing contemporary understandings of free will upon ancient discussions of moral responsibility is an anachronistic retrojection that impairs our ability to understand ancient thought.

Introduction

Stoicism is one of the most practiced, yet most seemingly paradoxical philosophies. On the one hand, it professes to offer practitioners control over their lives. By discerning what is inside of their control from what is outside of it, followers of Stoicism are afforded a better sense of where to direct their energy. On the other hand, Stoics argue that the world follows a causal order, often referred to as causal determinism. This chain of causality is driven by universal reason, or *logos,* and it is, at all times, unbroken. This raises an important question: How can individuals truly be responsible for their actions, and thus assume control over their lives, if every choice they make is part of an unbroken causal chain?

This dilemma connects to the broader philosophical concept known as compatibilism, which allows for moral responsibility in a determined world.

Compatibilists believe that "freedom is nothing more than an agent's ability to do what she wishes in the absence of impediments that would otherwise stand in her way."[1] In this way, Stoics contend that moral responsibility is compatible with a determined world because freedom does not require choices to occur spontaneously, that is, without an uncaused source.

However, the validity of compatibilism has been a matter of contention for thousands of years, and the inherent tension between moral responsibility and causal determinism has remained an important question in discussions of Stoicism. Even today, scholars engage in a lively debate regarding Stoic compatibilism.[2]

In this paper, I insert myself into this debate. I argue that Stoic compatibilism is tenable because it creates grounds for moral responsibility in a deterministic world by showing that people's actions are a consequence of their internal states regardless of whether they possess the ability to act otherwise.

I do this by first examining causal determinism and establishing an understanding of Stoic psychology (i.e., how actions are attributable to individuals). Then, I argue that Stoic compatibilism fits within Frankfurt's framework of compatibilism and creates reasonable grounds for moral responsibility in a deterministic universe.

Causal Determinism in Stoicism

Determinism has been an important facet of Stoicism since its inception, though the term itself does not appear in antiquity.[3] It is, most simply, the idea that all events, including human actions, are caused by preceding events and preexisting states. Famously, Chryssipus laid out the inchoate base for Stoic determinism as he argued that nothing can move without a cause.[4] Intuitively, we understand that if something is moving, then some other thing must have caused that object to begin moving. Using this same logic, Chryssipus argued that believing that our actions come about spontaneously would be the same as believing that objects could be the source of their own movement. In turn, he posited that every event, including every human action, comes about as a result of antecedent causes.

The Stoics believe that each set of antecedent causes may lead to only one future. Naturally, the unpredictability of the universe seems contrary to this sentiment. Not only do different people respond to the same stimuli in different ways, but the same person may also respond differently at various points in their lives. However, the Stoics believe that the human capacity for rationalism, or what drives the decisions they make,

1 Michael McKenna and D. Justin Coates, "Compatibilism," The Stanford Encyclopedia of Philosophy, eds. Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Spring 2024, https://plato.stanford.edu/entries/liberty-positive-negative/.

2 Dorothea Frede, "Stoic Determinism," in *The Cambridge Companion to the Stoics,* ed. Brad Inwood (Cambridge University Press, 2003), 179.

3 Frede, "Stoic Determinism," 180.

4 Suzanne Bobzien, *Determinism and Freedom in Stoic Philosophy* (Oxford University Press, 2001), 38.

is fundamentally a part of the causal chain. In other words, despite possessing the ability to rationally contemplate actions, human actions are as causally determined as every other part of reality. The Stoics believe that when humans act, they could not act in any other way than they do.[5]

Historically and contemporarily, the conclusion that all actions are predetermined or fated seems to imply that people cannot be held morally responsible for their actions. If someone could not have acted differently than they did, how can we blame them for their action? The Stoics hold that despite their deterministic worldview, moral responsibility remains possible. They argue that while human rationality is part of the causal chain, actions still stem from their actor, rather than from an external cause. The idea that human choices are attributable to us as agents, rather than to external causes, forms the backbone of Stoic compatibilism and creates rational grounds for responsibility despite humans lacking the capacity to act differently than they do.

Nature and That Which Depends on Us

Contemporary applications of ancient Stoicism place a major emphasis on a "dichotomy of control," or an understanding that some parts of the world are beyond human control while others are within it. While not explicitly an aspect of Stoicism, the dichotomy of control most closely relates to the notion of "that which depends on us."

Determinists and non-determinists alike recognize that many aspects of the world are often determined by causes beyond human control: the weather, genetics, the inevitability of death. The Stoics believe that these aspects of the world do not depend on us; however, they do form an important facet of Nature. The idea of Nature is perhaps the most important aspect of Stoicism. It encompasses all things and is driven by *Logos*, or the guiding principle of universal reason. Indeed, the key goal of Stoicism is famously imagined as aligning oneself with Nature.[6] But how can humans aim to align themselves with Nature if they themselves are part of it? And how does this contribute to an understanding of Stoic compatibilism?

The answer to both questions is found in the Stoic understanding of fate. A comprehensive understanding of the term requires far more time, space, and analysis than this paper can offer, but for the discussion at hand, only a basic understanding is necessary. Importantly, fate is both eternal and active. While fate is teleological in the sense that all events follow its order, it is also active in the sense that living, rational beings participate in its creation. Like actors in a play following a pre-written speech, our active participation in fate creates reality.

It is this understanding that defines how humans can and should live in accordance with Nature: "Connected with the eternal predetermination of all motions and states is the point that it is pointless to make attempts to influence or change one's fate, since at any time all future occurrents have been determined already, and are unchangeable."[7] Living

5 Tad Brennan, "Stoic Moral Psychology," in *The Cambridge Companion to the Stoics*, ed. Brad Inwood (Cambridge University Press, 2003), 265.

6 Lucius Annaeus Seneca, *Letters from a Stoic*, trans. Robin Campbell (Penguin Books, 1969), 37.

7 Bobzien, *Determinism and Freedom,* 49.

in accordance with Nature, then, is living without resistance to fate. It is recognizing that all things are as they should and will be.

Returning to the idea of that which depends on us, we can connect this understanding of fate to Stoic compatibilism. Chrysippus laid out that something depends on us if we are causally responsible for it occurring or not occurring.[8] While something may have been fated to happen, it still depended on us in the sense that we were an active agent in bringing it about (e.g., A cup of coffee may have been fated to spill, but it depended on us in the sense that we dropped it).

Hundreds of years after Chryssipus, Epictetus revisits the idea of that which depends on us, and he approaches the idea in a different light than earlier Stoics. Namely, he focuses on the sorts of things that depend on us and those that do not. He begins *The Enchiridion* by stating,

> Some things are in our control and others not. Things in our control are opinion, pursuit, desire, aversion, and, in a word, whatever are our own actions. Things not in our control are body, property, reputation, command, and, in one word, whatever are not our own actions. The things in our control are by nature free, unrestrained, unhindered; but those not in our control are weak, slavish, restrained, belonging to others.[9]

For our discussion of Stoic compatibilism, the most important part of this quote is the use of the word "free." Epictetus was the first Stoic to connect what depends on us with the Greek word for freedom, *ελευθερία*.[10] Today, we are said to be free when we possess the ability to make a choice–that is, we are free when we possess the ability to act in two different ways. As discussed, the Stoics fundamentally did not believe that we possessed this ability.

As such, we must consider the historical and linguistic context in which Epictetus uses the word free. Instead of using free in a personal sense, the Greeks used it in a political sense to mean free from tyranny. For Epictetus, acting freely meant acting unconstrained by external forces, not acting outside of causal determinism.[11]

The contrast between negative and positive freedom (or liberty) captures this difference. One source describes this difference, writing, "Negative liberty is the absence of obstacles, barriers or constraints. One has negative liberty to the extent that actions are available to one in this negative sense. Positive liberty is the possibility of acting—or the fact of acting—in such a way as to take control of one's life and realize one's fundamental purposes."[12] Given the Stoic (and especially Epictetus's) emphasis on recognizing what is outside of our control, negative freedom more aptly fits into the Stoic imagination.

---

8 Bobzien, *Determinism and Freedom,* 331.

9 Epictetus, *The Enchiridion*, trans. Elizabeth Carter, 135 C.E., https://classics.mit.edu/Epictetus/epicench.html.

10 Bobzien, *Determinism and Freedom,* 331.

11 Bobzien, *Determinism and Freedom,* 340.

12 Ian Carter, "Positive and Negative Liberty," The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta, Metaphysics Research Lab, Stanford University, Spring 2022, https://plato.stanford.edu/entries/liberty-positive-negative/.

Finally, that which depends on us is an important concept because it plays a crucial role in the Stoic understanding of moral responsibility. The things that are directly in our control—meaning those things that occur unconstrained by external physical forces—are subject to internal, rather than external causes. In his earlier quote, Epictetus laid the groundwork for an understanding of those things which are directly attributable to us as agents, which is fundamental to understanding the things for which we can be held morally responsible. However, without an understanding of Stoic psychology, it would still seem as though even those actions which depend on us would be attributable to external, rather than internal, causes since we were fated to act as we did. To understand why the Stoics believe our actions are our own, we must understand the role of assent in Stoic psychology.

Impressions and Assents

The Stoics spent a considerable amount of time explaining the dynamics of internal causes. This is often referred to as the psychological branch of Stoicism; however, given the metaphysical nature of this branch, discussions are hardly beyond the philosophical realm. Impressions and assents are the main features of Stoic psychology.

Impressions refer to "a certain alteration or change in the mind."[13] This is a broad definition because impressions refer to a broad category. They may be sights we witness, events we experience, ideas we conjure, or any other thing which calls for a response from us.[14] Importantly, impressions are thought to be outside of us; they do not depend on us. It is only through our assents that impressions lead to action.

Most simply, an assent is the acceptance of a particular impression that takes place between the reception of a stimulus and our response.[15] It is the moment we can consider how we should act, or whether the impression we receive is a true one. An example can be helpful here. Imagine you are driving down the freeway, and someone cuts you off. The impression that you receive here is that you have been cut off. This may bear with it the proposition that you should, in response, get angry. If you do get angry, the Stoics would say that this was not a result of the impression you received, but rather because of the assent to the impression. If you were a patient person, the proposition that you should get angry would have been one to which you would have withheld assent, that is one you would not have acted upon.

There are two implications of this discussion of assents with respect to moral responsibility and freedom. First, Stoics do not believe that events determine our actions; rather, our assents determine our actions. Marcus Aurelius writes in *Meditations*, "Take away thy opinion, and then there is taken away the complaint, 'I have been harmed.' Take

---

13 Tad Brennan, *The Stoic Life: Emotions, Duties, and Fate* (Oxford University Press, 2005), 53.

14 There is some disagreement about whether impressions are best understood to be external or internal (e.g., Brennan, *The Stoic Life*, 54). In this paper, I treat impressions as outside of oneself, even if it happens to be a thought that crosses the mind. In other words, impressions are not in the realm of things that depend on us.

15 Brennan, *The Stoic Life,* 52.

away the complaint, 'I have been harmed,' and the harm is taken away."[16] Here, Aurelius shows that events themselves do not carry value; rather, their value is something we add. This is to say that our assents are in our control, but impressions are not. This allows the Stoics to attribute actions to individuals rather than external events.

Second, the Stoic belief in freedom comes directly from our capacity for assent. One source describes assents as "the source of human freedom . . . something non-human animals do not have."[17] If we did not have the ability to assent or withhold assent, then there would be no way to understand our actions beyond the events that caused them. Indeed, Stoic compatibilism would become indefensible without our capacity for assent because events would fundamentally determine our actions. However, the capacity for assent allows our actions to be determined by us, rather than external events, since we can withhold assent from an impression. With this understanding of Stoic psychology, let us consider how Stoicism fits within the compatibilist framework of Frankfurt and how this creates rational grounds for moral responsibility.

Frankfurtian Compatibilism and Stoicism

As a reminder, compatibilism is the idea that humans still possess free will in a determined world.[18] To preface this discussion, debates of free will often concern definitional problems. This is to say that how we define free will is an important aspect of discussion. The more contemporary understanding of free will emerged around 500 CE, around the time of the emergence of Christian philosophy.[19] This notion of free will is one in which individuals possess the capacity, at any time, to act or not act in a certain way, regardless of their past or their beliefs (within the bounds of environmental constraints). To many, this is the prerequisite for our actions to be described as free: we could have acted differently than we did.

It would be notably anachronistic to attribute this notion of free will to discussions of the Stoics. Ancient discussions of free will do not have any interest in understanding if the same person with the same beliefs and desires in the same circumstances could possess the ability to do or not do a certain action.[20] Rather, the Stoics were concerned with the question of whether an agent, rather than "something else," was responsible for an action. Still, some contend that for an agent to be responsible for an action that they

16 Marcus Aurelius, *The Meditations*, trans. George Long, 167 C.E., https://classics.mit.edu/Antoninus/meditations.html, book four.

17 Brennan, "Stoic Moral Psychology," 52.

18 McKenna and Coates, "Compatibilism."

19 Timothy O'Connor and Christopher Franklin, "Free Will," The Stanford Encyclopedia of Philosophy, eds. Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Winter 2022, https://plato.stanford.edu/entries/freewill/.

20 Bobzien, *Determinism and Freedom,* as cited in Tad Brennan, "Fate and Free Will in Stoicism: A Discussion of Susanne Bobzien, Determinism and Freedom in Stoic Philosophy," in *Oxford Studies in Ancient Philosophy*, edited by David Sedley, Volume XXI (Oxford University Press, 2001), 264.

must have possessed the ability to act differently than they did; however, compatibilists reject this notion, arguing that individuals can still be said to have acted freely even if they were determined to act as they did.

In *Alternate Possibilities and Moral Responsibilities*, Harry Frankfurt defends this compatibilist understanding of moral responsibility. His discussion is highly applicable to the Stoic's understanding of compatibilism, and his framework can demonstrate how Stoic compatibilism is defensible. Frankfurt's work directly engages with the question of whether individuals must possess the ability to act differently than they did to be morally responsible for an action. He describes this as the principle of alternate possibilities, which he defined as an understanding of moral responsibility wherein "a person is morally responsible for what he has done only if he could have done otherwise."[21]

Frankfurt argues against the validity of this principle by asserting that there are cases where an individual may be guaranteed to do something without being compelled to do that thing. This means that they could not have acted otherwise, but they still acted freely. These are now known as "Frankfurt cases." One such example follows.

A man, Tim, goes to vote in a presidential election. He could vote for President A or President B; however, the government has covertly implanted a device into Tim's brain, such that if he tries to vote for President B, he will be compelled to vote for President A. However, Tim has no intention whatsoever to vote for President B. As he walks into the voting booth, he confidently checks the box next to President A and leaves. Is Tim morally responsible for his vote?[22]

While Tim certainly acted as he wished, he could not have acted otherwise. This points out a flaw in the principle of alternate possibilities. Despite acting without compulsion, the principle would hold that Tim is not morally responsible for his vote since he could not have acted otherwise. But since Tim voted for who he wished, the government's device had no impact on his vote whatsoever.

This sort of reasoning is supremely applicable to Stoic psychology. Returning to the idea of impressions and assents, as mentioned earlier, humans are different from non-humans because we possess the capacity for assent. However, humans do not possess the ability to have assented to impressions to which they withheld assent, nor could they have withheld assent to impressions to which they assented. In other words, our assents are as determined as any other part of the causal chain.[23] A simple example of this is that, given the education that you received, if you were presented with the impression that 2 + 2 = 3, you could not assent. Your past education, which influenced your faculty of reason, determined this.

Expanding on this, our total disposition to assent is known as our *prohairesis*.[24]

21 Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66, no. 23 (1969): 830.

22 Hank Green, "Compatibilism: Crash Course Philosophy #25," posted August 22, 2016, by Crash Course, YouTube, https://www.youtube.com/watch?v=KETTtiprINU, 3:50–4:55.

23 Brennan, "Stoic Moral Psychology," 293.

24 Brennan, "Fate and Free Will," 276.

This can be understood as a person's character and their internal reason. Whenever we assent to an impression, we do so based on our *prohairesis*. It is what makes the selfish man accept a bribe, the curious man ask a question, and the vicious person act without mercy. Our *prohairesis* is shaped by our inherent nature and our experiences, and it is governed by *Logos*.

The concept of *prohairesis* is essential to understanding how the Stoics maintain freedom and moral responsibility within a determined framework of actions. As discussed earlier, the Stoics conceive of freedom in a negative sense—as freedom from external constraints. Building on this, when the Stoics speak of freedom, they mean freedom from forces outside our internal judgment. As Epictetus suggests, acting freely involves transcending external obstacles. For example, if your car breaks down, you are not free to drive to the store; external circumstances prevent you from doing so. However, if you are presented with the impression of going to the store and your *prohairesis* determines that you will withhold assent to this impression, the Stoics would argue that you still acted freely because no external force determined your decision, even though you could not have acted otherwise.

To illustrate this further, consider the voting thought experiment. If you initially intend to vote for President B but are externally forced to vote for President A, according to Stoic principles, you are not acting freely because you were compelled by an external force. However, if you voted for President A based on your internal disposition—your *prohairesis*, which never had the intention of voting for President B—you acted freely, even though your actions could not have differed from what occurred.

This is all to say that when humans act, they do so according to their *prohairesis*. It is impossible for someone to want to act any differently than their internal state would allow them to, and as mentioned earlier, our internal state is guided by our sense of reason. We can never act against our own sense of reason. If someone were to act irrationally, the Stoics would not say that they acted against their own sense of reason; rather, they would say that this person's sense of reason was misguided. Epictetus explains this as he writes, "When any person harms you, or speaks badly of you, remember that he acts or speaks from a supposition of its being his duty. Now, it is not possible that he should follow what appears right to you, but what appears so to himself."[25] This is important to the Stoics' understanding of freedom.

Before stepping too far away from our discussion of Stoicism and Frankfurtian compatibilism, one scholar's take on this is worth examining. In his piece "Stoicism and Frankfurtian Compatibilism," László Bernáth disagrees that the Stoics' conception of free will is similar to Frankfurt's. He writes:

> It is misleading and ultimately false to say that Frankfurt's and the Stoics' conception of free will are the same or notably similar to each other. Frankfurt has a contrafactual analysis of free will that refers to a psychological ability which is specific to humans and exercised by most people in most cases. In contrast, the Stoic considers free will as an aim for everybody that is achieved only by the sage, who can choose the option that she regards as the best one every time.[26]

25 Epictetus, *The Enchiridion*.

26 László Bernáth, "Stoicism and Frankfurtian Compatibilism," *Elpis*

I would like to focus primarily on Bernáth's understanding of the Stoic conception of free will. In his argument, Bernáth conflates Stoic freedom with libertarian free will. I noted that it is anachronistic to ascribe this to the Stoics, and it remains a common mistake in some of the literature.

Bernáth attributes the idea of free will directly to Epictetus.[27] As discussed earlier, when Epictetus used a rough translation of freedom, he almost certainly did so in the political sense to refer to being free from tyranny.[28] Bobzien clearly delineates Epictetus's idea of freedom from an understanding that more closely resembles free will:

> In these contexts of politics and ethics, freedom is never the freedom to decide between alternative courses of actions, or the power to do otherwise, or causal indeterminedness; nor is it ever connected with a two-sided potestative concept of that which depends on us. It is always the freedom of an individual (or group of individuals) from certain external or internal determining factors, thus providing a sphere in which the individuals are masters of their own affairs.[29]

Freedom, then, for Epictetus is much closer to the traditional Stoic understanding of the goal of life: to align oneself with Nature.[30] As discussed, we are a part of this very Nature, so aligning ourselves with it requires us to play the role we were assigned. In other words, we are free in so far as we accept our role in Nature. We are free when we embrace the negative freedom we have and do not protest the positive freedom that we lack.

In this way, Bernáth is right to assert that the sage operates with a greater sense of freedom than others. The sage is the one who acts with perfect virtue, and one does this by aligning oneself with Nature, which is the only true source of freedom. However, Bernáth describes the sage as the one who can "choose the option that she regards as the best one every time," but according to the Stoics, everyone does this. Each person acts based on their own internal sense of reason, their *prohairesis*.[31]

Sages do act properly on every occasion, but they do this because their sense of reason is perfectly aligned with Nature, not because they have better control over choosing between right and wrong. In fact, as Bobzien describes, "Sages, *qua* being wise, will have usually only one option they can take."[32] Because the sage's sense of reason is perfected, their internal state, their *prohairesis*, can only ever assent to true impressions and withhold assent from false ones. The Stoics believe that freedom entails aligning oneself with nature, which the sage does, but this does not mean that the sage has more "free will" than anyone else.

The Cone and The Cylinder

Filozófiatudományi Folyóirat 11, no. 2 (2018): 67, https://doi.org/10.54310/elpis.2018.2.6.

27 Bernáth, "Stoicism and Frankfurtian Compatibilism," 69.

28 Bobzien, *Determinism and Freedom,* 338.

29 Bobzien, *Determinism and Freedom,* 338–39.

30 Seneca, *Letters from a Stoic,* 37

31 Bernáth, "Stoicism and Frankfurtian Compatibilism," 67.

32 Bobzien, *Determinism and Freedom,* 341.

A key objection to Stoic compatibilism is that it emphasizes freedom from external constraints while simultaneously maintaining that our *prohairesis* is causally determined. In other words, we are free because we act how we wish, but how we wish to act is not determined by us. It may seem as though the Stoics are hiding the problem of freedom elsewhere. Perhaps we assent to an impression based on our own nature, but how can we hold someone responsible for their own nature if it is causally determined? Sure, a murderer may have assented to the impression of killing someone, but if that was driven by a *prohairesis* that was externally shaped, then wouldn't it be the events that shaped his *prohairesis* that determined he would kill someone? A timeless example is helpful in explaining how the Stoics would respond.

Dating back to Chryssipus, the Stoics illustrated their notion of moral responsibility and freedom with an example involving a cone and a cylinder. Imagine there is a cylinder sitting next to a cone. Then, someone gives both a push. Once they begin moving, their path is a result of their nature.[33] This is to say that the cylinder rolls straight because it is a cylinder and the cone rolls in a circle because it is a cone. It is not that the cylinder is necessarily the cause of its own shape, but rather the fact that the cylinder is of its own shape that it acts as it does. In determining moral responsibility, the Stoics would not concern themselves with asking how the murderer became one, but rather, they would argue that he is one.

It is a long quote, but Bobzien expresses her defense of compatibilism, and in turn, moral responsibility, most clearly in her discussion of this example. She writes:

> The point of the analogy then is that there are cases in which the external antecedent causes are similar . . . but the effects differ noticeably in kind; hence it follows that the nature of the objects at which the effect takes place (geometrical bodies, human beings) must be responsible for the difference in effect. This point is quite different from some modern arguments: Chrysippus defends responsibility by arguing that someone else would do something else in the same situation or that it is in the range of possible human behavior to do something else; he does not argue that the same person could do something else in the same situation.[34]

The most important point to highlight here is that when Stoics seek to explain moral responsibility, they do not try to explain what led to a person acting as they did; rather, they seek to show that an action comes from an agent rather than from something else. The example shows that it is not the push that determines the path of the figures, but rather their shape. Just as the cone is responsible for its circular path, so too is our *prohairesis* responsible for our actions. When we praise or blame someone for an action (hold them morally responsible), we are assigning value to their *prohairesis* as it is the cause of the action. And, as Epictetus says, we are our *prohairesis.*

With respect to the discussion of freedom, this example is interesting. One would not say that the cone is free to roll in a straight path, but that is because of its shape. It is up to the figure how it rolls. In this way, the cone is only free to act in accordance with its own nature. While this appears like a limited understanding of freedom, this aligns with the Stoic view that true freedom lies in acting in accordance with nature. Our actions are said to be free because they come from us—from *our* nature—rather than from the external forces that are outside of us. To the Stoics, it does not matter if who we are is shaped by external factors—the actions we take are still ours, and this is the way in which the Stoics consider actions to be free.

33 Brennan, *The Stoic Life,* 254.

34 Bobzien, *Determinism and Freedom,* 268–69.

Conclusion

The Stoic understanding of freedom and moral responsibility is deeply complex, but it is tenable. All events occur within a framework of causal determinism driven by *logos*, including our actions. However, actions are up to us when we act beyond the external constraints of the world. This is due to our capacity for assent—the ability for us to consider how we should respond to external events (impressions). Whether we assent to an impression or not is part of the causal chain, and this is driven by our internal state called our *prohairesis.*

Humans do not possess the ability to act differently than they do; however, Frankfurt's defense of compatibilism reveals that so long as individuals act as they wish to act uncompelled by external forces, they are responsible for their actions. Unless physically compelled, humans always act as they wish to, which is determined by their *prohairesis,* and the Stoics argue that this is what freedom entails. To the Stoics, moral responsibility is not about asking how a person acquired a certain nature; rather, moral responsibility is about showing that a person's action came from them rather than from something else.

Each person has a different nature, a different *prohairesis*. Just as a cylinder is responsible for its straight path and a cone is responsible for its circular path, each person's internal state is responsible for the actions they take. Moral responsibility lies not in the freedom to choose contrary actions, but in our actions being our own. With the capacity for assent, external forces cannot be said to be responsible for our actions when it is our internal nature that decides how we act.

Works Cited

Aurelius, Marcus. *The Meditations*. Translated by George Long. 167 C.E. https://classics.mit.edu/Antoninus/meditations.html.

Bernáth, László. "Stoicism and Frankfurtian Compatibilism." *Elpis Filozófiatudományi Folyóirat* 11, no. 2 (2018): 67–81. https://doi.org/10.54310/elpis.2018.2.6.

Bobzien, Suzanne. *Determinism and Freedom in Stoic Philosophy*. Oxford University Press, 2001.

Brennan, Tad. "Fate and Free Will in Stoicism: A Discussion of Susanne Bobzien, Determinism and Freedom in Stoic Philosophy." In *Oxford Studies in Ancient Philosophy*, edited by David Sedley. Volume XXI. Oxford University Press, 2001.

Brennan, Tad. *The Stoic Life: Emotions, Duties, and Fate*. Oxford University Press, 2005.

Brennan, Tad. "Stoic Moral Psychology." In *The Cambridge Companion to the Stoics,* edited by Brad Inwood. Cambridge University Press, 2003.

Carter, Ian. "Positive and Negative Liberty." The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, Spring 2022. https://plato.stanford.edu/entries/liberty-positive-negative/.

Epictetus. *The Enchiridion*. Translated by Elizabeth Carter. 135 C.E. https://classics.mit.edu/Epictetus/epicench.html.
Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66, no. 23 (1969): 829–39.
Frede, Dorothea. "Stoic Determinism." In *The Cambridge Companion to the Stoics*, edited by Brad Inwood. Cambridge University Press, 2003.
Green, Hank. "Compatibilism: Crash Course Philosophy #25." Posted August 22, 2016, by Crash Course. YouTube. https://www.youtube.com/watch?v=KETTtiprINU.
McKenna, Michael, and D. Justin Coates. "Compatibilism." The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Spring 2024. https://plato.stanford.edu/entries/liberty-positive-negative/.
O'Connor, Timothy, and Christopher Franklin. "Free Will." The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Winter 2022. https://plato.stanford.edu/entries/freewill/.
Pigliucci, Massimo. "Stoicism." Internet Encyclopedia of Philosophy, accessed April 9, 2025. https://iep.utm.edu/stoicism/
Sanchez, Christopher. "The Evolution of Stoicism: An Overview of Prominent Features and Discussions in Modern Stoicism." Master's Thesis, West Chester University, 2024.
Seneca, Lucius Annaeus. *Letters from a Stoic*. Translated by Robin Campbell. Penguin Books, 1969.

Whom to Write to: Writing's Phantom Audience through Derrida

Roger Lin

Abstract: This essay problematizes the notion of "private writing" through the deconstruction of the diary, drawing on Jacques Derrida's analysis in Plato's Pharmacy, his commentary on Plato's Phaedrus, where Socrates critiques the contradictions between writing and memory. Derrida excavates the self-contradictoriness of the Greek word pharmakon, meaning both "cure" and "poison", and applies this duality to writing itself. Building on Derrida alongside Ludwig Wittgenstein's critique of private language and Sigmund Freud's work on Totem and Taboo, this essay investigates the dilemma between privacy and publicness in writing. While could be perceived as an intimate act, writing inevitably fetishizes its non-present audience, thereby always exposing it to the risk of misinterpretation. This tension is further amplified by fear: the fear of self-exposure, misreading, and the unintended publicization of private writing. Yet paradoxically, these fears do not deter writing but rather compel it: fears push individuals to document thoughts precisely out of the anxiety of exposure. Based on the discussion on private writing, this essay examines writing in a more generic sense through the example of the dedication page. The dedication page of a book implies a phenomenological conflict between "intended" and "unintended" readerships, as well as the delayed interaction between author and audience. While an author must fetishize their audience, writing always risks a discrepancy between those whom the author intends to invite and those who ultimately engage with the text. Yet, this gap is not a flaw or failure in writing but the very mechanism of how any interpretation could happen in the first place.

On Private Writing

Is private writing possible? But first, what is private writing? A diary serves as an example: a deeply personal record, a diary is often perceived as being written

to and for oneself alone. It hides the confessions of a high schooler agonizing over an unspoken crush or the venting of frustrations that cannot be voiced aloud so scribbled in shaky handwriting at midnight. One might also envision a diary that is not only hidden but also encoded—locked with a password, tucked away from prying eyes. Yet, for all its protections, it must remain findable and readable. A diary is not a tomb: it does not bury secrets; even if it does, its act of being hidden invites discovery. A tomb may be obscure, but it also compels a search for what lies beneath, and its silence gestures for revisitings. One must make sure of a diary's existence as a readable medium, but here is also the paradox: the privacy of a diary is always-already at risk of its very openness to being deciphered.

To radicalize the notion of private writing, one should consider the possibility of private language. Insofar as writing is the actualization of language, then considering private language resonates with the very conditions under which writing can be understood. In *Philosophical Investigations*, Ludwig Wittgenstein (1953) poses the question directly: "What about the language which describes my inner experiences and which only I myself can understand?"[1] Not coincidentally, Wittgenstein then turns to the **diary** as a medium to critique this notion of private language. In his thought experiment, Wittgenstein describes an attempt to keep a diary recording the recurrence of a particular sensation.[2] However, he then immediately challenges this attempt: *"But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation . . . for in this way I commit to memory the connection between the sign and sensation. . . .* But in the present case, *I have no criterion of correctness."*[3] Here, Wittgenstein suggests that if the act of recording a sensation is entirely private, namely without external verification or a shared system of meaning, there is no standard by which one can determine whether the notation is correct or not. The absence of an external framework prevents the diary entries from being intelligible and reduces them to arbitrary marks. Thus, even the diary, perceived as the most **private** form of writing, must still depend on a **communal space of understanding** to be intelligible. Writing, even when addressed to oneself, is never entirely private as it already presupposes a shared linguistic structure, as meaning cannot exist in isolation. This is what Wittgenstein calls the **"criterion of correctness":** for language to function, it must be based on a public system where meaning can be recognized and validated.

The issue, then, is not that writing transforms inner sensations into external signs. As Wittgenstein points out, sensations are already signs even before they are written. Writing does not mediate between the non-discursive and the discursive or the private and the public. It inhabits them already. Therefore, for writing to be readable by oneself, it thus must be readable and exposable to others, where precisely the very vulnerability of a diary lies: to write the diary already acknowledges its inherent readability. As Jacques Derrida (2000) writes in *Of Hospitality*, "There is no house without doors or windows."[4] For a house to be a house, it must be hospitable and allow the possibility of entrance. For writing to be writing, it must allow for the possibility of being read. Hence, the diary's discursive openness would not be obstructed by its physical enclosure: it entails a space that is at once closed off and inviting, private but exposed.

1 Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (Blackwell, 1953), §256, 98.

2 Wittgenstein, *Philosophical Investigations,* §258, 98.

3 Wittgenstein, *Philosophical Investigations*, §258, 98-99.

4 Jacques Derrida, *Of Hospitality,* trans. Rachel Bowlby (Stanford University Press, 2000), 61.

The Compulsion to Write

Now, the problem of private writing does not become clearer with closer examination; if anything, it only becomes more enigmatic: why do we feel compelled to write things down at all, especially when it concerns a will to privacy? In *Phaedrus*, Socrates critiques writing as a *pharmakon*—a substance that, like a drug, can act as both poison and cure—because it fosters forgetfulness; people come to rely on external symbols rather than cultivating internal memory and wisdom. Socrates himself refused to write, but Plato didn't:

> But when they came to letters, this, said Theuth, will make the Egyptians wiser and give them better memories; it is a specific both for the memory and for the wit. Thamus replied: O most ingenious Theuth, the parent or inventor of an art is not always the best judge of the utility or inutility of his own inventions to the users of them. And in this instance, you who are the father of letters, from a paternal love of your own children have been led to attribute to them a quality which they cannot have; for this discovery of yours will create forgetfulness in the learners' souls, because they will not use their memories; they will trust to the external written characters and not remember of themselves. The specific which you have discovered is an aid not to memory, but to reminiscence, and you give your disciples not truth, but only the semblance of truth; they will be hearers of many things and will have learned nothing; they will appear to be omniscient and will generally know nothing; they will be tiresome company, having the show of wisdom without the reality.[5]

In this passage from *Phaedrus,* which Derrida tackles in *Plato's Pharmacy*, Socrates recounts the Egyptian myth to Phaedrus as a warning that writing is a **poison** since it externalizes thought and makes knowledge vulnerable to loss and distortion. But it is also because Socrates' thoughts were written down—preserved by Plato—that his wisdom has endured for nowaday readers. In this sense, Plato himself acts as a ***pharmakon*** for Socrates: both a cure and a risk. Namely, by committing Socrates' words to writing, Plato guards his wisdom from being lost to time, yet he also distances them from the immediacy of dialogue.

However, the concern of writing a diary here is not the same as retaining knowledge. Without getting into a debate about what a "proper" diary is, private writing, as framed in this paper, functions as a confession of secrets. Then, why must a secret be performed in the first place? Why not keep it locked away in one's mind, like the remembered location of the diary or the password protecting it (perhaps itself noted in another hidden diary)? The Ancient Greeks did not fear writing itself; they feared forgetting: writing is the scapegoat of forgetting. They fear people "will trust to the external written characters and not remember of themselves," and that they would become "hearers of many things" yet

5 Plato, *Phaedrus,* trans. Benjamin Jowett (Dover Publications, 1999), 274e–275a.

"learn nothing."

But in the case of secret or private writing, the compulsion to write does not stem from the fear of forgetting but from the fear of writing itself. Writing should be held accountable here. So to speak, it is precisely the fear, the fear of writing's potential to be accessed by others—a fear entangled with the desire to write—the very possibility that a "private" secret might be revealed is what propels one to write it down in the first place. But secrecy resists simple forgetfulness; it needs a conflicting engagement. To keep a secret is to enact it, and writing becomes the medium through which secrecy is both preserved and made precarious. The act of inscribing secrecy externalizes what should remain internal, making it legible not only to memory but also to the possibility of its own betrayal. Writing becomes not a medium for memory but a stage for performing secrecy.

Freud's Taboo and Derrida's Secret

Private writing, particularly its secret and confessional form, finds its resonance with taboos, which operate through self-imposed prohibitions that lack justification but are nonetheless compelling. Sigmund Freud (1913) writes: "The taboo restrictions are different from religious or moral prohibitions. They are not traced to a commandment of a god but really they themselves impose their own prohibitions. . . . Though incomprehensible to us they are taken as a matter of course by those who are under their dominance."[6] Like taboos, secrets impose their own law: they must be kept, yet they demand recognition as secrets. Private writing functions as an act of both transgression and submission. It both defies the implicit prohibition against revealing a secret and, in doing so, affirms the secret's hold over the writer. Writing transforms secrecy into something tangible, making it vulnerable to exposure, but this very risk reinforces the secret's power. The paradox of secrecy, then, reflects that of taboo: the more one resists or attempts to break it, the more its force is confirmed. That is, the effort to protect it makes it feel more significant. This means that private writing does not merely record a secret, it makes a secret possible and empowers it through the urge to confess colliding with the fear of revelation, much like the way a taboo derives its authority from the very anxiety it produces.

