compos mentis

compos mentis

Table of Contents

Table of Contents (cont.)

compos mentis

How the Self-Serving Attributional Bias Affects Teacher Pedagogy

Natalie Anderson

Eastern Michigan University

ABSTRACT

In this paper, I argue that the Self-Serving Attributional Bias, which involves attributing success to internal features about oneself and failure to external factors, negatively affects the teaching practice known as "Closing the Loop" wherein teachers recognize issues in student learning and then adjust their approach to increase student learning. This practice requires teachers to recognize that the learning activity is to blame for both failure and success in student learning which the SSAB might prevent teachers from noticing. I detail what can be done to address this bias. I suggest that institutional structures and measures that are intended to serve as a check on biased reasoning would be an effective way to help reduce the negative impacts of the SSAB. I explore the possibility of increasing accountability in order to bypass the SSAB, but conclude that this practice would be counterproductive. Finally, I suggest that we need to eliminate the evaluative nature of standardized testing in favor of a localized model of assessment which would be a beneficial structure to help mitigate the negative effects of the SSAB on the teaching and learning environment.

In this paper, I argue that the Self-Serving Attributional Bias, which involves attributing success to internal features about oneself and failure to external factors, negatively affects the teaching practice known as "Closing the Loop" wherein teachers recognize issues in student learning and then adjust their approach to increase student learning. This practice requires teachers to recognize that the learning activity is to blame for both failure and success in student learning which the SSAB might prevent teachers from noticing. This becomes problematic when teachers attribute the failure of a lesson to deficits in the students rather than recognizing the parts of the lesson that are in their control which they could change to better teach their students.

I detail what can be done to address this bias. One of the solutions involves educating teachers about this bias so that they can avoid it. Unfortunately, we cannot rely on teachers to recognize and monitor this bias on their own. I suggest that institutional structures and measures that are intended to serve as a check on biased reasoning would be a superior way to help reduce the negative impacts of the SSAB. I explore the possibility of increasing accountability in order to bypass the SSAB, but conclude that this practice would be counterproductive. Finally, I suggest that we need to eliminate the evaluative nature of standardized testing in favor of a localized model of assessment which would be a beneficial structure to help mitigate the negative effects of the SSAB on the teaching and learning environment.

TEACHERS ARE SUSCEPTIBLE TO THE SELF-SERVING ATTRIBUTIONAL BIAS

What is the Self-Serving Attributional Bias?

The Self-Serving Attributional Bias (SSAB) describes our folk psychological tendency to attribute our success to stable features of ourselves that we can control (e.g. our own efforts, or qualities such as persistence and diligence) and attribute our failures to features that we cannot control (e.g. bad luck or bias against us) (Spaulding 2018, 49). By invoking this bias, we feel good about our successes and brush our failures off onto other factors.[1] One example of this is a

---

1. Shannon Spaulding suggests that this bias is a sign of healthy psychological functioning (Spaulding 2018, 49). The claims made in this paper are to demonstrate that this bias, although helpful in some cases, can be harmful in others.

student attributing their good grades to their hard work and their bad grades to bad luck or flaws in their teacher. Knowing that humans have this bias, we can use it to explain and predict what others will do in response to success and failure. Through this recognition, we also can reflect on our own thinking and change our approach if necessary. I will use this idea to demonstrate how we can predict what teachers might do in response to the failure of a learning activity which can negatively impact the continual improvement of pedagogic practices.

What is Closing the Loop?

Closing the loop is a common concept in teaching and learning. This entails a teacher deciding what they want their students to learn and designing a learning activity with this goal in mind. The teacher then implements that learning activity and provides an evaluative tool (e.g. a test, composition of an essay, etc.) which allows students to show their mastery of the learning target. Teachers then assess these completed assignments and from them gauge how well the learning activity helped students to meet the lesson objective. From this evaluation of data they adjust the next lesson accordingly to better address the learning target. This adjustment might be specifically for that particular lesson for the next iteration of the class, or different, future lessons in the same class depending on if it was the kind of activity that was problematic or just problems with the specific one done. This process has the goal of creating better learning activities which helps the teacher's practices improve over time (Arcario et al. 2013, 21). Closing the loop works across disciplines and at a variety of levels (i.e. elementary schools, high schools, or colleges). All teachers design learning activities and provide assessments where they assess whether or not students learned the content that they were supposed to. The challenge is then applying that data to adjust future lessons in a productive manner.

The Self-Serving Attributional Bias Could Hinder Closing the Loop

Although closing the loop is a good way for teachers to recognize problems in their approach and adjust their lessons accordingly, the Self-Serving Attributional Bias could prevent teachers from recognizing the deficiencies in their lessons. Teachers might attribute the failure of students to learn from a lesson to external factors such as the students not paying attention, not putting in effort, or the students not being smart enough to understand what they were trying to teach

them. This external attribution places the responsibility on the students when the teacher should recognize that they are responsible for fixing the broken activity. The SSAB sometimes prevents this recognition of responsibility.

The recognition of responsibility is hard for people because of the emotional component of the SSAB. Accepting responsibility comes with the psychological cost of blaming oneself for the failure. The shame that comes with failure causes people to become defensive, which gives rise to the SSAB. Spaulding suggests that the SSAB is actually a sign of healthy psychological functioning, so it's not something that should be completely eliminated (Spaulding 2018, 49). The problem is that sometimes this self-preservation strategy gets in the way of good teaching. We don't want teachers to blame themselves or consider themselves a failure, we want them to recognize their power to change the situation to better teach their students.

WHAT CAN BE DONE?

Teachers Could Educate Themselves to Avoid this Bias

One way for teachers to avoid this bias is for them to educate and monitor themselves in order to avoid it. Miranda Fricker (2003) and Jose Medina (2013) argue, independently, that people should train themselves to avoid biased reasoning and compensate for it in a variety of ways so that it doesn't affect the situation negatively (Fricker 2003, 154-173; Medina 2013). This idea suggests that whenever teachers are evaluating a learning activity and they catch themselves placing the responsibility of the failure on the students, they should, after learning about the bias, recognize this problematic reasoning and change their mindset to place the responsibility back on themselves to adjust the learning activity. Teachers should train themselves to be vigilant about their thinking processes and avoid the harmful ones in order to help eliminate the impact of the SSAB in their teaching.

In order to recognize the SSAB and train oneself to monitor it, there must first be education about the bias. If teachers know about this bias and when they are likely to invoke it, they can come to recognize and avoid it. This education can come from a variety of sources such as research and targeted professional development.

Problems with Educating Oneself to Avoid the SSAB

One problem with education being the solution to the SSAB problem is that it relies on teachers to monitor themselves after they learn about the bias. Spaulding suggests that education is not enough to help people avoid biased reasoning. She says that this strategy requires individuals to police themselves or to recognize and compensate for their biases (Spaulding 2018, 89). This is problematic due to another bias that humans have known as naïve realism. Naïve realism is the tendency to think of oneself as unbiased or simply perceiving the world how it is (Spaulding 2018, 89). People often fail to recognize biased reasoning in themselves because of their tendency to think of themselves as unbiased. Spaulding says that Fricker and Medina's solution might work in some cases, but it won't be enough to recognize a majority of biased reasoning. This is because it will just be chance whether a person catches that they are using biased reasoning and corrects it in a particular situation. Patterns of biased reasoning are likely to go unnoticed which is problematic and suggests that this solution won't solve the multitude of problems that we need the solution for the SSAB problem in teaching to solve.