Yet, by the same logic with which we deny the attainableness of absolute private writing, we find that pure publicity, or true openness, is similarly unattainable. Absolute public writing, like unconditional hospitality, is impossible. As Derrida describes in *Of Hospitality*, this impossibility of unconditional hospitality is an "insoluble antinomy, a non-dialectizable antinomy."[7] He frames it as unsolvable between the "law of unlimited hospitality"—the ideal of welcoming the other without limit—and the "laws" that impose conditions and boundaries on this welcome. These "laws (in the plural)," he explains, are "those rights and duties that are always conditioned and conditional."[8] The very gesture of inviting someone into a house is predicated on the fact that the house is private property.

Just as hospitality is conditioned by the limits of ownership, writing is conditioned by its medium, language, circulation, and so on. Even a work as ubiquitous as the Bible cannot reach every corner of the Earth. As Derrida hints, the notions of private writing and public writing are not opposites but rather two forms of exclusivity. Any form of writing must deal with the inherent dynamics of secrecy. Derrida (2001) writes, "There's no question, difference is not presented, or nature loves to hide. But such hiding is not the only secret. The professional secret, confessional secret, military secret, political secret, the secret police, the secret in novels, etc., all the semantics of the secret are possibilities that are more determinate than the general possibility to which you refer."[9] When one writes a secret to oneself, this act is not categorically different from sharing it with others, i.e., it still participates in the broader "semantics of the secret" that Derrida describes. Writing, whether private or public, must engage with secrecy as both an organizing principle and a destabilizing force.

This duality is further explained in Derrida's comparison of the secret to a gift. A gift, Derrida (2002) writes, is "not present, the gift of something that remains inaccessible, unpresentable, and, as a consequence, secret."[10] Like the secret, a gift ceases to exist as such when it is fully recognized or acknowledged. It lies in its inaccessibility, its deferred revelation. Both secrecy and gift organize by structuring what is known and unknown, yet they destabilize by always carrying along the possibility of revelation.

"What does it mean to share a secret?" Derrida asks. "It isn't a matter of knowing what the other knows, for Abraham doesn't know anything. It isn't a matter of sharing his faith. . . . To share a secret is not to know or to reveal the secret, it is to share we know not what: nothing that can be determined."[11] This framing shifts the understanding of secrecy from a possession of specific content to an act of sharing that resists determination. In Derrida's view, sharing a secret is not about transmitting knowledge or faith that is endangered by forgetfulness; it is about embracing the indeterminacy of what remains unspoken yet shapes the bond between those who engage with it. In other words, it is not what we know connecting us but the unknown or unspoken.

This indeterminacy of secrecy does not simply denote absence or negation. As pointed out earlier, to declare something "a secret" (maybe with an exclamation point, "a secret!") is to draw attention to its status as hidden. Derrida's reflections on the secret as an inaccessible gift are precisely to illustrate this: "The gift of something that remains inaccessible, unpresentable, and, as a consequence, secret. . . . A gift destined for recognition would immediately annul itself. The gift is the secret itself if the secret itself can be told."[12] Secrecy, then, must *appear as hidden* to function as a secret at all. Thus, secrecy cannot be reduced to either full disclosure or total silence. It is not simply a negation, as Derrida's critique continues: "There is a secret of denial and a denial of the secret. The secret, as such, as secret, separates and already institutes negativity; it is a negation that denies itself. It de-negates itself. . . . The enigma . . . is the sharing of the

6 Sigmund Freud, *Totem and Taboo: Resemblances Between the Mental Lives of Savages and Neurotics, trans. James Strachey (Routledge, 1913),* 12.

7 Derrida, *Of Hospitality*, 77.

8 Derrida, *Of Hospitality*, 77.

9 Jacques Derrida, *I Have a Taste for the Secret,* ed. Michel Jacob, trans. Giacomo Donis (Polity Press, 2001), 75.

10 Jacques Derrida, "Secrets of European Responsibility," in *Acts of Religion,* ed. Gil Anidjar (Routledge, 2002), 29–30.

11 Derrida, *I Have a Taste for the Secret*, 79–80.

12 Derrida, "Secrets of European Responsibility," 29–30.

secret. . . . Not only the sharing of the secret with the other . . . but also shared within itself . . . as it begins to be lost, to divulge itself, hence to dissimulate itself, as secret, in showing itself: dissimulating its dissimulation."[13] This means that secrecy is not just an epistemological problem, what is known or unknown, but a disruption of the categories that structure knowability or perhaps knowledge itself. It does not just negate but undoes itself in its own articulation, thereby resisting final determination. In this way, secrecy is part of Derrida's broader critique of presence: it unsettles the oppositions and boundaries between knowing and not-knowing, speech and silence, or affirmation and negation.

Performative Writing

The written word, like or as the secret, "dissimulates its dissimulation," both keeping and giving away. But to whom? A secret to whom? A gift to whom? A writing to whom? Writing is never self-contained. Even when writing *appears as private*, it *appears the private* by engaging an implicit other and shapes its content in anticipation of a reader. In this way, writing is always performative: its meaning is not simply conveyed through content but is possible through the way it negotiates its phantom audience. Thus, unlike speech, which typically presumes a present and immediate interlocutor, writing operates in a different temporal and relational dimension. Deprived of an immediate recipient, the writer must compensate for this absence by constructing an audience—or rather, **audiences**—that are simultaneously fetishized and deferred.

Derrida's *différance* suggests that meaning emerges not through fixed presence but through deferral and difference; that deferral operates not only between the play of signifiers but also between the moment of inscription and the moment of reception. The gap between author and reader, between intention and reception, prevents the immediate determination of meaning. The fetishized audience alleviates but also accretes this instability because it allows the writer to orient their words toward an imagined interlocutor even when the actual reader remains unknown or deferred. This audience is the *pharmakon* to the writer, both helpful and unstable. It provides coherence to the writing process that is ultimately unattainable, as every real reader will interpret the text differently from the idealized recipient the author had in mind. The absence of a direct audience, then, does not sever writing from relationality; instead, this temporal deferral of interaction makes the construction of the phantom audience indispensable.

Consider, again, the act of writing a diary, which is *prima facie* private and directed at oneself. Even here, the fetishized audience is present, whether it takes the form of the writer's future self or an abstract confidant. The diary's performativity lies in addressing an audience that is simultaneously intimate and abstract, present and absent. The fetishized audience of the diary enables expression and threatens exposure. The fear of being read, of misinterpretation, of betrayal, haunts the inscription of a diary. But this fear compels writing: the anxiety of being read is inseparable from the desire to be understood. Similarly, in "public" writing, such as essays or books, the author writes with a particular readership in mind, tailoring their arguments and language to this imagined group. However, as the text enters the world, iteration takes over, and the relationship between the text and its readers becomes unpredictable. The fetishized audience is replaced by actual readers who bring their own interpretations.

13 Jacques Derrida, "How to Avoid Speaking: Denials," in *Languages of the Unsayable,* ed. Sanford Budick and Wolfgang Iser (Stanford University Press, 1996), 95

Secret Being Told: The Dedication Page

This displacement is a conflict, one that becomes subtly visible when a reader encounters a book's dedication page and sees a name that is not their own. A phenomenological conflict resides in the ambiguous role of the dedication page. At first glance, it appears straightforward: the dedication is addressed to a specific individual, like a personal postcard. But this simplicity is deceptive. By publishing the book, the author extends the dedication beyond its intended recipient, inviting a broader audience to witness what seems like an intimate gesture. The reader, encountering this dedication, finds themselves in an ambiguous position—**not the intended recipient, yet drawn into the act as though it were meant for them**. The tension, then, is that the dedication, *prima facie* a private acknowledgment for a non-present subject, becomes a public performance. It pretends to be intimate, yet it is deliberately exposed to others. For the author, this public exposure transforms the gesture such that it is no longer just a thank-you to Husserl but to manipulate the gaze of unintended readers. The dedication page is thus neither purely private nor purely public; it is caught between.

As previously argued, writing is propelled by its **readability**, an author writes a secret precisely because it *can* be read but also fears its exposure for the same reason. The dedication page condenses it: it is both an intimate gesture and a public act. More broadly, writing requires a **fetishized, non-present audience**—a projection of "the other" who will eventually read the text. The dedication page exemplifies this necessity: it is a performance of intimacy directed at a specific other, but it is also a performance for the public—a tension that lies at the heart of all writing.

Consider Heidegger's dedication in *Being and Time*: "To Edmund Husserl, in friendship and admiration." As a reader encountering this dedication, I find myself in a paradoxical position: I am participating in a discourse that seems to have been crafted for someone else. Heidegger, as the host of this intellectual party, appears to extend his invitation solely to Husserl. Yet, by publishing the book, he simultaneously opens the doors to a wider, unintended audience. The intimacy implied by the dedication contrasts with the public aspect of the text. **If the book were truly intended only for Husserl, why publish it at all?** The dedication exposes the unresolved question of "whom" in writing, where the text's intended recipient is perpetually unsettled between specificity and generality, the imagined and the actual, and the known and the unknown.

This **"whom"** invites Derrida's critique of intention to return. Derrida argues that **writing, through its iterative nature, dissolves the original intention of the author**. A text's meaning does not rest in its author's intent but in its **capacity to be re-contextualized across time and space**. Thus, while the **insignificance of intention** may free the text from the author's authority, **it cannot dissolve the *différance* between the phantom audience and the real audience**. Meaning does not preexist this conflict; it emerges *through* it. ***Différance* is this conflict, the gap between the author's imagined audience and the audience who ultimately encounters the text**. It is through this conflict that meaning becomes possible.

This essay, too, must fetishize its audience, but in doing so, it distances itself from both the audience I imagine and the one that actually reads it. Writing projects a reader; this reader is always a step removed from reality. When the text is read, a conflict emerges through the discrepancy between the name on the dedication page and the name that the reader bears; even if they match, the gap persists. This conflict, this discrepancy, is where meaning becomes possible. Meaning is neither fixed in the author's intent nor fully contained within the reader's reception, and it jumps between the continuous withdrawal of meaning from a fixed presence that writing both relies on and resists. As such, this essay itself **embodies** the enigmatic conflict it seeks to explore. Derrida says, *"Writing did not 'enter' philosophy, it was already there,"* and so, too, was the **phantom audience.**[14] Whom then? The question falls to the writer and the reader of philosophy.

Works Cited

Derrida, Jacques. *I Have a Taste for the Secret*. Edited by Michel Jacob. Translated by Giacomo Donis. Polity Press, 2001.

Derrida, Jacques. *Of Hospitality*. Translated by Rachel Bowlby. Stanford University Press, 2000.

Derrida, Jacques. "Secrets of European Responsibility." In *Acts of Religion*, edited by Gil Anidjar. Routledge, 2002.

Derrida, Jacques. "How to Avoid Speaking: Denials." In *Languages of the Unsayable*, edited by Sanford Budick and Wolfgang Iser. Stanford University Press, 1996.

Freud, Sigmund. *Totem and Taboo: Resemblances Between the Mental Lives of Savages and Neurotics*. Translated by James Strachey. Routledge, 1913.

Plato. *Phaedrus*. Translated by Benjamin Jowett. Dover Publications, 1999.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G. E. M. Anscombe. Blackwell, 1953.

14 Derrida, *I Have a Taste for the Secret*, 8

How Is Mindreading Possible: A Contextual Analysis of Theories of Mind

Xuning Gao

Abstract: The question of how we successfully interpret and predict other people's minds without having access to them—commonly referred to as "mindreading"—has fueled the ongoing debate between theory-theory and simulation theory. Theory-theory argues that mindreading relies on folk psychological frameworks, while simulation theory suggests it is based on imaginative projection. Yet, the debate is intractable and this impasse has paved the way for pluralism—the idea that multiple theories can be valid, each offering accurate depictions of people's mental states in different contexts. This paper advances a pluralist perspective, proposing that the applicability of each theory depends on two key factors: the familiarity of mindreaders—individuals who interpret others' minds without direct access—with the person being predicted and with the situation in which mindreading occurs. By categorizing mindreading scenarios into four types based on these two dimensions, I demonstrate how theory-theory and simulation theory can be contextually integrated. Three of the four scenarios allow successful mindreading, where individuals rely on either theory-theory, simulation theory, or a combination of both, depending on the strengths and limitations of each approach. In the fourth scenario, mindreading is not possible due to insufficient information. This categorical analysis shows that neither mindreading theory is universally applicable, but both play crucial roles in different contexts. By emphasizing the specific contextual applicability of each theory, this paper offers a more nuanced understanding of mindreading under a pluralism account, bridges the gap between the two competing theories, and provides a framework that recognizes the strengths of both approaches in predicting mental states.

Introduction

How do we successfully interpret and predict other people's thoughts and emotions when the only mind we truly access is our own? Although this task of mindreading seems logically unachievable, the constant accuracy of predictions about others' minds plays a crucial role in facilitating our daily interactions. Various theories have been developed to explain our success in mindreading, with simulation theory and theory-theory being the most well-regarded theories. Theory-theory argues that we rely on folk psychological theories to infer others' mental states. Folk psychological theories here refer to theories that people generate using concepts like 'desire" and "belief," which interpret, predict, and explain the behaviors or feelings of others. Simulation theory, in contrast, claims that we understand others by imaginatively projecting ourselves into their situations, simulating how others would feel or act based on our own potential reaction.

The proponents of the two theories remain fundamentally at odds with each other, leading to the rise of a new pluralist view on mindreading. Pluralists argue that multiple theories of mindreading can coexist, each suited to a different context in diverse human interactions. Advancing this pluralistic perspective, this paper examines the applicability of theory-theory and simulation theory, proposing that their use depends on two dimensions: the mindreader's familiarity with the individual being interpreted and the context of the situation.

In this paper, I first introduce the debate between these two dominant theories to provide the context for my view, outline the framework in which we are situated, and raise the question we seek to answer. Then, I present the pluralist perspective in this debate and provide evidence supporting the possibility of cooperation between the two theories. Next, I explore different categories and explain which theories should be applied in specific thought experiments. Finally, I address potential objections and offer responses.

Theory-Theory and Simulation Theory at a Glance

The debate between theory-theory and simulation theory centers on the question: "What sort of process subserves our capacity to predict, explain, and interpret other people?"[1] Theory-theorists and simulation theorists offer distinct perspectives on this issue. One of the most widely discussed arguments of theory-theory involves Alison Gopnik's examination of the false-belief task, which demonstrates how children's ability to predict others' beliefs evolves between ages three and five is based on their attainment of folk psychological theories.[2] For simulationists, philosophers like Goldman and Gordon have received considerable attention, proposing that people predict others' minds and behaviors by imaginatively "putting themselves in the other's shoes."[3] In this section, I introduce the basic ideas of mainstream theory-theorists and simulation theorists with the help of the false-belief experiment.

Theory-theory (mainly by Gopnik, Wellman, Stich, and Nichols) in a nutshell is the idea that people utilize folk psychological theories to predict people's mental states. In their 1994 article, Gopnik and Wellman compare children's generation of folk psychological theories with scientists' development of scientific theories (i.e. the atomic theory), saying that "scientific theory change and conceptual change in childhood are both the product of human minds trying to understand the world around them" even though children's theories are much less formally and rigorously tested.[4] In other words, people attain folk psychological theories by generalizing their past experiences into straightforward formulas such as "If x then y."[5] Gopnik and Wellman then extend a theory-theorist perspective to explain the false-belief task.[6] Before further illustrating the false-belief experiment from the theory-theorists' perspective, I will first outline the relevant experiments related to desires, beliefs, and predictions in their original form.

To start with a desire-based task by Woolley and Wellman—a simpler version of the false-belief task featuring only desire—children aged two, three, and five are told a story about a person and asked to predict what the person would do next, examining their understanding of the relationship between desires and human actions.[7] For example, the children are told that "Tom wants to drink water" and are asked to predict Tom's subsequent actions. Across all ages, most children understand the desire-action relationship and predict that Tom will go drink water. The task becomes more challenging when beliefs are introduced. In a study by Karen Bartsch and Henry Wellman, children aged three and four were presented with a scenario involving a person's simple action and asked to

---

1 Shaun Nichols and Stephen P. Stich, "Folk Psychology: Simulation or Tacit Theory?" *Philosophical Issues* 3 (1993): 240.

2 Gopnik, Alison, and Henry M. Wellman. "The Theory Theory." In *Mapping the Mind: Domain Specificity in Cognition and Culture*, edited by Lawrence A. Hirschfeld and Susan A. Gelman (Cambridge University Press, 1994).

3 Robert M. Gordon, "Folk Psychology as Simulation," *Mind & Language* 1, no. 2 (1986): 162.

4 Gopnik and Wellman, "The Theory Theory," 258.

5 Note that one of the reasons why folk psychological theories are classified as "theories" is that they are generalized and may apply not only to an individual, but to a group of people.

6 Gopnik begins by comparing scientific theory change in adults with conceptual change in children. She argues that, despite differences in cognitive ability and methodological rigor, these types of change share "deep similarities between the underlying cognitive mechanisms involved in the epistemological endeavors of childhood and of science" (Gopnik and Wellman, "The Theory Theory," 259). According to Gopnik, key features of scientific theories—such as falsifiability, the inclusion of abstract entities, causal explanations, systematic coherence, and more—should also apply to children's understanding of the mind. In this view, children's predictive abilities arise from their possession of folk psychology theories. Children *with* differing theories are likely to interpret evidence distinctively, leading them to different conclusions about others' mental states.

7 Henry M. Wellman and Jacqueline D. Woolley, "From Simple Desires to Ordinary Beliefs: The Early Development of Everyday Psychology," *Cognition* 35, no. 3 (1990): 245–75.

interpret that action.[8] When asked, for example, "Why is Amy searching for her dog under the sofa instead of outside the house?" both three- and four-year-olds generally respond with explanations hybrid with belief terms and desire terms, such as "Amy *wants* her dog and *thinks* it is under the sofa," with four-year-olds demonstrating better accuracy.[9] Wellman and Woolley's 1990 study further explores children's recognition of belief and desire by performing similar experiments featuring either belief or desire performances on two-year-olds. The results suggest the concept of belief is not fully accessible to two-year-olds, and that "many two-year-olds pass desire tasks but fail comparable belief tasks."[10] Between the ages of two and four, children's understanding of the concept of belief appears to grow in direct proportion to their age.

The experiment that compares three-year-olds' and five-year-olds' understanding of the concept of belief, which is when the false belief task comes into play, is much more intriguing: "Children may see a candy box that has pencils instead of candy inside. They are asked what someone else [who just walks into the room] will think is inside the box. Three-year-olds typically say "pencils." They behave as if there is a simple and reliable causal link between the real state of affairs in the world, and our mental states about it."[11] While three-year-olds can talk about beliefs, they struggle to accurately predict others' beliefs as shown in the experiment. By contrast, five-year-olds successfully answer "candy" instead of "pencil," demonstrating a more sophisticated understanding of beliefs as different representations of the world instead of direct perceptions. The false-belief task further illustrates the direct correlation between maturation and the understanding of others' minds.

Theory-theorists explain this correlation by linking children's development of understanding to shifts in their acquisition of folk psychological theories at different ages. They argue that two-year-olds, who lack a concept of belief, operate on a desire-perception theory: "If an agent desires x, and sees that x exists, he will do things to get x."[12] By applying this folk psychological theory, two-year-olds can infer that Tom will drink from a water fountain if he is thirsty and perceives one.

By the age of three, children's responses shift to incorporate beliefs, marking a theoretical evolution. At this stage, children begin to understand fictional mental states such as beliefs, dreams, and images, and they form a new theory within the desire-perception framework. This new theory, still heavily influenced by the simple desire-perception view, remains nonrepresentational—just as one's perception has a direct causal link with reality, three-year-olds also think there exist direct causal connections between beliefs and real-world objects.[13] For instance, in the false-belief task where children are shown a candy box that contains a pencil, three-year-olds mostly conclude that someone

8 Karen Bartsch and Henry Wellman, "Young Children's Attribution of Action to Beliefs and Desires," *Child Development* 60, no. 4 (1989): 946–64.
9 Bartsch and Wellman, "Young Children's Attribution."
10 Wellman and Woolley, "From Simple Desires to Ordinary Beliefs," 267.
11 Gopnik and Wellman, "The Theory Theory," 266.
12 Gopnik and Wellman, "The Theory Theory," 265.
13 Gopnik and Wellman, "The Theory Theory," 265.

seeing a candy box would think it contains a pencil due to their possession of the theory that there is a direct causal link between tangible reality (what is actually inside the candy box) and belief (thought of what is inside the candy box).

By age five, children understand beliefs as representations, explaining their success in false-belief tasks. They realize that people believe what looks like the truth to them rather than being directly aware of the truth of the world. Such a realization marks a shift in children's minds toward representational beliefs similar to those of adults. This shift also transforms their non-representational theory of belief—which assumes a direct causal connection between belief and reality—into a representational theory. In the candy-pencil false-belief task, five-year-olds tend to respond with "candy" instead of "pencil," indicating their grasp of representational mental states. A representational theory for five-year-olds might take the form: "There exists an object P, which appears as p1 from an outside perspective. Person A observes from the outside and thus perceives P as p1." Providing a coherent explanation for the false-belief task, theory-theorists suggest that children's predictive responses evolve due to the development of their folk psychology theories, and people make predictions through applying folk psychology theories.

Nonetheless, theory-theory is "not the only game in town."[14] Simulation theorists like Goldman and Gordon argue that prediction does not require such folk psychological theories—people do not need to generalize their experience into theories to predict or explain others' behavior. Gordon claims the decisions that people make are "often products of practical reasoning,"[15] or in Goldman's words "decision-making mechanism,"[16] without clarifying the composition of such a mechanism. Since simulation theorists presuppose that practical reasoning is the underlying basis for all individuals' decision-making, people generate similar actions when having similar beliefs and desires. Thus, the predictors imagine being in others' positions and use their practical reasoning to predict or explain the predicted person's behaviors.

A simple way of simulating a person is "to answer the question, "What would I do in that person's situation?""[17] When we are trying to predict what we will do in an imaginary scenario, we put ourselves in a pretend world with pretend intentions, which allows our practical reasoning (or decision-making mechanism) to work offline (i.e., working without making actual decisions in real life) and predict what we will do if such an imaginary event happens. For instance, if I want to know what I would do if my cat went missing, I put myself into a modified world where I have lost my cat. My practical reasoning will then tell me that, in such a possible world, I would want to call my friend to come over and help me look for the cat in the neighborhood, which is a prediction of my own mental state if I had actually lost my cat. In self-prediction, we modify certain parameters while keeping others constant. In the cat example, we adjust our desires—the

14 Nichols and Stich, "Folk Psychology," 239.
15 Gordon, "Folk Psychology as Simulation," 160.
16 Alvin I. Goldman, "Conceptualizing Simulation Theory," in *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (Oxford University Press, 2006), 27.
17 Gordon, "Folk Psychology as Simulation," 161.

intention to find my lost cat—while retaining our beliefs, like that "my friends can help me." Similarly, when predicting others' actions, we engage in a process of "quarantining [our] own idiosyncratic desires and beliefs"[18] to fully adopt their perspective, which involves putting ourselves in their shoes and incorporating their specific desires and beliefs rather than our own.[19] If I want to predict what Marielle will think or do after she receives a sapphire necklace as a birthday present, I would imagine that I am Marielle and have her beliefs and desires. I know that Marielle loves the color blue and jewelry, but she is not aware that I am giving her a present on her birthday. Pretending to have such desires and beliefs, I imagine that I would feel very surprised and happy when seeing the birthday gift in this pretend world, which are the emotions I attribute to Marielle. Simulation theorists claim people do not predict through a rigorous, systematic format in which we strictly follow explicit rules that generalize the behaviors of a large group of people, rather, their predictions are individually focused and shaped by the predictors' past experience with the predicted. To return to the false-belief experiment, typical simulation theorists would explain the false-belief task by saying that the change from failure at the age of three to success at the age of five is due to the completeness of the mental ability. Children at the age of three simply do not have the mental capacity to simulate the other's perspective fully. While at the age of five, their capacity develops and allows them to simulate a pretend world from the other's perspective, which is why they answer "candy" instead of "pencil."[20]

To end this section, a helpful analogy can be made to compare the two theories. Stich and Nichols compare an engineer predicting how a "plane will behave at a certain speed"[21] to a person mindreading how an individual will behave or feel in a situation. There are two ways of figuring out how a plane will perform after taking off. The first way is to use physical and mathematical functions to calculate the precise data points, and the other is to imagine the plane taking off based on the engineer's experience. The first approach aligns with theory-theory, that people utilize theories that work for all to make predictions, while the second aligns with simulation theory, which asserts that people predict by envisioning themselves in similar situations as the predicted individuals. These two theories seem to be fundamentally at odds with each other. Now the question is, can we reconcile these two seemingly opposing theories of mind? If so, how?

18 Goldman, "Conceptualizing Simulation Theory," 30.

19 I understand there are differences between Gordon's and Goldman's versions of simulation theory. However, this paper does not delve into these distinctions. as the focus is on utilizing the core principles of simulation theory and theory-theory to construct a pluralistic framework. Due to space constraints of this paper, Goldman and Gordon are treated here as proponents of the mainstream simulation theory.

20 In this paper, I focus on the arguments presented by mainstream theory-theorists and simulation theorists. While I acknowledge the existence of alternative perspectives, such as Jane Heal's co-cognition, which could offer grounds to reconsider my argument, I do not address these views due to the limited scope of this paper.

21 Nichols and Stich, "Folk Psychology," 229.

Distinguishing Ways of Predictions

Many philosophers adhere to a dichotomy, viewing theory-theory and simulation theory as incompatible frameworks such that only one of them is true. However, this approach has proven unproductive, as neither theory can definitively refute the other. Instead, the debate often results in each theory talking past the other, reinforcing their incompatibility rather than advancing the conversation. The failure to resolve the debate comes from an underlying assumption that one theory must dominate all mindreading contexts, while in fact, each theory has its own strengths and weaknesses. Pluralism offers a solution by recognizing that the applicability of each theory depends on context, allowing us to move beyond the dichotomy and embrace both approaches. Theory-theory is useful for drawing quick conclusions in familiar situations with generalized theories that apply to all, while simulation theory helps with getting a more specific, nuanced understanding of an individual's mental states. The complementary strengths of these theories demonstrate their collaborative potential, showing that mindreading is not a one-size-fits-all process, but rather a dynamic skill that adapts to various contexts.

Pluralists such as Jane Suilin Lavelle and Shannon Spaulding claim there are many ways to answer the question "How do we predict people's minds" depending on the context. These different ways are all valid interpretations of human behaviors because they are applied in multiple contexts for different purposes. Following pluralism, I propose that theory-theory and simulation theory can complement each other in context-specific ways. This section specifies the problem we are addressing, provides evidence for the compatibility of the theories, and analyzes how they can collaborate.

In addressing the ongoing debate between theory-theorists and simulation theorists, it is crucial to consider the scope in which mindreading is applied. As Lavelle points out in her book *Mindreading and Social Cognition*, traditional theorists of mind such as Stich and Nichols take mindreading to be ubiquitous, in a sense that we both "engage in mindreading for mundane chores . . . [and] in loftier endeavors like trying to glean Descartes's reasons for thinking that many ideas are innate."[22] Lavelle refers to this traditional view as "the ubiquity principle," which holds that mindreading is a constant, pervasive part of human interaction. Nonetheless, following the pluralistic perspective, I argue that it is superfluous and impractical for people to attribute mental states to everyone in every single daily interaction. Many interactions occur without conscious awareness of others' mental states, as these processes are so habitual that we seldom pause to analyze or think about them. When we are checking out at a convenience store, we do not bother to attribute mental states to the cashier. When the cashier says "19.99," we don't need to analyze their beliefs and desires, imagine ourselves in their position, or apply a theory to interpret their words and predict their intentions. Instead, we focus solely on the number we hear. Our cultural traditions, customs, and constant visits to supermarkets assured us that, in this context, the number the cashier states is simply the price we need to pay for our items. Since such scenarios happen constantly, we develop a mental shortcut that bypasses the need to consider the speaker's beliefs and desires to understand their intent. Instead, we respond automatically to the words we hear without

22 Shaun Nichols and Stephen P. Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Clarendon Press, 2003), 2.

referring to the speaker's mental state. Therefore, our reading of mind does not happen in mundane scenarios, but rather only in the scenarios that catch us by surprise or force us to think about other people's minds. We do not mindread the cashier when they say, "How are you?" or "Your total is $19.99," but when they start to act in a way that is absurd within the context, for example, when they burst into tears or start praying, people will be confused and utilize their ability of mindreading. In brief, mindreading does not happen on a daily basis. It is not a mundane or ubiquitous activity—people only mindread when something unexpected takes place. Some might argue that mindreading now sounds more like explaining or interpreting behaviors rather than predicting their mental states in hypothetical scenarios. To clarify, the term "mindreading" in this paper does not merely refer to forecasting others' mental states and behaviors, it also encompasses the explanation of off-script behaviors by referencing the agents' mental states. Having such an interpretive role, mindreading allows us to provide reasoning behind people's behavior (i.e., the cashier dancing or crying), and use the reasoning as evidence of people's beliefs and desires for predictions in the future.

My argument here aligns with the *script theory* raised by Nelson and Gruendel, they introduced the concept of a *script* in their 1979 paper as a typical sequence of events within a specific context or routine such as "going to bed, attending a birthday party, eating at a restaurant, taking a bath, or having a telephone conversation."[23] Though their experiment investigates the scripted content of children's conversation, their findings can be applied more broadly, that a large part of people's daily interactions are scripted. Adults also endorse commonly acknowledged routines across multiple contexts that, according to Lavelle, shape our expectations and "limit the amount of information one needs to process in a particular situation."[24] Script theorists claim that the script provides a framework and guides what we should focus on in a given context and what we can ignore. Since a large amount of people's daily interactions with each other follow the script, they are able to find the corresponding meaning of each other's actions for each script, which saves them from extensive mindreading and attributing mental states to each other. When following a script, predictions become mechanical without referencing the mental states of others. Thus, unlike the traditional view that supports the ubiquity principle of mindreading, real-world interactions do not always require attribution of mental states to people we are interacting with. The debate between theory-theory and simulation theory arises only when events deviate from the expected script or are entirely devoid of a script. It is only when people go completely off-script that they are compelled to see them as intentional beings with feelings and mental states rather than as scripted characters who interact with us mechanically.

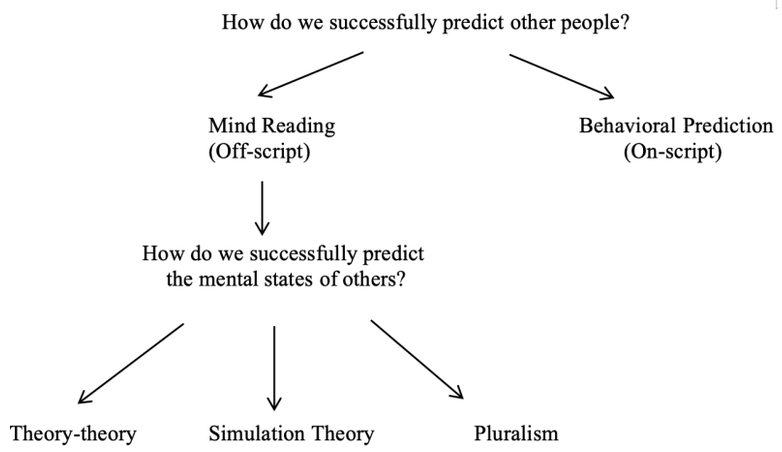
With this in mind, we can outline the problem at hand more clearly. There are two ways to answer the question, "How do we successfully predict other people?" The first is to predict according to the script and current actions and perform a behavioral prediction that only concerns their behavior. For example, we predict that the next move

23 Katherine Nelson and Janice M. Gruendel, "At Morning It's Lunchtime: A Scriptal View of Children's Dialogues," *Discourse Processes* 2, no. 2 (1979): 78.

24 Jane Suilin Lavelle, *Mindreading and Social Cognition* (Cambridge University Press, 2022), 36.

of the cashier will be to give me my receipt, because I asked for it. The second way is to perform a mental prediction, or mindreading. When people we are interacting with get off the script, we have no script to rely on and need mindreading to explain or predict their move (see *Diagram No. 1*). We think of other people's minds and mindread only when an off-script behavior takes place or when we are forced to think about other people's mental states, and the circumstances where people feel the necessity of ascribing mental states varies by person. The following discussions of the term "mindreading," therefore, implies the appearance of off-script behaviors.

The question for the new mindreading then is: "How do we successfully predict the mental state of others?" Several different answers can be considered as the answer to this question. There are the mainstream accounts—theory-theory and simulation theory, and there is the pluralist account—that we have different theories for different contexts and circumstances we are in. The distinction between mindreading and behavioral prediction clarified above is central to how we interpret the positions of theory-theorists and simulation theorists, as well as for generating thought experiments later on in this paper. In the following sections, I will challenge the traditional view and expand on pluralism to demonstrate how theory-theory and simulation theory can work together in a complementary way for different scenarios in mindreading.



(Diagram No. 1)

Many theorists of mind, even the traditional ones, have made the argument that theory-theory and simulation theory do not entirely contradict each other. Simulation theorist Goldman specifies in the footnote of his 2009 paper that his previous book *Simulating Minds,* "does not claim that all mindreading is executed by simulation. It leaves room for theory-based mind reading, and hence [his] overall approach is a simulation-theory hybrid."[25] Moreover, another simulation theorist Jane Heal offers a less ambitious version of simulation theory in her book *Mind, Reason and Imagination*. She points out that "[simulationists] do not yet have enough grip on how that project might

25 Alvin I. Goldman, "Mirroring, Simulating and Mindreading," *Mind & Language* 24, no. 2 (2009): 24.

actually work out in detail to have any confidence that we are working in terms of the right architecture."[26] This modest simulationist view acknowledges the limitations and complexity of the mindreading project. Even though they are significant defenders of simulation theory, neither Goldman nor Heal claims it is the only method people use to make predictions. Simulationists are not the only ones who take a step back on the battlefield, though the most prominent theory-theorists like Stich and Nichols are still holding the line. Peter Carruthers, another defender of theory-theory, sketched a limited view of simulation similar to Heal's version of simulation theory that a person can simulate the inferential role of a concept in others' minds relying on their grasp of the same concept. Carruthers claims that this limited simulation theory is something that "a theory-theorist should have no principled objection to."[27] His proposal allows the collaboration between the theory- and simulation theory that keeps the theoretical framework from theory-theory, and adds in advantages simulation has, such as "fine-grained predictions and explanations of thoughts, feelings, and actions of other people."[28]

Even though many traditional theorists of mind accepted the collaboration between the two theories, they still want to argue that their own theory is the dominant one instead of accepting the pluralist account that these two and more theories are equally explanatory for people's prediction behavior at different times for different purposes. To put it in Lavelle's words, they are still disagreeing upon the "how" question of mindreading, ignoring the other elements of the mindreading debate such as "what," "why," and "when." From a pluralist standpoint, the goal isn't solely to understand *how* people mindread. Instead, pluralists recognize the variations of theories among scenarios and cultures, emphasizing the value of exploring how mindreading varies in different contexts. Since the context determines which theory is applied, questions about "what," "why," and "when" we mindread are deeply connected to "how" we mindread. Pluralists view the two theories less as contradictory frameworks for explaining how people predict and more as different tools among a box of tools that individuals use for prediction in various circumstances. Although pluralism suggests that there is no single approach to mindreading and that individuals use different methods depending on the situation, pluralistic arguments focus more on various kinds of mindreadings while rarely addressing the specifics of when or where people apply either simulation theory or theory-theory—a gap that the following analysis aims to fill.