Spaulding offers a second problem for the self-monitoring solution to bias: when people reach cognitive load they rarely invoke newly learned strategies. Worse, novel strategies are not invoked when the situation is so familiar to individuals that they are running on familiar, implicit scripts. Because of this, Spaulding suggests that although there are effective ways to educate oneself, this alone won't be enough to avoid the problematic reasoning that comes with biases such as the SSAB (Spaulding 2018, 91).

Spaulding suggests that the best way to solve the problem of biased reasoning is to implement structures and institutional measures that are intended to serve as a check on biased reasoning. These structures take the responsibility of monitoring biased reasoning off of teachers who are already cognitively overloaded and puts the responsibility on an external, objective structure. The intention of this is to prevent the problems that come along with self-policing one's own biased tendencies (Spaulding 2018, 91-92).

Increasing Accountability as an Institutional Measure to
Reduce the SSAB

One pragmatic solution to the SSAB problem in teachers is increasing accountability in the form of standardizing materials used in schools. Historically, there have been two options in order to raise the standards in teaching. The first option is to improve the level of education among teachers and the second is to "establish bureaucratic control whereby supervisors ensure that performance standards were achieved" (Evans, Lester, & Broemmel 2010, 183). The second is the one that is most widely utilized and with this idea came the inclusion of standardized testing to increase accountability. Spaulding's concern about relying on education, as the first option suggests, means that we should favor the second one, since the standardization of materials is a structure that might be beneficial to fixing problems in teaching.[2]

Raising accountability by implementing standardized testing could be thought to be a useful way to monitor the SSAB. When students don't do well on standardized tests, or a particular aspect of the tests, teachers would feel pressured and obligated from external sources to adjust their approaches due to the penalties that could be invoked should their students continue to do poorly. This could help them to become more reflective on their failed learning activities as to why they actually failed to produce adequate student learning. It would also allow them to reflect on what is in their control about the lesson design or their teaching methods in general, thus bypassing their tendency towards biased reasoning.

Problems with Increasing Accountability

(1) Policy Makers Are Susceptible to the SSAB. Along with the push for accountability in the form of standardized testing came an increased focus on what teachers and schools needed to do to improve. If schools were failing to improve their test scores, policy makers placed the blame on teachers and principals. If schools failed to implement stringent measures in order to improve their students'

---

2. This representation of standardization is intended to be the most charitable version of the intentions of policy makers. I note that this might not be the most accurate interpretation of policy makers' intentions and behaviors, but this is the most charitable way that it might be justified.

learning, there would be detrimental consequences such as reduced funding or being shut down (Ravitch 2014, 3).

Placing the blame on teachers and schools for their students' failure to do well on the tests that the policy makers mandate demonstrates the SSAB in the policy makers. Whenever decisions were being made to implement even more standardized testing, it was because the blame for the failure of schools was placed on teachers and principals (Ravitch 2014, 3-4). What policy makers didn't take into account was the harmful nature of standardized testing. Not only is this form of testing filled with implicit biases that favor middle to upper class students (Popham 1999, 8-15), it also increases stress in teachers and schools, making them less effective. In addition to this, it takes away time from valuable learning experiences and focuses on test prep which is mistaken because students would learn more if they were given more of these things, not less.[3] The responsibility for the failure of the students to achieve proficiency on standardized tests should fall on policy makers, but this is not what happens. If policy makers didn't externalize the failure of students onto teachers, then perhaps they would recognize that there are problems with the policies that they are invoking. That is, if policy makers were able to recognize their own implicit SSAB, we may have a better system of accountability all around. This is not happening, and students, teachers, and schools are suffering.

(2) Teachers Are Susceptible to the SSAB Once Again. One of the big problems with using standardized tests for accountability purposes is that even if they did help teachers bypass their SSAB in their creation of lessons, it causes a much more problematic SSAB in them. Imposing an external standard on teachers invites their SSAB, since all problems with student learning could be associated to the tests, and much of this responsibility is correctly externalized. Because of all of the problems with standardized testing, it is easy to see how teachers might feel as though these tests are responsible for many problems in teaching. Whenever the tests are themselves embedded with biased reasoning, curriculum is mandated, and their abilities are put into question, teachers will certainly resort to placing the responsibility of their failures onto the external standardized tests, whether justified or not.

---

3. These problems with standardized testing are widely acknowledged. I provide the detailed arguments in an expanded version of this paper, but due to space constraints, they are not included here.

Because of this new SSAB in teachers, any chance of them recognizing their failure to close the loop in their teaching will be overshadowed by their newfound loathing of the learning "experience" that they are now forced to endure. Although much of the externalizing done by teachers due to standardized tests is rightfully done, some of the problems with student learning are the teachers' responsibility to fix and the SSAB could prevent them from recognizing those things. If teachers blame the tests for their students' failure to perform or learn from their teaching methods, they aren't internalizing the need to improve student learning and are likely to externalize even more than if they weren't being held accountable for their students test scores.

HOW TO REDUCE BIASES IN THE TEACHING AND LEARNING ENVIRONMENT

Longino's Theory of Objectivity

Helen Longino, in her book *Science As Social Knowledge,* offers her theory of what makes a science objective. This theory of objectivity is offered specifically for scientific practices, but it is also useful for other kinds of endeavors. She offers four criteria that a science must meet to be considered objective and she says that methods of inquiry are objective to the degree by which they allow for transformative critique (Longino 1990, 76). A field of inquiry must not just listen to criticism, it must also apply this criticism in productive ways to make itself better. Her four criteria are (1) there must be recognized avenues for criticism, (2) there must be shared standards among individuals working on a project, (3) there must be community response to criticism, and (4) there must be equality of intellectual authority (Longino 1990, 76-79).

Longino's first criteria for objectivity requires that there be some way for presenting ideas so that they can then be critically evaluated by others. She suggests that public forums such as journals and conferences are crucial to this practice as this allows ideas to be shared to a wider audience, thus allowing for more ways in which an idea could be put up to scrutiny. The more people working on an idea, the more objective it will be (Longino 1990, 76). The second criteria requires that the members in a community abide by the same rules and standards. This allows for people with different viewpoints the opportunity to say useful things about the proposed ideas since there is a similar foundation from which they are

working from (Longino 1990, 77). The third criteria is the community who presents the idea should remain attentive to the criticisms offered by other people and communities and they should work to dislodge their complaints. This work might produce a change in the theory, or it might produce a better supported theory. This criteria requires that people are listening to others and working to make their theories better in response to the critique (Longino 1990, 78). Longino's fourth criteria for objectivity is that there should be equal intellectual authority among qualified practitioners. This means that those who have power shouldn't automatically win the fight and it requires that those voices which are often marginalized be heard and taken seriously (Longino 1990, 78).

These criteria can be met to different degrees by different sciences, and the more that a community abides by these criteria, the more objective that the ideas from this community will be. For standardized testing to be viewed as an objective way to judge teachers and schools and improve education, the community which put this forward ought to abide by Longino's criteria as well. If they do not, then we cannot consider it an objective measure. The standardized testing paradigm, including the design of tests, interpretation of results, and so on, fails to meet even the most minimal requirements for being objective set out by Longino, therefore it should not be considered as an improvement to education without major improvements.

The most important reason that standardized testing fails to meet Longino's criteria is that educators aren't included in the decision making about standardized testing even though they are the ones who the tests and policies impact. Educators are experts in student learning (i.e. the ones who have the greatest experience and training in measuring student learning), and they are constantly arguing for less testing by citing the harmful effects that it has on student learning. Although this is the case, they are not being listened to, rather the policy makers remain steadfast in their intense accountability approach. This demonstrates that the policy makers are failing to give equal intellectual authority to qualified people and also that they are not responsive to the critiques that they are getting regarding their tests and polices.

Although by utilizing Longino's theory we could improve the implementation of standardized testing, this doesn't mean that it should be the tool selected in order to improve education for students and/or reduce problematic SSAB on the part of teachers. If we started following Longino's suggestions regarding the kind

of community of practice required for objectivity, the amount of criticism on the practice would likely lead to a reform which would reduce or eliminate the use of standardized testing, especially the use of them as an evaluation of teachers and schools. Since these two things are likely the case, rather than considering further what Longino can do to make standardized testing more objective, I will move up a level and consider how Longino could make all of the decisions made about education more objective.

How "Longino Style" Objectivity Could Improve the Teaching and Learning Environment

Longino's theory of objectivity demonstrates problems with the way decisions about education are currently being made. The biggest problem is that educators are not being taken seriously in decision making, which is harming them and inadvertently harming students. The denigration of the profession is causing unnecessary harm and increases the chances that seasoned educators will leave the profession due to the lack of respect they are receiving. Newer educators are sometimes deciding that teaching is not for them. As well, due to the demanding and demoralizing nature of the profession, new and potential teachers sometimes shy away from completion or using their degree. I suggest that these things would not be as pervasive if Longino's theory were utilized when making decisions about education.

Currently, policy makers utilize their power to push their agenda and sometimes educators might be listened to, but it is obvious that this listening isn't widespread enough to enact any meaningful change in the policy makers' decisions. Arguments against standardized testing have been numerous, yet the amount of testing and accountability continues to increase. This can be attributed to the SSAB found in policy makers.

Although Longino's theory isn't currently being utilized, if it were, one can see how this could improve the education system for all stakeholders. The first thing that needs to be done is remove the emphasis of the evaluative nature of standardized tests. This is in line with the complaints from educators and demonstrates Longino's requirement that the decisions about education be open to transformative critique. If standardized tests are included at all, they should be used merely as an assessment for the teachers' and schools' use in order to target areas where their students need improvement. This practice should not include

any kind of punishment for teachers, rather they should work in groups of teachers from the school to brainstorm and work to implement better teaching practices. If teachers participate in the development of teaching and learning practices and the evaluation thereof, it is not (as) possible for them to externalize the results of student performance. That is, it minimizes the opportunity to evoke the SSAB.

Standardized testing could even be completely eliminated in favor of a localized assessment plan where the educators within the school districts determine how to assess their students' learning and then from that assessment data, develop and implement an improved approach to target those areas which need improving. This local approach would allow educators to focus on the needs of the students within their schools without requiring that these students fit into a generic model of what a student their age should be. Those children who live in poverty need different things than those students who live in wealth and a localized assessment and improvement model would allow for this to be taken into account. This model would provide a way to monitor the SSAB in teachers with regard to lesson design.

If everyone adhered to a standard of social objectivity in decisions about education, many of the problems in the teaching and learning environment would be solved, or at least improved. With communities of practice, teachers would recognize when their SSAB is getting in the way of improvement of their pedagogy, since they would have a community who is working together to recognize actual reasons for student failure. This collaborative endeavor would prove beneficial for test creators (should we keep standardized tests around) because their implicit biases would be easier to take notice of and change their approach to the creation of test questions. Finally, policy makers, if they listen to those who are in the classroom every day, would be able to implement policies which would actually improve the schools in the country like they claim that standardized tests mixed with intense accountability, would do. Obviously implementing Longino's standards of objectivity would minimize only some of the current problems with the educational system, but I think that we should start there.

CONCLUSION

Longino provides us with a new way to think about decision making in education and I suggested in this paper that this model could help mitigate the

SSAB in teachers and policy makers. Although I only addressed one type of bias, I believe that Longino's model would be usful for many different things within the field of education. There is still further research that could be done to address how else Longino's model could improve the teaching and learning environment and this is just the first of many projects that could be done around this idea.

REFERENCES

Arcario, Paul, Bret Eynon, Marisa Klages, and Bernard A. Polnariev. 2013. "Closing the Loop: How We Better Serve Our Students through a Comprehensive Assessment Process." *Metropolitan Universities* 24 (2).

Fricker, Miranda. 2003. "Epistemic Justice and a Role for Virtue in the Politics of Knowing." *Metaphilosophy* 34.

Evans, Katherine R. Jessica Lester, and Amy D. Broemmel. 2010. "Talking Back to Scripted Curricula: a critical performance ethnography with teachers' collective narratives." *Power and Education* (2).

Longino, Helen E. 1990. *Science as Social Knowledge.* Princeton University Press.

Medina, Jose. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. Oxford: Oxford University Press.

Popham, James W. 1999. "Why Standardized Tests Don't Measure Educational Quality." *Educational Leadership: Using Standards and Assessments* 56 (6).

Ravitch, Diane. 2014. *Reign of Error.* Vintage.

Spaulding, Shannon. 2018. *How We Understand Others: Philosophy and Social Cognition.* Routledge.

compos mentis

Murdoch's Vision: Original-M vs. Brain-of-M

Betty Barnett
Vanderbilt University

ABSTRACT

In her book, *The Sovereignty of Good*, Iris Murdoch argues that the moral philosophy of her day overemphasizes, to a fault, the role of choice in ethics. Instead, she says, vision—which is the moral perception of, or attention toward, things other than oneself—ought to be the region of focus in ethical inquiry. However, although the vision that Murdoch describes is of supreme importance ethically, it is not for the reasons that she claims. Rather than vision being an isolated object of ethical importance, as she sees it, its salience is completely dependent on the potential it holds to affect conscious beings through influencing action.

In her book, *The Sovereignty of Good*, Iris Murdoch argues that the moral philosophy of her day overemphasizes, to a fault, the role of choice in ethics. Instead, she says, *vision*—which is the moral perception of, or *attention* toward, things other than oneself—ought to be the region of focus in ethical inquiry. However, although the vision that Murdoch describes is of supreme importance ethically, it is not for the reasons that she claims. Rather than vision being an isolated object of ethical importance, as she sees it, its salience is completely dependent on the potential it holds to affect conscious beings through influencing action.