Collaborating Theories of Mind

Building on pluralism, I propose that simulation theory and theory-theory can function together not merely as one being a supplement for another, but are equally useful as distinct tools suited to different contexts due to their differing requirements for background knowledge for the target of mindreading. One major difference between the implications of the two theories is their acquisition of information. Theory-theory is often considered "knowledge-rich" because its application relies on using folk psychological theories to predict others' actions, which requires predictors to have the knowledge of theories. In contrast, simulation theory is viewed as "knowledge-poor," as it depends on imagining oneself in another's position, which requires minimal background knowledge about folk psychology. Just as the theories differ in being either knowledge rich or knowledge poor, they are more suited to certain types of scenarios corresponding to their features. Given their distinct characteristics, I argue that people apply different theories based on two contextual elements: (1) how familiar we are with the situation we're trying to predict, and (2) how well we know the person we are predicting. To give precise definitions, a person is familiar to me when I know the uniqueness of their beliefs and desires, especially the parts that are similar to or differ from generalized or stereotypical knowledge of a social group that they belong to. The word "situation" here implies specific contexts the predicted person is in, such as nationality, disease, and any factors external to their minds that affect our prediction. Nonetheless, it is important to keep in mind that since these situations all concern mindreading, they are all either off-script scenarios or circumstances in which people are forced to refer to other people's minds. For example, a familiar situation might be something that we have encountered several times, such as when a person receives his check and immediately shouts.[29] A severe example of an unfamiliar situation might be when a person discovers that her friend has secretly been campaigning for the president of their country. In the following paragraphs, I will evaluate four scenarios in detail, providing specific examples to support my argument that three of them qualify as mindreading scenarios, each involving the application of a different theory or theories. It is worth noticing that the examples discussed in this paper all focus on the mindreading of an individual's emotions and feelings for simplicity and consistency, while the argument can extend to all other mental states such as desires, beliefs, and decision-making.

(1) *Familiar Person with Unfamiliar Situation (FPUS)*

By mixing and matching these two dimensions, the four distinct scenarios in mindreading are: predicting a familiar person with an unfamiliar situation, an unfamiliar person with a familiar situation, a familiar person with a familiar situation, and an unfamiliar person with an unfamiliar situation (see *Graph No.2*). I argue that in each scenario, people tend to apply different theories of mindreading. To start with the first scenario, if a person has never encountered, seen, or heard of the situation they are in, it is by definition impossible for them to have a generalized folk-psychology theory and predict the mental states of others with it. Simulation theory is more useful here since it does not rely on pre-existing folk psychological theories. Instead, we simulate the mental

26 Jane Heal, *Mind, Reason and Imagination* (Cambridge University Press, 2003), 35.

27 Peter Carruthers, "Simulation and Self-Knowledge: A Defence of Theory-Theory," in *Theories of Theories of Mind*, edited by Peter Carruthers and Peter K. Smith (Cambridge University Press, 1996), 25.

28 Carruthers, "Simulation and Self-Knowledge," 25.

29 Note that a familiar situation happens when the attribution of mental states is necessary, yet the same situation has been repeatedly encountered in various ways. The event is not as mechanical as scripted behaviors but also not a brand new, extremely shocking circumstance as in unfamiliar situations.

state of a familiar person, using our past experiences of their reactions to predict future response. Moreover, as we are familiar with the predicted person, we are better equipped to attribute beliefs and desires to them and predict their mental states more accurately. This familiarity with the person allows us to refine our predictions by drawing on the historical tendency of their typical reactions or decision-making.

For example, if Amy is a five-year-old kid who wants to predict how her 30-year-old cousin Betty would feel when Betty's mom passes away, she tends to simulate Betty's mind instead of applying theories to her.[30] Suppose the "death of someone she knows" is a situation that Amy has never encountered in her life, and nor has anyone talked to her about the very notion of death. It is then impossible for her to develop a folk psychological theory in her mind about what death is and how it affects people's mental states. She lacks the developed theories that adults often have, such as those linking death to sadness, sorrow, and suffering. Betty tells Amy that after someone dies, you will never be able to meet them again. With no theory in this unfamiliar situation, Amy's only way to mindread Betty's feelings when her mother dies is by putting herself in Betty's shoes and imagining how she would feel if she could never see her own mother again. Amy knows that both she and Betty love their mothers deeply, and she can imagine the sadness and loneliness she would feel if her own mother passed away. By performing such a simulation with her knowledge of the situation Betty is in, Amy predicts by projecting her own emotions onto Betty that she might have a similar experience instead of utilizing a folk psychological theory such as, "If a person lost their mother, they tend to feel sad and cry." As a result, predicting a familiar person's reaction to an unfamiliar situation can be done without relying on a formal theory. Simulation theory brings us more successful predictions compared to theory-theory in the case of FPUS.

(2) *Unfamiliar Person with Familiar Situation (UPFS)*

When predicting an unfamiliar person in a familiar situation, theory-theory is most effective, as we apply general folk psychological theories based on prior experience with similar situations despite limited knowledge of the individual. Since we have encountered similar situations before, we have generalized folk psychological theories and are knowledge-rich in the context we are predicting. Yet, our limited knowledge of a stranger's mental states makes it difficult to accurately simulate their feelings or even build an imaginary world from their perspective. By drawing on general theories we've developed about different groups with specific traits, we can apply these insights to individuals within those groups and accurately mindread without needing to fully adopt their perspective. It is particularly useful when predicting the actions of strangers, whom we know little about and whose social group we can only infer from their appearances or behaviors.

For instance, if my friend asked me the question when we walked past an old European lady on the street, "What would the lady think after she sees a black cat walking past her?" I can barely relate to the old lady since we are strangers and I do not know

30 Death of a parent is an off-script event to Amy since it is not something that she can understand by mechanically interacting with her cousin. People relate to the relatives of the dead and naturally refer to their mental states.

her beliefs and desires at all.[31] It would be difficult for me to simulate and construct an imagined world from her perspective. Fortunately, I have a folk psychological theory that "If a European sees a black cat crossing their path, they may feel uneasy, viewing it as a sign of bad luck." I can apply this folk psychological theory to the old lady since she belongs to the social group mentioned in the theory, and predict her mental states without knowing her beliefs and desires. From this example, we can conclude that using theory-theory's approach to mind reading can be an effective and practical method when predicting the behavior of unfamiliar people in familiar situations. We use generalized theory to mindread when simulation is not available.

Critics might object that this scenario (UPFS) should not be considered as mindreading because theory-theory is less specific than simulation theory. Mindreading in this case is predicting the behavior of an entire social group rather than mindreading a particular individual, which diminishes both its particularity and accuracy. However, this concern is not warranted. While accuracy is certainly a key strength of any mindreading theory, a theory with relatively low accuracy does not mean it is rarely used in practice. Even simulation theory cannot guarantee one hundred percent accuracy in mindreading. In the UPFS scenario, theory-theory (predicting based on the typical mindset of a social group) is exactly what people rely on, and is the best option available. Without theory-theory, we would be left to guess the preferences and beliefs of an unknown individual, which introduces a vast array of possibilities and leads to no clear conclusions. Simulation theory cannot be effectively put into use in this scenario since we are not familiar with the predicted person. Therefore, while theory-theory may have comparatively lower accuracy than simulation theory in other scenarios, it is still the dominant (and perhaps the only) approach that enables our successful mindreading in this case.[32]

(3) *Familiar Person with Familiar Situation (FPFS)*

In familiar situations with familiar people, both theory-theory and simulation theory are at play. Familiar situations allow us to have a generalized theory while our familiarity with the person gives us prior knowledge about their unique beliefs and desires. Thus, we use general theories to quickly draw a conclusion while using simulation to make predictions more precisely about the individual we are predicting through our previous experience of their specific reactions and feelings, which explains why our predictions of familiar people and familiar cases tend to be more accurate.

For example, Amy from the first scenario is in her thirties and receives the message that her uncle, Betty's father, is very ill and about to pass away. Shaken by the shocking news, Amy tries to predict how Betty might feel about her father's passing. Since Amy is now a grown-up and is a lot more familiar with the situation of people she knows passing

31 This is not so much an off-script behavior, but it is a scenario in which I am forced to refer to other people's mental state because I am asked to do so. Therefore, it should be considered as a mindreading scenario.

32 The theories discussed and categorized in this paper do not guarantee success in mindreading; rather, they offer frameworks for how people engage in mindreading, without addressing the accuracy of the process.

away, she will quickly come up with a conclusion that Betty will be very sad according to the folk-psychology theory she generalized—"People generally feel sad when their parents pass away." However, Amy's prediction does not stop with the generalized theory as it did in the previous UPFS scenario. Instead, since Amy knows how much Betty loves her father and has constantly seen their interactions, Amy can further understand why Betty is sad and knows the objects and activities that would possibly trigger her sadness (i.e., crafts that Betty and her dad made together when she was young). When mindreading a familiar person, the prediction of feelings is much more nuanced and concise than mindreading an unfamiliar one, even though both types of mindreading can be considered successful. Since you know the person well, you're able to understand their perspective more deeply. In this scenario, we use theory-theory to reach a general conclusion, and simulation theory to predict one's feelings with much more detail. The two theories work together to bring us accurate predictions of the minds of people we know.

(4) *Unfamiliar Person with Unfamiliar Situation (UPUS)*

Lastly, it is almost impossible to predict an unfamiliar person in an unfamiliar situation because we completely lost the ground for any theory to be useful. Since it is an unfamiliar situation, people do not have a pre-existing theory readily available to apply. Moreover, people can barely simulate the individual they are predicting since they know nothing about their background mental states. There is no stable basis for the predictor to begin their mindreading. The result of mindreading in this scenario would likely be, "I don't know what that would be like." An example for UPUS might be reading in the newspaper that a man from a savage tribe is about to be sent to Saturn, and you are asked to mindread his feelings when he arrives there for the first time.[33] The man from the savage tribe is an individual who has completely different customs and habits from you, and you have never heard of someone being on Saturn. Thus, there is no way to tell whether he will be happy or sad since you have no reference point for where to begin. As there is no ground for the two theories to build upon, the only answer for UPUS is "I don't know." It is arguably possible to develop folk psychological theories about the savage after learning more about them, however, when we have a developed theory, they should already be considered a familiar situation to us. In comparison to the UPUS case, if I were to predict my friend's feelings when they arrived on Saturn (FPUS), I could reach a serious prediction since I know her desires and beliefs well. Through my prior knowledge, she loves space and especially Saturn, I can accurately mindread what she will feel when she arrives on Saturn using simulation theory. However, under the UPUS scenario, there is too little information that theory and simulation cannot reach serious predictive conclusions that allow us to reference and interact with the predicted individual accordingly.

33 Not all off-script cases can be successfully mindread. The example here is clearly an abnormal case in which we tend to refer to a person's mind. However, there is too little information for us to successfully mindread.

Person (P)/ Situation (S)	Familiar P	Unfamiliar P
Familiar S	ST and T-T	T-T
Unfamiliar S	ST	Unable to predict

(Table No.2: Four Scenarios for Contextual Mindreading)

Views Against Collaborative Theories

My pluralistic argument entails certain limitations on the simulation theory, particularly that it is not the dominant theory people use when dealing with the UPFS scenario. Simulation theorists like Goldman might argue that simulation is not limited by the familiarity of people, and we can still use simulation theory to predict people's mind when we do not know them. Prediction does not differ according to familiarity, instead, simulationists claim that we apply one single method for all circumstances—we imagine a world from the perspective of the person we are predicting using their desires and beliefs.[34] Going back to the example raised for UPFS, simulationists might claim that we can simply picture ourselves in the old European lady's shoes using general human knowledge and mindread by imagining how we will react when seeing the black cat to ascribe the same decision to the lady.

However, the problem with the case Goldman is making here is that we cannot attain the old lady's beliefs and desires without using a theory of mind. Since we do not know this lady beforehand, the only way for us to have reasonable assumptions about their beliefs and desires is to induce through the situation they are in. In order to get from the information in a situation to an individual's potential mental state while largely maintaining accuracy, we must apply a theory instead of using imagination. Imagining what the lady will do when sees a black cat without a starting point has so many possible results and is getting us nowhere. Instead, knowing that this old lady is European, we are able to apply the theory, "If a European sees a black cat crossing their path, they consider it as a sign of bad luck and want to avoid it," to get the conclusion that the lady might decide to step back and avoid the cat. The only way to draw trust-worthy conclusions about someone's mental state in an unfamiliar situation is to induce through our experiences with similar situations or the stereotypes we hold. Therefore, in order to have enough information about an individual's beliefs and desires to simulate their perspective, we need to apply theory-theory. Goldman might further argue that theory-theory here is just the way we gain information for performing simulation, while simulation is still the main theory we use.[35] Yet, applying folk psychological theories is still the crucial step that we must take under the UPFS scenario to make simulation possible. Moreover, theory-theory also seems to be a more direct route to get to the results of prediction compared to first applying theory-theory and then performing simulation given that we are already using

34 Goldman, "Conceptualizing Simulation Theory," 27–30.

35 Goldman, "Conceptualizing Simulation Theory," 44.

theories in this scenario. The indispensability and convenience of using theory-theory are why I identify it as the dominant theory for UPFS.

Another possible objection is a general one concerning the pluralistic nature of my argument. Since my view is grounded in a pluralistic response to the problem of prediction, it is important to address significant objections to pluralism that may also apply to my argument. The most immediate and urgent attack on the general pluralist account might be the commonality problem. The commonality problem arises when we define something as "A is B or C." For B and C to both belong to A, they must share a common feature. If such a feature exists, it replaces the disjunction as the definition of A. Yet, without a shared feature, A cannot coherently include both B and C.

The problem is more pressing when it comes to my proposal on mindreading. My definition of mindreading can be expressed as "Mindreading is 'ST for FPUS,' or 'ST and T-T for FPFS,' or 'T-T for UPFS.'" Having disjunction as a definition, I should encounter the same problem of commonality other pluralistic statements had. Critics might argue that these three scenarios are too distinct to belong under a single concept of mindreading. If there is no unifying feature, mindreading risks incoherence. Conversely, if I identify a commonality, mindreading will simply reduce to that feature, undermining my pluralist framework.

However, this critique misunderstands the nature of my pluralism and does not threaten the pluralism that I propose in this paper. By comparing the two statements, we can identify several crucial distinctions between my mindreading statement and the statement of defining A as B or C. An example of the former is the grue problem introduced by Nelson Goodman.[36] Grue is defined as either "being green before time t" or "blue after time t," an artificial grouping of unrelated properties. The statement answers the question, "What is grue?" However, definitions for green or blue are completely different. Grue lacks coherence because it forces unrelated concepts (green and blue) into a single category at the same conceptual level, rather than a higher-level category such as "color" that logically encompasses them. The connective, "or," in this definition creates a disjunction that lacks intrinsic logic and is ultimately arbitrary.

By contrast, my statement treats mindreading as a higher-level category unified by its purpose: understanding and interpreting others' mental states. It does not force unrelated scenarios into an artificial unity. ST, T-T, and their combinations are distinct methods, but they are unified because they share a common goal of mindreading. Nonetheless, this commonality in purpose does not dissolve the distinctions between the methods. The disjunctions in my statement reflect the diversity of methods, which entails the contextual adaptability of mindreading instead of a lack of coherence. Thus, the commonality problem does not undermine my pluralism, as the shared purpose of mindreading provides the necessary unity without collapsing distinctions between methods. Since "to mindread" operates at a higher conceptual level than the individual processes (ST, T-T, and their combinations), the critique that applies to the grue problem does not apply to my pluralist account of mindreading.

36 Nelson Goodman, "The New Riddle of Induction," in *The Pragmatism Reader: From Peirce Through the Present*, edited by Robert B. Talisse and Scott F. Aikin (Princeton University Press, 2011).

To summarize this section, simulationists might object that their theory can still function when we are not familiar with the person we are predicting. However, I point out that one's beliefs and desires can only be obtained through the familiar situation and general theories for unfamiliar individuals. This shows that theory-theory is crucial and irreplaceable under the UPFS scenario and should be the dominant theory people use. Furthermore, to answer the commonality problem, mindreading is a higher-order category that encompasses multiple valid processes (ST, T-T, and combinations thereof), each serving a shared purpose: understanding others' mental states, which defines mindreading. Yet, finding the definition of mindreading is not what we are looking for in this paper. Instead, I focus on when and how we engage in mindreading. My answer deals with the general inquiries about mindreading but does not seek to define it in a similar manner as the grue statement. This distinction ensures that my pluralistic proposal avoids the incoherence of the commonality problem while maintaining the flexibility needed to account for the complexity of mindreading.

Returning to the False-Belief Task

It is time to revisit the first section and connect the four scenarios to the well-discussed false-belief experiment. I argue against theory-theory that the cognitive differences observed between children at ages two, three, and five stem from biological development instead of the obtainment of folk psychology theories. Around the age of five, children's biological functions are mature enough to enable them to recognize differences in people's viewpoints. The ability to differentiate perspectives is critical for the mindreading behavior across the three mindreading scenarios. The recognition of this fact does not fall within either dimension (familiarity with people or situations) but instead serves as a precondition for mindreading to occur. Theory-theorists suggest that this change of understanding results from children at the age of five having acquired more folk psychological theories. According to this view, repeated experience familiarizes children with mindreading tasks, accelerating their development of awareness of different perspectives. Theory-theorists are taking an empirical bet that simply increasing a child's exposure to mindreading situations would accelerate the ability to obtain theories that differentiate perspectives before the biological readiness around the age of five. I challenge this claim and argue that the difference is rooted in biological maturation with the simulation theorists, but with more complex reasons.

Around the age of two, children's cognitive processes can be classified as a special case within the fourth non-mindreading category in *Table No. 2* in Section III—unfamiliar people and unfamiliar situations. At this age, children are not aware of the beliefs of others or representational folk psychological theories. Yet, this case is not that easy to get out of like any other examples in the fourth category, at least not by simply informing the children what theory they should use or changing the person they are predicting. Their unawareness of any mental state of others is not due to their unfamiliarity with them, but because of their underdevelopment in the biological brain. Changing the individual they are mindreading to their guardians would not help with the results. Similarly, children's unfamiliarity with situations is also not because they have not yet experienced them before—they are asked to mindread in experiments several times—but due to their lack

of the biological capacity to apply the theory and acknowledge the uniqueness of each person's mind. Children's brains are still developing when they are three-years-old and reach full mental capacity regarding desires and beliefs at the age of five.

In all, this paper explores the enduring question of how people predict each other's mental states and behaviors, proposing a pluralist framework that integrates theory-theory and simulation theory. By categorizing mindreading scenarios based on familiarity with individuals and situations, I argue that these theories are not mutually exclusive but contextually collaborative. Simulation theory dominates in predicting familiar individuals in unfamiliar situations and using personal knowledge of individuals to simulate while theory-theory is more efficient when predicting unfamiliar individuals in familiar contexts, relying on generalized folk psychological theories. In situations in which both the individual and the context are familiar, the theories work together, enabling more detailed and accurate predictions. This paper advances the pluralist argument by providing concrete scenarios and categorizations, addressing the potential triviality in the application of pluralism to real-life cases, and laying the groundwork for a more nuanced discussion on which theory should be applied in specific contexts to enable successful mindreading. The success of human prediction lies in our ability to flexibly employ different theories depending on the context, a skill essential for navigating diverse social situations. These insights not only advance philosophical debates but also hold practical implications for fields such as psychology. Future research could build on this framework by exploring additional mindreading theories and their application across diverse cultural and situational contexts, deepening our understanding of this fundamental human capacity.

Works Cited

Bartsch, Karen, and Henry Wellman. "Young Children's Attribution of Action to Beliefs and Desires." *Child Development* 60, no. 4 (1989): 946–64.

Carruthers, Peter. "Simulation and Self-Knowledge: A Defence of Theory-Theory." In *Theories of Theories of Mind*, edited by Peter Carruthers and Peter K. Smith. Cambridge University Press, 1996.

Goldman, Alvin I. "Conceptualizing Simulation Theory." In *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, 2006.

Goldman, Alvin I. "Mirroring, Simulating and Mindreading." *Mind & Language* 24, no. 2 (2009): 235–52.

Goodman, Nelson. "The New Riddle of Induction." In *The Pragmatism Reader: From Peirce Through the Present,* edited by Robert B. Talisse and Scott F. Aikin. Princeton University Press, 2011.

Gopnik, Alison, and Henry M. Wellman. "The Theory Theory." In *Mapping the Mind: Domain Specificity in Cognition and Culture,* edited by Lawrence A. Hirschfeld and Susan A. Gelman. Cambridge University Press, 1994.

Gordon, Robert M. "Folk Psychology as Simulation." *Mind & Language* 1, no. 2 (1986): 158–71.

Heal, Jane. *Mind, Reason and Imagination*. Cambridge University Press, 2003.

Lavelle, Jane Suilin. *Mindreading and Social Cognition*. Cambridge University Press, 2022.

Nelson, Katherine, and Janice M. Gruendel. "At Morning It's Lunchtime: A Scriptal View of Children's Dialogues." *Discourse Processes* 2, no. 2 (1979): 73–94.

Nichols, Shaun, and Stephen P. Stich. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Clarendon Press, 2003.

Nichols, Shaun, and Stephen P. Stich. "Folk Psychology: Simulation or Tacit Theory?" *Philosophical Issues* 3 (1993): 225–70.

Wellman, Henry M., and Jacqueline D. Woolley. "From Simple Desires to Ordinary Beliefs: The Early Development of Everyday Psychology." *Cognition* 35, no. 3 (1990): 245–75.

Black Nihilism, Existentialism, and Political Struggle

Allan Cao

Abstract: This paper explores Calvin Warren's conception of black nihilism as a potentially legitimate response to the prevalence of antiblack violence. I discuss the notion of antiblackness in an American context and the applicability of black nihilism to this context. I address the basis of the existentialist dismissal of nihilism in general and demonstrate black nihilism's rejection of the existentialist foundation. Specifically, I analyze Warren's usage of Heidegger's metaphysics as the foundation for his conception of antiblack violence and its result, the inevitability of black suffering. I consider Devonya Havis's response to antiblack violence as a challenge to this inevitability and propose Havis's Black Vernacular Phenomena and Performative Utterance as a form of meaning-creation that addresses the metaphysical foundations of Warren's black nihilism. I argue that Havis's Black Vernacular Philosophy is a stepping stone towards destabilizing the antiblack foundations of metaphysics while also following in Warren's skepticism of the rhetoric surrounding action, progress, and liberation.

Introduction

Calvin L. Warren's essay titled "Black Nihilism and the Politics of Hope" is centered around the defense of black nihilism against the dismissive viewpoint that believes that the nihilistic attitude is a problem that should be solved and addressed. For Warren, black nihilism should be understood as one of the possible responses to antiblack violence, specifically in the American historical context. It involves a movement of spiritual hope away from the assumed necessity of political investment, concrete action, and progress.[1] The essay discusses the philosophical foundations of the black nihilist position, placing it in conversation with relevant existentialist views and theories of black thought.

My goal is to analyze the philosophical moves that Warren makes in his defense of black nihilism in relation to the claims of some existentialist authors. Specifically, many existentialists assert that nihilism is an undesirable, indeed dangerous mindset to have, and that the creation of particularized meaning is possible, necessary, and actionable. I consider the specific relevance of Beauvoir's existentialist view of action and struggle to black nihilist theorizations of antiblackness in the American context and apply the existentialist logic of action and struggle to Fanon's theorization of black struggle and his own existentialist views. I then detail Warren's specific use and criticisms of Heidegger that ground his theories around political progress, metaphysics, and antiblackness. Along these lines, I consider Devonya Havis's use of Levinas as a link to Warren's use of Heidegger to compare their two perspectives of meaning-creation. Ultimately, I argue that one can view Devonya Havis's Black Vernacular Phenomena and Performative Utterance as a form of reconfigured spiritual hope that is not nihilist but still destabilizes the metaphysical foundations of the Political that Warren criticizes.[2]

Antiblackness in the American Context

In order to discuss black nihilism, one must understand how Warren conceptualizes antiblackness and how he relates it to epistemic and metaphysical conditions. Warren is thinking within the specific historical context of the United States, although he asserts that the analysis is a gateway for thinking about antiblackness globally. More specifically, Warren believes that the transatlantic slave trade was not merely a historical event that came and went but irreversibly imposed a condition of black suffering that continues to structure society.

Warren locates the Political as the site of black suffering because it is simultaneously where black suffering occurs and is channeled for a notion of betterment. Warren defines the Political as that which "constitutes an episteme of metaphysics, as a way to think being through a particular set of predispositions—progress, bio-futurity, change, betterment, and so forth. The 'political' (the uncapitalized 'p') dockets the *programmatic* effort

1 It is my intention that Warren's philosophy be used as a criticism of existentialist philosophy.

2 It is important to clarify my position. It is difficult to make value judgments on a topic like this because of my positionality. I am not black, nor is my positionality particularly "blackened." Furthermore, Warren and other black philosophers criticize how academia may commodify and objectify black suffering, turning it into an object of analysis or mere intellectual reflection. I am also indebted to the "Queering Masculinities" discussion group's conversation around Tommy Curry's *The Man-Not: Race, Class, Genre, and the Dilemmas of Black Manhood*. It gave me important perspective into what black suffering entails and the material conditions surrounding it, and what a healthy philosophical discourse looks like when discussing topics such as these. As put forth by the group, "What happens when social projection becomes so powerful that morality is no longer operable?" In what way am I capable of making value judgments of the positions and attitudes of others when I do not have even a fraction of the experience they have?

to materialize metaphysical sensibilities."[3] He identifies a litany of political programs and perspectives that demonstrate this view of black suffering and the impossibility of progress. He specifically cites the political failings of the Reconstruction Era (such as poll taxes, grandfather clauses, literacy tests, and extra-legal and legal violence), the convict leasing system sidestepping the 13th Amendment, and Martin Luther King, Jr.'s centering of black suffering and sacrifice as part of the black struggle for progress. In contemporary times, Warren cites the numerous murders of black people as signifiers of how the logic of the Political is haunted by the destruction and suffering of black bodies, as well as to highlight the failings of the Political to ameliorate the various injustices that black people face. Furthermore, Warren makes reference to a congregant from the Civil Rights era specifically stating that "we already heard that and tried that. Nothing has changed."[4] In this way, Warren critiques the Politics of Hope, which is how spiritual hope is necessarily placed within the political realm insofar as there is a political hope or the possibility of a could-be temporality that involves emancipation.[5] Warren seeks to reclaim this spiritual hope through the black nihilist position, arguing that the politics of hope reconfigures hope in a way where black people begin to invest in an impossible political project. Beyond his empirical evaluation of the American historical context, Warren also advances a theoretical approach to describing the contemporary condition of black suffering, using both to describe the impossibility of a political solution.

Black Nihilism and Political Movements

Broadly, the "threat" of nihilism is one that can be construed as global, in the sense that anyone may have an attitude of nihilism towards their circumstances. Warren asserts that Nietzsche's notion of nihilism is as "a particular crisis of modernity. The universal narratives and grounds of legitimation that once secured meaning for the modern world had lost integrity."[6] If political institutions both domestically and globally are thought of as being grounded in modernity, then in any case, one's attitude may be nihilistic towards their institutions or material conditions. For example, Tommy Curry in *The Man-Not: Race, Class, Genre, and the Dilemmas of Black Manhood* agrees with Warren. The fundamental problem that faces blackness is that "the Black man is not a normative subject; there is no 'should' question that does not imply his death. . . . The problem posed to the thinker aiming to describe and animate the Black man as a political subject capable of political life and social participation . . . [is that] he is asked to respect laws that do not respect him. . . . The consequence of this 'nonbeing' is dire, as there are no historical patterns of rationality or ethicality that project Black male existence into futurity."[7] While Curry explicitly identifies this problem with the black man's experience, Warren generalizes this to all black people. At face-value, the problem that is identified is the same. Black people are expected to invest their spiritual hope into political or existential programs that promise progress and liberation, while these emancipatory goals do not even account for blackness itself.

A core assertion that Warren makes in his theorization of black nihilism is that "the Political and anti-blackness are inseparable and mutually constitutive. The utopian vision of a 'not-yet-social order' that purges anti-blackness from its core provides a promise without relief—its only answer to the immediacy of black suffering is to keep struggling."[8] For Warren, the action and inaction are placed in a dichotomy where action always entails black struggle and inaction always entails black suffering. Furthermore, the black struggle is the only response to black suffering, and black struggle also always entails black suffering. This formulation of the Political effectively locks in the condition of black suffering. A forthcoming discussion of the existentialist view on action and struggle will help inform Warren's assertion that black suffering is inevitable. Black nihilism is the refusal to engage in the Political, to reclaim the hope for an end to metaphysical violence by disrupting the action/inaction dichotomy.

Broadly, this theorization of the inseparable nature of the Political and antiblackness is considered Afropessimist, which considers contemporary society and black existence to be continually structured by slavery. The nuances between Warren's theory of black nihilism and Afropessimism are beyond the scope of this paper, but it is worth discussing the broader context of other political thought that will elucidate black nihilism's position among other theories.

Among these include political movements centered around black joy as an act of resistance, as well as Afrofuturism, which concerns itself with the reimagination of black futures, identity, and agency, often through cultural aesthetics. These movements converge and diverge in various theoretical and practical moments. For example, Afrofuturists also seek to destabilize traditional notions of linearity and temporality,[9] like the black nihilist's critique of the Political use of the progress of betterment. Similarly, Black Joy acknowledges how past events and narratives may devalue black and brown individuals today.[10] However, these movements offer different responses to these conditions. The nuances of these responses are out of the scope of this paper, but a question that these diverse perspectives raise is how it is possible for similar views on a set of circumstances can lead to divergent responses, such as the Black Joy movement's emphasis on survival and development as opposed to black nihilism's notion of black-sacrifice.

The goal, then, is to analyze the philosophical moves in Warren's argument, which

3 Calvin L. Warren, "Black Nihilism and the Politics of Hope," *The New Centennial Review* 15, no. 1 (2015): 246, https://doi.org/10.14321/crnewcentrevi.15.1.0215.arren. Emphasis in original.

4 Warren, "Black Nihilism," 216–17, 228.

5 Warren, "Black Nihilism," 222.

6 Warren, "Black Nihilism," 224.

7 Tommy J. Curry, *The Man-Not: Race, Class, Genre, and the Dilemmas of Black Manhood* (Temple University Press, 2017), 189.

8 Warren, "Black Nihilism," 233.

9 Daylanne English, "Afrofuturism," Oxford Bibliographies, last modified October 24, 2024. https://www.oxfordbibliographies.com/display/document/obo-9780190221911/obo-9780190221911-0004.xml.

10 Elaine Nichols, "Black Joy: Resistance, Resilience and Reclamation," accessed February 24, 2025, https://nmaahc.si.edu/explore/stories/black-joy-resistance-resilience-and-reclamation.

will illuminate both what it means for him to reclaim spiritual investment, hope, into an antiblack political program and how black nihilism interacts with alternative political movements. In this case, Devonya Havis demonstrates how her use of Levinas aligns and deviates in very specific, demarcated ways from Warren's use of Heidegger that make these interactions explicit. Furthermore, one can trace black nihilism as a response to the European form of existentialism, which emphasizes the importance of action and struggle. Additionally, Frantz Fanon's political and existentialist work sheds light on how antiblackness and humanism may relate to action and political struggle. Investigating these traces of existentialism and humanism will develop a better picture of the black nihilist's critique of action, struggle, and suffering.

The Black Nihilist's Critique of Existentialism

An inspiration for Warren's defense of black nihilism comes from a conversation he has with a black woman on a train, where he was struck by her attitude towards her skeptical non-engagement with the Political.

> She began to talk about the government shutdown. "They don't care anything about us, you know," she said. "We elect these people into office, we vote for them, and they watch black people suffer and have no intentions of doing anything about it." I shook my head in agreement and listened intently. "I'm going to stop voting, and supporting this process; why should I keep doing this and our people continue to suffer," she said. I looked at her and said, "I don't know ma'am; I just don't understand it myself." She then laughed and thanked me for listening to her—as if our conversation were somewhat cathartic. "You know, people think you're crazy when you say things like this," she said giving me a wink. "Yes they do," I said. "But I am a *free* woman," she emphasized "and I won't go back." Shocked, I smiled at her, and she winked at me; at that moment I realized that her wisdom and courage penetrated my mind and demanded answers.[11]

There are two important qualities to this conversation. The first is the observation that "people think you're crazy when you say things like this." The second is the cathartic and freeing effect of the conversation itself. For Warren, the importance of the first quality is that it indicates there is an opposition to nihilism, viewing it as a "'disease of the soul' that produces callousness, meaninglessness, and masochism," wherein attitudes of nihilism are "pathological" and "immature" responses to one's circumstances.[12] The second quality will be explored later as a justification for the black nihilist position.

Many existentialist philosophers share a similar attitude in their opposition to nihilism. Simone de Beauvoir, in *The Ethics of Ambiguity*, asserts that the problem with the nihilist is that they do not recognize "the importance of that universal, absolute end which freedom itself is."[13] For Beauvoir, freedom is the end towards which one must always be oriented, and this freedom must be devoted towards the freedom of others as well. To understand Warren's objection to this view of nihilism, it is necessary to analyze

11 Warren, "Black Nihilism," 245. Emphasis in original.

12 Warren, "Black Nihilism," 224.

13 Simone de Beauvoir, *The Ethics of Ambiguity* (Kensington Publishing, 2002), 57–58.

Beauvoir's view of freedom, ambiguity, and the liberation from oppression. Beauvoir asserts that "freedom realizes itself only by engaging itself in the world: to such an extent that man's project toward freedom is embodied for him in definite acts of behavior."[14] This is to say that to put forth one's freedom, one must act. Action becomes privileged for Beauvoir when she argues that only through action can one be free. Furthermore, Beauvoir says this about action: "through his own struggle he must seek to serve the universal cause of freedom."[15] In this quote, Beauvoir explicitly references the need to struggle for freedom, that this struggle must always be oriented towards the removal of oppression for others. Notice here the emphasis on struggle; in the face of immediate suffering and oppression, it is only through struggle for oneself and others that one can be free.

Warren criticizes this obligation that Beauvoir puts forth. This commitment to action, to a project with an end in which the world is bettered, is exactly how a "logic of struggle, then, perpetuates black suffering by placing relief in an unattainable future, a future that offers nothing more than an exploitative reproduction of its own means of existence. Struggle, action, work, and labor are caught in a political metaphysics that depends on black-death."[16] This logic of struggle towards an unattainable future finds its roots both in Warren's empirical evaluation of the American historical context as well as his philosophical foundation in Heidegger and its application to the Political. More specifically, the Politics of Hope involved in the logic of the Political configures black political action towards an impossible object that ignores the historicity and metaphysical formation of political progress.[17] It is important to note that this analysis of action and existentialism is not meant to singularly critique the work of Beauvoir, nor is it to set her philosophy up as the token example of existentialism. However, insofar that the existentialist project asserts that one must always create their own meaning and that the thrownness of a situation ensures that one must always contend with themselves, then struggle and action are the ways in which one asserts themselves into the world from an existentialist perspective.

The Black Nihilist and Fanon

In Warren's essay, he draws upon much of Fanon's analysis of the black experience. Furthermore, Fanon's emphasis on the psychological impact of systems of colonialism in *Black Skin, White Masks* serve as an example of the historical analysis of antiblackness that Warren uses. In line with Warren's analysis of the paradoxical nature of the Political, Fanon finds that for "the black Frenchman the situation is unbearable. Unsure whether the white man considers him as consciousness in-itself-for-itself, he is constantly preoccupied with detecting resistance, opposition, and contestation."[18] Warren sees Fanon as correctly

14 Beauvoir, *Ethics of Ambiguity*, 78.

15 Beauvoir, *Ethics of Ambiguity,* 89–90.

16 Warren, "Black Nihilism," 233.

17 Warren, "Black Nihilism," 221.

18 Frantz Fanon, *Black Skin, White Masks,* trans. Richard Philcox (Grove Press, 2008), 196–97.

identifying the dilemma in which black people find themselves, that they are expected to always be struggling, to invest themselves into the political system in which past generations fought so intensely to be included. In the conclusion of *Black Skin, White Masks*, Fanon shares his sentiment that he no longer wishes to be chained by history, by the wrongs of white people, and by the suffering of black people. However, Warren would take issue with the orientation of his sentiment. For Fanon, "if the question once arose for me about showing solidarity with a given past, it was because I was committed to myself and my fellow man, to fight with all my life and all my strength so that never again would people be enslaved on this earth."[19] Fanon privileges action and struggle even in his more existential work. In this way, he is oriented towards a "utopian" dream where black people are no longer constantly existing in relation to white destiny. Warren argues that this orientation is not only harmful but also implicates the destruction of the world itself. The same argument against Beauvoir applies here, where the expectation of struggling to solve the immediacy of black suffering is problematic. The caveat here is that this commitment to struggle only exists to an extent. Fanon's personal acceptance of a racialized history was done to form a politically efficacious response to colonialism.[20] In other words, I assert that Fanon's philosophical move is two-fold. He believes that political response is important only in the way of creating the conditions in which a new, non-Western form of humanism would be possible. It is this belief that is the foundation of his existentialist position. He emphasizes the psychological condition of colonialism as justification for the need to dismantle the system to resolve the racism associated with it.