Murdoch classifies those philosophers who tout an erroneously choice-exclusive view of ethics as *existentialist-behaviorist-utilitarians* (EBUs). In *The Sovereignty of Good*, she designates a single philosopher, Stuart Hampshire, to serve as the prime example of an EBU, drawing from his work in order to critique the EBUs' views on moral agency. According to her, EBUs like Hampshire are wrong in claiming that "'good' is a function of the will," and hence that ethics wholly revolves around choice (Murdoch 2013, 4). This mindset leads to a form of moral evaluation that "is and can only be concerned with public acts," entirely neglecting one's inner state, which Murdoch holds to be vital (Murdoch 2013, 9).

This failure to factor the internal life into ethical considerations, however, is intentional on the part of EBUs, and stems from another core tenet of theirs: the *genetic analysis* of mental concepts, which is the notion that "what identifies [] emotion[, for example,] is the presence not of a particular private object [in the mind], but of some typical outward behavior pattern" (Murdoch 2013, 13). This perspective, Murdoch explains, leads to an unrealistically narrow conception of the individual's inner life, in which all that exists there is an isolated will; likewise, it lends credence to the idea that, again, freedom pertains merely to choice. By ignoring the inner activity of vision—in which one, over time, develops a view of one's environment—EBUs therefore hold that moral meaning exists only when in union with action.

After articulating her empirical, philosophical, and moral objections to the EBU philosophy, then, Murdoch proceeds to present an analogy, seeking to illustrate why her view of human freedom is a superior one. In this famous thought experiment of hers, a mother, M, "feels hostility" toward her daughter-in-law, D, due to her perception that D is pert, rude, juvenile, and so on (Murdoch 2013, 16). Despite M's negative feelings toward D, however, M "behaves beautifully" toward

her, "not allowing her real opinion" to show through in any way (Murdoch 2013, 17). Eventually, D immigrates elsewhere, and M begins to question her long-held opinions concerning D. Wholly independent of any action or influence on the part of D, then, M changes her view of her daughter-in-law, now construing the many traits she used to abhor as endearing. All the while, however, M's outward behavior has remained consistently positive, without even a trace of animosity directed toward D.

In relaying this tale, Murdoch attempts to show that M's conceptual shift, despite not being a matter of observable behavior, is a morally meaningful activity, and that, accordingly, ethical value can sometimes be unearthed in vision alone. This interpretation is problematic, though. Consider, for example, a novel continuation of this same story: after perceiving D in a new light, M dies. Since D already left the continent some years before, M never got a chance to see D again, and so never expressed her inward transformation of opinion concerning D to anyone. Without ever having exhibited any observable behavior indicating her shift in perception, therefore, this change in M is not morally meaningful. If, however, M had communicated her newfound feelings to another individual, A, then I would agree that M's shift was morally meaningful since her confession would have, at the very least, affected A, and, quite possibly, D, as well, assuming an open channel of communication between A and D. However, even in this instance, M's relaying of her feelings would have constituted an empirically observable action, and hence would no longer count as strictly internal in nature.

To best understand why this is the case, then—why M's completely silent change in vision, followed by her abrupt death, lacks ethical import—one must consider yet another take on the M&D example. In this dystopian version of the analogy, M's brain, BoM, resides in a self-maintaining vat, which has been forever abandoned by scientists. BoM, harvested in the year 2150, retains the same consciousness it had when it was still inside of M's skull, minus, of course, incoming sensory perception. With all of its memories still intact, then, it is free—for the rest of time—to form and re-form opinions concerning D. Given its utter disconnect from any other forms of consciousness, however, no moral meaning could possibly be assigned to BoM's endless musings. This amoral status can be explained by the fact that BoM, being confined to its solitary vat, cannot ever again effect change in the world (assuming, of course, that any hope of technical progress—of BoM's someday accessing a conduit of information—is in vain).

The lack of agency exemplified here is fully attributable to the one difference between BoM and original-M (OG-M): BoM lacks a body, an outlet for action, whereas OG-M does not. Even if BoM were to be rediscovered, and then carried around in its vat—a move that, were BoM to be aware of, would be against its unable-to-be-articulated wishes—perhaps even to D's own apartment, BoM still would not be capable of effecting change. Its carrier, in this particular case, would be effecting change, prompting an emotional reaction from D upon seeing her mother-in-law's voiceless brain, for instance, but this and any other outcomes would be entirely unrelated to BoM, since BoM did not choose to be carried about in its vat.

How, then, does OG-M—a woman who keeps her thoughts on D entirely within her head—differ from BoM ethically? In other words, why is it that OG-M's opinions of D matter, whereas BoM's do not? The answer to this question, it seems, exposes the primary flaw in Murdoch's reasoning on the ethical primacy of vision: it certainly appears ostensibly as if OG-M's change in attitude toward D matters morally. However, in light of the aforementioned comparison between OG-M and BoM, it becomes clear that this change in attitude is meaningful solely because of its potential to effect change. If, as in BoM's situation, this potential is entirely eliminated, then the ethical meaning of the attitude change is lost, as well.

In this same vein, then, it goes to say that if even a remote chance existed that BoM could someday be synced with a computer—and therefore finally able to communicate again, to effect change in the world—then its *vision* would certainly matter. The ethical salience of BoM's musings would rest on the potential that they hold to dictate future action—in this particular case, action in the form of communication. Likewise, the only reason that OG-M's thoughts matter is because of the potential that *they* carry. OG-M's change of attitude, even if left unexpressed for a time, never loses its potential—and, by extension, its moral meaning—so long as she, the messenger, is alive and breathing. The hypotheticals, the ways in which her attitude could effect change—leading her to treat others, including D, with more kindness, for instance, or leading her to express her opinions, thereby influencing someone else—do not cease to exist when she decides to keep things to herself. Rather, because she is a living human, her active vision will always carry ethical meaning, since it is impossible to know whether, at any future point in her life, this vision will influence her enactment of agency.

Murdoch, however, seeks to separate the cause and its effect: she wants vision to matter always, not because of its potential, but rather because of its intrinsic value. This explains why Murdoch would likely argue that OG-M's forever unexpressed change of attitude toward D, sealed by her own death, is morally meaningful. She sees OG-M's vision as meaningful *per se*, whereas I see it as meaningful in a contingent sense. OG-M's change in vision is, in some ways, like Schrödinger's Cat: it both matters and does not matter, and exists in this state so long as it has the potential to come in contact with the external world. If and when this change in vision does touch the world, its wavefunction finally collapses into a concrete state of value. If the change in vision does not touch the world, though—if, as with OG-M, its potential dies in conjunction with the agent herself—then its contingent value disappears. Altogether, then, these two counterexamples demonstrate that vision is important only insofar as it is capable of effecting change in the world.

REFERENCES

Murdoch, Iris. 2013. *The Sovereignty of Good*. New York: Routledge.

compos mentis

Coherent Infinitism as a Hybrid Model of Epistemic Justification

Luke Batt

Calvin University

ABSTRACT

In this paper I will be arguing for a new theory of justification as a hybrid of coherentism and infinitism which I shall call Coherent Infinitism (CI), with the goal of showing that we can have justified beliefs. This paper will be divided into sections. I will begin with a basic layout of both coherentism and infinitism as separate theories of justification. In the subsequent section major objections to both these theories will be analyzed, which make them implausible by themselves. The third section of this paper will focus on the nature of justification and the properties of the justification relation between epistemic beliefs. The next section will examine the basic justificatory structure of CI combining elements of the two aforementioned theories and integrating the findings of section three to show how we can have justified beliefs. Lastly, I shall investigate possible objections to CI not including skepticism and attempt to provide answers to them.

INTRODUCTION

In this paper I will be arguing for a new theory of justification as a hybrid of coherentism and infinitism which I shall call Coherent Infinitism (CI), with the goal of showing that we can have justified beliefs. This paper will be divided into sections. I will begin with a basic layout of both coherentism and infinitism as separate theories of justification. In the subsequent section major objections to both these theories will be analyzed, which make them implausible by themselves. The third section of this paper will focus on the nature of justification and the properties of the justification relation between epistemic beliefs. The next section will examine the basic justificatory structure of CI combining elements of the two aforementioned theories and integrating the findings of section three to show how we can have justified beliefs. Lastly, I shall investigate possible objections to CI not including skepticism[1] and attempt to provide answers to them.

SECTION 1: COHERENTISM AND INFINITISM
AS THEORIES OF JUSTIFICATION

Before any serious conversation about justification can be had, we must first discuss the Regress Argument. This argument claims that all beliefs stem from branched doxastic chains which are non-circular, and do not go on *ad infinitum* (meaning each chain is finite) and therefore, each chain must terminate in an immediately justified belief.[2] The infinite regress is typically seen as a defense of foundationalism which is the dominant view of justification among epistemologists. However, many philosophers who reject foundationalism have developed competing theories of justification, all of which get their start from rejecting a particular premise of the Regress Argument.

The theory of coherentism rejects the premise that doxastic chains are non-circular and therefore, cannot justify themselves. With this understanding, beliefs are considered to be part of circular doxastic chains which can justify themselves. While most serious coherentists would never endorse vicious circularity such as

---

1. To adequately engage with possible objections offered by skepticism would require a separate more focused paper, as such I shall not be speaking directly to these arguments in section five.

2. This version of the regress argument is paraphrased from a more complete version typically used to defend foundationalism.

"p therefore p,[3]" they would endorse something of this sort: A is justified by B, and B is justified by C and D, and D is justified by E which is in turn justified by A. In this way, A can be seen to indirectly justify itself. Different premises are then added to this inherent sort of circularity to support and expand the theory. Lawrence BonJour in his 1999 paper called *The Dialect of Foundationalism and Coherentism* outlines a particularly influential coherentist theory in opposition to the prevailing foundationalist theory. However, this theory is riddles with issues and hence is largely considered to have failed in gaining widespread support (Olsson 2017).

In 2014 Catherine Elgin presented a more plausible version of coherentism in her paper titled *Non-Foundationalist Epistemology: Holism, Coherence, and Tenability* (for the remainder of this paper all references to coherentism shall refer to the version presented here). In this paper, Elgin images all beliefs to be a part of a coherent set which is defined as:

(i)   s is a coherent set (CS) of S´s beliefs if and only if s is suitably comprehensive, consistent, cotenable and supportive.[4]

However, while Elgin does think a CS is necessary for justification, it alone is not sufficient. In this way we can imagine justification on a scale. The more a set of beliefs fits the criteria of (i) and therefore coherence increases, so too does our justification. Elgin uses the example of gathering several testimonies from unreliable sources. In this situation each testimony would be judged as unreliable. If, however, each made similar statements, (i.e. were coherent) they would all have a higher degree of justification when considered as a set. (Elgin 2014).

However, even after increasing the degree of justification we still do not have a justified belief, because all the sources could have agreed to share the same false testimony. Elgin argues that for a CS to be justified, its truth must be its best explanation. This means that if upon closer examination, all the sources´ statements fit into a CS, and the best explanation of them all fitting together and being suitably comprehensive, consistent etc. is that they are all true, then, we have justified belief. Moreover, if the CS is justified, then justification is also

---

3.   An exception to this would be the belief "I am believing something" which does appear to justify itself in this way.

4.   This is paraphrased from *Non-Foundationalist Epistemology: Holism, Coherence, and Tenability.*

conferred from the whole to each individual belief in that set. Note that the whole confers justification to the parts, but the reverse is not true. Each individual belief is, therefore, said to be initially tenable.

Let us look at an example of how Elgin's system might help us arrive at the belief:

(ii)   *It is Fall in Michigan.*

We may have an array of initially tenable beliefs: we see the leaves are changing from green to red, the corn and apples are becoming ripe and are ready to pick, and the weather has begun to get colder. Suppose we have certain beliefs about the time of year and have memories of people telling us about fall and the months in which it occurs. These beliefs can then be put together in a coherent set as described in (i), gaining coherence and a higher degree of justification. Then we would consider their truth as being the best explanation. Meaning that given our CS, it is more likely that (ii) is true, rather than the season being winter or even some skeptical scenario in which we are deceived into thinking it is only fall-like. Therefore, according to Elgin with our CS and truth as the best explanation, we are justified in believing (ii) and, by extension, are each other belief within our CS.

Having looked at coherentism as presented in its strongest form, we can now turn our attention to infinitism. In reply to the Regress Argument, infinitists deny the premise that doxastic chains are finite, arguing instead that they can indeed go on *ad infinitum*. Peter Klein is the only well-known staunch defender of infinitism. He lays out his defense of the theory in his 1999 paper- *Human Knowledge and the Infinite Regress of Ideas*. An important assumption that Klein makes in this paper is that we do not have immediately justified beliefs. Instead, he claims that only a reason can justify a belief. Thus, our doxastic chains must be infinite. In an attempt to prove his assertion, Klein states two important principles, the first being the principle of avoiding circularity (PAC).

(iii)   PAC: For all x, if a person, S, has justification for x, then for all y, if y is in the evidential ancestry[5] of x for S, then x is not in the evidential ancestry of y for S.

---

5.   Meaning the links in the chains of reasons that support beliefs

In so many words, Klein asserts that beliefs cannot justify themselves in a circular (or question-begging) manner. Recall the example of circularity endorsed by coherentists above on page 3. Klein would argue that A cannot be used to justify E, because E is in the evidential ancestry of A. In his paper, Klein does not defend PAC, instead, stating that it is an "obvious presupposition of good reasoning" (Klein 1999).

By stating this, Klein also seems to assume that justification is an anti-symmetrical relation, meaning that for a given x which has a relationship R to y, y cannot have that same relationship R to x. This holds true for relations such as the "taller than" relation, e.g., if Greg is taller than Rick, then Rick cannot be taller than Greg. By assuming justification to be similar to the "taller than" relation circularity also appears to be avoided and thus appeals to many people.

The second important principle Klein uses to defend his theory is the principle of avoiding arbitrariness (PAA).

> (iv)   PAA: For all x, if a person, S has a justification for x, then there is some reason, r_1, available to S for x; and there is some reason, r_2, available to S for r_1; etc. and there is no last reason in the series.

In order to fully appreciate this principle, we must first analyze what Klein means by an available reason. While Klein himself talks of a reason as being objectively and subjectively available, he himself claims there are many accounts by which these two conditions can be met. I will be using the following:

> (v)   A reason r is an available reason to S for believing p, if and only if r makes p more epistemically probably for S, and S would affirm r if asked[6].