The Black Nihilist's Position on Metaphysics

The previous section's analysis of the existentialism in the backdrop of black nihilism leads to the question of how black nihilism is different from traditional nihilism. Additionally, it brings up important points of analysis regarding black existentialists, such as Havis and Fanon. In what ways do Havis and Fanon align with the fundamental problem of antiblackness that Warren describes? Do they implicitly or explicitly criticize positions of nihilism? Do their solutions to antiblack violence fit within Warren's theoretical framework?

The key feature of Warren's black nihilism is the introduction of Heidegger's criticism of metaphysics. I argue that the explanatory power of Warren's use of Heidegger is what puts forth antiblackness as this insurmountable obstacle to the black nihilist. The pertinent question here for Warren is "can we think at all without anti-blackness?"[21] Here, we should draw upon Heidegger's notion of anxiety and Nothing and Warren's analysis of Heidegger's critique of metaphysics.

Warren simultaneously draws upon Heidegger's understanding of metaphysics from "What is Metaphysics?" while also critiquing it. Heidegger's analysis begins with the Nothing and anxiety. For Heidegger, there is no specific object of anxiety. Rather, it is the Nothing that structures Being through the feeling of anxiety. In other words, anxiety does not involve being anxious about something, but rather about no-thing (Nothing). When one is anxious, Being is held out into the Nothing, as Dasein. Dasein is the specific term that Heidegger uses to describe human beings as "being-in-the-world," in which one encounters the Nothing. This interaction with the Nothing is essential for the meaning-creation of Dasein.[22]

For Warren, the Political's encounter with antiblackness is analogous to Being's encounter with Nothing.

> For the black nihilist, anti-blackness *is* metaphysics. It is the system of thought and organization of existence that structures the relationship between object/subject, human/animal, rational/irrational, and free/enslaved—essentially, the categories that constitute the field of Ontology. Thus, the social rationalization, loss of individuality, economic expansionism, and technocratic domination that both Vattimo and Heidegger analyze actually depend on anti-blackness. Metaphysics, then, is unthinkable without anti-blackness. Neither Heidegger nor Vattimo explores this aspect of Being's oblivion—it is the literal destruction of black bodies that provide the psychic, economic, and philosophical resources for modernity to objectify, forget, and ultimately obliterate Being (nonmetaphysical Being).[23]

In this way, the epistemological categories on which modernity depends, the way morality and thought are rationalized, and indeed, the "organization of existence" itself is reliant on continued antiblackness. In order for society as we know it today to continue to exist, the destruction of black bodies is necessary. Society sustains itself on black bodies to ground "its institutions, economic systems, environment, theologies, philosophies."[24] For example, black bodies are used as capital via incarceration.[25] Furthermore, society uses antiblackness to self-actualize, where the black body is the vehicle for social betterment, and similarly sustains the logic of modernity and metaphysics.[26]

Warren focuses on the interaction with the Nothing and its importance for meaning creation but locates the function of the Nothing within antiblackness. By doing so, he questions the ability for the black-as-object to act as Dasein, or in other words, have access to meaning creation. For black subjectivity, then, there is no access to anxiety, no access to the Nothing that structures being, because black bodies themselves are rendered as *nothing*. Black subjectivity has no meaning-creation because the condition of antiblackness prevents them from doing so. This is their ontological death. Indeed, Curry

19 Fanon, *Black Skin, White Masks*, 202.

20 Robert Bernasconi, "Situating Frantz Fanon's Account of Black Experience," in *Situating Existentialism: Key Texts in Contex*t, ed. Jonathan Judaken and Robert Bernasconi (Columbia University Press, 2012), http://www.jstor.org/stable/10.7312/juda14774.15, 353.

21 Warren, "Black Nihilism," 237.

22 Martin Heidegger, "What Is Metaphysics?" in *Basic Writings: From Being and Time (1927) to the Task of Thinking (1964),* ed. David Farrell Krell (Harper & Row, 1977), 92–93.

23 Warren, "Black Nihilism," 237. Emphasis in original.

24 Warren, "Black Nihilism," 239.

25 Warren, "Black Nihilism," 243.

26 Warren, "Black Nihilism," 218.

asks whether hope can "even be justified in a white-supremacist society that murders Blacks to maintain its social order."[27] Warren continues to say that it's not only the case that metaphysics is predicated on antiblack violence, but it also relies on black suffering and struggle itself to sustain itself. "The struggle is presented as a spiritual virtue, and the spiritual concept of hope is contaminated with the prerogatives of a political order."[28] It is here that we can refer back to the critique of existentialism. Political metaphysics relies not only on black suffering, but also black *struggling* to uphold civil society. With this in mind, the existentialist solution for meaning creation becomes poisoned by its reliance on antiblackness to even make sense of the notion of "meaning creation."

To go further, Warren's criticism of the Political does not just extend to the traditional modes of governance that we see in the contemporary United States. It also applies to the notion of progress and liberation itself. Warren asserts that "it is a discourse of hope, a politics of hope that advances the belief that we can weaken metaphysics and reduce suffering, violence, and pain. When it comes to black suffering, however, we are compelled to hold up the mirror of historicity and inquire about the possibilities of emancipation for the black-as-object."[29] Warren draws upon Heidegger to claim that the nexus of antiblackness is now beyond level of human existence.[30] In other words, black people have been "thrown" into a situation where the effects of the transatlantic slave trade are no longer considered merely an event, but a conditioning of existence itself that has been tied directly to blackness. Because of this, black people do not have access to *Dasein* because *Dasein* itself is reliant on antiblackness in its "movement toward Being."[31] This analysis is caught up in the fundamental question that Warren poses for existentialists and the advocates of the Political. Here, he criticizes not only the political institutions of modernity, but also the discourse surrounding the possibility of liberation for existentialists. In other words, existentialists assume that liberation is possible, but for the black nihilist, it isn't. It is Warren's unique usage of Heidegger that allows him to put forth this argument. Liberation does not exist for the black-as-object because metaphysics itself and the way metaphysics conceives of liberation fundamentally treats black people as *nothing*. Furthermore, because metaphysics is so dependent on antiblack violence, to imagine an end to antiblack violence is to imagine an end to metaphysics, and thus the end of the world itself.[32] As such, fundamental projects of life and the creation of meaning both individually and universally are predicated on black suffering and thus are inaccessible and inconceivable for the black body.

The Black Nihilist and Havis

27 Curry, *Man-Not*, 181.
28 Warren, "Black Nihilism," 229.
29 Warren, "Black Nihilism," 238.
30 Parisa Vaziri, "Black Execration: A Review of Calvin Warren, Ontological Terror: Blackness, Nihilism, and Emancipation," *Postmodern Culture* 29, no. 3 (2019): https://doi.org/10.1353/pmc.2019.0009.
31 Vaziri, "Black Execration."
32 Warren, "Black Nihilism," 238–39.

The question of the possibility of liberation and meaning creation is approached differently by Havis's *Performative Utterance*. It is her reference to the possibility of liberation as "the move from *Performative Utterance* to Black Philosophy, which seeks Black liberation," opening "up an interrogation of the stakes of 'truth' and 'liberation' with respect to the management of Black bodies" which Warren continues to problematize.[33] At the same time, Havis demonstrates an understanding of black suffering that coincides with Warren. Havis is sympathetic to the notion of the impossibility of traditional political progress. She asserts that "denial concerning how impossible it is to produce the required proof for universal Black humanity continues to fuel even more rigorous attempts to ascertain redemption through letters."[34] Both Havis and Warren criticize the notion of black liberation as a re-appropriation of humanism. In this case, Havis and Warren are speaking of a European form of humanism that originates in Enlightenment-period conceptions of reason and the self as a basis for defining humanity, with Warren focusing more on the political recognition that humanism grants. They view this humanism as one of the justifications that allowed for the transatlantic slave trade. In this case, the logic of the Political goes beyond Fanon's understanding of the black man's destiny as the white man's destiny. The formation of a black humanism is untenable for Warren and Havis. For Warren, humanism involves a metaphysical violence that revolves around the use of a humanist grammar and political discourses of betterment that colonize the spiritual practice of hope to sustain an antiblack organization of existence.[35] For Havis, this focus on humanism is bankrupted by Modernity and Western paradigms of what it means to be human, insofar as the focus on humanism has traces of a crypto-Enlightenment tendency that treats the Black as an object of knowledge. Havis asserts that this definition of reason renders the Black subhuman and that a humanistic enterprise is bankrupt because it is "unable to produce a Black reason" in the context of the paradigms of Western philosophy like the Enlightenment.

In light of these problems with humanism, Havis identifies the Black Vernacular Phenomena as a shift away from humanism in which a Performative Utterance disrupts Western binaries that are sourced from metaphysical organization. She begins this analysis with an excerpt from *Alice in Wonderland* and describing the "play" that occurs in Wonderland.

> The adventure of wonderland is the "*play*" that springs forth in an active indeterminacy without the pursuit of resolution. *Black Vernacular Phenomena*, and the specific modalities that share its character, create a "wonderland" in which what seemed obvious and given fades into perpetual traces upending the given and pointing toward the illusion of mastery. In *Black Vernacular Phenomena* we come face-to-face with an otherwise, a beyond which, that exceeds our capacity to be thought. It is an asking that cannot be completely answered, a rhythm that exceeds the utility of mechanical time. These *traces*, noted in a peculiar excess,

33 Devonya N. Havis, *Creating a Black Vernacular Philosophy* (Lexington Books, 2023), 11–12. Emphasis in original.
34 Havis, *Creating a Black Vernacular Philosophy*, 14.
35 Warren, "Black Nihilism," 242–43.

mark an ever-present gap between thought and its containment in concept.[36]

There are many points where Havis aligns with Warren when she refers to the importance of "play." Throughout her project, she refers to the "play" in various different contexts alongside a litany of other terms that she describes as unable to be contained by concept. Broadly, these terms have their foundation within Levinas's concept of alterity, so it is worth giving some examples of how Havis herself uses the word "play" and how it relates to Levinas as a theoretical foundation. The "play" refers to a dynamic process that is open to possibility, or an *otherwise*. For example, the adventure of wonderland engages in "play" because the concrete notions that one might take as given, are actually dynamic in their meaning. As such, "play" explores these multiple meanings through *alterity*. "Play" also occurs between the gaps in traditional hierarchical limits that do not fully encompass reality. It is worth briefly grounding this within the philosophical foundation of Levinas. For Levinas, the relationship of the self and the "Other" and alterity is privileged instead of a concrete notion of the self. Furthermore, the relationship with alterity has a structuring effect that allows for possibility beyond the self, including sociality. It is precisely this theorization of alterity and Havis's notion of a black Other that demonstrates the difference in philosophical foundation between Havis and Warren, and Levinas and Heidegger accordingly. Whereas Heidegger is concerned with the Nothing as structuring Being via Anxiety, Levinas uses the Other as alterity to demonstrate that the recognition of the unknowable other allows spaces for possibility. That is, one does not need to be oriented towards Nothing but rather exploring differences and possibilities in the Other. This ontological distinction has political ramifications when we relate it back to "play." The "play's" "active indeterminacy without the pursuit of resolution,"[37] is something that would disrupt the Political because there is no notion of completeness and no promise for betterment for Havis. Black Vernacular Phenomena also is unable to be coded by traditional Western metaphysics, as it "exceeds our capacity to be thought."[38] There is no linear temporality that forces blackness to give answers for antiblack futurity because there is "a rhythm that exceeds the utility of mechanical time."[39] In these ways, Havis seems to adequately respond to the oppressiveness of the Political by placing the metaphysical problem with Levinas instead of Heidegger. By doing so, she is able to reimagine blackness by "playing" in the gap, in alterity, instead of Warren's ontologizing of the black condition itself through Heidegger.

The question, then, is whether Havis's Performative Utterance adequately removes itself from the ontological problems that Warren puts forth. Havis's Performative Utterance is a disruption to Western philosophy's binary oppositions. Here, it is essential to point out their common verbiage of metaphysical "residue" and epistemological meaning. Havis asserts "that in the *performance* there is a distinction between the *Black* and *Blackness* which entails a distance, a tracing of what could have been there, but has now passed away leaving residue in its wake. *Performative Utterance* is the *play* emerging in the

36 Havis, *Creating a Black Vernacular Philosophy*, 7. Emphasis in original.
37 Havis, *Creating a Black Vernacular Philosophy*, 7.
38 Havis, *Creating a Black Vernacular Philosophy*, 7.
39 Havis, *Creating a Black Vernacular Philosophy*, 7.

observation that vision is biased, that meanings are multiple."[40] For Havis, the Black is a reified and essentialized notion of what it means to be black, whereas Blackness is the open-ended notion of this meaning,[41] indicating a gap between a definite characterization of the Black and dynamic sense of Blackness. The analysis of the gap between the Black and Blackness is not found in Warren. Havis finds that the performance is what "plays" in this gap, while Warren does not find a gap, but a paradox. Warren's paradox is society's destruction of the black body while positing the black body as a requirement. As such, any "performance" from the black body is rendered pathological for blackness itself.

Havis's theorization of the gap between the Black and Blackness is important because within it lies the possibility of escaping the social projection of Modernity. To make sense of this social projection, Curry provides a framework of study. "The Man-Not grows from the incongruity I observed between what theory claims to explain and the *actual* existence of Black men and boys—an *actual* reality that remains excluded from its purview."[42] In other words, the Black male is "projected" onto as a disease to society, similar to Warren's analysis, but there is a gap between the projected Blackness and the existence of the black body. Curry's analysis focuses on the contradictions that produce this gap, then uses the Man-Not as coined by Sylvia Wynter. Havis focuses on what to do with the gap. Havis finds that "the vernacular *play* is an oscillation between what *the Black* is presumed to be (i.e., not a lady, not articulate, not clean) and Blackness as a dynamic experience that clears the way for possibility in the practice of interrogation."[43] In this way, the Performative Utterance's play in the gap between the Black and Blackness is destabilizing to traditional Western metaphysics and categories. It does so by refusing to be static, to always be changing in ways that render the meaning it creates unintelligible to Modernity. If we translate Performative Utterance to the verbiage of Warren, then Performative Utterance would be a form of anti-grammar to the language of metaphysics, something that sits in the contradictions that the West has created for antiblackness. If we look back at the "multiple meanings" that arise from Performative Utterances, then it follows that it combats the "translation of this anti-grammar into a system of understanding that is designed to exclude it. This tension between grammatical exclusion and compulsory inclusion is part of the violence of captivity."[44] Warren finds the contradiction to be violent, but for Havis, it is generative in ways that antiblack epistemology cannot even capture. The social projection, or the collection of antiblack expectations that are put on the black body, is rejected in the Performative Utterance. Havis describes Nina Simone and Thelonious Monk as rejecting the logic of the Political and playing oneself respectively, which rupture the very way that the Black is inscribed in the black body via an invocation of Black Difference.[45] Here, Havis draws upon both Fanon and Derrida to ground her argument. Havis replaces Fanon's analysis of the epidermalization of an inferiority complex with

40 Havis, *Creating Black Vernacular Philosophy*, 12. Emphasis in original.
41 Havis, *Creating Black Vernacular Philosophy*, 4.
42 Curry, *Man-Not*, 7. Emphasis in original.
43 Havis, *Creating Black Vernacular Philosophy,* 23.
44 Warren, "Black Nihilism," 241.
45 Havis, *Creating a Black Vernacular Philosophy*, 23. Emphasis in original.

the epidermalization of the Black. Then, Havis defines Black Difference as "a dynamic process of engaging difference, *alterity*, and playing in and through those differences,"[46] which destabilizes the static concept of the self, i.e., playing oneself. This playing oneself is irreducible to Western understandings of ontology and reflects the function of engaging with alterity to disrupt and look beyond the binary oppositions and traditional conceptions of the "*self*, liberation, revolution, and ethics."[47] Black Difference, then, is a vehicle for engaging in alterity in ways that deconstruct the meaning of concepts.

In these ways, Havis presents a convincing way of *doing* philosophy. There is an interplay of pessimistic attitudes towards the Political with optimistic attitudes towards rethinking liberation, solidarity, and action. It is here that Havis and Warren once again depart. For Havis, "the dialogic quality of these performances place one always in a context of discursivity where the greater appeal to the individual comes from the Other. These vernacular performances characterize the *Black Difference*. *Performance* allows direct expression or indirect commentary and serves to motivate participants to action."[48] Havis finds herself once again at the nexus of action. Even after a complete rethinking of what it means to act, the Performative Utterance still situates itself in terms of the other. Action is towards the betterment of material conditions, which for Warren will never come to fruition. Warren's critique of metaphysical violence does not stop at the logic of the Political, which Havis deconstructs. The black nihilist's position is that "putting an end to metaphysics will also put an end to the world itself—this is the nihilism that the black nihilist must theorize through. . . . The black nihilist has as little faith in the metaphysical reorganization of society through anarchy than he does in traditional forms of political existence."[49] Even the notion of action itself is part and parcel of the biopolitical ordering. As such, "the nihilist does not promise redress within the structure of the political, for this is impossible, but offers, instead, *rejection* of the political as a spiritual practice itself."[50] Warren's stance here is radical and uncompromising. By placing the black body as the no-thing that Heidegger theorizes for being, "it is impossible to end metaphysics without ending blackness, and the black nihilist will never be able to withdraw from the Political completely without a certain death-drive or being-toward-death."[51] Reading this in relation to Havis and Fanon, there is no universal humanism, nor is there a movement from Performative Utterance, to Black Vernacular Phenomenon, to black philosophy that can adequately "liberate" the black body from the destruction that Modernity necessitates without already being towards death. In other words, the unfortunate situation that the Black finds itself in is that it must orient itself towards its own death in order to free itself from death.

Curiously, Warren laments that there is a "lack of a coherent grammar to articulate these dilemmas."[52] We know that Fanon's grammar involves, in the end, a universal humanism that assuredly fails in Warren's view. But it seems that Havis's view formulates a grammar, Performative Utterance, in which these dilemmas are able to be articulated. "Recalling in these lines the ongoing tension between the *proper* and the *Black*, Simone's performance generates the critical trace as *critique*. The listener experiences, insofar as it is possible, a phenomenon that cannot be captured in concept or word—the Black Vernacular that at once states what appears and in the same moment turns it around to refashion it anew."[53] Furthermore, "The silent commentary which leaves the critical trace is performed by Simone in what she never says but implies. This is called signifying. As Ellison writes, ""signifying" here meaning, in the unwritten dictionary of American Negro usage, "rhetorical understatements."""[54] Here we already see that Warren's critique of signifying as something that is always a Western metaphysical organization is itself being destabilized by the ambiguity that Havis embraces via the Performative Utterance. In this way, is Havis's approach to Black Philosophy, while still rooted in liberation and betterment, an alternative to black nihilism? In other words, does the end to antiblackness necessarily mean the end of the world? Is black nihilism the only way to rescue hope from political investment? If it is true that Havis's Performative Utterance provides a way to disrupt antiblack epistemology, then the notion of "progress" that is unintelligible by Western metaphysical organizations is indeed viable at least for the individual. Because of Havis's difference in the metaphysical theorization of antiblackness, she can conceive of "solutions" to antiblack violence that still fit within the theoretical framework that Warren provides. This does not delegitimize black nihilism as a position but rather puts forth a possible alternative to direct spiritual hope. While Warren objects to even using the term "solution," Havis nevertheless represents an optimistic orientation towards antiblack violence.

The Consideration and Clarification of Black Nihilism

Calvin Warren's response to the prevalence of antiblack violence and the perceived lack of political progress culminates in a critique of the Politics of Hope and the theorization of black nihilism that retrieves the spiritual concept of hope. This analysis is done in the American context, where the contested historicity of movements, progress, and political solutions reveal the necessity to clarify the philosophical foundations of black nihilism. This is because alternative political movements, such as Black Joy as resistance and Afrofuturism, demonstrate both converging and diverging viewpoints on the condition of black suffering as well as the possibility for emancipation.

To clarify how Warren defends black nihilism from those who criticize it as a pathological reaction to antiblackness, I argued that tracing the philosophical foundations and lineages of various thinkers provide insight into how Warren treats antiblackness and the impossibility of progress. This tracing involved using existentialists to understand how Warren conceptualizes black suffering as inevitable in the context of black struggle, action, meaning-creation, and political betterment. Then, I discussed Fanon to show how Warren

46 Havis, *Creating a Black Vernacular Philosophy*, 15. Emphasis in original.
47 Havis, *Creating a Black Vernacular Philosophy*, 9.
48 Havis, *Creating a Black Vernacular Philosophy*, 95.
49 Warren, "Black Nihilism," 244.
50 Warren, "Black Nihilism," 229. Emphasis in original.
51 Warren, "Black Nihilism," 244.
52 Warren, "Black Nihilism," 245.
53 Havis, *Creating a Black Vernacular Philosophy*, 24. Emphasis in original.
54 Havis, *Creating a Black Vernacular Philosophy*, 23.

was both influenced by and critical of the political praxis and existentialist tendencies of Fanon's analysis. It demonstrated the impact of colonialism and the transatlantic slave trade on Warren's understanding of the contemporary conditions of black suffering, as well as how Warren differs from Fanon's own understanding of black struggle for political progress.

After discussing the role of action and struggle in Beauvoir's existentialism and applying it in the context of Blackness with Fanon, we then discussed Warren's Heideggerian influence and criticisms. In short, Warren uses Heidegger's analysis of the Nothing and Dasein as a foundation for understanding how antiblackness structures the fabric of metaphysics, including the Political and society in general. This provided a line of comparison to Devonya Havis's Performative Utterance through her own use of Levinas as a philosophical foundation. This foundation was the theorization of the "Other" as a Black Other. This, alongside the theorization of the Performative Utterance as capable of destabilizing traditional Western metaphysical concepts like the self and linear temporality via the "play," demonstrated the capacity to conceive of Havis as a possible alternative to black nihilism without outright pathologizing or rejecting it. This was because Havis's use of Levinas and Warren's use of Heidegger provided the philosophical foundation to draw distinctions between the two arguments on the possibility of meaning-creation and a world without antiblackness while also providing a common ground in order keep them in conversation with one another.

Finally, the goal of this essay was not to definitively decide on the "correct" approach to understanding antiblackness, but rather to illuminate the different ways in which people may conceive of it. The aim is for one to walk away from this with a clearer understanding of how competing perspectives on the possibility of change and the conditions of black suffering came about. By grounding these perspectives within the relevant veins of philosophical thought, it is my hope that the reader can better interpret the claims that black nihilism makes in order to seriously consider its arguments when thinking of the nature of political struggle. The essay also serves as an example of this consideration by putting Havis's Performative Utterance in conversation with the black nihilist without an immediate rejection of its perspective.

Works Cited

Beauvoir, Simone de. *The Ethics of Ambiguity*. Kensington Publishing, 2002.

Bernasconi, Robert. "Situating Frantz Fanon's Account of Black Experience." In **Situating** *Existentialism: Key Texts in Context*, edited by Jonathan Judaken and Robert Bernasconi. Columbia University Press, 2012. http://www.jstor.org/stable/10.7312/juda14774.15.

Curry, Tommy J. *The Man-Not: Race, Class, Genre, and the Dilemmas of Black Manhood*. Temple University Press, 2017.

English, Daylanne. "Afrofuturism." Oxford Bibliographies, last modified October 24, 2024. https://www.oxfordbibliographies.com/display/document/obo-9780190221911/obo-9780190221911-0004.xml.

Fanon, Frantz. *Black Skin, White Masks*. Translated by Richard Philcox. Grove Press, 2008.

Havis, Devonya N. *Creating a Black Vernacular Philosophy*. Lexington Books, 2023.

Heidegger, Martin. "What Is Metaphysics?" In *Basic Writings: From Being and Time (1927) to the Task of Thinking (1964)*, edited by David Farrell Krell. Harper & Row, 1977.

Nichols, Elaine. "Black Joy: Resistance, Resilience and Reclamation." Accessed February 24, 2025. https://nmaahc.si.edu/explore/stories/black-joy-resistance-resilience-and-reclamation.

Vaziri, Parisa. "Black Execration: A Review of Calvin Warren, *Ontological Terror: Blackness, Nihilism, and Emancipation*." *Postmodern Culture* 29, no. 3 (2019): https://doi.org/10.1353/pmc.2019.0009.

Warren, Calvin L. "Black Nihilism and the Politics of Hope." *The New Centennial Review* 15, no. 1 (2015): 215–48. https://doi.org/10.14321/crnewcentrevi.15.1.0215.

Sculptor vs. Detective: An Analysis of Gender Identification

Kabeer Haider

Abstract: This paper analyzes whether we can shape our gender identity. This is presented through an analysis of the opposing sculptor and detective accounts of gender identity. The sculptor account presents that we have an active role in shaping our gender identity, and the detective account posits that we have a passive role regarding our gender identity. I argue that both are wrong and in doing so present that to capture the experiences of those in the trans community requires moving away from a model of gender identity as a determinate thing towards the probabilistic cloud model of gender identity, where it is not some determinate thing, but rather a cloud of possible identities that they feel, to varying degrees, reflects them. The rest of the paper involves explaining the nuances of the model and examining its various upshots.

Introduction

In this paper, I argue that current accounts of what it is to have a gender identity and our relation to gender identity fail because they present gender as too much of a determinate *thing*. Opposingly, we should imagine something that is more of a "probabilistic cloud" but is not exactly determinate where gender identity exists, similar to the Quantum model of the electron. In service of this argument, I will present two accounts of gender identity and the relation we stand in to it, the sculptor account and the detective account, and argue that they both fail because they treat gender identity as too much of a determinate *thing*. It should be noted that these are not themselves accounts but rather are a cluster of multiple different accounts. Despite this, the core distinction between the accounts on the relation we stand in to our gender identity is the same, where the sculptor cluster of accounts posits gender identity as something we can have causal power over, and the detective cluster of accounts the opposite. I will then go on to explain why my probabilistic cloud model of gender identity is explanatorily virtuous. In §2 I will present the case of Alex and Abhinav and analyze their experiences with regard to the sculptor account and the detective account, and how they present issues with both accounts, as well as evaluate these accounts more broadly. In §3 I will argue that both of these views fail because they cannot account for the experiences of both Alex and Abhinav. I will also present that the sculptor and detective accounts' treatment of gender as a determinate thing is undesirable for independent reasons, and I will argue instead that we should think of gender identity as something that is indeterminate but does exist in a given space, a probabilistic cloud.

Sculptor Account

This account of the relation we stand to our gender identity posits that we have an active role in shaping our gender identity.[1] In this analogy, we are the sculptors of our gender identity, with gender experiences—that is, the experiences that we judge to have an important gender content to us and that make us feel a certain way that we feel to be connected to our gender identification, such as being referred to by correct pronouns, wearing clothing that aligns with one's gender identity, etc.—being the clay used in that sculpture.[2]

Take the case of Alex, who for the past 20 years has been confident in her gender identity as a trans woman. She has undergone social transition, dresses in a stereotypically feminine way, refers to herself by female pronouns, etc. Now imagine that one day she looks in the mirror and finds that she feels more androgynous. Say they now start experimenting with gender-neutral pronouns, gender-neutral presentation, etc. Coming to terms with this, Alex may rule that their gender identity is non-binary, but they were not mistaken about their gender identity in the past. Alex always approached their gender identity in good faith, attempting to build an understanding of their gender that matches their understanding of themselves. Prior to looking in the mirror that day, Alex had taken the sum total of their gender experiences and from that sculpted the sculpture that was her gender identity. After receiving this new experience, like fresh clay, she changed her sculpture. It is important to note that while Alex was passive in looking in the mirror and feeling themselves to be more androgynous than not, they were active in making their understanding of their gender identity aligning with their conception of themselves. This involves both changing their expression and outward behaviour to appear and present as non-binary, as well as recognizing the change of their internal mental state.

This kind of active relation that we stand in regarding our gender identity is reflected in the self-identification account of gender identity. This account presents that our gender identity is the product of the phenomenological synthesis of our gender subjectivities, the sum total of our gender experiences, where the phenomenological synthesis takes the gender subjectivities and synthesizes them into the whole that is our gender identity.[3]

The upshot of this account is that phenomenological synthesis allows those with the same gender subjectivities to still have different gender identities.[4] Take Alexis,

1 Florence Ashley, "What Is It Like to Have a Gender Identity?" *Mind* 132, no. 528 (2023): 1060.

2 Ashley, "What Is It Like to Have a Gender Identity?" 1057

3 Ashley, "What Is It Like to Have a Gender Identity?" 1054.

4 Ashley, "What Is It Like to Have a Gender Identity?" 1060.

who has essentially the same history of gender subjectivities as Alex and formulates her gender identity in a very similar way yet never goes on to change the way she views her gender identity. The value of this is that it provides an account of gender that is not purely deterministic from gender experiences and is consilient with the diversity of identity in the trans community. Put another way, gender identity is *under determined* purely by gender experiences, with phenomenological synthesis piecing together the gender identity into a coherent whole.

One of the most insightful applications of this account is for those who identify as gender fluid. Those of a gender fluid identity may not feel especially committed to any one gender identity, finding themselves oscillating between different identities, or just unable to confine themselves to one. Applying this to the analogy, it would be as if the sculptor regularly changes the sculpture while using the same clay. Whether the individual feels more masculine or feminine or non-binary is shaped by the individual, without necessarily having some new gender experience causing this change. Individuals may choose not to make a sculpture at all, which is a way that we can understand agender conceptions of gender identity.

The drawbacks of this active account of the relation that we stand in to our gender identities is that it runs into conflict with the view that those of a particular gender identity, even if they came to the realization that they were of that gender identity later, are and have always been that gender identity. Take Abhinav, who at 25 had the epiphany that, despite being assigned female at birth, he is in fact male. Abhinav identified as being female prior to this moment, not realizing that he was actually male and just conforming with the gender he was socialized as. Abhinav does not consider himself to have *become* male so much as realized that he is. Looking back on all his life experiences, thinking back to those moments where he was discouraged from playing with dolls, from wearing Kurta instead of Lehenga at weddings, cutting his hair short, etc. These experiences are evidence to him that his gender identity is male. Under the sculptor/self-identification account, Abhinav only *became* male at the point when he performed a phenomenological synthesis of his gender subjectivities at this epiphanic moment. Prior to this, he was still synthesizing his gender subjectivities in such a way that held he had a male gender identity.

The sculptor's account may attempt to get around this by arguing that he may have been performing a phenomenological synthesis of his gender subjectivities prior to his epiphany, but they were not veridical, they were wrong. However, by presenting that there could be falsity in the phenomenological syntheses of our gender subjectivities, the account loses what makes it valuable in the first place, which is that it makes it such that we shape our own gender identities, without regard to some standard in mind. The sculptor account may attempt to get around this by presenting that the sculpting was occurring subconsciously, however it does not seem intuitive to say that we were actively involved in shaping our identity if we never actively thought about doing that. It would be like saying the sculptor thoughtfully sculpted while sleepwalking. The key is not just causal agency but *conscious* causal agency.

If the self-identification account is to accept that we can be wrong about our gender identities, that implies that we can also be right about them. However, in so doing it presents that there is a fact of the matter as to what our gender identity is, independent

of what we create, when the gender identity should just be what is created.

The sculptor accounts struggle in that they cannot simultaneously capture the experiences of Alex and Abhinav, experiences that are common in the trans community. The account is better at explaining Alex's case than Abhinav's. The following account exhibits the opposite yet will still be flawed on account of its inability to capture the experiences of both individuals.

Detective Account

The detective account sees us standing in a passive position to our gender identity.[5] The detective view presents that we do have a gender identity "in us," and it is through an evaluation of our gender experiences that we come to judge which of those experiences most aligns with us. In the analogy, gender identities are lined up like suspects, and gender experiences serve as the evidence for the case, from which we determine what gender identity most closely aligns with those experiences.

This account of the relation we stand to our gender identity is particularly powerful at explaining the cases of individuals who later in life judge that their gender identity is something different, and they have in fact been that identity their whole lives. This is exemplified in the case of Abhinav, presented earlier.

The benefit of this account is that it offers a view that presents that our gender identities are a fundamental part of our identity. When we try to understand our gender identities, we are not constructing a gender identity so much as coming to better grips with it.

This account is in line with much of the way that cisgender individuals, as well as many transgender individuals, relate to their own gender identity. For cisgender individuals, many have a very passive understanding of their gender identity, not actively shaping it, rather it just being a part of who they are. It could be that for these individuals, their gender identity is something they do not give much thought to, given their socialization, but even under this situation, there is a passive position held to one's gender identity. Using the language of the detective account, it is the easiest case ever. All the evidence points to one and only one suspect such that it is frankly pointless considering any other. Among transgender individuals, many view the case as similarly easy, where it is just so incredibly obvious to them that they are the gender that they profess to be, and in fact have always been that gender identity.

As well, the trans community has a term to refer to those who are trans yet who are not currently aware of it: eggs. Their true gender identity is suppressed within the shell, sometimes forcefully, and "hatching" is realizing oneself to be trans. This term implies that while someone is not aware that they are trans, that does not change the fact that they are. It also presents that the true gender identity is actually within them, not something that is created.

This account also permits the possibility of being wrong about our gender identity. Under this account, for S to falsely judge that G is their gender identity is for S to judge a gender identity that is incongruent to their experiences to be theirs or perhaps judge that

5 Rach Cosker-Rowland, "Gender Identity: The Subjective Fit Account," *Philosophical Studies* 181 (2024): 2708.

which is not a gender experience to be such and influence their judgment of what is their gender identity. To say that there is incongruence in the attitudes towards the gendered experiences and the gender identity is to say that the way that S feels towards those gender experiences does not line up with what would be expected given their gender identity. This can be most clearly seen in the case of gender dysphoria, where there is a feeling of incongruence between how S feels about their gender experiences and the gender identity that S is socialized as or thinks themselves to be. Putting this in the language of the detective account, S could make the wrong judgement about the case from the evidence that they have or falsely take something to be evidence is (The possibility of the latter is something even Ashley admits).[6]

Being wrong about a self-assessment of one's gender identity amounts to them having gendered experiences, and upon having those experiences, coming to some judgement about their gender identity. Although, that judgement is one that is not actually congruent with their view of themselves, with this incongruence usually stemming from a sense of gender dysphoria. The case of Abhinav from earlier exemplifies being wrong about one's gender identity, where he falsely judges his gender identity to be female until he has this epiphany at 25.

One of the accounts of gender identity that broadly mirrors the detective account is the subjective fit account put forward by Cosker-Rowland. This account presents that for a subject S to have a gender identity G is for S to feel that G "fits" their gender experiences and view of themselves.[7] Under this account, our gender identity involves taking our gender experiences and evaluating whether they fit our own view of ourselves and our gender identity. This account aligns with the detective account because there is some gender identity that we judge to be the one that is most consilient with our gender experiences, like the detective judging the suspect whose testimony aligns most closely with the evidence of being the one who committed the crime.

The issue with this account is that it seems odd to say that our gender identity is just some inherent part of our nature that just exists, waiting for us to discover it. Gender theorists have historically taken gender to be something of a social construct, being formed out of the context of societies and societal structures.[8] Moreover, no two societies have had the exact same concept of gender. Moreover, several societies have had some concept of a third gender. With all of this social influence, it seems odd to say that there exists some gender identity that just is present "in us" as part of our identity. Even if we are not to take a social view of gender, this account is still subject to similar problems. Psychological accounts of gender identity may appear promising, but they fail in capturing the breadth of experiences that characterise what it is for one to have a gender identity. The brain structures of trans women are closer to those of the brain structures of cis women than they are cis men, but that refers to a neuroanatomical similarity between these groups (hardly deterministic about the complex views someone may have towards

6 Ashley, "What Is It Like to Have a Gender Identity?" 1065.

7 Cosker-Rowland, "Gender Identity: The Subjective Fit Account," 2701.

8 J.T. Ton, "Judith Butler's Notion of Gender Performativity: To What Extent Does Gender Performativity Exclude a Stable Gender Identity?" (Bachelor's thesis, Utrecht University, 2018), 8.

their gender identity) and how they may feel about individual gender experiences.[9]

It may be argued in defense of this account that our gender identity is partly shaped and determined by our societal influences, leading to our gender identities adapting to match that of the context they are in. In terms of the detective account, in different societies there may be a completely different set of suspects. However, in the case of Abhinav, his gender identity is somewhat of a foregone conclusion. Even if there were different suspects, that would not change the fact that the evidence most closely matches whichever suspect it does. The social influences in this instance affect *what* gender identities are deemed socially permissible, not what Abhinav's gender identity actually *is*.

What this analysis of the psychological and social views of what gender is meant to demonstrate is that regardless of what kind of stance we want to commit ourselves to on the nature of gender itself, prior to even gender identification, the detective account struggles to capture the full totality of what it means to have a gender identity per those accounts of what gender itself is.

Away from Gender as a Determinate Thing

Both the sculptor and detective accounts presented earlier demonstrate clear strengths and weaknesses. The sculptor account captures cases where gender identities can change and puts agency of our gender identity into our own hands, yet it struggles to explain cases of constant gender identity and recognizing one's gender identity to have been there for their whole life despite only realizing later in life, like the case of Abhinav. The detective account captures those who feel themselves to have always been the gender identity they are, as well as being consilient with a view that we can be wrong with evaluating our gender experiences as such. However, it struggles to explain those who feel that they change and affect their gender identity, like the case of Alex.

These accounts appear to be at odds, but they both share a fundamental attribute: they treat gender as a determinate *thing*. To take such a view is to say that a gender identity *X* is something discrete at some time that we can determine. Putting this into the language of the accounts, the sculptor account treats gender identity as a determinate thing, that being the sculpture, and the detective accounts also treats gender identity as a determinate thing, that being the culprit.

Gender being treated as a determinate thing is something that I take to be undesirable for independent reasons. By making gender identity something determinate, these theories fail to capture the complexity of those who feel like no terms could capture the whole of their gender identity. This is something that even self-identification accounts of gender identity acknowledge as a problem.[10] A determinate conception of gender identity ignores much of the nuance that they themselves associate with their gender identity and the concept of gender more broadly.

I propose that we change the way we think about gender from being a determinate thing, to a view of gender as a "probabilistic cloud" within which gender identities exist. This model avoids the independent concerns of determinate gender identity flagged earlier

9 *C.E Roselli, "Neurobiology of Gender Identity and Sexual Orientation," Journal of Neuroendocrinology 30, no. 7 (2018): 5.*

10 Ashley, "What Is It Like to Have a Gender Identity?" 1066.

(and by extension issues present in both the sculptor and detective accounts given they both take gender to be a determinate thing) and captures the experiences of both Alex and Abhinav.

This approach draws from physics, specifically the developments of quantum mechanics. Prior to the formulation of quantum mechanics, the most accurate physical view of the nature of the atom was the Bohr model. This model shows electrons in stable orbits around their nucleus. However, the contributions of Schrödinger demonstrated that electrons do not follow stable orbits around the nucleus but instead can be found in a probabilistic cloud around the nucleus. This means that the electron exists within that probabilistic cloud, as opposed to obeying one stable orbit around the nucleus, and its behaviour is more like a smear (for a visual representation of this refer to Appendix A).

Under this view, gender identity does exist, but is indeterminate at one exact instance, as it exists within the probabilistic cloud.

This probabilistic cloud has a size, shape, and location, and these are all subject to change.

The cloud size refers to the identities that an individual could conceivably accept. Someone who is gender fluid may have a very large probabilistic cloud, capturing male, female, non-binary, and other gender identities, whereas someone who is, say, a cisgender or binary trans person may very well have a very small probabilistic cloud, capturing only, say female gender identities.

A change in size can refer to an expansion or contraction and involves a change in the range of gender identities that fall within the probabilistic cloud. Say an individual judges that they are female but, as they get older, feel they are more androgynous than they originally thought. In line with this, she might start using she/they pronouns or even they/she pronouns, presenting more androgynously, etc. A contraction in the size of the probabilistic cloud refers to the exact opposite.

Location designates the gender identities that fall within the probabilistic cloud. For Abhinav, his probabilistic cloud is located in a region that is broadly labelled as male. A change in location denotes when the probabilistic cloud of possible gender identities translates from one "location" to another. This is exemplified in the case of Alex. The location of the probabilistic cloud of their gender identity was before around female gender identities but later moved to more androgynous identities. It also could have moved somewhere else.

This account leaves open the potential for being wrong about our gender identities. In the case of Abhinav, prior to his epiphany, he judged his gender identity to be something that it was not. Putting this in terms of the probabilistic cloud model, he believed his gender identity to be located somewhere it was not. He previously believed that it was around broadly female-associated gender identities, but through reflection came to realize that he was wrong in that assessment. Furthermore, he finds that his gender identity was always located around more male-associated gender identities than it was around female-associated gender identities.

This account allows us to keep the commonsense gender markers of male, female, non-binary, etc., that are classed as the "determinate things" in the sculptor and detective accounts. To do so, we label certain regions of where the probabilistic cloud can exist as being female, certain regions as male, etc. As such, Abhinav taking "male" to be his gender identity amounts to saying "the location of the probabilistic cloud of my gender identity is in the male region." The boundaries of this region can depend on the individual, societal forces, etc. These are of course broad characterizations, and there is debate on where those lines would be drawn, but suffice it to say that this way of understanding gender identity markers allows for common-sensical interpretations of what gender identity is and the broad range of what our gender identities can be.

There are also those whose gender identity cannot be neatly slotted into one gender identity marker or another. In the probabilistic cloud account, this would amount to the identity being somewhere between gender identities or landing somewhere outside what is broadly labelled. There is not some deep metaphysical significance about the labelling; it is rather a pragmatic approach to gender identity that allows for easy access to social facts about the individuals in question.

Another complexity that the account can capture is the difference between internal conceptions of gender identity and the external expression of one's gender identity. Ashley gives a case of a friend of hers who uses different pronouns depending on the situation she is in, using certain pronouns when in professional settings and certain other pronouns when outside of them.[11] A way to conceptualize this in line with the model is to understand this in terms of complexities emerging from different axes our gender identity can manifest on. Here the axes in question are internal conceptions of one's gender identity and the way gender identity manifests in terms of outer expression. The model gives us multiple potential instructive answers to this end. It could be that Sam just has two probability clouds, which are active or not depending on her situation. It could be that she just has a large gender identity cloud that stays mainly the same on the axis of internal self-perception but changes at external self-perception.

This is not the subject of this paper, but the question of whether or not someone is something because of their outward expression of such a trait is a fascinating one in the philosophy of identity, and the intuitions can seemingly be quite different depending on the identity marker. It may be intuitive to label someone whose acts outwardly affect right-wing social change as a conservative, even though they may internally identify as a liberal. On this account, what it means to have a political belief amounts to the consequences of a particular set of behaviours. This intuition may not be shared with gender, as accounts of gender mostly stress internal conceptions of gender identity as being of prime importance. Both the sculptor and detective accounts emphasize an internal process of understanding gender identity, with other expression being somewhat irrelevant to someone's gender identity. However, this is not a facet of all versions of gender identity. For example, the norm-relevancy account (which is arguably part of the detective account cluster of views) emphasizes the influence that our external behaviour has on our gender identity, with our gender identity essentially being the ways in which we engage with our external environment and the social norms that exist in that environment.[12]

Returning to the issue of gender identity markers, these markers are similar to the Bohr model, in that while they do not reflect the entirety of the experience of having

11 Ashley, "What Is It Like to Have a Gender Identity?" 1063.

12 Cosker-Rowland, "Gender Identity: The Subjective Fit Account," 2713.

a gender identity, they are good enough at conveying it. Similarly, the Bohr model is often good enough to do chemistry, even though the Quantum model is technically more accurate.

The problem with previous accounts of gender identity is their ability to get the totality of gender experiences to align with a particular gender identity marker. This is similar to using the Bohr model to capture the totality of the behaviour of the electron. Of course, gender markers are going to fail at this task. They are like the Bohr model: they are often good enough, but they are not able to capture the totality of the truth of gender identity, and that is fine, so long as the appropriate model is used in the situations that call for it. Both the sculptor and detective account fall victim to this, where they both take gender to be a determinate thing, profoundly simplifying the complexity that can come from gender identity. The probabilistic cloud model of gender identity presents a view of gender identity that starts from the gender experiences and understands gender identity markers as roughly attempting to categorize those experiences.

Conclusion

This paper is centrally concerned with whether we can have causal powers over our gender identity. It is theoretically possible but does not necessarily have to be the case. In the case of Alex, gender identity is something that can change. They can actively push that change in the direction that they feel best resembles them, by both changing their internal and external behaviour to better reflect their own self-conceptualization of their gender identity. In the case of Abhinav, his gender identity is just fixed, even though he misattributed it for much of his life. We can change the location and size of our gender identity cloud, but where our gender identity is within that cloud is indeterminate. Our gender identity clouds essentially determine what gender identities and attitudes toward gender identity we feel reflect our self-understanding, and how our gender identity actually manifests at a particular moment is explicable by an actualization of a probabilistic value within the gender identity cloud. It is not completely illusory, and it is more likely to be certain identities than others, but it is still indeterminate and is closer to a smear of different identities and responses to gender experiences than a single, determinate gender identity.

In this paper, I have argued for the probabilistic cloud model of gender identity by presenting the sculptor and detective accounts of gender identity and the relation we stand in to it, showing why both fail at capturing basic testimony of trans individuals, and presenting how my own account does. I discussed why a view of gender identity that presents it as something that is determinate is undesirable for independent reasons and how my indeterminate account of gender avoids these problems and captures experiences paradigmatic to both the sculptor and detective account. I also presented how my account allows us to keep our intuitive notions of gender identity. This model involves treating gender as an indeterminate cloud of gender identities, around which an individual's gender identity is located. I present that in developing this view of gender, we can invoke the concepts of "size" and "location" to describe the potential gender identities of individuals. "Size" refers to the range of possible gender identities that a particular person's gender identity can be. Larger gender identity clouds indicate something like gender fluidity, whereas smaller gender identity clouds usually would correspond with cisgender individuals' view of their gender identity, with their responses to gender experiences being largely similar. "Location" here refers to the actual identity taken on by the individual, where a particular location could correspond with a female gender identity, for example. This paper concluded by arguing a middle position between the sculptor and detective account, presenting that we can have causal powers over the location and size of our gender identity clouds, but since our gender identity is indeterminate, we cannot control exactly where our gender identity is going to be at any given moment. We can have control over the overall location of the cloud, that being the range of possible gender identities we feel to some extent reflect who we are and what matches our self-conception of ourselves, but we cannot control where in that cloud our gender identity will manifest itself.

Works Cited

Ashley, Florence. "What Is It Like to Have a Gender Identity?" *Mind* 132, no. 528 (2023): 1053–1073.

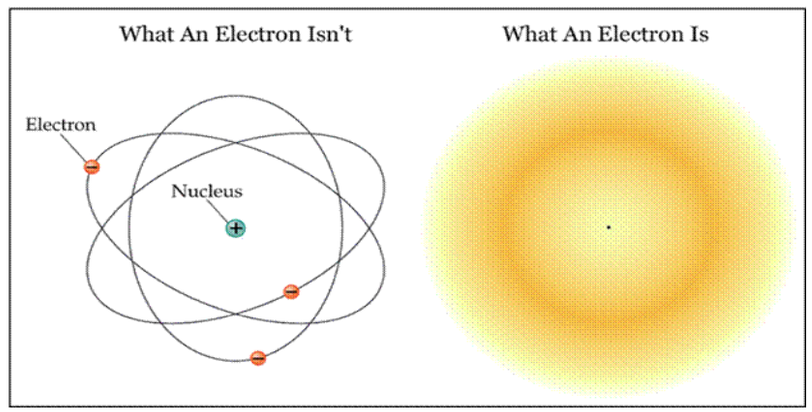
Cosker-Rowland, Rach. "Gender Identity: The Subjective Fit Account." *Philosophical Studies* 181 (2024): 2701–736.

Roselli, C.E. "Neurobiology of Gender Identity and Sexual Orientation." *Journal of Neuroendocrinology* 30, no. 7 (2018): https://doi.org/10.1111/jne.12562.

Ton, J.T. "Judith Butler's Notion of Gender Performativity: To What Extent Does Gender Performativity Exclude a Stable Gender Identity?" Bachelor's thesis, Utrecht University, 2018, https://studenttheses.uu.nl/bitstream/handle/20.500.12932/30880/JTTON_BScThesis_Judith_Butler_Gender_Performativity_FINAL.pdf?sequence=2&isAllowed=y.

Weinberg College, Northwestern University. "Plate 1: What An Electron Is." Accessed March 7, 2025. https://faculty.wcas.northwestern.edu/infocom/The%20Website/plates/Plate%201.html.

Appendix:



How Does Pain Fit into the Social Model of Disability?

Sophia Decherney

Introduction

The Social Model of Disability, though its conception varies across theorists, claims that people are disabled not only by impairment and difference, but also, by the barriers constructed by society. However, theorists such as Elizabeth Barnes and Susan Wendell believe pain poses a special problem for social theorists. Insofar as pain, even with the absence of social barriers, detracts from an individual's quality of life. If so, Social Model theorists are left to evaluate: is it possible to fit pain into the Social Model of Disability? And, if so, would it be a worthwhile pursuit?

In this paper, I first describe how the Social Model departs from the previously accepted medical model of disability. I then explore several critiques of the Social Model that show how it fails to incorporate the experience of chronic pain sufficiently. I present three strategies with which the Social Model responds to those critiques: breaking away from a literal interpretation of the model, distinguishing chronic pain from other forms of disability, and considering how the experience of pain might be alleviated through the removal of societal barriers. Lastly, I investigate the usefulness of the Social Model for conceptualizing pain by applying it to the controversial case of assisted suicide. I conclude that the Social Model of Disability is a useful tool for theorizing chronic pain because it offers important objections to the idea that assisted suicide offers autonomy and choice to a chronic pain experiencer. In the final analysis, I demonstrate that although the Social Model has theoretical difficulties in modeling pain, the case study of assisted suicide offers good reasons to work through those difficulties.

What is the Social Model of Disability?

The goal is to assess the efficacy of the Social Model of Disability for chronic pain experiencers through evaluating the ethics of assisted suicide. However, to do so, one must first understand what adherents to that model believe about health and disability. Developed in the 1970s, the Social Model of Disability emerged in opposition to the medical model, which views health and disability through the lens of the natural sciences. The medical model, which defines disability as "independent of our specific, contingent social categories and practices . . . social or personal values," is based on "statistical normalcy, adaptive fitness, and biological function."[1] By contrast, the Social Model defines disability through a barriers-based approach, which considers the environmental, social, cultural, and political factors that impede and, in doing so, disable non-normative bodies.[2] Unlike the medical model, which emphasizes individual impairment in defining disability, the Social Model is societal in scope. Its theorists contend that "disability definitions are not rationally determined but socially constructed" by the "social meanings individuals attach to particular physical and mental impairments."[3]

Many Social Model theorists take the stance that disability and impairment are constituted by societal barriers, and since "barriers can be removed [and] if you remove the barrier then you [can] remove the disability."[4] Whereas medical model theorists believe that "difficulties faced by disabled people are a direct result of their individual impairments or lack or loss of functioning," Social Model theorists contend that disability is created by a society built for non-disabled people. Accordingly, disability can be defined as a type of "social oppression."[5] Thus, proponents of the Social Model claim that a reconstruction of society to fit disabled and chronically ill individuals will lift this social oppression and, if not eliminating disability entirely, will at least eliminate the need for the category of disability.[6] However, critics of the Social Model object that it does not fit all disabilities, including, but not limited to, chronic pain.

Critique of the Social Model: No Room for Pain

One major critique of the Social Model of Disability is its failure to address the issue of pain. In a literal reading of the barriers-based approach, if the removal of barriers leads to the removal of disability, then removing barriers should also eliminate pain from someone living with chronic pain. However, it seems unlikely that the dismantling of societal barriers could totally eliminate the experience of pain for persons with chronic

---

1 Quill Kukla, "Medicalization, 'Normal Function,' and the Definition of Health," in *The Routledge Companion to Bioethics*, eds. John D. Arras, Elizabeth Fenton, and Quill Kukla, 1st ed. (Routledge, 2014), 515–16.

2 Tom Shakespeare, "The Social Model of Disability," in *The Disability Studies Reader*, ed. Lennard J. Davis, 2nd ed. (Routledge, 2006), 197.

3 Gary L. Albrecht and Judith A. Levy, "Constructing Disabilities as Social Problems," in *Cross National Rehabilitation Policies: A Sociological Perspective,* ed. Gary L. Albrecht (Sage Publications, 1981), 14. This paper evaluates the usefulness of the definition provided by the Social Model of disability and so mentions of "disability" going forward can be assumed to take on the Social Model definition.

4 "The Social Model of Disability," Welsh Government Services and Information, accessed April 27, 2024, https://www.gov.wales/sites/default/files/publications/2023-06/atisn17524doc7.pdf.

5 Sally French and John Swain, "Changing Relationships for Promoting Health," in *Tidy's Physiotherapy,* ed. Stuart Porter, 15th ed. (Churchill Livingston, 2013), 190.

6 Michael Oliver, *The Politics of Disablement: Critical Texts in Social Work and the Welfare State* (Palgrave, 1990), 78–81.

pain.[7] Cara Jones raises this critique of Social Model in her article "The Pain of Endo Existence: Toward a Feminist Disability Studies Reading of Endometriosis" (2016). She points out that while one symptom of endometriosis, infertility, is widely addressed by Social Model theorists, another symptom of endometriosis, chronic pain, is "largely under-theorized" precisely because it is harder to explain by a barriers approach.[8] Society might no longer consider infertility disabling if it ceased to value fertility as a norm of the female experience (after all, men who are unable to carry a pregnancy are not considered disabled); by comparison, a life devoid of pain seems less like a socially constructed value than it does a universal good.

Emily Rogers offers a similar critique in her article "Recursive Debility," which follows a patient activist group that fought to have their painful condition identified as a chronic illness (ME/CFS). Rogers argues that it is necessary to recognize bodily impairment, such as pain, alongside socio-political aspects of disability in order to understand the interdependent facets of the disabled/chronically ill experience which a purely Social Model of Disability might obscure.[9] Indeed, critics of the Social Model caution that when it comes to impairments such as pain, Social Model theorists "are in danger of assuming that impairment has no part at all in determining our experiences."[10] Even critics who acknowledge that social factors play a large part in the destabilization of people living with chronic pain argue that the Social Model offers too simplistic an explanation by presenting impairment as "irrelevant, neutral and, sometimes, positive, but never, ever as the quandary it really is."[11] In fact, by conceptualizing disability as a condition caused primarily by societal factors and by including only those disabilities clearly delimited by social barriers, the Social Model risks excluding painful conditions entirely.[12]

Theories of Pain Under the Social Model

Studies tackling the issue of how to incorporate chronic and acute pain into the Social Model framework can be categorized into three major approaches. All three address the problematic notion that removing social barriers should eradicate pain. The first approach argues that the removal of barriers will eradicate not disability in a literal sense

7 Liz Crow, "Including All of Our Lives: Renewing the Social Model of Disability," in *Equality, Participation, and Inclusion: Diverse Perspectives,* eds. Jonathan Rix, Melanie Nind, Kieron Sheehy, Katy Simmons, and Christopher Walsh, 2nd ed. (Routledge, 2010), 134–37.

8 Cara Jones, "The Pain of Endo Existence: Toward a Feminist Disability Studies Reading of Endometriosis," *Hypatia* 31, no. 3 (2016): 55, https://www.jstor.org/stable/44076492. Endometriosis is a uterine condition with symptoms including painful periods and infertility.

9 Emily Lim Rogers, "Recursive Debility: Symptoms, Patient Activism, and the Incomplete Medicalization of ME/CFS," *Medical Anthropology Quarterly* 36, no. 3 (2022): 412, https://doi.org/10.1111/maq.12701.

10 Crow, "Including All Our Lives," 126.

11 Crow, "Including All Our Lives," 126.

12 Crow, "Including All Our Lives," 136.

but rather the need to specify disability as a discrete category of the human condition. A second approach separates chronic pain from other disabilities by creating a special framework for impairments or unhealthy disabilities.[13] A third approach contends that the removal of social barriers will actually diminish and, in extreme cases, totally eradicate the experience of pain.

Approach One: Breaking Away from the Literal Sense

This first approach to the problem of pain in the Social Model of Disability asserts that, when it comes to pain, the removal of barriers should not be taken literally.[14] Indeed, as some Social Model theorists suggest, the removal of barriers can be expected to eradicate not the physical experience of pain but rather the disruption that chronic pain causes. It stands to reason, therefore, that society should dismantle the physical and ideological obstacles preventing chronic pain experiencers from participating fully in society. In doing so, it would not only reduce the limitations on those living with pain but also the need to categorize people who experience chronic pain.[15]

In his book *The Polio Paradox* (2002), for example, Dr. Richard Bruno suggests that it is not the experience of the sensation of pain that makes life difficult for those disabled by Post-Polio Syndrome (PPS) but rather a society that promotes norms incompatible with PPS existence.[16] Such social factors also contribute to the underdiagnosis and undertreatment of PPS, by perpetuating, for instance, the cultural prejudice that young people—the most common age group affected at the time of his book—do not experience chronic pain but merely seek drugs.[17] Moreover, according to Bruno, many other global (though by no means universal) social norms make chronic pain debilitating in our modern society: constantly working at the upper limits of one's ability instead of at a sustainable level, using every available assistive device at all times, internalizing the idea that pain is necessary for advancement (no pain, no gain); overdoing rehabilitation exercises; and pushing through uncomfortable temperatures.[18] If these practices were to be denormalized, he argues, PPS patients would continue to experience pain but would no longer be considered disabled.[19] In this view, remodeling society's values and practices neither cures pain nor does away with deviations from the "normal body"; instead, it accounts for these deviations, thereby abolishing the need for the category of disability. By suggesting problems with breaking away from the literal definition of disability, Bruno creates the preconditions for a unique consideration of pain within the social model.

Approach Two: Separating Out Chronic Pain from Other Disabilities

Susan Wendell, in her article "Unhealthy Disabled: Treating Chronic Illnesses as

13 Susan Wendell, "Unhealthy Disabled: Treating Chronic Illnesses as Disabilities," *Hypatia* 16, no. 4 (2001): 25–28, https://doi.org/10.1353/hyp.2001.0062.

14 Elizabeth Barnes, "The Hysteria Accusation," *Aeon*, June 20, 2020, https://aeon.co/essays/womens-pain-it-seems-is-hysterical-until-proven-otherwise.

15 Richard L. Bruno, *The Polio Paradox* (Grand Central Publishing, 2024), 5.

16 Bruno, *The Polio Paradox*, 3–5.

17 Bruno, *The Polio Paradox*, 32.

18 Bruno, *The Polio Paradox,* 36–40.

19 Bruno, *The Polio Paradox,* 51.

Disabilities" (2001) posits that there are actually two categories of disabled people who ought to be treated differently under the Social Model. The first is "healthy disabled" people who have impairments or normative deviations but are not actually "sick." People in this category might include individuals who are deaf, autistic, or wheelchair-bound. Some characteristics of this type are that their disability does not lower their life expectancy and that their impairments are fairly static rather than deteriorating or fluctuating (although someone's status as healthy disabled may change throughout their lifetime).[20] The Social Model deals well with situations such as these as removing the barrier removes the negative experiences of the disability, i.e., if everyone knew sign language, there would be few negative consequences of being deaf.

On the other hand, those with chronic pain fit into a group Wendell categorizes as "unhealthy disabled."[21] This category includes terminal conditions as well as any condition that would have a negative impact on the life of the individual even if society was built with them in mind. The Social Model is applied less effectively to this category because "impairment in itself can be a negative, painful experience."[22] Wendell cautions against applying the message "remove the barriers that have been erected arbitrarily against our participation, and we will perform as well as anyone else" to people in the unhealthy disabled category.[23] Instead, she claims that this aspect of the Social Model may only work for those in the category of healthy disability.[24] Thus, by separating the unhealthy and healthy disabled, she illustrates how a Social Model of Disability needs to be differently tailored to each of the two groups.[25]

Approach Three: Literally Reducing the Pain Experience

A third group of Social Model theorists believe that removing barriers does, in fact, reduce or remove the physical experience of pain. In the weaker version of this view, societal barriers do not cause the physical sensation of pain as much as they contribute to the experience of that sensation as debilitating. For example, women, people of color, and younger people are far less likely to receive adequate care for pain. As a result, pain plays a larger role in their lives, and because they are inadequately treated, they experience far more debilitation and impairment from that pain in the long term.[26] Moreover, many societal factors cause or exacerbate pain, such as stress, hygiene, and access to consistent and comfortable sleeping conditions. If these barriers were to be removed, experiencers of pain would suffer less or may not experience pain in the first place.

In this theory, deconstructing the social barriers will not obliterate the experience of pain, but it will remove those factors that make pain worse and prevent pain experiencers from receiving adequate care.[27] It might also, in particular cases, prevent painful conditions

20 Wendell, "Unhealthy Disabled," 19.
21 Wendell, "Unhealthy Disabled," 18.
22 Crow, "Including All Our Lives," 134.
23 Wendell, "Unhealthy Disabled," 27.
24 Wendell, "Unhealthy Disabled," 28.
25 Wendell, "Unhealthy Disabled, 30–32.
26 Barnes, "The Hysteria Accusation."
27 Barnes, "The Hysteria Accusation."

from developing in the first place, removing chronic pain entirely.[28] In a stronger iteration of this claim, some theorists go so far as to argue that removing social barriers would wholly abolish the sensory experience of pain. These theorists point out that although we currently have technology that can completely eliminate pain, prejudices regarding morally correct ways to experience pain prevent doctors from prioritizing pain mitigation in treatment plans.[29] Without promoting any one particular theory, I submit that the strong arguments outlined above indicate that there is potential for theorists to create a space for pain within the theoretical framework of the Social Model. However, the question remains whether there is a good reason to adopt the Social Model for pain.

Contributions of the Social Model to Pain: A Case Study of Assisted Suicide

One way to evaluate the usefulness of a Social Model of pain is by assessing its contributions to a major controversy within pain ethics: assisted suicide. Assisted suicide, physician-induced patient-consented euthanasia, has been on the rise across the world. Social Model theorists offer four main objections to assisted suicide for chronic pain experiencers: it represents an individualist vs. societal approach to the body, it misleadingly presents patients' motivations for suicide, gives a false sense of choice, and, lastly, relies on cultural narratives that offer merely the illusion of agency as well as harmful moral ideas of what constitutes a good death.

Focus on the Individual Body

The Social Model of Disability argues that viewing pain as an individual misfortune rather than a condition impacted by political and societal conditions causes people to view chronic pain as a tragedy. According to some Social Model theorists, assisted suicide reinforces this tragic view of pain by suggesting that the death of a human in pain is better than their continued living experience of pain. By contrast, they argue that when barriers such as social isolation, impossible work expectations, etc., are successfully removed, a life of pain no longer means a life of suffering.[30] And yet, the rhetoric around assisted suicide reinforces the view that people in chronic pain would prefer death (more than they would prefer to have fulfilling jobs, equality, and higher standards of life, etc.) by portraying a life of pain as a tragedy worse than death. Ultimately, the option of assisted suicide prevents society from taking steps to alleviate pain and improve the quality of life. Proponents of this idea posit that if pain were seen as a societal rather than individual problem, as the Social Model suggests, then a person in pain who requests death would signal the failure of society to support that person, not an individual tragedy.[31]

Questioning Motivations: Is Pain Really the Problem?

While physicians and ethicists cite the unbearable suffering of chronic pain as a

28 Barnes, "The Hysteria Accusation."
29 Felicia Nimue Ackerman, "'I Support the Right to Die. You Go First': Bias and Physician-Assisted Suicide," in *Palgrave Handbook of Philosophy and Public Policy*, ed. David Boonin (Palgrave Macmillan, 2018).
30 Understanding suffering as a lowered quality of life.
31 Martha Stoddard Holmes, "Pain," in *Keywords for Disability Studies*, eds. Rachel Adams, Benjamin Reiss, and David Serlin (New York University Press, 2015), 133–34.

primary justification for assisted suicide, the social modeling of assisted suicide reveals a more complex set of rationales for chronic pain experiencers who seek death. In fact, according to the data, pain is not a top consideration in requests for assisted suicide by people with chronic pain. Rather, they cite social isolation and economic dependency/poverty, both societal-level barriers.[32] Additionally, there is no objective experience of pain that physicians can reference when determining the quality of life of a chronic pain experiencer, so there is no objective way to link a pain level to decreased quality of life.[33]

The assumption that living with pain is inevitably much worse than living a "normal" life overestimates the difficulty faced by people with impairments, and incorrectly identifies pain as the main cause of their difficulties. Indeed, patient reports show that when political, economic, and societal barriers are removed, the self-reported quality of life of a chronic pain experiencer falls well within the normal range of people without chronic pain.[34] In summation, by legalizing euthanasia for those experiencing chronic pain on the basis of their quality of life, physicians and lawmakers miss the larger societal issues contributing to the poor quality of life of disabled individuals. A

False Choices and Illusions of Agency

Another way that the Social Model can be evaluated is in its ability to counter normative arguments in favor of assisted suicide for pain experiencers. A contested argument for assisted suicide as an option for chronic pain experiencers is that it affords autonomy through choice.[35] Social Model theorists counter this argument by pointing out ways in which societal factors make assisted suicide a non-autonomous choice. For example, the lack of high-quality end-of-life care for those without the means to afford expensive private hospital treatment creates an environment for pain experiences in which death is indeed preferable.[36] Similarly, there is persuasive evidence that the pervasive cultural narrative about a life of pain—namely, that it is unbearable—affects the brain's neuroplasticity. That is, when a physician tells a patient that their pain is unbearable, the physician limits the patient's natural ability to adjust to the pain. In short, merely the narrative of a "chronic pain sufferer" can coerce the sufferer of pain to make a decision in favor of assisted suicide.[37]

Similarly powerful cultural ideas, such as the idea that a "good death" is one

32 Carol J. Gill, "Depression in the Context of Disability and the 'Right to Die,'" *Theoretical Medicine and Bioethics* 25 (2004): 186, https://link.springer.com/article/10.1023/B:META.0000040058.24814.54.

33 Sara Goering, "'You Say You're Happy, but...': Contested Quality of Life Judgments in Bioethics and Disability Studies," *Journal of Bioethical Inquiry* 5 (2008): 132, https://link.springer.com/article/10.1007/s11673-007-9076-z.

34 Goering, "You Say You're Happy," 125–26.

35 Joel Anderson and Warren Lux, "Accurate Self-Assessment, Autonomous Ignorance, and the Appreciation of Disability," *Philosophy, Psychiatry, and Psychology* 11, no. 4 (2004): 309–12.

36 Harold Braswell, "Euthanasia," in *Keywords for Disability Studies*, eds. Rachel Adams, Benjamin Reiss, and David Serlin (New York University Press, 2015), 80.

37 ***Anderson*** and Lux, "Accurate Self-Assessment."

without pain, or the idea that death is better than dependence, can persuade experiencers of pain that assisted suicide is the morally desirable option, further limiting the so-called autonomy of assisted suicide.[38] Lastly, the Social Model contextualizes support for assisted suicide. If a non-disabled person were suicidal, they would typically be deemed irrational, and their suicidal ideation would be treated. By contrast, as Social Model theorists explain, people in chronic pain who request assisted suicide are deemed rational by physicians, politicians, and the general public.[39] This societal context exposes the flaws in the argument that assisted suicide confers autonomy on those with chronic pain since it is only preferable when outside parties agree that the pain experiencer "rationally" desires death. In this way, the Social Model's framework undermines the assumption that physicians and policymakers can justify assisted suicide for chronic pain experiencers by measuring "objective" levels of pain. Rather, quality of life is determined not simply by the experience of pain but also by the societal contexts in which pain is experienced. Thus, when the Social Model is adapted to account for chronic pain, it usefully problematizes the promotion of assisted suicide for pain experiencers.

Conclusion: We Can and Should Incorporate Pain into the Social Model of Disability

To be sure, the Social Model has clear problems with the way that it conceptualizes chronic pain, particularly its claim that removing barriers will remove disabilities. However, the Social Model of Disability and the lived experience of pain need not be at odds with each other if the Social Model is modified to incorporate pain better—whether by breaking away from a literal interpretation of the model, separating out chronic pain, and theorizing the model as reducing the pain experience. While I am not advocating one specific solution, I do maintain that there is good reason to accept that chronic pain *can* be incorporated into the model. Moreover, I have proven good motivation exists *to* incorporate the Social Model into narratives of pain. The Social Model has proven useful in weighing important issues within the study of chronic pain, adding key insights into problems of individualism, motivation, and agency to the conversation on assisted suicide. While the Social Model does not offer an answer to how or if to promote assisted suicide, these insights illustrate the usefulness of the Social Model for conceptualizing pain. Thus, theorists of assisted suicide, as well as pain theorists more generally, can and should incorporate pain into the Social Model of Disability.

Works Cited

Ackerman, Felicia Nimue. "'I Support the Right to Die. You Go First': Bias and Physician-Assisted Suicide." In *Palgrave Handbook of Philosophy and Public Policy,* edited by David Boonin. Palgrave Macmillan, 2018.

Albrecht, Gary L., and Judith A. Levy. "Constructing Disabilities as Social Problems." In *Cross National Rehabilitation Policies: A Sociological Perspective*, edited by Gary L. Albrecht. Sage Publications, 1981.