The last important piece to understanding PAA, is to note why there must be an infinite chain of available reasons. Klein argues that if there is not, then we must arbitrarily stop the doxastic chain at a particular belief. It then seems highly subjective as to where an individual chain ends, and what would be considered immediately justified beliefs, something which Klein thinks should be avoided. Combining PAC and PAA, Klein concludes that justification must involve non-

---

6.   Here I would like to thank and credit Jeff Snapper with this definition of available reason, which I find to suitable and concisely meet the two criteria presented by Klein in his 1999 paper

circular chains of belief, with each belief being justified by a reason with no arbitrary stopping point, meaning the chains continue *ad infinitum*.

To get a firm grasp on infinitism let us once again consider (ii) above. The infinitist would claim they have infinitely many reasons available to believe (ii). They might even begin by listing the same propositions we did above. If questioned further regarding "the leaves are changing color" the infinitist may reply with something about how changes in temperature cause color change due to different processes inside the leaf, which she learned from testimony, and which he trusts etc. This sort of reasoning would continue in this way. Thus, the infinitist would claim to be justified in believing (ii).

SECTION 2: IMPORTANT OBJECTIONS TO COHERENTISM AND INFINITISM

The theories I presented above each have several important objections to them, making either seem rather implausible by itself. While the inherent circularity of coherentism does worry some epistemologists, as long as vicious circularity (as mentioned above) is excluded, this worry is not as significant as many others. The first major objection is often called the "coherent fairy tale," which argues that if justification only requires beliefs to be coherent and comprehensive with each other, then people would be justified in believing that suitably written fictions, such as Alexandre Dumas´s *The Count of Monte Cristo* are true.

Another objection to coherentism is this: by claiming justification requires truth as its "best" explanation, it appears to be highly subjective. For example, if Sadie and Ben were presented with information about climate change from various sources, they may disagree about which truth provides the best explanation for justification. Let us assume Sadie believes climate change is caused by human beings, while Ben believes climate change is caused by God as a punishment. In this case both use a different "best" explanation based on perceived truth to interpret the same body of information.

Coherentism also argues that all our beliefs are revisable, meaning nothing is unchangeable. Van Cleve raises an objection to this assertion. He claims there are some beliefs which are not revisable, and if they were, our world would look very different. Here Van Cleve is concerned primarily with fundamental logical laws, such as modus ponens or the law of non-contradiction. Additionally, even if Van

Cleve is wrong, this raises another objection: coherentism relies on a coherent set of beliefs to lend to justification. This holistic view requires all beliefs in the set to be mutually supportive. However, since our beliefs are subject to revision, a change in one belief may threaten the integrity of the whole web, by not mutually supporting seemingly unrelated beliefs in a different part of the web.

We shall now turn our attention to the problems facing infinitism. The first objection is against PAA, rejecting the claim that only a reason can justify a belief. Reliabilism and foundationalism will argue that having a reason be required to justify a belief is assuming that one must show they are justified. Rather, they would contend that we can be justified in a certain belief without showing it, by way of immediately justified beliefs. Take, for example, the fact I am writing with my pen. According to reliabilism/foundationalism, I need not provide a reason (and therefore show I am justified) to in fact be justified. My own sense perceptions (my pen-ish perceptions) immediately justify my belief that I am writing with a pen.

Another objection that is raised against PAA is called meta-justification. The idea behind this objection is that beliefs which have some property F are probably true, or at least acceptable. As such, they need not be shown to be true with an additional reason. An example of such a property F may be "I felt it with my hand." In which case, most things one feel with his hand are probably true and she needs not provide additional reasons. Another example may be testimony from a particularly trustworthy individual[7], such as a parent or doctor.

The finite mind objection raised against infinitism claims that, given the nature of our mind, doxastic chains cannot continue *ad infinitum*. Human minds are finite and, therefore, cannot consciously believe infinitely many things. However, according to infinitism, infinitely many beliefs must be available for justification. Since we clearly have many justified beliefs at any given moment, infinitism appears to be false. This is because it is not possible for a finite mind to hold infinitely many reasons for multiple justified beliefs at a particular point in time.

Another objection raised PAC, seems to assume the justification relation is anti-symmetrical. However, there are pairs of beliefs which appear to justify each other, thereby showing the relation is not, in fact anti-symmetrical. Take for example the beliefs God exists and angels exist. If we are justified in believing one, it appears that belief, then, would justify the other. For if we are justified if

---

7. This would of course depend heavily on the individual and how much they truly trust the testimony.

believing God exists, it would stand to reason that we are also justified in believing He created the angels. Likewise, if we are justified in believing angels exist as heavenly beings, we would be justified in believing their heavenly creator also exists. Therefore, it seems there are at least some instances of justification which are not anti-symmetrical.

A final objection against infinitism combines a temporal concern and a worry that justification is too easy. Let us say, for instance that I am standing outside at ten in the morning, and I see a car go past me at 100mph. Why do I not need something in that moment to justify my belief about the car? If all I need is an infinite chain of reasons, how could I have the time to justify my belief at ten if the chain of justification is infinite? If this concern is combined with Klein´s definition of available reason (v), justification becomes too easy. Take, for example, the belief "I am a genius." This is a belief that is not intuitively justified. However, according to infinitism one could very easily be justified in holding such a belief. She needs only begin to list off some premises: "I can speak three languages" or "I am good at organic chemistry" "I know 100 -1 and 99-1" etc. Then he might claim that infinitely many reasons are available if she were to sit down and think about it. Thus, making him justified in her belief. In this way the infinite chain seems to require little in the way of actualization or articulation of reasons to justify a belief.

SECTION 3: THE NATURE OF JUSTIFICATION AND ITS RELATION TO EPISTEMIC BELIEFS

Having laid out the groundwork for both infinitism and coherentism, there is one more matter to which we must attend before we can articulate the structure of CI. We must first define justification and what it means to say we have justified beliefs. Additionally, we must analyze and define the properties of the relation of justification between beliefs, as both of these concepts will undergird the justificatory structure of CI and play a major role in its overall structure.

Firstly, I shall argue that all justification is inferential, meaning that justification for a belief can only come from a reason. This is to deny we have immediately (or non-inferentially) justified beliefs. In stating this I must contend with two problems: various emotional states and basic sense perceptions. I would like to begin with a simple observation, that we can (and often do) live by things which we are not justified in believing. In response to the objection raised above, I would also like

to state that having a reason does not require us to articulate it to others. We can have a reason for a belief without showing it.

We as individuals have complex webs of beliefs,[8] which are built from the time we are born and never stop growing and developing. Beliefs become justified in so far as justification is conferred to them, as it emerges from the structure of the web as a whole. As such, each new piece of incoming information or "incoming belief" will have, as Elgin argues "initial tenability." This means the belief is there, but we are not yet justified in believing it, until the web is developed enough to include the concepts needed. Let us now examine how this sort of tenability and emergent justification works in each of the cases mentioned above.

I shall first begin with emotional states. A necessary distinction to establish is the difference between the act of "being angry" and "belief about being angry." A person may be in a particular emotional state, but that does not mean she is justified in believing he is in such a state. A baby may be "angry," but would not be justified in believing it is indeed angry. This is because its cognitive web has not yet developed the concepts needed to justify such a belief (namely the concept of anger). Likewise, when I was a teenager, I did not know what it felt like to be in love or to feel heartbreak. At the time I knew I felt something but had not yet developed my web to include such emotions. So, while I did feel those emotions, I was not justified in believing I felt them.