Anderson, Joel, and Warren Lux. "Accurate Self-Assessment, Autonomous Ignorance, and the Appreciation of Disability." *Philosophy, Psychiatry, and Psychology* 11,

38 Braswell, "Euthanasia," 79.

39 Braswell, "Euthanasia," 80.

no. 4 (2004): 309–12.
Barnes, Elizabeth. "The Hysteria Accusation." *Aeon,* June 20, 2020, https://aeon.co/essays/womens-pain-it-seems-is-hysterical-until-proven-otherwise.
Berger, Dan. "Recovering the History and Historiography of the Disability Rights Movement through Paul Longmore." *Journal of Civil and Human Rights* 6, no. 2 (2020): 78–80. https://doi.org/10.5406/jcivihumarigh.6.2.0078.
Braswell, Harold. "Euthanasia." In *Keywords for Disability Studies,* edited by Rachel Adams, Benjamin Reiss, and David Serlin. New York University Press, 2015.
Bruno, Richard L. *The Polio Paradox*. Grand Central Publishing, 2024.
Crow, Liz. "Including All of Our Lives: Renewing the Social Model of Disability." In *Equality, Participation, and Inclusion: Diverse Perspectives,* edited by Jonathan Rix, Melanie Nind, Kieron Sheehy, Katy Simmons, and Christopher Walsh. 2nd ed. Routledge, 2010.
French, Sally and John Swain. "Changing Relationships for Promoting Health." *In Tidy's Physiotherapy*, edited by Stuart Porter. 15th ed. Churchill Livingston, 2013.
Gill, Carol J. "Depression in the Context of Disability and the 'Right to Die.'" *Theoretical Medicine and Bioethics* 25 (2004): 171–98. https://link.springer.com/article/10.1023/B:META.0000040058.24814.54.
Goering, Sara. "'You Say You're Happy, but...': Contested Quality of Life Judgments in Bioethics and Disability Studies." *Journal of Bioethical Inquiry* 5 (2008): 125–35. https://link.springer.com/article/10.1007/s11673-007-9076-z.
Jones, Cara. "The Pain of Endo Existence: Toward a Feminist Disability Studies Reading of Endometriosis." *Hypatia* 31, no. 3 (2016): 554–71. https://www.jstor.org/stable/44076492.
Kukla, Quill. "Medicalization, 'Normal Function,' and the Definition of Health." In *The Routledge Companion to Bioethics,* edited by John D. Arras, Elizabeth Fenton, and Quill Kukla. 1st ed. Routledge, 2014.
Oliver, Michael. *The Politics of Disablement: Critical Texts in Social Work and the Welfare State.* Palgrave, 1990.
Rogers, Emily Lim. "Recursive Debility: Symptoms, Patient Activism, and the Incomplete Medicalization of ME/CFS." *Medical Anthropology Quarterly* 36, no. 3 (2022): 412–28. https://doi.org/10.1111/maq.12701.
Shakespeare, Tom. "The Social Model of Disability." In *The Disability Studies Reader*, edited by Lennard J. Davis. 2nd ed. Routledge, 2006.
Stoddard Holmes, Martha. "Pain." In *Keywords for Disability Studies,* edited by Rachel Adams, Benjamin Reiss, and David Serlin. New York University Press, 2015.
"The Social Model of Disability." Welsh Government Services and Information, accessed April 27, 2024. https://www.gov.wales/sites/default/files/publications/2023-06/atisn17524doc7.pdf.
Wendell, Susan. "Unhealthy Disabled: Treating Chronic Illnesses as Disabilities." *Hypatia* 16, no. 4 (2001): 17–33. https://doi.org/10.1353/hyp.2001.0062.
Zola, Irving Kenneth. "Self, Identity and the Naming Question: Reflections on the Language of Disability." *Social Science & Medicine* 36, no. 2 (1993): 167–73. https://doi.org/10.1016/0277-9536(93)90208-l.=

"Style" and "The Historical Sense:" Historiography and Self-knowledge in Nietzsche and T.S. Eliot

Brendan Murphy

Abstract: Nietzsche's view of the possibility of self-knowledge, a convoluted and much-debated topic in Nietzsche scholarship, in continental philosophy generally, and in contemporary analytic discussions of self-knowledge, finds its most clear presentation in his middle and mature works. Connecting his philosophy of history to his discussions of self-knowledge provides a compelling account of self-knowledge through the "doing" of history. This paper analyzes the philosophical models of history found in Nietzsche's works and in T.S. Eliot's "Tradition and the Individual Talent." I first argue that Nietzsche's historical model is primarily aesthetic, with an emphasis on the role of history in the development of a form of self-knowledge. Further, I argue that Eliot's conception of the poetic work and the poet is a concrete example of the development of historical self-knowledge. Finally, I suggest that Eliot's poet is an aesthetic embodiment of Nietzsche's often quoted "style" motif. Eliot's concept of the poet and the historiography he indirectly models through his theory of literary criticism provides an example of Nietzsche's person of "style," allowing philosophers to gain greater insight into Nietzsche's conception of the self.

Nietzsche's *On the Genealogy of Morality* (G.M.) is an intensely historical work. For Nietzsche, the problems of morality, Christianity, and asceticism can be traced historically, which allows the philosopher to identify origins, manifestations, and present purposes. A historical project such as Nietzsche's genealogy necessitates a historiography: a theoretical structure that allows the philosopher or the historian to account the development of ideas through past events. Historiography, as I refer to it here, roughly follows the standard definition found in philosophical literature. Tucker defines it as "[w]hat historians write, about past events, about history."[1] In this paper,

1 Aviezer Tucker, ed., *A Companion to the Philosophy of History and*

historiography refers to the body of speech (from Nietzsche and Eliot) about the past and speech about the past is typically structural. Thus, historiography will refer to the *way* or *method* that we talk about the past.

In G.M., history is primarily an accounting of the development of ideas. Nietzsche approaches the accounting of intellectual history by developing a creative historiography based upon the individual will, or more specifically, "the will to power."[2] His approach (strongly influenced by Nietzsche's reading of Schopenhauer) allows him to speak about the past as a development of the will to power through time.

T.S. Eliot, in his 1919 critical essay "Tradition and the Individual Talent," operates under a similar historiography, or way of speaking about the past, to Nietzsche's. His essay is a work of literary theory, focused on developing an aesthetic model for understanding the value and creation of poetry. While their arguments seem superficially distinct and, in some ways, even opposed, both Nietzsche and Eliot view history as primarily creative, with its source within the individual will. In fact, both utilize the term "historical sense" to describe this creative force (per Kaufmann's *Beyond Good and Evil* [B.G.E.] translation).[3]

Eliot's examination of the poet's place in historical contexts mirrors Nietzsche's concept of history as an aesthetic phenomenon produced through the doing of history. In his examination of the role of the poet within a historical context and discourse, Eliot argues that the poet exerts influence on past works. Similarly, Nietzsche sees history as a creation of the present, with the objectivity of past events always created and influenced by one's own experiences and perspectives (perspectival knowledge). Eliot's arguments about the creation of a work of art provide insight into Nietzsche's view of the constructed life, a life of "style," as he expresses a desirable life in *The Gay Science* (G.S.).[4] Nietzsche's life of "style" is a life of artistic creation, a life formed in the creation of history itself.

Ultimately, I will argue that because of the remarkable similarities between Eliot's aesthetic historiography of tradition and Nietzsche's genealogy, Eliot's case study of the poet provides us with a model for understanding Nietzsche's historical self-knowledge as demonstrated by the person of style. To do so, I will frame the two thinkers' conceptions of history, showing that, contrary to possible superficial impressions, Eliot's historiography mirrors Nietzsche's. Next, I will show that Nietzsche's conceptualization of history is aesthetic and ingrained with Nietzsche's view of a character of "style." I will further argue that Eliot's understanding of the work of art as influential on past works helps to concretely evaluate what a life of style may look like, furthering Nietzsche's project.

Historiographies:

Here, I will lay out the theories of history presented by Nietzsche and Eliot, showing

Historiography (John Wiley & Sons, Incorporated, 2009), xii.

2 Friedrich Nietzsche, *On the Genealogy of Morality,* trans. Maudemarie Clark and Alan J. Swensen (Hackett Publishing Company, Inc., 1998), 2:§12.

3 Friedrich Nietzsche, "Beyond Good and Evil," in *Basic Writings of Nietzsche,* trans. Walter Kaufmann (New York: Modern Library, 1992), §224; T.S. Eliot, "Tradition and the Individual Talent," *Perspecta* 19 (1982): 37, JSTOR.

4 Friedrich Nietzsche, *The Gay Science,* trans. Walter Kaufmann (Random House, 1974), §290.

their similarities and the possibility for a hybridization of the aesthetic subjects that, for both figures, are at once products and creators of their historical contexts. In 1916, Eliot published a book review in *The International Journal of Ethics* that reviewed a book on Nietzsche's philosophy. The review is not particularly complementary of Nietzsche, with Eliot's primary criticism resting in Nietzsche's lack of systematic treatment of traditional philosophical subjects. Eliot expresses his sentiment regarding Nietzsche's lack of a systematic philosophy: "Sometimes he holds that the mind alters things; sometimes that there is no other nature than that we know."[5] The most revealing part of the review, however, Eliot leaves us at the end, in which he laments the author's lack of discussion about Nietzsche's philosophy of art. Eliot writes, "we regret the omission of any account of Nietzsche's views on art, with the interesting pessimism with respect to the future of art evinced in Human, All-too-Human (sic)."[6] Eliot's observations are informative to our own treatment of Nietzsche: his philosophy of art is nuanced and deserving of critical engagement. Eliot's "Tradition and the Individual Talent," first published three years after his book review, provides excellent comparative material for such a study.

In the essay, Eliot makes an argument for a history of art in which the work of art is impersonal, a result of its context and influences. The author of a poem acts as a stimulant for creation to take place in response to material already present in the author's environment. Eliot compares an author to a piece of platinum that causes a chemical reaction to take place between two chemicals, "oxygen and sulphur dioxide."[7] The platinum does not change in any way during the reaction, nor does it end up in the product; rather, it only causes the reaction to take place.

Eliot's poet, like the platinum in his metaphor, only causes creation to take place as a result of preceding and surrounding influences. Creation, for Eliot, is a historical product as much as a contemporary one, with a work of art proceeding from its predecessors. He notes, "the poet cannot reach this impersonality without surrendering himself wholly to the work to be done. And he is not likely to know what is to be done unless he lives in what is not merely the present, but the present moment of the past."[8] The moment of creation, then, is "the present moment of the past," a living history in which what *was* actively shapes what *is* and what *will be*.

However, the past, or as Eliot frames it, tradition, is not an immediately accessible resource for the poet. He argues that one needs "the historical sense," "a sense of the timeless as well as of the temporal and of the timeless and of the temporal together. . . . And it is at the same time what makes a writer more acutely conscious of his place in time, of his own contemporaneity."[9] For a poet to write historically, she must be historically minded. Eliot describes what one could term a temporal geography, a space that an author must actively inhabit as a city finds itself on a map, in relation to all that surrounds it. The

5 T.S. Eliot, "Book Review of *The Philosophy of Nietzsche* by A. Wolf," *The International Journal of Ethics* 26, no. 3 (1916): 427.

6 Eliot, "Book Review of *The Philosophy of Nietzsche* by A. Wolf," 427.

7 Eliot, "Tradition and the Individual Talent," 39.

8 Eliot, "Tradition and the Individual Talent," 42.

9 Eliot, "Tradition and the Individual Talent," 37.

larger the map's scale, the greater one's historical sense will be.

The poet's awareness of her temporal position leads to what Eliot terms "an escape from personality," suggesting an apparent contradiction, but which is resolved when understood in the light of closer reading of Nietzsche's historical model.[10] The poet recognizes that she is writing impersonally, writing not as a lone individual, but instead as the product of tradition. In many respects, this seems to contradict many of Nietzsche's claims about the role of individuality in the development of strength. However, much of this apparent contradiction finds its source in an unnuanced interpretation of Nietzsche's own historiography, and further, in his framing of the development of an individual.

Nietzsche's conception about self-knowledge, is, like Eliot's, impersonal and situated rather in a broader context. In G.M. 1 Nietzsche discusses the possibility of self-knowledge. He says, "We remain of necessity strangers to ourselves, we do not understand ourselves, we *must* mistake ourselves, for us the maxim reads to all eternity: 'each is furthest from himself.'"[11] Self-knowledge, according to Nietzsche, is epistemically inaccessible. Internally contained self-knowledge is not a possibility. The genealogical project of G.M. relies on a non-internal conception of the self. Instead, Nietzsche argues that we must look to tradition, to the process of intellectual and moral development through time, that we can understand ourselves, and eventually, to create identity as an aesthetic phenomenon. He goes further in B.G.E., arguing that contemporaneous knowledge of greatness (of character and aesthetic achievement) is also inaccessible.[12]

How then, could history itself be possible for Nietzsche? If one cannot evaluate what has occurred in the past, then a historical sense seems to be outside of human experience. Nietzsche addresses this problem (in a similar move to Eliot's temporal geography) by describing the individual as a being in time. He says, "The greatest events and thoughts . . . are comprehended last: the generations that are contemporaneous with them do not *experience* such events—they live right past them."[13] Typically, people live *past* history; instead of identifying contemporaneous events as individual events, they experience its flow as life itself, the present moment. Comprehension occurs "last," as a product of comprehension, or as Nietzsche's German reads, *Begreifen*, which can also be translated as realization or recognition.[14] One recognizes the past, and in so doing, is enabled to make aesthetic judgements.

In *The Genealogy of Morality*, which is itself a historical project that relies on one's recognition and interpretation of the historical event, Nietzsche expands on his view of the role of history, arguing that it is dynamic and organic, constantly changing based upon new orderings or curations. In 2:12, we find a brief genealogy of punishment. In explaining his genealogical methodology for examining punishment, Nietzsche says, "the entire history of a 'thing,' an organ, a practice can be a continuous sign-chain of ever new interpretations and arrangements, whose causes need not be connected even among themselves."[15] History, then, is not merely the progression of a continuous chain of events leading to the present. Rather, it is something constructed through interpretation and re-interpretation, the product of contemporaneous realization and recognition. As we interpret the events of the past, they become something new, helping to frame current experience. Nietzsche's genealogy, then, is a reflection on and realization of how interpretation has created the historical record.

While I focus here on the role of history in Nietzsche's later works, many philosophers have spent more time discussing the role of history in his earlier works, especially *The Birth of Tragedy* and "On the Advantage and Disadvantage of History for Life." Robert Doran, writing about both works, notes, "there is nevertheless a common project (and one that Nietzsche would never completely abandon): that of an existentialist historiography under the aegis of art."[16] This "existentialist historiography" continues in his later works with his focus on the formation of the individual as a temporal, and more importantly, temporally aware being. The place of the individual is always historically situated, even without reflection and recognition of the unique events of the past.

While much has been written on Nietzsche's theory of history and its creation, few have examined it in conversation with Eliot. In "Modern Monuments: T.S. Eliot, Nietzsche, and the Problem of History," however, John Zilcosky reads Nietzsche's essay "On the Uses and Disadvantages of History for Life" in conversation with "Tradition and the Individual Talent," making a compelling argument that Eliot's concept of the preservation of tradition is compatible with and informative to Nietzsche's view of history. Zilcosky restricts his analysis to evaluate only literary creation, comparing the two author's aesthetic models in conversation with continental thinkers such as Walter Benjamin and Georges Bataille. His analysis, which, like Doran's, focuses on "History and the Disservice of Life," is the only focused academic treatment of Nietzsche's historiography in conversation with Eliot, making it a valuable resource for our analysis.

In his treatment, Zilcosky explains that Eliot conceptualizes history as accumulative, yet also tightly ordered and curated.[17] For example, Eliot suggests "that the whole of the literature of Europe from Homer and within it the whole of the literature of his own country has a simultaneous existence and composes a simultaneous order."[18] History, Eliot says, finds itself gathered and bound together in the individual poet, making the past at once contemporaneous and historical. Thus, Eliot's creation of poetry is as dynamic as it is historical, with the recombination and chemical restructuring of substances, or in other words, of historical events. It is not pure accumulation, as Eliot himself posits, but rather a dynamic system—much like Nietzsche's historical model.

Nietzsche, according to Zilcosky, similarly does not see history as purely

10 Eliot, "Tradition and the Individual Talent," 42.
11 Nietszche, *On the Genealogy of Morality,* preface, §1.
12 Nietzsche, "Beyond Good and Evil," §285.
13 Nietzsche, "Beyond Good and Evil," §285.
14 Friedrich Nietzsche, *Jenseits von Gut und Bose* (Project Gutenberg, 2005); *LEO German-English Dictionary,* s.v. "Begreifen," 2024.
15 Nietzsche, *On the Genealogy of Morality*, 2:§12.
16 Robert Doran, "Nietzsche: Utility, Aesthetics, History," *Comparative Literature Studies* 37, no. 3 (2000): 321, JSTOR.
17 John Zilcosky, "Modern Monuments: T.S. Eliot, Nietzsche, and the Problem of History," *Journal of Modern Literature* 29, no. 1 (2005): 29, JSTOR.
18 Eliot, "Tradition and the Individual Talent," 37.

accumulative, but instead as corrosive, leading to a consistent loss of the past in favor of the mental representation of events. In "History and the Disservice of Life," Nietzsche uses allegories of monuments, chains, and mountain peaks to describe this decay. Zilcosky notes, "History is always a history of misrepresentation and of mourning. The past is either forgotten, or, if it is remembered, it is 'damaged' in the process."[19] History, then, is something that changes. This echoes Eliot's view of the simultaneous existence of the past and the present. For example, one may have eaten an apple yesterday, but if it is not relevant to her experience the next day, she will likely not experience the event historically. However, if her current experience (say, eating another apple) causes her to reflect or reconsider the previous event, then that event and her conception of its causes become historical. She frames her experience in reference to the historical event, shifting her understanding of other, related events.

Despite the key difference between their accumulative and corrosive models, both thinkers frame history in a similar way, with each event or literary work influenced by the works that follow it. Hannah Sullivan summarizes this vision of history in Eliot's historiography: "Eliot saw history as a method of epistemic organization rather than a line connecting the past to the present."[20] Sullivan insightfully identifies the role of the poet and the historian as "epistemic organization," the end of which, according to Nietzsche, is comprehension (or *Begreifen*). "Epistemic organization," an act of historical creation or creativity, enables aesthetic judgement (for Nietzsche) and the poetic work (for Eliot). Thus, their understandings of history differ only in history's possibility of dynamism, or important change, not in its mode of creation or the aesthetic effects (ultimately, the development of a historical self-knowledge).

When evaluated together, Nietzsche's and Eliot's historiographies help to frame a conception of history as a phenomenon with its source in the individual acting as an intermediary between past events or works and the present. The historian (or the poet) synthesizes the past into the present. Nietzsche's "genealogy" and Eliot's "tradition" represent philosophical and literary approaches to historiography that perform the synthesizing historical work of the poet. This work is aesthetic, focused on reinterpretation of the past and its relevance to the individual. Through this work, the individual crafts herself as a product of what she "survey[s]" within herself.[21] The work of history is the work of crafting a self, of creativity in the midst of context. In Nietzsche's later works, I will argue, this crafting is the work of the person of "style."

"Style," Self-Knowledge, and the Poet:

I will now utilize the development of an aesthetic, historical self-knowledge in Nietzsche and Eliot's works to analyze Nietzsche's concept of "style." I will show that Eliot's figure of the poet effectively demonstrates the capacity of style, better than Nietzsche himself does in his mature works. Style is an important theme throughout all of Nietzsche's works. According to Mark Alfano's forthcoming digital humanities study of style in Nietzsche's works, Nietzsche refers to "style" over 100 times in various contexts.

19 Zilcosky, "Modern Monuments," 28–29.

20 Hannah Sullivan, *The Work of Revision* (Harvard University Press, 2013), 144.

21 Nietzsche, *The Gay Science,* §290.

An aesthetic characterization, Nietzsche applies the concept in various scenarios. Oftentimes, as in G.S. 99, Nietzsche criticizes or praises artists for their sense of style or their aesthetic contributions. In G.S. 290, titled "*What is* needful," Nietzsche argues that style is "practiced by those who survey all the strengths and weaknesses of their nature and then fit them into an artistic plan until every one of them appears as art and reason and even weaknesses dot the eye." The stylish person is an artist, one who is self-aware, at least to the extent that she crafts an "artistic plan" based upon her "strengths and weaknesses."

This presents an issue, however, because of Nietzsche's insistence that self-knowledge is not possible. Paul Katsafanas discusses Nietzsche's arguments regarding self-knowledge in comparison with Kant. He says, "If each action were motivated by one or several desires and affects, then it would be easy enough to identify them. However, Nietzsche believes that the etiology of our actions is far more complex than this."[22] Self-knowledge lies beyond the capacity of typical human epistemology in the same way that non-subject oriented history does —a hypothetical history, or "the view from nowhere," as Thomas Nagel puts it.[23] Nietzsche notes while discussing his genealogy of the "origin and purpose of punishment," "However well one has grasped the *utility* of some physiological organ, one has still not comprehended anything regarding its genesis."[24] How then, if one is unable to develop self-knowledge or historical knowledge through etiological analysis, can she develop the style that Nietzsche so highly praises?

History, it seems, is a pathway to self-knowledge for Nietzsche. In other words, greatness, taste, or aesthetic appreciation are products of our "historical sense."[25] History allows us to develop an image of where we came from, creating an identity we can rely upon that does not depend on a personality-based self-knowledge, but rather in the impersonality of history. This is not to say, however, that history is purely impersonal. Rather, the doing of history creates personality, or as Eliot may frame it, the poetic content of a poet's work.

Nietzsche's genealogical project traces the past to understand the present not in terms of direct causes-and-effects, but rather, in terms of interpretation and re-interpretation. As quoted earlier, "a practice can be a continuous sign-chain of ever new interpretations and arrangements."[26] For example, take a woman reflecting on a childhood experience being bitten by a dog multiple times, once at age twenty, once at age forty, and once at age sixty. At age twenty, she has a fear of large dogs, which she connects to her childhood experience. At age forty, she has overcome her fear of dogs and looks back on the experience with humor, reflecting on her family's Cocker-Spaniel and the initial trepidation each of her children experienced learning to live with the dog. At age sixty,

22 Paul Katsafanas, "4. Kant and Nietzsche on Self-Knowledge," in *Nietzsche and the Problem of Subjectivity,* ed. Joao Constancio (De Gruyter, 2015), 117.

23 Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1989), 1; see Nagel 4 for his description of Nietzsche's critique.

24 Nietzsche, *On the Genealogy of Morality*, 2:§12.

25 Nietzsche, "Beyond Good and Evil," §224.

26 Nietzsche, On the Genealogy of Morality, 2:§12.

she remembers the death of her dog with melancholy, remembering the previous periods of reflection.

Although the emotions of each reflective experience can be related to the previous experiences, the woman finds that she understands her previous experiences differently upon reflection. Each experience makes room for the next, changing the structure of the historical narrative. As we see, she gains a form of self-knowledge with each reflection, but that self-knowledge is reliant upon the structure of past events, not only her current experiences. Each event allows her to re-situate herself within a historical record, essentially creating more history. In Eliot's poetic model, each poem builds upon the next and re-shapes its past, forcing every past poem in the tradition to re-situate in response. The work of history, then, is as poetic as it is sedimentary, as curatorial as it is archival.

Nietzsche does not explore the theoretical intricacies of self-knowledge; instead, he provides a case study: *On the Genealogy of Morality*. G.M.'s project is one of gaining self-knowledge through the development of the "historical sense." Throughout, Nietzsche identifies the interpretations and re-interpretations that have led to our current understanding of morality and "the ascetic ideal."[27] For example, he reviews a hypothesized broad re-interpretation of the purpose of suffering at the end of the study: "The meaninglessness of suffering, not the suffering itself, was the curse that thus far lay stretched out over humanity—*and the ascetic ideal offered it a meaning*!"[28] Nietzsche's genealogy seeks to trace the interpretation of meanings, acting out his historiography.

Nietzsche does not clearly provide examples for the role of style and its incumbent self-knowledge within the individual or the artistic tradition, but Eliot's poetic model provides just such an example. While he often references individuals and art (for example, Richard Wagner and his music in G.S. 99), we do not receive the same kind of sustained treatment that he offers through intellectual history in G.M. Because of his lack of systematic treatment, we are left to pick out bits and pieces scattered throughout the early, mature, and late works.

Eliot's poet, as described in his chemistry metaphor, is a catalyst for artistic creation because of her awareness of tradition. One's awareness of tradition relies on her "historical sense," her ability to draw upon the works that float in the contextual waters. As Eliot argues, "No poet, no artist of any art, has his complete meaning alone. His significance, his appreciation is the appreciation of his relation to the dead poets and artists."[29] The poet is a figure who is located as much in the past as in the present, finding her identity caught up in reflection of other poets in history. In other words, she is working through genealogy.

At the beginning of his essay, Eliot references another discipline focused on the study of the past: archaeology. He says, "You can hardly make the word [tradition] agreeable to English ears without this comfortable reference to the reassuring science of archaeology."[30] Digging up the past is the key to the poet's work. Without this digging, the

27 Nietzsche, *On the Genealogy of Morality*, 3:§28.
28 Nietzsche, *On the Genealogy of Morality*, 3:§28.
29 Eliot, "Tradition and the Individual Talent," 37.
30 Eliot, "Tradition and the Individual Talent," 36.

poet is only "personal," a poet without style.[31] Personality, the poet alone in the world, is of no interest to Eliot nor Nietzsche. The poet is one who displays strength, a product of living with style.[32] Eliot writes, "There is a great deal, in the writing of poetry, which must be conscious and deliberate. In fact, the bad poet is usually unconscious where he ought to be conscious, and conscious where he ought to be unconscious."[33] Here, Eliot argues for a similar limitation to self-knowledge as Nietzsche, suggesting that the best poets maintain an awareness of what they ought to know (tradition) and a lack of awareness about details unimportant to their historical narrative (personality).

Style, when applied to Eliot's framework, can be aesthetically understood as the proper application of tradition to poetry. It becomes a historical phenomenon, always contextualized within a tradition. Just as Nietzsche explores the genealogy of morality, Eliot argues for the aesthetic genealogy of the poet. The poet, then, can be compared with the intellectual historian, both operating under frameworks of tradition that find their source in the seeking for self-knowledge, which Eliot may characterize as an archaeological search. The creation of the self-knowing individual begins in historical creativity.

Conclusion:

Nietzsche's person of "style," not clearly defined in his own works, becomes more clearly understood when viewed aesthetically through Eliot's notion of the poet. The poet—also the historian, chemical catalyst, and archaeologist—is one who allows the past to live contemporaneously with the present. Nietzsche's philosopher seems to take a comparable, artistic role. All that has happened before is material for new creation, for new moral "horizon[s]."[34] Although Nietzsche appears to seek the destruction of the past through the death of God, his project consistently shows itself to be one of reevaluation and new recognition (*Begreifen*), allowing for the philosopher to engage with the world more honestly as a self-aware individual. This self-awareness does not arise purely from self-reflection; rather, it is a product of "the historical sense."[35]

Nietzsche's historiography, which involves temporal exertion between the past and the present, is one which is necessarily aesthetic and, to borrow Eliot's imagery, poetic. By analyzing Nietzsche's notion of history as the creation of self-knowledge, we find that "style" is the primary hallmark of the individual. This style is aesthetic and is meant to be historically evaluated, just as Nietzsche demonstrates in his numerous aphorisms discussing art, politics, and law (all of which he frames within historical contexts). History, for Nietzsche, results in the creation of the individual, and style is the proper evaluation and putting-on of history.

Works Cited

"Begreifen." *LEO German-English Dictionary*, 2024.

31 Eliot, "Tradition and the Individual Talent," 42.
32 Nietzsche, *The Gay Science,* §290.
33 Eliot, "Tradition and the Individual Talent," 42.
34 Nietzsche, *The Gay Science,* §124.
35 Nietzsche, "Beyond Good and Evil," §224; Eliot, "Tradition and the Individual Talent," 37.

Doran, Robert. "Nietzsche: Utility, Aesthetics, History." *Comparative Literature Studies* 37, no. 3 (2000): 321–43. JSTOR.
Eliot, T.S. "Book Review of *The Philosophy of Nietzsche* by A. Wolf." *The International Journal of Ethics* 26, no. 3 (1916): 426–27.
Eliot, T. S. "Tradition and the Individual Talent." *Perspecta* 19 (1982): 36–42. JSTOR.
Katsafanas, Paul. "4. Kant and Nietzsche on Self-Knowledge." In *Nietzsche and the Problem of Subjectivity,* edited by João Constâncio. De Gruyter, 2015.
Nagel, Thomas. *The View from Nowhere*. Oxford University Press, 1989.
Nietzsche, Friedrich. "Beyond Good and Evil." In *Basic Writings of Nietzsche*, translated by Walter Kaufmann. The Modern Library, 1992.
Nietzsche, Friedrich. *Jenseits von Gut und Böse*. Project Gutenberg, 2005.
Nietzsche, Friedrich. *On the Genealogy of Morality*. Translated by Maudemarie Clark and Alan J. Swensen. Hackett Publishing Company, Inc., 1998.
Nietzsche, Friedrich. "On the Uses and Disadvantages of History for Life." In *Untimely Meditations*, translated by R.J. Hollingdale. Cambridge University Press, 1983.
Nietzsche, Friedrich. *The Birth of Tragedy*. Translated by Duncan Large. Oxford University Press, 2000.
Nietzsche, Friedrich. *The Gay Science*. Translated by Walter Kaufmann. Random House, Inc., 1974.
Sullivan, Hannah. *The Work of Revision*. Harvard University Press, 2013.
Tucker, Aviezer, ed. *A Companion to the Philosophy of History and Historiography*. John Wiley & Sons, Incorporated, 2009.
Zilcosky, John. "Modern Monuments: T. S. Eliot, Nietzsche, and the Problem of History." *Journal of Modern Literature*, no. 1 (2005): 21–33. JSTOR.

A Refugee's Right to Citizenship

Carol Michelle Beyda

Abstract: This paper discusses the extent of duties that are morally owed to refugees in ethical immigration practices. I begin by explaining the humanitarian account often used to ground obligations to refugees and present this approach's limitations. After rejecting this inadequate method, I draw from Rebecca Buxton's writing on the connection between membership and the claimability of basic rights as well as Felix Bender's work on political standing and public autonomy. The original breakdown of membership between an individual and their state, which gives rise to refugeehood, must be remedied through citizenship. I morally justify the duty to provide refugees with citizenship by discussing Kantian ethics and the nature of basic rights. My argument demonstrates that the moral obligations owed to refugees must be determined through a political approach, addressing their broken-down membership in their country of origin, and the loss of basic rights that come with it.

Introduction & Thesis

Discourse on ethical responses to migrants naturally considers the need to help refugees, but what is the extent of the moral duties owed to refugees? Surely, their basic rights must be securely restored, but what is the best way to achieve this goal? To answer this question, I will analyze the core factors of refugeehood to determine what is morally owed to refugees in order to securely restore their basic rights.

In this paper, I argue that refugees must be provided with citizenship in order to remedy their loss of membership and basic rights. Membership refers to the political standing necessary in a given state for one to securely exercise their basic rights to live free of threat, fear, and violence. I will begin by laying out the condition of refugeehood in terms of membership and access to basic rights. I will then analyze a common approach philosophers utilize to ground obligations to refugees:

the humanitarian approach. This path establishes the need to tend to refugees simply because they are in need of assistance. I will be criticizing this approach by describing its limitations when assessing obligations to refugees, as it does not recognize the extent of the harm and active violation which they are faced with. Because of these limitations, the humanitarian approach does not motivate obligations or actions which can adequately rectify the core harms of refugeehood. As an alternative to the humanitarian approach, I will discuss the perspectives of Rebecca Buxton and Felix Bender, drawing connections between membership, political standing, and claimability of basic rights. Using these frameworks, I will demonstrate how the core vulnerability of refugees lies in their lack of membership in a political community, which is necessary for them to securely regain their basic rights. I will then discuss the differences between various forms of membership, including permanent residency and temporary protected status, and ultimately provide reasons for citizenship being the most secure and effective way to remedy a refugee's lack of membership and claims for basic rights.

The United Nations 1951 Refugee Convention officially defined refugees to the international community as "owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country."[1] This definition has been vastly criticized for its limiting and narrow makeup, but it crucially focuses on the breakdown of the relationship between the refugee and the government of their state.[2] While I will not adopt the entirety of this definition, my argument will rely specifically on the core aspects of refugeehood which the definition introduces—the breakdown in the political relationship between an individual and their state. This paper will understand refugees to be those who experience this relational breakdown and can no longer rely on their state to protect their basic rights. In understanding this central viewpoint, a political approach, which focuses on the relationship between an individual and their state, will be key to determining obligations to refugees.

There are around 25 million refugees today, who on average are likely to maintain refugee status for seventeen years.[3] In other words, these individuals live in a state of temporary permanence, unable to integrate, work, or settle permanently in another country.[4] They are not able to live safely in their home country, and they have not yet received adequate protection from another country.[5] Often times, this means they are hosted in urban refugee camps with desperate conditions, enduring torture, forced labor, and lack of basic conditions of human dignity.[6]

Most of the world's refugees are displaced from Syria, Ukraine, Afghanistan, Venezuela, South Sudan, or Myanmar, and have experienced active government repression, military invasion, arbitrary detention or violence, sexual or gender-based violence, etc.[7] Further, less than 3% of these refugees are able to return home and 78% have been in exile from their home country for over 5 years.[8] These individuals have endured active violations to their basic rights, (to live free of threat, fear, and violence,) and have not yet been able to regain those basic rights to live a safe and secure lifestyle. There are varying approaches to defining refugeehood and determining who should be able to claim refugee status, as well as extensive discourse on the moral duties owed to refugees. I will argue that a political approach rather than a humanitarian approach is most effective in encompassing the vulnerabilities of refugees and for grounding the duties which they are morally owed. Instead of analyzing the general harms placed on refugees in the humanitarian approach, a political approach emphasizes the breakdown between an individual and their state in order to determine what they are owed. Whereas the humanitarian approach may require states to provide a refugee with basic rights, a political approach can go a step further and claim that refugees are owed citizenship in another state, as it recognizes a vital aspect of refugeehood which the humanitarian approach does not. Further, I will argue that citizenship is necessary to remedy one's lost membership in a state and ensure secure access to basic rights on account of their political standing, rather than the discretion of the state.

Engaging with Philosophical Literature

The Humanitarian Approach

The humanitarian approach is principally grounded by the belief that we are obligated to help those in need.[9] To apply this principle toward refugees is to see them as subjects of humanitarian concern, individuals whose states are no longer able to protect them and therefore must turn to the international community for some form of protection. In this approach, states are obligated to provide rescue of refugees, meaning they must save them from harm. In Betts and Collier's *Refuge: Rethinking Refugee Policy in a Changing World*, they describe how the humanitarian approach to refugees "ground[s] the ethical case for helping refugees in the duty of rescue because it is a moral norm that is widely accepted and it is all that is needed."[10] In other words, obligations towards others are grounded in the fact that persons are in need of assistance and there are others who can provide it to them (these others have a duty to rescue the persons in harm). In this account,

1 David Miller, *Strangers in Our Midst: The Political Philosophy of Immigration* (Harvard University Press, 2016), 78, http://www.jstor.org/stable/j.ctvjf9z4w.
2 Miller, *Strangers in Our Midst,* 78.
3 Serena Parekh, *No Refuge: Ethics and the Global Refugee Crisis* (Oxford University Press, 2020), 5.
4 Parekh, *No Refuge, 5.*
5 Bradley Hillier-Smith, *The Ethics of State Responses to Refugees* (Routledge, 2024), 135.
6 Parekh, *No Refuge, 5.*
7 Hillier-Smith, *The Ethics of State Responses to Refugees*, 129.
 Hillier-Smith, *The Ethics of State Responses to Refugees*, 133.
9 Rebecca Buxton, "The Duty to Naturalise Refugees." *Critical Review of International Social and Political Philosophy* 26, no. 7 (2023): 1122, https://doi.org/10.1080/13698230.2021.1880198.
10 Paul Collier and Alexander Betts, *Refuge: Rethinking Refugee Policy in a Changing World* (Oxford University Press, 2017), 99.

moral duties are based in sympathy, compassion, and generosity—those in need must be helped because of our shared common humanity.[11]

The humanitarian approach is intuitively shown with Peter Singer's famous moral thought experiment, in which a child is drowning in a shallow pond and a stranger who walks by must be morally obligated to jump in and save the child, even if their clothes could get wet. His thought experiment exemplifies the essence of the humanitarian approach, the stranger has a duty to rescue the child from harm simply because they are in a position which allows them to do so.

Singer's thought experiment draws a parallel between the moral obligation that a hypothetical stranger has to save a drowning child to broader moral obligations persons or states have to save others from harm if they are in a position to do so, regardless of proximity. He asserts that "if it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it."[12] Betts and Collier further expand on Singer's theory, stating that our duty of rescue to refugees is to restore their circumstances as near to normalcy as it is practically possible for us to do. Because of an "appeal to our common humanity" and a "raw compassion for human condition," the aim must be to restore refugees' basic rights to live free of fear and violence.[13] They note the special obligations of states to their particular political communities yet contend that general obligations to those outside these political communities are needed as well. They maintain that the boundaries of the moral community go beyond those of political communities.[14]

Though the humanitarian approach is popular among philosophers such as Joseph Carens and Bradley Hillier-Smith, who argue that duties are owed to refugees on account of humanitarianism, I will recognize the approach only as a starting point in grounding duties to refugees. While acknowledging the compassionate and altruistic principles put forth in the humanitarian approach, I identify the ways in which it is insufficient in rectifying refugees' complex and vulnerable conditions. The humanitarian approach may require states to provide refugees with basic rights and remedy their general harm, but it does not address the political breakdown of membership which refugees experience, and which can only be remedied with new citizenship. Unlike the humanitarian approach, my argument will highlight the political breakdown between an individual and their state, which will then elucidate the proper duties owed to refugees to remedy this breakdown.