A similar line of reasoning can be used to argue against sense perceptions. Once again, since justification emerges from the structure of our web, it must be developed enough to encompass the belief and confer justification upon it. In this way, we may think of each sense perception, as with emotions, as initially tenable. I know that I see something or feel something, but neither of those beliefs could non-inferentially justify the belief "I am typing." Rather, some part of our web must help us make sense of "typing" and what that means. Let us say, for example, we were to give someone from the 16th century a cell phone. She would be justified in believing that she is talking and that she can hear someone, but she would not be justified in believing that she is talking on a cell phone.

As babies and toddlers, I think we have many initially tenable beliefs, which are not justified. Yet, as we grow and begin to construct our web, justification emerges. Without that piece we simply believe we feel or see something[9]. As

---

8.   This idea will be better established in section 4

9.   This could be taken even further as babies do not really understand what it means to "see" or

adults, this sort of background is so deeply ingrained that it seems obvious, and it may even seem silly to say what I explained above. Because our web is well developed enough that we know when we see something or something, we are able to properly identify it.

Now, we must address the objection that, in the cases above, it seems we must "show "we are justified. Yet, I do not think this is necessarily true. In this case, I think to show is nearly synonymous with articulate, and having a reason does not seem to be the same as articulating that reason. Our reason is our own deep cognitive webs which provide the underlying structure, which is not always articulable. Nevertheless, there is a reason for our beliefs. As I argued above, each belief needs to be justified by a reason. So now, we turn to what this justificatory relation between beliefs looks like. I believe this relation to be non-transitive, non-symmetric (without being anti-symmetric), and what I shall call "minimally circular."

To be transitive is to have such a relation as "greater than" meaning if x is greater than y and y is greater than z, then x is greater than z by a transitive property. However, I do not hold justification to be a relation having such a property. For example, let us imagine a disease called ROVID20. Let us assume I am justified in believing that ROVID20 is deadly. We then use this belief to justify the belief that ROVID20 can kill people, which we then use to justify the belief that some people may have died from ROVID20. If the justification relation is a transitive property, then we would be able to say "my justified belief that ROVID20 is a deadly virus justifies my belief that some people may have died from it. However, this seems like an abrupt jump without the second premise. We could imagine that ROVID20 is a deadly disease for sharks, but not humans, or that it is caused by a virus that was synthesized and known to be deadly but is sealed in a test tube and therefore not introduced to the public. It then appears that justification is a non-transitive relation.

The justificatory relation is also non-symmetrical, meaning it is not the case that for any two beliefs x and y, that the same relation must exist between x and y and y and x. It is important to note this is not synonymous with anti-symmetry, which claims that for every x and y, if x has a given relation to y, then y does not have that relation to x. An example of such a relation is *loves*. Luka can love Renee, and Renee can love Luka (in this case the relation holds in both ways), but Luka

---

"feel" anything, they act more on survival instinct.

can love Shawn and Shawn does not have to love Luka. In this second instance, a particular x and y do not have the same relation, but in the first example, for a different x and y, that relation does hold.

These two properties build into the last one, which I call minimal circularity. In any sort of theory in which coherentism is used, as a web or something of the like, in which beliefs rely on mutual support, there is at least some degree of circularity inherent. However, it need not be vicious in that a belief justifies itself. For example, let us consider the flat earther Isabelle. When asked how she knows the earth is flat, she launches into an explanation about infinite planes, non-existent gravity, and light reflections, which make the horizon only appear rounded[10]. However, upon further questioning regarding how she knows these things, she states "because the earth is flat." It would seem then, that the explanation of the premises comes from the conclusion itself, thereby creating a viciously circular web of beliefs.

A non-viciously circular web of beliefs must have what I call minimal circularity. Let us consider the economist Pearce. When asked why he believes capitalism is superior to many other economic systems, he begins by stating certain properties of each system, combining certain elements and propagating different outcomes, drawbacks etc. for each system. In this example Pearce's beliefs exist in a web and mutually support each other. However, if justification is non-transitive, being justified in knowing facts about capitalism and about socialism does not necessarily lend justificatory power to larger conclusions about either system. Likewise, if justification is non-symmetrical, then certain beliefs cannot justify themselves and must instead be considered as part of the whole. In this way beliefs can exist in the same web (or sub-web as we shall see) but need not be direct reasons for justifying other beliefs.

SECTION 4: COHERENT INFINITISM: BASIC GROUNDWORKS

Having established my views on justification, I will now lay out the basic foundations of CI. As I mentioned in section 3, each individual has a unique and complex web of beliefs, which grows and becomes more complex as we do, stretching and growing infinitely. In order to better explain the particular justificatory structure of CI, how justification emerges as the result of the particular

---

10. These ideas were paraphrased from The Flat Earth Society FAQ page.

structure of the web, and how all of our beliefs can be justified I will be introducing and using four terms: the center web, the inner web, the outer web, and sub webs.

Each individual web of beliefs begins at the center web. This web can be imagined as taking the shape of a circle and holds all of our most central beliefs, without which none of our other beliefs could be possible. Examples of beliefs found in this circle may include, I can trust my perceptions, I have a body, I am a being, our "sensus divinitatis", or even that God exists. All of these beliefs exist in a tight web mutually supporting each other and lend explanatory and justificatory power to one other. However, each of the beliefs stated above has its own sub-web, consisting of all the evidence[11] we have for that belief.

To better understand the importance of these sub-webs, let us take a specific example from above; "I can trust my perceptions." In this particular case, evidence would consist of all the times we saw something and were able to trust in it, i.e., all the times I have seen this item or that, this evidence may also include testimony from others agreeing with what we see and can easily stretch on to infinity, as we see infinity many things, or even the same thing again and again. However, we need not continue *ad infinitum*. Rather, we need only continue until we have established a coherent set, as defined by Elgin,[12] while remembering we could provide even more reason if needed. Yet, the sort of infinite reasons may start to look quite similar, at which point the difference between 1000 and 1001 reasons means little. This coherent set built from a number of reasons from our infinite chain to create the sub-web gives us reason to believe a center belief, such as "I can trust my perceptions." However, full justification does not come until the belief is considered in respect to and mutually supported by the other center beliefs.

The inner web is dependent upon and sprawls outwards from the center web. This is where the webs begin to differ more significantly from individual to individual. The beliefs found in the inner web shape how we view and interact with the world and others, playing a major role in our lives. Beliefs here may be, for example, I am a human, God created the world, I cannot breathe underwater or laws of logic to name a few. Each of these beliefs themselves have sub-webs as

---

11. This can refer a variety of different things, such as testimonies, various sense perceptions etc. which we believe to be evidence, whether true or false.

12. See Elgin´s 2014 paper mentioned above.

defined above. Again, these inner beliefs only attain full justification after being mutually supported not only by the rest of the inner web, but also the center web.