Further Understanding Refugeehood

Basic Rights

Before we can entirely understand the limitations of the humanitarian approach in addressing moral duties towards refugees, we must begin by understanding what exactly basic rights, refugeehood, and membership entail. In Shue's account of basic rights, he breaks down the definition and what these rights require. Fundamentally, for an individual to have a basic moral right, they must be "in a position to make demands of others."[15] Further, that right must be able to be demanded without feelings of embarrassment or shame, as there are good reasons for the demands to be granted. For an individual to actually enjoy and exercise their basic rights, they must have secure access to those rights. This secure access requires the guarantee of social institutions to ensure that the right holder can actually exercise their basic rights and enjoy their substance. Shue gives the following example to demonstrate this claim: "[i]f [individuals] do not have guarantees that they can assemble in security, they have not been provided with assembly as a right. They must assemble and merely hope for the best, because a standard threat to assembling security has not been dealt with."[16] Without security from social institutions, especially in cases where individuals can not enjoy the substance of the right on their own, persons are "readily open to coercion and intimidation through threats of [the] deprivation."[17] In this way, refugees do not have basic rights that they can securely access and exercise.

Membership and Claimability of Basic Rights

Refugees face severe harm from subjection to human rights violations throughout their displacement. The term "human rights" originally refers to the unconditional and intrinsic moral worth of all human beings, simply in virtue of being human.[18] These rights must be respected regardless of political, legal, or societal implications or recognition.[19] Nonetheless, human rights are commonly defined by their political function in global politics today, as states become the primary suppliers of their citizens' human rights. Though human rights originally refer to universal truths, it is more reasonable to understand human rights in their political role, as that is the way they function today. As primary duty-bearers, states have the duty to protect their citizens' human rights as basic rights within their society. When a state initially violates their citizens' rights, they are abandoning their function as primary duty-bearers for "protecting and upholding the human rights of their respective inhabitants." [20] As a result, individuals become refugees and remain in these conditions, unable to receive protection and access to their rights, until they are granted citizenship elsewhere.

In further analyzing the extent and nature of refugees' conditions, we can connect the core harms and vulnerabilities of refugeehood directly to their loss of membership and resultant political standing in their country of origin. An individual loses their membership in a state when they no longer have the ability to securely access their basic rights on account of their equal political standing in that state. According to Rebecca Buxton, the basic condition of refugeehood is the break-down of an individual's membership and

11 Collier and Betts, *Refuge,* 100.

12 Peter Singer, "Famine, Affluence, and Morality," *Philosophy & Public Affairs* 1, no. 3 (1972): 231, http://www.jstor.org/stable/2265052.

13 Collier and Betts, *Refuge,*100.

14 Collier and Betts, *Refuge,*100.

15 Henry Shue. *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy: 40th Anniversary Edition* (Princeton University Press, 2020), http://doi.org/10.2307/j.ctvqsdnkw.

16 Shue, *Basic Rights,* 27.

17 Shue, *Basic Rights,* 26.

18 Hillier-Smith, *The Ethics of State Responses to Refugees*, 138.

19 Hillier-Smith, *The Ethics of State Responses to Refugees*, 139.

20 Hillier-Smith, *The Ethics of State Responses to Refugees*, 204.

relationship with their state.[21] Refugees do not gain their status from a singular violation of their rights, but from ongoing instability in the relationship with their state, and in turn their inability to securely access their basic rights. When a state does not recognize and uphold an individual's basic rights, that individual is no longer able to maintain an ongoing and stable relationship with their state. In this way, even if an individual remains a citizen of that state, their membership becomes compromised when their basic rights are not recognized and upheld. Moreover, membership requires a state to recognize an individual's entitlements to rights, it is purely focused on the relationship between state and individual as opposed to feelings of personal affiliation.[22] In this way, membership does not only depend on the possession of formal rights, but instead on the "ability to claim certain rights as a matter of political standing."[23]

Buxton argues that because of the way in which basic rights are distributed by states to their members, individuals must be members of a community to claim their basic rights on a matter of standing as opposed to charity. Without a secure standing within a state in which an individual and their rights are securely recognized, refugees become reliant on "the good will of states to uphold [their basic rights]."[24] Even if states do provide refugees with basic rights, they remain at the discretion of the state, with the possibility of them being taken away at any given moment. In these cases, as well as cases where rights are not provided at all, refugees remain in a position of structural vulnerability; they lack functioning membership within a state where they can claim their basic rights solely on the basis of their standing. In other words, they cannot ensure secure access to their rights, as is the case for recognized citizens. As I will establish later in the paper, this secure standing in which individuals can exercise their rights as a matter of standing can only be achieved through formal citizenship.

For refugees to have their basic rights, they must be in positions where they are able to securely claim those rights: equal, formal, and legal membership, or citizenship. An individual without formal citizenship may be given basic rights and opportunities, but they do not have the security needed to exercise those rights based solely on their standing. Without proper standing and security in society, refugees remain in positions where their ability to exercise their basic rights is ultimately up to the discretion and charity of the state. Namely, basic rights are not substantive if they are not able to be securely accessed by the right-holders. Thus, refugees must be granted citizenship in order to securely claim their basic rights from a state and to have them be both respected and protected. The formal status of citizenship allows for refugees to be recognized as equal members and therefore make relevant claims to the state to protect their rights.

Political Standing and Public Autonomy

In his writing on refugeehood, Felix Bender emphasizes the role political oppression plays in categorizing the condition of refugees. Similarly to Buxton, he argues that the main harm refugees face is lack of public autonomy and legal-political standing once they are faced

21 Buxton, "The Duty to Naturalise Refugees," 1121.
22 Buxton, "The Duty to Naturalise Refugees," 1128
23 Buxton, "The Duty to Naturalise Refugees," 1128.
24 Buxton, "The Duty to Naturalise Refugees," 1125.

with active threats or violations.[25] Public autonomy can be understood as the counterpart of private autonomy; in other words, both public and private autonomy are what make up an individual's personal autonomy. The two coexist and depend on each other to allow for the freedoms of a moral agent. As private autonomy refers to an individual's ability to choose what to make for dinner alone in one's home, public autonomy refers to an individual's ability to generate claims and defend themselves within the public sphere of their state. This public autonomy is expressed by an individual's legal-political standing, or equivalently, status, which "formally recognizes a person as morally equal by imbuing her with fundamental legal equality and political rights."[26] As the primary duty-bearers of people's rights, states must provide individuals with the equal standing that gives them the ability to access those legal and political rights. For a state to properly fulfill their duty as right-bearers to their people, they must issue legal and political rights on a matter of equal standing rather than by their own discretion, in order for the rights to be securely accessed. Similarly to Buxton's claimability argument, Bender asserts the necessity of standing for effective maintenance of basic rights. In lacking public autonomy and the ability to defend themselves from rights violations, refugees lose the ability to change the conditions which govern their lives. In other words, refugees are deprived of the standing needed within a state to effectively protect themselves against violations to their basic rights and autonomy.

With a nuanced understanding of refugeehood, we can further understand how the humanitarian approach is far too general to establish moral duties towards refugees. The general duty of rescue that this approach puts forward is not able to ground duties to refugees which address their demolished membership. Thus, the duties owed to refugees cannot be grounded and approached simply on the basis of a broad fear of harm, but rather on the breakdown of membership and political standing within their home state, which compromises their public autonomy, and in turn, their ability to securely claim and access their basic rights.

Critique of the Humanitarian Approach

The generality of the humanitarian approach creates limitations for responding to the severity of refugee conditions. Because the approach is grounded in a general duty to help those who are in need, it fails to distinguish between varying types of harms and the particular case of refugees, in which they are deprived of membership and political standing. Without identifying the specific political breakdown which refugees experience, the emphasis cannot be on remedying their membership and political standing. This specific harm must be recognized in order to properly establish the obligations required to remedy it. Because the humanitarian approach is focused on grounding duties solely based on altruism, generosity, and compassion, it cannot adequately address the political harms refugees face and the duties they are owed.

The humanitarian account introduces a significant challenge in the way we might understand refugeehood: the over-inclusivity of individuals in claiming refugee status.

25 Felix Bender, "Refugees: The Politically Oppressed," *Philosophy & Social Criticism* 47, no. 5 (2021): 624, https://doi.org/10.1177/0191453720931924.
26 Bender, "Refugees: The Politically Oppressed," 624.

Over-inclusivity may include cases in which individuals face extremely severe threats and harms yet have the legal or institutional abilities within their state to amend their situations. In these cases, individuals are presented with options to prevent or remedy their harms without fleeing their state and seeking international protection, which therefore implies their stable relationship with their state. When individuals are still able to receive help from their state institutions, their relationships with their states are not fully compromised, and they cannot claim refugee status.

When the "threat of serious harm functions as the central determining factor for claims to refugee status," other crucial aspects of refugeehood, such as political membership and public autonomy, are not recognized.[27] There is no question of whether or not refugees endure serious harms, yet the distinct facets of their harms, including lack of membership and political standing, are vital for properly assessing duties owed to them. As Bender describes, there is a significant difference between individuals who are harmed yet are able to remain in control of their remedies and those who are harmed with no options other than flight. [28] To further understand this distinction, take individuals who experience environmental disasters in their home state. According to the humanitarian approach, these individuals can be categorized as refugees because of the harm they are faced with. However, individuals in these conditions still stand in a stable relationship with their state and can seek political protection without having to flee. The general harm in this situation is not sufficient for claiming refugee status, "[w]hat matters is the legal-political disenfranchisement of a person that leaves her to her fate with no recourse other than flight." [29]

Now, we must assess the proper course of action and the moral duties owed to refugees who lack public autonomy, political standing, membership, and in turn, a secure way of claiming their basic rights. With the core of the situation being break-down of membership, it only makes sense for refugees to be provided with membership in a community that can protect their rights. But does any form of membership become sufficient for refugees? Does it matter whether it takes the form of temporary protected status, permanent residency, visa status, or citizenship? I argue that it absolutely does, and that formal citizenship in a destination country is morally required for refugees to remedy their loss of membership and secure basic rights.

There are a multitude of positions philosophers take when considering which states are morally obligated to provide these necessary duties to refugees. They consider many questions, such as: Which states are obligated to help these refugees? How can we allocate responsibility among states to care for them? While these questions are all immensely valuable to the study and understanding of refugees, I will not be elaborating on them, as this paper is focused on the extent of the duties owed to refugees rather than who exactly is obligated to provide them. In the following section I will support the claim for a refugee's right to citizenship and demonstrate the limitations present in any other form of membership.

27 Bender, "Refugees: The Politically Oppressed," 621.
28 Bender, "Refugees: The Politically Oppressed," 622.
29 Bender, "Refugees: The Politically Oppressed," 622.

The Argument for Citizenship

Positions such as temporary protected status or permanent residency might appear to be adequate solutions to remedy conditions of refugeehood, yet upon further analysis this is not the case. This type of status may be adequate for individuals who face harm in their home state, yet do not fall into the category of refugee. Buxton holds that "those who are pushed across a border for a very short amount of time and still maintain a stable relationship with their state" are not eligible for new membership.[30] For example, an individual might be forced to flee their home state because of war. In this case, the relationship between individuals and their state might not necessarily be broken or unstable, and they can sustain their public autonomy and membership when the war is over. For these individuals, temporary protected status or permanent residency would be a sufficient remedy as opposed to those with a more fundamental breakdown in membership and public autonomy.[31]

Nonetheless, most refugees do face the harms of losing membership in their home states, as they are devoid of state recognition of their rights based on their standing. In these cases, options such as temporary protected status may provide individuals with basic rights and opportunities, although they would still be deprived of the secure standing which grounds those rights. In other words, "the claimability of rights in some communit[ies] are still the most secure when one is considered a member in equal standing with her co-members."[32] Without the formal and legal status of citizenship, this utmost security to access and exercise basic rights cannot be realized.

In the simplest sense, citizenship is a legal status that is a requirement for a modern state to recognize someone as a member of a political community.[33] The legal status of citizenship allows for individuals to have complete membership in that state, with their rights and standing remaining protected. Citizenship serves as a way to divide responsibility for persons' protected rights.[34] States recognize citizens as individuals whom they are responsible for and must protect under their own laws.[35] Further, citizenship allows for an individual's formal legal membership within a state, a membership which presents individuals with both standing and rights and enables greater consideration from the state.[36]

It is important to draw the distinction between membership and citizenship here, as the two are not synonymous. Rather, membership is a feature of citizenship which

30 Buxton, "The Duty to Naturalise Refugees," 1133.
31 Buxton, "The Duty to Naturalise Refugees," 1133.
32 Buxton, "The Duty to Naturalise refugees," 1134.
33 Joseph H. Carens, *The Ethics of Immigration* (Oxford University Press, 2013), Chapter 2.
34 Daniel Sharp, "Immigration, Naturalization, and the Purpose of Citizenship," *Pacific Philosophical Quarterly* 104, no. 2 (2023): 409, http://doi.org/10.1111/papq.12428.
35 Sharp, "Immigration, Naturalization, and the Purpose of Citizenship," 410.
36 Sharp, "Immigration, Naturalization, and the Purpose of Citizenship," 422.

provides formal recognition and political standing within a state. Though refugees do not automatically and formally lose their citizenship when they become refugees, many refugees are "de facto" stateless, in that their state is not able to ensure protection of their basic rights.[37] In these cases, citizenship remains while membership, referring to the political protection of both standing and rights, is lost. One can have citizenship but still lack membership, yet one cannot have complete membership without citizenship. While it is possible for non-citizens of a state to have their rights respected and even protected, the respect and protection would come from a place of charity and discretion rather than standing in society. In these situations, an individual's rights might be protected, but that protection is not secure. The political standing that comes with membership is necessary for rights to be protected in a secure way without having them stripped away at the discretion of the state. As Buxton describes, an individual without political standing cannot effectively make claims on the state. That individual "ceases to enjoy the membership which that citizenship is meant to formally recognise."[38] Persons with permanent residency or temporary protected status may be able to hold the same sets of rights as citizens, whether it is the civil right to free speech, the political right to vote, the social right to healthcare, or more. Yet, even for individuals who are given extensive features of citizens, such as the same set of rights, they do not hold the same equal status. Not only does the status aspect of citizenship allow for legal recognition of basic rights from the state, but it also holds social significance. Daniel Sharp describes the problematic hierarchies that are created when individuals share the same sets of rights but not status.[39] Citizenship holds a social meaning which "shapes people's self-conception and serves as a source of esteem and identity."[40] The single status that comes with citizenship is greatly valued by individuals, as they gain a public and elevated form of political agency and standing in a state.[41] In other words, citizenship status is publicly recognized as equal standing both legally and socially. That being the case, the formal status of citizenship is the necessary first step for a refugee to regain political standing and exercise their public autonomy to make effective claims upon the state. Even if a refugee still holds formal citizenship in the country they have fled from, a "functioning citizenship" must be issued to terminate refugee status.[42]

Moral Justification for Citizenship

Kantian Ethics

Having established the significance of membership for securely claiming basic rights based on standing, I will now demonstrate further moral justification for providing refugees with citizenship by applying Kantian Ethics. Kant's moral philosophy is centered around individuals as rational agents with their own autonomy, or the right to self-govern

37 Buxton, "The Duty to Naturalise Refugees," 1127.
38 Buxton, "The Duty to Naturalise Refugees," 1129.
39 Sharp, "Immigration, Naturalization, and the Purpose of Citizenship," 431.
40 Sharp, "Immigration, Naturalization, and the Purpose of Citizenship," 412.
41 Sharp, "Immigration, Naturalization, and the Purpose of Citizenship," 431.
42 Buxton, "The Duty to Naturalise Refugees," 1127.

one's life.[43] He develops his moral principles in what he calls the formula of humanity, which significantly asserts that individuals must be treated as ends in themselves, rather than used as a means to get to some other end. [44] To treat someone as an end in themselves is to recognize and respect the absolute and unconditional worth each person holds. On the other hand, to use someone merely as a means to some other end is to respect someone only at the discretion of one's will. With this core principle, Kant discusses the duties that arise in relation to these principles. One who "transgresses the rights of human beings [and] intend[s] to make use of the person of others merely as a means " violates a duty owed to rational agents.[45] Kant describes both perfect and imperfect duties. Perfect duties being specific prohibited actions which cannot be violated, and imperfect duties being general ends to strive towards. When other states aid refugees in any way other than granting them citizenship, they are using them as a means to an end. By providing them with rights and aid only at their own discretion, rather than allowing them the security that comes with citizenship, states are violating aspects of refugees' autonomy and therefore their own perfect duty. With citizenship as the most secure form of membership needed to claim equal political standing and basic rights, refugees must have the opportunity to reclaim citizenship in a new state and exercise their public autonomy. Kantian Ethics serves as an inviolable moral justification for a refugee's right to citizenship in order to retrieve the membership and equal standing necessary for public agency.

Objection & Response

Feasibility & Desirability Objection

One might object to my argument on account of feasibility and desirability. Some could argue that providing refugees with citizenship is not a realistic approach to rectifying conditions of refugees. Most states are already hesitant to admit refugees without the obligation of providing them citizenship; many have formed techniques of exclusion in the form of visa controls or carrier sanctions to hinder the overflow of refugees seeking admittance.[46] Some states and refugees may not want the temporality of refugeehood to go away. For states, it would be easier for them to simply send refugees back to their home states once it is safe to do so rather than providing them with equal status of citizenship. Similarly, refugees might not want to abandon their hopes of returning to their home countries and integrate into a new state.

However, providing citizenship to refugees does not prevent them from one day remedying their relationship with their home state and returning safely. For example, some states allow dual citizenship, so refugees can securely live as citizens in a new state and still return to their home state as citizens one day. In states which do not allow dual

43 Immanuel Kant, *Groundwork for the Metaphysics of Morals* (Oxford University Press, 1785).
44 Kant, *Groundwork for the Metaphysics of Morals*, 37.
45 Kant, *Groundwork for the Metaphysics of Morals*, 38.

46 Carens, *The Ethics of Immigration*, 209.

citizenship, individuals can decide to revoke their new citizenship if they ever decide to return home and remedy their relationship with their home state. But, while their citizenship in their country of origin is not functioning properly and protecting their basic rights, equal membership and political standing in another state is necessary for refugees to securely access their basic rights.

For states, refugees can be looked at as an investment. Though individuals might need time to assimilate before their contribution becomes productive, they will eventually become active agents able to work more effectively towards their own personal goals as well as possibly contributing to state goals. Once states allow the conditions for refugees to work and take part in the economy, refugees will be able to exercise their potential, absorbing and mastering new skills and resilience.[47]According to the World Economic Forum and the IMF, "additional spending by EU states on asylum seekers and refugees boosted EU GDP by 0.09% in 2016 and by 0.13% in 2017," and "every one euro invested in refugees assistance yields nearly two euros in economic benefits within five years." [48] Therefore, providing refugees with citizenship and restoring their political membership and standing should actually incentivize states to issue refugees citizenship. Once given citizenship, new members will be able to integrate more easily and eventually have the opportunity and the public autonomy to give back to their new state, both politically and economically.

Conclusion

Throughout this paper, I described the harms and vulnerabilities which categorize refugeehood—the loss of membership, political standing, public autonomy, and the ability to claim basic rights. I discussed the general limitations of the humanitarian approach in determining the moral duties owed to refugees, as it inadequately recognizes the roots and results of the harms which refugees are faced with. I argued that citizenship must be given to refugees to rectify their loss of membership and unequal political standing. Because of the lack of security and equal standing, forms of membership such as temporary protected status or permanent residency are not sufficient for refugees living in states of temporary permanence. Secure access to basic rights lies in the legal and political status of citizenship, in which they can be recognized both socially and politically as equal members of society.

Works Cited

Bender, Felix. "Refugees: The Politically Oppressed." *Philosophy & Social Criticism* 47, no. 5 (2021): 615–33. https://doi.org/10.1177/0191453720931924.

Buxton, Rebecca. "The Duty to Naturalise Refugees." *Critical Review of International Social and Political Philosophy* 26, no. 7 (2023): 1119–139. https://doi.org/10.1080/13698230.2021.1880198.

Carens, Joseph H. *The Ethics of Immigration*. Oxford University Press, 2013.

Clements, Kelly T. "Refugees Mean Business. This is Why Investing in Them Pays Off," *World Economic Forum*, January 20, 2025. https://www.weforum.org/stories/2025/01/refugees-business-why-investing-in-them-pays-off/.

Collier, Paul and Alexander Betts. *Refuge: Rethinking Refugee Policy in a Changing World*. Oxford University Press, 2017. http://ebookcentral.proquest.com/lib/northeastern-ebooks/detail.action?docID=505097.

Hillier-Smith, Bradley. *The Ethics of State Responses to Refugees*. Routledge, 2024.

Kant, Immanuel. *Groundwork for the Metaphysics of Morals*. Oxford University Press, 1785.

Miller, David. *Strangers in our Midst: The Political Philosophy of Immigration.* Harvard University, 2016. Press. http://www.jstor.org/stable/j.ctvjf9z4w.

Sharp, Daniel. "Immigration, Naturalization, and the Purpose of Citizenship." *Pacific Philosophical Quarterly* 104, no. 2 (2023): 408–41. http://doi.org/10.1111/papq.12428.

Shue, Henry. *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy: 40th Anniversary Edition*. Princeton University Press, 2020. http://doi.org/10.2307/j.ctvqsdnkw.

Singer, Peter. "Famine, Affluence, and Morality." *Philosophy & Public Affairs* 1, no. 3 (1972): 229–43. http://www.jstor.org/stable/2265052.

47 Kelly T. Clements, "Refugees Mean Business. This is Why Investing in Them Pays Off," *World Economic Forum*, January 20, 2025, https://www.weforum.org/stories/2025/01/refugees-business-why-investing-in-them-pays-off/.

48 Clements, "Refugees Mean Business. This is Why Investing in Them Pays Off."

Sensing Blackness, or: the (Racial) Difference of the Same

Daniel Zheng

Abstract: This paper stages an encounter between the notion of différance in Jacques Derrida or Jean-Luc Nancy and various theorizations of blackness by Christina Sharpe, Achille Mbembe, and David Marriott. I suggest that différance offers a useful framework for understanding the claim to "ontological blackness" in Sharpe and Marriott that troubles the boundary between, on the one hand, the essentializing logic of racial difference in its colonial guise, and on the other, the universal humanism in Mbembe which erases blackness as merely a non-ontological historical fantasy. Différance allows one to treat racial difference as, in Nancy's words, the "difference of the same," escaping the reductive logic of both difference and universalism. Throughout the paper, I trace a line of *différance* across the way that Sharpe and Marriott explore blackness. For Sharpe, this emerges from her description of "anagrammatical" or "annotated" blackness, where blackness is defined by its shifting definition and interpretation that always defers any essentialized meaning, and thus where it forms an "excess" close to Nancy's understanding of the excess or "suspended step" of sense. I then take this argument one step further through a reading of a Derridean logic in the work of David Marriott, who contends that blackness, properly speaking, can never appear as anything but a negation. In this way, blackness operates through precisely its place of non-presence or perpetual erasure and deferral, and *différance* offers a helpful corrective in "making sense" of this blackness that can never appear. I therefore contend that *différance* opens toward a properly liberatory logic of blackness that follows its own tracks and cannot be essentialized into "universalism" of "difference" since it never truly arrives as such.

Introduction

Near the end of his *Critique of Black Reason,* Achille Mbembe seems to identify a contradiction between two of his central interlocutors, Aimé Césaire and Frantz Fanon: "What did he [Césaire] mean by 'Black' (Nègre), this return to the name that Frantz Fanon, in *Black Skin White Masks*, said was only a fiction?"[1] Or, to put it in other words, if thinkers like Fanon highlighted how racial difference was merely a fantasy imposed upon black people utilized to separate and subjugate them, then why did black discourse remain so invested in maintaining this concept of "blackness?"

Mbembe's answer to this dilemma is that a project like *Négritude* remains a fundamentally *historical* one, seeking to recuperate the concept of blackness as merely a stepping stone on the way toward the reality of an absolute and universal humanism. Césaire and other thinkers of blackness used the term, Mbembe explains, as a method that "embraces the fiction"[2] of race, but does so exclusively historically and projects itself toward a future without the need for its fantasy. "Positive difference," in Mbembe's terms, is thus "fundamentally an orientation toward the future,"[3] a future in which we all share the world as equal humans defined by our essential and inherent humanity.

But implied by this orientation toward the future is a neatness wherein blackness, or racial difference writ large, is something that exists only to be expunged later. As a historical project on the side of both oppressor and oppressed, the endpoint of Mbembe's thought is the notion that one aspires, in the final analysis, to only and always *forget blackness*. But to be as practical as possible, does this smoothing over of difference in favor of a universal humanity not inherently re-open the door for racist discourse? That is to say, is it wise to orient oneself towards a dispensing of the category of a liberatory blackness given, well, the entirety of the history that Mbembe systematically outlines in his book?

David Marriott, in his review of *Critique of Black Reason*, puts this problematic in philosophical terms, asking whether a European humanism, which from the beginning assumed the inhumanity of blackness, can ever be the solution to anti-blackness: "it seems more plausible to say that the limits of critique have always already begun, were always already twisted and distorted by the desire to see blackness as a primordial difference within the human, and thus cannot be, and never could be, overcome in the humanistic ways Mbembe suggests."[4] Marriott thinks that Mbembe crucially misreads Fanon, in that black thought for Fanon must always "*refuse* everything that Europe has to offer—a point that Fanon insisted on when he said that black invention must go beyond the mere mimicry of European humanism's promissory structure."[5] Christina Sharpe expresses a similar sentiment when she remarks that "I am not interested in rescuing Black being(s)

1 Achille Mbembe, *Critique of Black Reason* (Duke University Press, 2017), 159.
2 Mbembe, *Critique of Black Reason,* 89.
3 Mbembe, *Critique of Black Reason,* 94.
4 David Marriott, "The Becoming-Black of the World?: On Achille *Mbembe's Critique of Black Reason*," *Radical Philosophy* 202 (2018): 66.
5 Marriott, "The Becoming-Black of the World?" 71.

for the category of the 'Human' . . . [that] continue to produce our fast and slow deaths."[6]

Debates about Fanon aside, we may say, without doubt, that Mbembe's conception of blackness has little to do with any ontology of racial difference, even though white colonial narratives declared blackness a fundamentally *ontological* lack by "comparing humans to animals"[7] and describing black subjects as "trapped in a lesser form of being."[8] We can see this in Mbembe's description of the "becoming-black of the world," where blackness, in today's merging between "animism and capitalism,"[9] is generalized and spread out across the world, "institutionalized as a new norm of existence and expanded to the entire planet."[10] Blackness thus becomes severed from its racial origin in Mbembe, representing a newly capacious quality of being defined by the technological and economic structure of our present. But again: does moving the concept of blackness away from a concept of race not inherently brush over the very real and continuing racial disparities of today?

The purpose of this paper is not to adjudicate over biological conceptions or narratives of racial difference, which often leads to overly reductive or misconstrued terrain.[11] Instead, it seeks to inquire into the explicit *ontologies* of blackness expressed in the discourse of other black writers and thinkers who treat blackness as something more substantive than Mbembe's future-oriented disavowal of the concept. Sharpe, for example, shares Mbembe's concern of blackness as something historically constructed, but suggests, crucially, that the "annotative" practices of Black artists and thinkers in the wake of this history carve out a new *ontology* of blackness: "Annotation appears like that asterisk, which is itself an annotation mark, that marks the trans*formation into ontological blackness."[12] A quick glance at other book titles by various black thinkers reveals a similar trend, revolving around the intersection between, in various configurations, "Blackness" and "Being."[13]

The point of this long-winded introduction is to suggest the intervention of this paper—that the application of the theories of Jean-Luc Nancy or Jacques Derrida might offer a corrective in navigating (or "deconstructing," perhaps) between the dichotomy of the white fantasy of race and its essentializing difference on one side, and Mbembe's universal humanism on the other. To speak, then, of the too-obvious pun in words: this essay proposes not a discussion of racial difference but of racial *différance*. Specifically, the framework of *différance*—Jacques Derrida's description of how "every concept is inscribed in a chain or in a system within which it refers to the other, to other concepts, by means of the systematic play of difference"[14]—offers a way to understand racial difference as neither subsumed into a universal or homogenous unity nor able to be essentialized into straightforwardly racist discourse. I suggest that this framework resonates closely with the work of black thinkers like Sharpe, Marriott, and Selamawit Terrefe who hold onto a conception of blackness as never static, never complete, or perhaps even never present. That is to say, blackness thought in its difference and its deferral can never fully arrive or fully appear—and in this way, exists as a liberatory *excess*, a (non)being unable to be put back as a static term into the set or framework it came from.

A "sketchy summary of an elucidation of différ*a*nce"[15] (and Blackness)

"Différance," writes Jacques Derrida, "is neither a word nor a concept,"[16] but "rather the possibility of conceptuality."[17] As a concept, a constructed term (by all parties—the colonial project but also Césaire, Sharpe and so on), blackness is, if we are to take Derrida at his (non)concept, always inherently composed through the play of *différance*—that structure where "the movement of signification is possible only if each so-called 'present' element, each element appearing on the scene of presence, is related to something other than itself."[18] This, of course, may be a fundamental, if basic, conclusion that underlies the points that follow in the rest of this paper: whatever blackness "is," is constructed discursively through interactions with other words or concepts. And as such, blackness can never be a static term—its "meaning," if one can even speak of something like that, is constantly re-negotiating itself within the conceptual chains of *différance*. But the general ambiguity of this claim feels inherently unworkable, so let's continue forward to a more specific engagement between the two ideas emerging here in the encounter between *différance* and Christina Sharpe's idea of "black annotation."[19]

Sharpe's conception of blackness, "annotated," highlights that the purpose of annotated blackness is to be "anagrammatical." More explicitly, it's "anagrammatical in the literal sense as when 'a word, phrase, or name is formed by rearranging the letters of another.'"[20] As an anagrammatical structure, blackness thus always offers new interpretations: "grammatical gender falls away and new meanings proliferate . . . a secret message is discovered through the rearranging of the letters of a text."[21] Sharpe's point is that blackness is never static—following Hortense Spillers, she suggests that black bodies, when reduced to commodities, become "bodies emptied of self-interpretation,

---

6 Christina Sharpe, *In the Wake: On Blackness and Being* (Duke University Press, 2016), 116.
7 Mbembe, *Critique of Black Reason,* 30.
8 Mbembe, *Critique of Black Reason,* 17.
9 Mbembe, *Critique of Black Reason,* 4.
10 Mbembe, *Critique of Black Reason,* 6.
11 See Kim TallBear's article "Genomic Articulations of Indigeneity" on this point.
12 Sharpe, *In the Wake*, 116.
13 Sharpe's book is subtitled "On Blackness and Being," David Marriott's book *Whither Fanon?* is subtitled "Studies in the Blackness of Being," Alexander Weheliye is writing on "Schwarz-Sein," and so on.
14 Jacques Derrida, "Différance," in *Margins of Philosophy,* trans. Alan Bass (University of Chicago Press, 1982), 11.
15 This is a reusing of a gesture from Jean-Luc Nancy, *The Sense of the World* (University of Minnesota Press, 1997), 36.
16 Derrida, "Différance," 8.
17 Derrida, "Différance," 11.
18 Derrida, "Différance," 13.
19 Sharpe, *In the Wake*, 77.
20 Sharpe, *In the Wake*, 76.
21 Sharpe, *In the Wake*, 76.

into which new interpretations could be placed,"[22] re-formed as an "index of violability and also potentiality."[23] As an anagrammatical structure, then, the ungendered black body engages in a structure similar to *différance*, where its meaning is not fixed, where letters are rearranged, and thus where the signification of blackness slips away and always defers any "ultimate" meaning. Anagrammatical blackness therefore takes, in Derrida's words, "a detour that suspends the accomplishment or fulfillment of 'desire' or 'will,'"[24] that thus suspends the "mechanism of fantasy"[25] that turns black bodies into commodities—instead carving out a new conception of blackness that indefinitely postpones the answer to the simple question demanded incessantly: "What is blackness?"

For Sharpe, blackness thus is better represented as a figure of *excess* or "signifying surplus,"[26] which she attempts to render using the term "Trans*."[27] For Sharpe, the Trans* is a way to visualize how "Black has always been that excess,"[28] that the signifying surplus and excess contained within the term "Black" once again holds a double move between a destruction of "normal" meaning and an opening to new interpretations. The signifying surplus, or excess of blackness, exemplifies for Sharpe "the failure of words and concepts to hold in and on Black flesh,"[29] that process in which "meaning slides, signification slips, when words like *child*, *girl*, *mother*, and *boy* abut blackness."[30] These terms, she highlights, are rendered distant from their former meaning when placed alongside the uber-signifier of blackness, such that "*girl* doesn't mean 'girl' but, for example, 'prostitute' or 'felon.'"[31] But again, the double-move: the capaciousness of this excess of blackness which contains "annotations" (prostitute, felon) that underlie it, allows for the insertion of new annotations, new conceptions of blackness, a process of "redaction and annotation toward seeing and reading otherwise."[32]

Blackness, therefore, in Hortense Spillers's words, is defined by its "enforced state of breach,"[33] where it can be invaded or subtended at any moment by an excess of both subjugating force and liberatory annotation. Is this not also the necessary implication of *différance*? That is to say, an "enforced state of breach" reflects Derrida's logic of the trace—that every word, concept or state contains echoes and traces of what it does not mean, where each term is breached in this way with the non-present trace of its excess or difference.[34] Derrida writes, "There is no breach without difference and no difference without trace."[35] We might say that this uncontainable excess that seems to define blackness for Sharpe exposes a logic of the trace inherent in its formulation, where blackness is supplemented, from the start, with a dual-sided violent and liberatory surfeit that subtends it. And again, then, when Sharpe argues for something like the "annotational" project of "reading and seeing something in excess of what is caught in the frame,"[36] we hear the echoes of Derrida's famous "there is nothing outside of the text," where, when blackness is concerned, the excess outside of the frame is precisely what *must be used* by black thinkers to understand it. To read blackness "in excess of what is caught in the frame" is to refuse the view of the potentate which delimits the black subject as cut off cleanly by an inscribed border of commodification and denial of humanity. It is, in the words of Toni Morrison, quoted by Sharpe, a way "to carve away the accretions of deceit, blindness, ignorance, paralysis, and sheer malevolence embedded in raced language so that other kinds of perception were not only available but were inevitable."[37] A reading of excess as *inevitability* is one that accounts for the trace, the supplement, and the entire logic of *différance* which underlies the co-construction of blackness with an innumerable and excessive slate of annotations of both violation and potential. In "the wake," the presence of blackness is defined by its deferred presence, an excess that imposes itself such that it is always still coming, still arriving.

If, as marked earlier, the Trans* for Sharpe posits or exemplifies this "trans*formation into ontological blackness,"[38] we might say that the move from a conceptual blackness to an ontological blackness draws our course from the essentially Derridean terrain of meaning and signification closer to the logic of Jean-Luc Nancy's ontological conception of sense.[39] For Nancy, rather than merely refer to constructions of meaning, *sense*[40] itself is structured through the logic of *différance*: "Being senses itself

22 Sharpe, *In the Wake*, 29.
23 Sharpe, *In the Wake*, 75.
24 Derrida, "Différance," 8.
25 Mbembe, *Critique of Black Reason,* 114.
26 Sharpe, *In the Wake*, 80.
27 Sharpe, *In the Wake*, 30.
28 Sharpe, *In the Wake*, 30.
29 Sharpe, *In the Wake*, 123.
30 Sharpe, *In the Wake*, 80.
31 Sharpe, *In the Wake*, 77.
32 Sharpe, *In the Wake*, 117.
33 Sharpe, *In the Wake*, 77.
34 What I'm leaving out here is the crucial difference made in the term "enforced," which hints toward a central element of violence of the term "enforced," which I will try to get to later in the section on David Marriott.
35 Derrida, "Différance," 18.
36 Sharpe, *In the Wake*, 117.
37 Sharpe, *In the Wake*, 117.
38 Sharpe, *In the Wake*, 116.
39 This is probably not so precise, Derrida himself would argue that différance is the precondition for any ontology, but nonetheless I think the general point holds (Nancy is more interested in différance and being, Derrida more meaning).
40 For Nancy, "sense" is to be understood as the very phenomenological idea of "sensation" (or as an opening to the world, hence its "archi-significance"), but this "sense" also encapsulates or traverses all of "the five senses, the sense of direction, common sense, semantic sense, divinatory sense, sentiment, moral sense, practical sense, aesthetic sense" (*The Sense of the World,* 15). One way to understand Nancy's use of the term is to note that the book begins by describing a "crisis of sense" (2) spurred on by the collapse of two previous concepts of sense—myth/religion (given sense) and nihilism (loss of sense). From this crisis, Nancy reaffirms sense hesitantly by noting the banal statement that "it is *necessary indeed* that there be something like a sense of the

deferring and differing. It senses itself or knows itself to be differing and different."[41] Between the dual poles of myth or religion (given sense) and the abyss of nihilism (a loss of sense), Nancy wants to suggest instead that sense should be understood as spacing and *différance*, a sense always spaced out between singularities, or what he repeatedly refers to as "the difference of the same,"[42] a sense that is always ongoing and deferred rather than completed. This sense, then, structured through the play of *différance*, is one of "nonsignificance or archi-significance,"[43] where its orientation toward the world takes place before signifier or signified come into play.[44]

We might suggest that this nonsignifying logic of Nancy's description of sense and its occupation of the "suspended step" mirrors Sharpe's understanding of ontological blackness as trans*, as if blackness exists precisely at the nonsignifying point of sense (for Sharpe, this is the asterisk) that the step of thought passes over. That is to say, in a thought of blackness, the step of thought does not pass straight into a signification of blackness ("blackness is ___") but rather remains too suspended over the "sense" of blackness, which is the very condition of possibility for any signification at all. This delicate ontology suggests that racial difference might be inscribed (or *exscribed*) into the moment of existence—the primordial opening toward the world or the "irreducible strangeness of each one of these touches to the other."[45] This understanding of blackness at this moment of sense prior to signification might highlight why Sharpe remains invested in a notion of blackness as "that excess." A blackness explored through the "absolute excess of sense"[46] is in the same breath one that is always *more* than what can be turned into signification—an annotated excess that comes always *before* the subsequently essentialized formulations of "what blackness is," *before* the boundaries are marked out in the dehumanizing frame.

The asterisk in Sharpe, this void of blackness wherein "meaning slides, signification slips," therefore occupies something like Nancy's description of the asignifying sense of

world (or *some* sense in the world)" (7)—tentatively, then, going back to sense in all of its meanings, but also "simply" to the very basic idea of being an opening to the world, to occupy this infinitely slight gap between myth and nihilism.

41 Nancy, *The Sense of the World*, 35.

42 Nancy, *The Sense of the World,* 8.

43 Nancy, *The Sense of the World,* 9.

44 Différance is relevant to sense because it poses the question of "spacing" (the spaces between signs, overly simply put)—if sense is understood not merely phenomenologically, but as structured too via this spacing of singularities, then this is how it occupies this "infinitely slight" place. Différance implies a constant breakdown of significations and forming of new ones, which Nancy wants to think at the heart of sense and nonsense: "[the common space of sense] is nothing but the limit that separates and mixes at once the insignificance that arises out of the pulverization of significations and the nonsignificance or archi-significance encountered by the need of being-toward-the-world" (*The Sense of the World,* 8).

45 Jean-Luc Nancy, *Being Singular Plural* (Stanford University Press, 2000), 6.

46 Nancy, *The Sense of the World,* 24.

the monolith in Kubrick's *2001: A Space Odyssey*, a structure that offers "something like an instance or indication of sense" because it refuses to signify in its "smooth, hard, surface," existing only as a "presence of an absence," while still interpellating through its "signal, call, or intimation."[47] Like the monolith, the asterisk in Sharpe's trans* calls out to the reader, but it does so *before* it signifies or suggests a signifying absence in its opaque formal orthography. Sharpe writes, "The asterisk after the prefix 'trans' holds the place open for thinking."[48] It functions as a "means to mark the ways the slave and the Black occupy what Saidiya Hartman calls the 'position of the unthought.'"[49] In other words, the asterisk of Trans* occupies in Nancy's framework the suspended step of sense, as if it expresses in punctuation "*the step of thought suspended over this sense that has already touched us*."[50] In this way, Hartman's "unthought" both "initiates and terminates" any thought of blackness, at once begging to be thought (*what does this * mean?*) and at once remaining, without fail, still necessarily unthought. We are touched by something we might call "blackness," or perhaps when we attempt to formulate something like what blackness is, we remain posterior to "*this sense that has already touched us*."[51]

If in having suggested that blackness takes its place (perhaps by necessity) within Nancy's transcendental structure of sense, we find ourselves stuck at a mirror of Nancy's "there is no longer anything here to be said," where blackness lies in its ultimate essence away from, behind, before, or at the very opening of the possibility of signification, and where thus there is no use speaking about it.[52] Facing this philosophical impasse, one should re-stress, above all, that, where blackness is concerned, many people have many (terrible) things to say. What I mean is that the disappearance of blackness to a philosophically mythical locus at the very opening of signification would play, we might note, rather well into the hands of racist discourse that claims that what is "black" absolutely does not signify, that is denoted by an absolute nonbeing. And, in a rather similar way, Fanon's famous "*Tiens, un Négre!*" in which the black subject is subjectified, offers something akin to a racialized interpellation which occurs, too, *prior* to any signification.[53] This is Marriott's point in his essay "Blackness: N'est Pas?" where in the "*Tiens*" episode "the signifier (and not the sign nor the gaze) acts as the unconscious confirmation and reminder that one is *already* racially subjected."[54] What I want to note here is the term "already" in Marriott's reading of Fanon, the very same *already* as Nancy's "the sense that has *already* touched us." As soon as the black subject is able to formulate the encounter as signification, they are *already* touched by their having-been-constructed as racial subject, such that

47 Nancy, *The Sense of the World,* 39.

48 Sharpe, *In the Wake,* 30.

49 Sharpe, *In the Wake,* 30.

50 Nancy, *The Sense of the World,* 11.

51 Nancy, *The Sense of the World,* 11.

52 Nancy, *The Sense of the World,* 79.

53 Frantz Fanon, *Black Skin, White Masks* (Pluto Press, 2008), 82.

54 David Marriott, "Blackness, N'est Pas?" *Propter Nos* 4 (2020): 30. Emphasis mine.

"the discovery of one's racial difference is always a belated discovery."[55] The belated or deferred temporality of the moment of sense here does not, then, offer a pathway between the dualism of essentialized difference and universal humanism, but merely reinscribes the ideology of racial difference into the structure of sense. In Fanon's terms, this is the too-late discovery of an already internal "historico-racial schema"[56] *beneath* the body, sketched by "the other, the white man, who had woven me out of a thousand details, anecdotes, stories"[57]—a racist subjectification already inscribed always in advance of its passage into legible meaning that parallels Nancy's structure of sense.

Where to go, then, from here? In a moment where he prescribes error to his text, or perhaps a moment when his discourse "can reveal itself to be sick,"[58] Nancy writes that his "sketchy summary of an elucidation of différance is still a bit too phenomenological-constitutive."[59] In our terms, we might say that this "sensuous" interpellating movement of blackness that precedes signification is, similarly, a bit too phenomenological, which attempts to express *différance* through a positive presence that itself is already compromised in the racial-phenomenological analysis of Fanon. Rather, as Nancy continues, "The *coming* (but is there 'the' coming, and not rather a 'to come' that comes without allowing of substantialization?) demands something else."[60] To render blackness expressed as the positive term "blackness" is thus to substantialize what a proper logic of *différance* would maintain as a blackness-to-come rather than "a" or "the" blackness. To speak of *différance* is after all, for Derrida, not a question of positive terms but of negative ones: "a difference generally implies positive terms between which the difference is set up; but in language there are only differences *without positive terms*."[61] To continue tracing these parallels between blackness and *différance*, we should move past Sharpe's positive conception of blackness[62] (blackness *is* anagrammatical, *is* annotated, and so on) to Marriott's more radical formulation in "Blackness: N'est Pas?" where he suggests that blackness can instead only ever be defined always as negation, nonbeing, such that "blackness cannot be uttered without at once being echoed by a voice that is not: n'est pas."[63]

(This) Blackness which is Not

Marriott's point, drawing off Fanon's "le Noir n'est pas un homme,"[64] is that blackness is always structured as this *n'est pas* or form of negation. Blackness as such can

55 Marriott, "Blackness, N'est Pas?" 31.
56 Fanon, *Black Skin,* 84.
57 Fanon, *Black Skin,* 84.
58 Nancy, *The Sense of the World,* 80.
59 Nancy, *The Sense of the World,* 36.
60 Nancy, *The Sense of the World,* 36.
61 Derrida, "Différance," 11.
62 To be fair, Sharpe does often use a term like blackness's non/being to highlight the dual sided perspective of this, although I would still suggest that even the use of the slash highlights an attempt to substantialize.
63 Marriott, "Blackness, N'est Pas?" 27.
64 Marriott, "Blackness, N'est Pas?" 27.

therefore never *appear* as anything but negativity, a statement of which it is not, such that black subjects "have to assume a being that is not in order to be recognized as subjects."[65] As soon as "blackness" is substantialized, we might say, it is struck with the inversion that it "is not," or it is struck with this negation prior to the level of substantiation, such that it can never be a positive term. As Marriott writes, "to make it [blackness] into an *is* rather than an is not, is always preceded (structurally and epistemologically, says Fanon) by the moment of negrophobic interpellation."[66] This non-being is inscribed not merely at the level of signification, then, but on a structural level deeper, something closer to that of sense, where "*the black is made to be non-being rather than an inauthentic failure to be*."[67] Which is to say, blackness is on one level structurally *forced into being*, "being-made-to-be"[68]—but only such that it can inevitably be assimilated to, paradoxically, the *negated* inscription that this being is always and forever non-being.

Of course, the Derridean would respond: but this is the structure of all meaning, or all ontology even, that there is only the play of "differences without positive terms." But the difference that makes the difference here is that *this difference* (blackness) is struck each time with violence when it comes—a violence that is both literal and discursive. At the level of discourse, this violence occurs in the implicit assertion that, where race is concerned, whiteness is the positive term par excellence, and blackness is condemned to the structure of its negation. Marriott thus explains: "blackness has no articulation, for even its difference is borrowed; the result is a language whose idiom is that of a *n'est pas*."[69] Under this logic, whiteness—which is indexed always to the basic and universal ontology of what it means to "be human"—does have a substantial articulation of its own that does not depend on its difference from difference to function. Thus, as Selamawit Terrefe, riffing off Marriott, puts it, "The incapacity of its [blackness's] signification within and beyond the paradigms of meaning and measure persist violently, expiating psychic blows indistinguishable from those that punctuate matter."[70] Within this framework, blackness is precisely defined by its failure to signify or exist phenomenally.

If the solution to this situation seems like it should be something akin to Mbembe's humanism, where one simply places this non-being as a historical condition to be overcome via a re-granting of the "being" of humanity to the black subject, Marriott would caution against the very construction of this universality. Rather, his proposition is that this negation is inscribed into the very structure of the "universal," such that blackness itself cannot take any place within a universal—it is conceived of as the very negative of the universal as such. Thus, blackness remains for Marriott "incomprehensible to both reason and ontology as traditionally understood (by which I mean: the white rhetoric of

65 Marriott, "Blackness, N'est Pas?" 31.
66 Marriott, "Blackness, N'est Pas?" 41.
67 Marriott, "Blackness, N'est Pas?" 34.
68 Marriott, "Blackness, N'est Pas?" 34.
69 Marriott, "Blackness, N'est Pas?" 31.
70 Selamawit Terrefe, "Black Surfeit, Black Grief: Not of Meaning nor Measure." *Political Theology* (2023): 2.

universality).”[71] This is why Marriott is suspicious of Mbembe’s conclusion, because in his gesture to forget blackness, Mbembe proposes a scenario in which there can only be a singular universal (the human) that the black subject will always still be the negative, the non-being of. In a similar way, Marriott thus also argues that blackness can never properly speaking be described via a phenomenology because its “existence” is always effaced upon arrival.

As an inscribed state of negation, Marriott suggests that any attempt to speak about blackness is always struck with a fundamental aporia where it can never appear without being erased or dragged away into fantasy or universality. And so any possibility of a black invention (or anagram, or annotation, in Sharpe’s terminology) has to reckon with the presupposition that blackness “cannot be placed or be made actually present as such.”[72] It is thus by necessity in this very location of blackness’s *erasure* that one searches for it: “We must try to hear, within this silence, the yet to be understood experience of blackness, in order to pose *the question of its being anew, before its reading can crystallize around the question of what it is, or, as Fanon conceives of it, is not*.”[73]

Of course, the connection here is that this conception of blackness as unable to take presence, a blackness that must *be* without arriving at “what it is,” reminds one strongly of the logic of sense and *différance*. Beyond the deferred meaning of Sharpe’s blackness, Marriott’s blackness takes this comparison one step further in that it does not ever appear; rather, it is always deferring itself, never coming, never arriving. Nancy’s clarification of a non-phenomenological conception of sense is thus precisely an opening toward something like a way to understand or speak about what Marriott’s blackness could be, where it is structured by its endless deferral and thus (non)presence: “the coming into presence of being takes place precisely as nonarrival of presence.”[74] In this way, Nancy’s thinking offers a way to make sense of how Marriott’s understanding of a liberatory blackness “makes sense” only in its positioning at the very *edge of sense*, or at its positioning at the perpetually withdrawing excess of sense. As Marriott writes, “the conclusion that race has no immediate knowledge, and can only be approached as the experience of its infinite deferment, is the object of this book.”[75]

To reach something like an understanding of what blackness is (not), one must contend with a conception of blackness that constantly erases itself even as it is produced. But for Marriott, this situation leads toward the same double move that Sharpe and Césaire made, where there is a transformation from a constricting structure into a liberatory potential. A blackness that perpetually effaces itself is also one that is unable to be subsumed back into universalism because it never arrives at a substantial form as such. In the same breath, then, a blackness that perpetually effaces itself is one that perpetually invents itself, becoming an “endless transvaluation, at least in so far as Fanon can say that

71 Marriott, “Blackness, N’est Pas?” 32.
72 David Marriott, *Whither Fanon? Studies in the Blackness of Being* (Stanford University Press, 2018), 122.
73 Marriott, “Blackness, N’est Pas?” 34.
74 Nancy, *The Sense of the World,* 35.
75 Marriott, *Whither Fanon,* ix–x.

in the world through which he travels ‘I am endlessly creating myself.’”[76] This blackness is thus a “spectrality of black undecidability,”[77] or what appears to be, for us, a Derridean ghost in the sense of its always non-present presence.

Différance, then, offers a method of reading or liberating blackness within its struggle of erasure through its always-postponed nature. Rendered through the logic of a non-present presence, blackness in this sense can never appear as a positive term (and thus be subsumed into universality), instead taking the structure that Nancy describes: “The coming is infinite: it does not get finished with coming.”[78] We might say, the “transvaluation” of blackness is infinite, it does not get finished with creating itself. Or, it is a positioning of racial difference at the very edge of the excess of sense and its opening onto “nothing,” where a blackness that must be endlessly invented from its perpetual erasure is also “open on nothing but its own distance from nothing, within nothing, its ‘something’ having been thrown there from nowhere to nowhere,”[79] the nonarrival of the nothingness that blackness is (not).

Marriott describes this method of writing blackness as a “future imperfect” tense: “a moment of inventiveness whose introduction necessarily never arrives and does not stop arriving, and whose destination cannot be foreseen, or anticipated, but only repeatedly traveled, and is therefore not future at all.”[80] In this moment, Marriott, describes, nearly explicitly, the temporality of *différance*, where the inventiveness, or perhaps the *coming*, of blackness, is always arriving but never arrives. But, in this same sentence, I think he also begins to open another line of Nancy’s thought, namely, the thought of a destination that is not an end but merely one to be “repeatedly traveled.” That is to say, he begins to open up something like a *world* via Nancy’s logic of the confines of the world where “we touch our limits on all sides.”[81] This black inventiveness is something *radically open* to a world, but open only along the edges of its own differing/deferring tracks—open, in this sense, along the same path as Nancy’s “universe of which the unity is nothing but unicity, open, distended, diffracted, slowed down, differed, and deferred within itself.”[82] It is thus a formulation where blackness is constituted finally without reference to whiteness or even humanity, but toward the traveling of its own tracks, and thus toward the confines of a world. Oriented but never arriving into the world, following its repeatedly traveled tracks, black invention “gives itself in one stroke, in being *toward the world*, as the arrival of being in advance of itself—differing/deferring—in advance in one stroke, each time, all the way to the confines of the world, present already where it is headed, where it is not.”[83]

Between the dual poles of racial difference and universal humanism, the world of

76 Marriott, *Whither Fanon,* 204.
77 Marriott, “Blackness, N’est Pas?” 39.
78 Nancy, *The Sense of the World,* 35.
79 Nancy, *The Sense of the World,* 37.
80 Marriott, *Whither Fanon,* 284.
81 Nancy, *The Sense of the World,* 40.
82 Nancy, *The Sense of the World,* 37.
83 Nancy, *The Sense of the World,* 154.

a properly liberatory blackness takes shape through a *différance* that navigates the ever-closing space between the two—where blackness, in its continuously differing/deferring invention, never arrives as such (and thus can never be essentialized, commodified, violated), and at the same time never becomes a universal, heading along its own traces or tracks to the confines and limits that it always touches. Occupying the infinitely slight crevice between the abysses on each side, what blackness "accomplishes as a revelation at each and every moment is repeatable only as its difference from itself."[84] This blackness is not, then, a humanism, because it refuses to fall so neatly into a category of the human of which it will always be a negative foil. Yet, if it is a form of difference, this is a difference that can only be thought as a difference, or *différance,* from itself. Differing and deferring from *itself*, racial difference here is rendered, as Nancy likes to say, as merely "the difference of the same."[85] Which is to pose, in other words, that this blackness is also by necessity *both* a humanism and a difference, where each side contains traces of the other, where, like with sense and world, humanism and difference make each other make sense. The task then, of black invention, is precisely this—to pull apart the border between humanism and difference, to inhabit the very limit between the two, in order to devise a blackness that is properly liberatory, but which necessarily remains a blackness-to-come.

Works Cited

Derrida, Jacques. "Différance." *Margins of Philosophy*. Translated by Alan Bass. University of Chicago Press, 1982.

Fanon, Frantz. *Black Skin, White Masks.* Pluto Press, 2008.

Marriott, David. "The Becoming-Black of the World?: On Achille Mbembe's *Critique of Black Reason.*" *Radical Philosophy* 202 (2018): 62–71.

Marriott, David. *Whither Fanon? Studies in the Blackness of Being.* Stanford University Press, 2018.

Marriott, David. "Blackness: N'est Pas?" *Propter Nos* 4 (2020): 27–51.

Mbembe, Achille. *Critique of Black Reason.* Translated by Laurent Dubois. Duke University Press, 2017.

Nancy, Jean-Luc. *The Sense of the World*. University of Minnesota Press, 1997.

Nancy, Jean-Luc. *Being Singular Plural*. Stanford University Press, 2000.

Sharpe, Christina. *In the Wake: On Blackness and Being*. Duke University Press, 2016.

TallBear, Kim. "Genomic Articulations of Indigeneity." *Social Studies of Science* 43, no. 4 (2013): 509–33.

Terrefe, Selamawit D. "Black Surfeit, Black Grief: Not of Meaning nor Measure." *Political Theology* (2023): 1–11. Web.

84 Marriott, *Whither Fanon,* ix–x.

85 Nancy, *The Sense of the World,* 8.

Outsourcing the Self: Corporate Control and Existential Responsibility in the Digital Age

Ander Petri-Hidalgo

Abstract: In an era where the targeted advertising industry generates a yearly 270 billion dollars in the United States alone, tracking has become ubiquitous yet largely invisible. This paper analyzes how modern self-tracking technologies, primarily controlled by mega-corporations, facilitate an escape from existential responsibility while creating new forms of alienation. Drawing from Marx's theory of alienation, Sartre's concept of bad faith, and Russell's critique of passive leisure, I argue that the technological conquest of the epistemic self undermines authentic decision-making. It also contributes to a profound power imbalance between individuals and corporations, and incentivizes broader societal passivity.

Introduction

In this paper, we will consider the epistemic self as the collection of knowledge relating to the individual. We will see how individuals in the modern world are constantly engaged in the production and self-extraction of such knowledge even if they are not aware of its contents. The vast majority of data collected on and by the individual is in fact hidden from said individual. Our analysis of self-tracking will therefore focus on this hidden epistemic self. The majority of this self-data is used by mega-corporations to make choices for you and influence your environment, behavior and beliefs. The result is that we become ever less in control of ourselves.

The aim of this paper is to apply three critical philosophical frameworks—Marxist, existentialist, and Russellian—to analyze how self-tracking technologies in their modern form facilitate, incentivize and manufacture bad faith, alienation and passivity with major detriment to both the individual and society. We will start by analyzing the nature of self-data, providing the necessary conceptual groundwork, and then dive into the philosophical frameworks.

The Nature of Self-Data

Let us analyze the current realities of self-tracking. While we live in the most quantified era of human history where big data reigns supreme, the self-tracking most people consciously engage in is fairly limited. This is especially clear when one contrasts it with the wealth of hidden data that is collected on individuals in the contemporary world. While a typical individual might think they only self-track fitness, sleep and other metrics, they in fact track much more. Data generated by and on the individual includes location tracking, browsing history, online behavior, clicking tendencies, device type, time tracking, attention spans, hobbies and interests, restaurant reservations and more.

According to Elea Feit, "Most companies are collecting data . . . on all the places that they touch customers in the normal course of doing business."[1] This data is either hidden or presented in a form far better suited to large statistical inference models than to human analysis. This is to say that much of the surveillance data pertaining to marketing preferences, time spent on webpages and more is not open to the user. Additionally, the data that is available—through services like Google's takeout and other "download your data" schemes—is not easily decipherable because of its scale. A human cannot read gigabytes of data, but a machine will churn through it with the end goal of optimization and growth. It is in the interest of companies to gather as much data as possible as they are much better equipped to analyze it. For the consumer, limited resources and time ensure that she will not be able to independently glean meaningful insights from most data that has been collected on her. Tracking of the self in its modern ubiquitous form is then not created for the benefit of the consumer, as she is neither privy to its contents nor active in its analysis.

Moreover, this self-tracking infrastructure creates a profound power imbalance between individuals and corporations. The accumulation of vast quantities of personal data in corporate hands represents not just an information asymmetry but a genuine power differential. Large technology companies leverage this data to predict and influence behavior, effectively exercising a form of control that individuals cannot meaningfully resist on their own. As this data accumulation accelerates, corporate power grows exponentially while individual agency diminishes proportionally. This dynamic bears a striking resemblance to what Marx observed in industrial capitalism where the means of production were concentrated in even fewer hands. Except now, it is the means of prediction and influence that are being consolidated.

Conceptual Groundwork

Let us begin by defining some relevant existential terms. One aspect of existentialism is the critique of the tendency to let "others decide our lives for us."[2] For an existentialist, this need to conform constitutes "a manifestation of inauthenticity or self-deception because it shows how we are unwilling or unable to face up to the freedom and contingency of our condition."[3] We are often inclined to deny either the real constraints of physics, society and ability that make us feel powerless, or our ultimate, radical freedom to choose our actions. This freedom can be seen as a so-called burden of responsibility, as it means that we are ultimately responsible (perhaps even to blame) for our actions. To have bad faith from an existentialist point of view is to deny this reality. In this paper, I focus on bad faith as the series of "strategies by which we deny our existential freedom and our moral responsibility."[4] By this definition, anything that leads us to become ever more subordinate to external agents is an instrument of bad faith. The Stanford Encyclopedia of Philosophy explains:

> When the existentialist refers to feelings of "nausea" (Sartre), "absurdity" (Camus), "anxiety" (Kierkegaard), "guilt" (Heidegger), or "mystery" (Marcel) they are describing uncanny affects that have the power to shake us out of our complacency, where the secure and familiar world breaks apart and collapses, and we are forced to confront the question of existence.[5]

Thus the strategies of bad faith are strategies to avoid feelings of nausea, mystery and absurdity brought on by the uncomfortable nature of reality. We will see that tracking of the epistemic self provides us with refuge to offload our decision-making in order to absolve ourselves of this burden of responsibility. Furthermore, we will see that this inauthenticity and passivity is liable to create a world with reduced critical thinking and diminished social, political and economic engagement. Therefore, the blind march towards inauthenticity is negative from an existentialist individualist perspective and a societal one.

From a Marxist perspective, alienation occurs when individuals become estranged from aspects of their humanity that should rightfully belong to them. Marx identified several forms of alienation under capitalism, but two are particularly relevant to our analysis. First, "immediate producers are separated from the product of their labour; they create a product that they neither own nor control, indeed, which comes to dominate them."[6] Second, "immediate producers are separated from their own human nature; for instance, the human capacities for community and for free, conscious, and creative work, are both frustrated by contemporary capitalist relations."[7] These forms of alienation involve a problematic separation between a subject and object that fundamentally belong

---

1 Wharton, "Your Data Is Shared and Sold... What's Being Done About It?" Knowledge at Wharton, 2019, https://knowledge.wharton.upenn.edu/article/data-shared-sold-whats-done/.

2 Kevin Aho, "Existentialism," The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Summer 2023, https://plato.stanford.edu/archives/sum2023/entries/existentialism/.

3 Aho, "Existentialism."

4 Debra Bergoffen and Megan Burke, "Simone de Beauvoir," The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Fall 2024, https://plato.stanford.edu/archives/fall2024/entries/beauvoir/.

5 Aho, "Existentialism."

6 Jonathan Wolff and David Leopold, "Karl Marx," The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta, Metaphysics Research Lab, Stanford University, last modified December 21, 2020, https://plato.stanford.edu/entries/marx/.

7 Wolff and Leopold, "Karl Marx."

together.

Russell's critique of passivity provides our third theoretical pillar. In "In Praise of Idleness," Russell challenges the moralization of work, arguing that "a great deal of harm is being done in the modern world by belief in the virtuousness of work."[8] He observes that technological advances should have liberated humanity from excessive toil, allowing leisure to be "evenly distributed throughout the community" rather than "the prerogative of small privileged classes." Yet instead, we find ourselves trapped between overwork and passive consumption. For Russell, the increasing passivity of leisure—"seeing cinemas, watching football matches, listening to the radio, and so on"—represents a troubling social development. Within this framework, the tracking of the epistemic self can be understood as an intensification of the passive tendencies Russell identified. For Russell, active, creative, and socially engaged use of leisure forms the basis for genuine individuality and democratic participation.

Authenticity Analysis: Two Types of Burden Relinquishing

There are two main ways in which self-tracking enables technologies to allow us to relinquish the burden of our existentialist responsibility. The first is by reducing the need for self-knowledge. The second is by allowing us to offload value judgment determination—freeing us from the burden of constructing our own moral values. Though we will restrict ourselves to an analysis of inauthenticity and the resulting impacts on the individual and society, there are interesting models of cognition and selfhood. For example, the Extended Mind Theory or Daniel Dennet's theory of the self as a center of narrative gravity, where the delegation of cognition can be seen as a process of cyborgification. Though we do not have the space in this paper, these models offer a powerful direction for further analysis by revealing how the offloading of cognition may not only lead to inauthenticity, but could also have major implications for the philosophical self.

The first way that self-tracking enables technologies to facilitate bad faith is by reducing the need for self-reflection and self-awareness. Though self-tracking could theoretically have great positive effects on self-knowledge, in reality most self-tracking data is being used to reduce the amount of self-knowledge that one actually needs in day to day life. Since an app can take over understanding one's sleep patterns, or fitness levels, it is possible to forgo the challenging work of genuine self-understanding. Moreover, algorithmic recommendation systems like those deployed by amazon and tiktok relieve us of the burden of reflecting on what we want to buy or what we are interested in seeing. In some cases, the algorithm can be said to know the person better than they know themselves. In effect, when external agents keep track of self-metrics for us, we do not need to know ourselves intimately. Furthermore, since these algorithmic nudges predispose us to certain choices and incentivize impulsive decisions over reflective ones, we spend less and less time on each decision. In combination, these two effects allow us to run away from the realities of our condition. Thus, we effectively observe our pre-processed data rather than engage in active becoming – we are shielded from actively engaging in life and from feeling nausea and absurdity.

Through the passivity and self-alienation that comes from viewing ourselves through

8 Bertrand Russell, *In Praise of Idleness and Other Essays* (Routledge, 2004), 3.

quantification, we are never forced to "confront the question of existence" and as such, can continue living in bad faith.

The second major way that self-tracking technologies lead to bad faith is through removing the need to construct our own moral values and make authentic ethical judgments. In every prescription and decision lies an inherent normative judgment. When we are prescribed the content that we interact with, the routes we take to travel and what we spend our time on, we no longer have to come to our own moral and ethical stances. The question of why we do things, what we should do and what is important is outsourced and ignored. The difficult work of self-determination is replaced by technological prescriptions that present themselves as objective, neutral guidelines. This gives substantial power to the companies that can make use of self-data, leading to the proliferation and entrenchment of the very neoliberal values that brought these companies to the top. In this way, self-tracking technology can allow the corporatocracy to take over the complex normative task of defining a "good" life, what bodies should look like, how productivity should be measured, and what constitutes personal wellness. This technological mediation of value judgments represents a profound abdication of personal responsibility. By increasingly allowing algorithms to determine the parameters of our ethical and personal choices, we distance ourselves from the fundamental existential task of creating meaning. We trade the difficult process of authentic self-determination for the comfort of pre-processed guidance, effectively neutralizing the transformative potential of ethical reflection.

Marxist Self-Alienation

Let us focus on the alienation that occurs when individuals become estranged from the products of their labor and from their fundamental human nature. In the context of self-tracking, this theory takes on new dimensions. Our digital activities constitute a form of unwaged labor that generates valuable data commodities. When we browse websites, use apps, or wear tracking devices, we produce data that is harvested by corporations without compensation. This is alienation from the product of one's labor—we create the data but cannot access, understand, or control it. The goods we construct are thus estranged from us. This extraction of self-data through hidden labor is de facto mandatory—it is the price of entry to the digital world and occurs within a profound power asymmetry. To avoid this dynamic is only possible by refusal to participate in the digital world.

This power imbalance creates a vicious cycle: tech companies start with advantages that let them collect our data, then use that data to cement more control. As they gather more of our information, we become more dependent on their services, steadily widening the gap between their growing power and our diminishing ability to resist this extraction. This is further complicated by our last Marxist concern. The alienation is not just from labor but from our epistemic self—our self-knowledge becomes expropriated, transformed into proprietary algorithms and predictive models that reflect us back to ourselves in fragmented, commodified forms. We experience a profound ontological reversal where part of our "self" is taken outside of us and housed in data centers.

Further, the vast accumulation of self-tracking data greatly widens the power imbalance between consumers and corporations. Companies like Google, Amazon, and Meta not only collect detailed data about our lives but use this information to actively shape

our choices, desires, and even our conception of the "good life." The corporate entities that control our self-tracking infrastructure are not neutral custodians of information but active shapers of social reality.

Passivity and Social Consequences

In 1935 Bertrand Russell lamented the increased passiveness with which the modern man spent his idle time: "The pleasures of urban populations have become mainly passive: seeing cinemas, watching football matches, listening to the radio and so on."[9] With the advent of self-tracking which allows for ever more personalized forms of entertainment, this trend of passivity has only increased. In the above section we saw how this passivity is intimately tied to inauthenticity with major detriment to the individual. Now, we will focus on the societal impacts of self-tracking technologies particularly as it relates to accelerating and continuing the march towards passivity and mindless consumption.

Knowledge of the self in the form of vast amounts of preference data can be seen to enable the modern phenomenon where algorithms optimize content for maximum engagement rather than substance or value. By tracking precisely what captures and holds attention, companies can deliver an endless stream of brief, dopamine-triggering content that gradually erodes attention spans and critical thinking abilities. This represents not just individual cognitive deterioration but a broader societal shift away from deep engagement and critical analysis.

The result is a potential manifestation of surveillance capitalism, where detailed behavioral tracking enables unprecedented manipulation of political behavior. Rather than having citizens actively participate in political dialogue and independently form political judgments, we allow our tracked data to increasingly determine our political exposure and, ultimately, our political reality. This represents a collective manifestation of bad faith, where society abdicates its responsibility for active political engagement in favor of algorithmically curated political experiences. The implications for democracy are profound—as citizens become increasingly passive participants, the foundation of democratic society is undermined.

Conclusion

When we descend into inauthenticity enabled by the tracking of the epistemic self, we lose out on the intangible and unquantifiable. Put differently, we lose "what counts but cannot be counted."[10] On the individual scale, this manifests itself as a descent into bad faith. On the societal scale, the result is social stagnation and a culture of passivity, where large corporations accumulate leverage and influence. This can be seen as part of the larger trend towards passive leisure described by Russell.

The problem of self-tracking and data abuse is fundamentally one of collective action. Because consumers are fragmented and there are no mechanisms for collective bargaining or collective decision-making, the power rests with large corporations. By herself, the individual is fundamentally powerless in comparison. Only through collective action, most easily achieved through policy reforms, can this change.

If individuals are to reclaim authentic selfhood in the digital age, they must first reclaim meaningful access to their self-data. This requires both technical access to the data they generate and the capacity to meaningfully interpret it—neither of which currently exists for most users. Increased transparency is necessary to raise awareness and facilitate structures and systems to ultimately allow individuals agency over their data. How this fundamental shift in data ownership will occur remains to be seen, but only through such reform can self-tracking technologies be transformed from instruments of corporate power into tools for authentic self-understanding and genuine autonomy.

9 Russell, *In Praise of Idleness,* 13.

10 Mireille Hildebrandt, "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning," *Theoretical Inquiries in Law* 20, no. 1 (2019): 96, https://doi.org/10.1515/til-2019-0004.

Works Cited

Aho, Kevin. "Existentialism." The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Summer 2023. https://plato.stanford.edu/archives/sum2023/entries/existentialism/.

Bergoffen, Debra, and Megan Burke. "Simone de Beauvoir." The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta and Uri Nodelman, Metaphysics Research Lab, Stanford University, Fall 2024. https://plato.stanford.edu/archives/fall2024/entries/beauvoir/.

Hildebrandt, Mireille. "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning." *Theoretical Inquiries in Law* 20, no. 1 (2019): 83–121. https://doi.org/10.1515/til-2019-0004.

Russell, Bertrand. *In Praise of Idleness and Other Essays*. Routledge, 2004.

Statista. "Online Advertising." 2023. https://www.statista.com/topics/1176/online-advertising/.

Wharton. "Your Data Is Shared and Sold... What's Being Done About It?" Knowledge at Wharton, 2019. https://knowledge.wharton.upenn.edu/article/data-shared-sold-whats-done/.

Wolff, Jonathan, and David Leopold. "Karl Marx." In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University, last modified December 21, 2020. https://plato.stanford.edu/entries/marx/.

Author Biographies

Zach Weis is a major in Philosophy & Religion and in Political Science. He is in
ed in ancient Greek philosophy and has a particular interest in the Greek Stoics.

Roger Lin is a senior Philosophy major at Haverford College.

Xuning Gao is a Philosophy student at Kenyon College with interests in epistem
philosophy of mind, and metaphysics.

Allan Cao is a senior undergraduate student at Case Western Reserve University
ing a dual degree in Philosophy and Computer Science. His academic interests in
network biology and the intersections of critical social/political philosophy and te
ogy.

Kabeer Haider is a second-year Philosophy student at University of Toronto. Th
very interested in analytic queer philosophy.

Sophia Decherney is a senior studying philosophy at Brown University. Her area
study include Philosophy of Disability, Stoic Ethics, and Metaphysics of Pain.

Brendan Murphy is a junior at BYU studying Philosophy and English.

Carol Michelle Beyda is a pre-law student at Northeastern University following
ticular interest in immigration law and ethics. She has worked with non-profit org
tions promoting global justice and has used her experiences to forward her philos
research.

Daniel Zheng is a senior at Brown University studying Modern Culture & Media
English. He is interested in critical theory, film studies, and cybernetics.

Ander Petri-Hidalgo is a Brown University Mathematics and Philosophy studen