The last layer of an individual´s web of beliefs is the outer web. This web is perhaps the most complicated and variant, being highly individualized. There are infinitely many outer webs possible, each focusing on something particular. Examples of outer webs may include, philosophy, chemistry, languages, cars. These webs may, themselves, branch off even further, to things such as electrochemistry, organic chemistry and biochemistry. There are a few important characteristics to note about outer webs. They all rely on both the inner and the center webs for any sense of meaning or justification. Certain beliefs may appear in more than one web and need not have sub-webs to build up a degree of justification. Rather, justification is built by the very sprawling nature of the webs themselves with full justification, coming only from the particular outerweb as a whole fitting, being mutually supported by the combined structure and in proper relation to the inner and center webs.

Two final considerations must be made regarding CI as a theory of justification. I stated above that the inner and center webs must be structured and mutually supported in the right sort of way to arrive at full justification. To this end I am inclined to agree with Elgin that coherence is necessary for this, but not sufficient. I think truth as a best explanation is also necessary for full justification for any given belief. Also, I would like to note that these webs are by no means stagnate, and that any particular belief may change, and, depending on its location, may have a huge impact on our overall justification for any given belief. An example of this would be if a theist were to become an atheist, God´s existence as (a belief which lives at least in the inner web) would be ripped out. After this, the atheist must now begin to reconstruct the many aspects of the web which were damaged by the removal of this belief. Conversely, a belief, which was part of an outer web, may become more important and move into the inner web. I personally have experienced this movement more I have studied philosophy, as my worldview becomes increasing shaped by my philosophical studies.

While we have already answered many of the objections listed in section 2 with the structure of CI as well as considerations of the nature of justification from section 3, there are a few less obvious ones I would like to mention explicitly. Elgin argues that all ideas are reversible, even those which dwell in the inner and center webs. To rationalize this point, we need only look at history. For many years

people thought the earth was the center of the universe. When it was discovered that it was not, many people were forced to begin reconstructing many of their views about the world.

Another key objection answered by CI is the finite minds objection against infinitism. Klein´s initial response to this objection was to state that there need only be infinitely many reasons available to have justification. This weakening of available reasons seemed to make justification too easy, as was seen in section 2. My response to this objection begins as Klein´s does but adds another condition. I agree that infinitely many reasons are available, but to simply say "if I had enough time, I am sure I could come up with more reasons" seems irresponsible and indeed makes justification too easy. The condition that I add is that all the reasons we do state, out of the infinitely many possible, must be comprehensive, cotenable, mutually supportive and consistent. While it may seem arbitrary where we stop, as I stated above when discussing sub-webs, as the listing continues on, the reasons begin to sound very similar to each other and are no longer significantly unique. For instance, the testimony of one friend is not much different from the five others stated. In this way, the set defined by 100 testimonies is not so different from the set defined by 1000 testimonies, if each one is not so different from the last. So, while infinitely many reasons are available, we need only articulate as many as are significantly unique to build a coherent set, which is possible for a finite mind.

SECTION 5: POSSIBLE OBJECTIONS TO COHERENT INFINITISM NOT INCLUDING SKEPTICISM

I will now address three possible objections to CI as I have laid out above. The first objection may arise against my argument for justification being inferential. An objection may be posed such as this: is a baby not justified in its basic sense perceptions, such as seeing its mother or drinking milk? My answer to this question would be no it is not. It believes it sees someone, or consumes something, but it has no idea what it sees or consumes. Its webs of belief have not yet developed enough to lead to full justification. However, this objection extends beyond just babies to everyone who has sense perceptions.

Take, for example a teenager who is walking down the hallway at his school.[13] He seems to believe non-inferentially that he is walking in a hallway without going

---

13.   I would like to thank Jeff Snapper for helping bring up this important objection.

through an argument in his head to arrive at that conclusion. Yet, I would maintain that there is a level of inference still occurring in this example. First, he must believe that he can trust his own sense perceptions, which as you will recall was a belief in our central web. For if he did not believe he could trust this sense he would not believe he were walking in a hall. Additionally, if his web was not developed enough to hold beliefs about hallways, how could he believe they were indeed walking in a hallway as opposed to a room? So, it seems our perceptual beliefs do require a reason to be justified, namely that we believe (and are justified in believing) we can trust our sense perceptions. This inference, of course, needs not develop into an argument every time we have a sense perception, for we already hold the belief, which is justified by being a part of our central web.

In section 4, I agreed with Elgin that, while justification emerges from the structure of our web of beliefs, this alone is not sufficient. Rather, we must also add the truth as being the best explanation, which once again invites the objection of this term being too subjective. In order to respond to this objection, I would like to note some important parallels between our method of justification and the scientific method. Let us consider a scientist who has just formulated a hypothesis. This hypothesis is formed on the basis of many years of past research with many of the properties and terms well defined. This hypothesis would then be tested using rather rigorous methods including lots of control groups to isolate particular variables. The scientist would then look at the evidence gathered, consider counterevidence, and competing explanations, finally arriving at a proposed truth as the best explanation of the body of evidence to justify the hypothesis.

Let us now consider how we form a belief which is analogous to the scientist´s hypothesis above. While our belief formation does not have the same precision of defined properties and other terms, we do not always start from nothing. Oftentimes, we inherit a history of testimonies and other forms of evidence, which, depending upon how much we trust the sources, we may use as the basis of our belief formation. We then begin the process of trying to justify our belief. While the scientist had many control groups at his disposal to isolate a single variable, we have no such measures in our own justification process. Rather, we have some limits as to what beliefs may be justified. To this effect, beliefs must fit into our coherent set, thereby receiving justification from the truth of the whole. These limits look slightly different depending upon in which web we are operating; justification of inner and outer beliefs looks different (recall section 4). Finally, just

as the scientist, we too look at the information gathered, consider counters, limits, and competing explanations until we too arrive at a truth perceived to provide the best explanation.

Of course, a glaring problem in both of the accounts I have given above is that "best explanation" still appears to be very much subjective in either case, and so the objection remains unanswered. However, I do not think it is possible to completely escape subjectivity. Partly due to the fact that everyone´s webs of belief are unique as I discussed in section 4. As such, belief formation, as fitting into a coherent set, will look slightly different for each individual. However, I do not believe that this is as much of a problem as it might seem, as long as beliefs and justification for those beliefs are formed in the correct way. Considering also the fact that all beliefs are revisable (just as in science) and that we are constantly reevaluating the truth we use to provide the best explanation based on our web of beliefs, this limited level of subjectivity seems acceptable.

The fin objection addresses the concern that the bar for justification has been too highly set, by requiring reasons for justification and some sort of understanding. Tt then seems justification is too difficult to obtain and actually comes after knowledge, rather than being a central part of the process. My response to this objection is that this is not the case. As I have laid it out, only babies and young toddlers are truly excluded from justifications, and this seems of little consequence. As people in these categories act on primitive measures and still lack advanced cognitive function. Furthermore, even they can still have degrees of justification, as they begin to gather evidence and build their webs; they merely lack complete justification. In response to the second part about justification presupposing knowledge, I would argue that this also is not the case. Justification does not require that someone have full knowledge in a strict epistemological sense. Rather, I argue something weaker, with only a belief and an understanding of that belief being required for justification. With these considerations, it appears that all of our beliefs, each possess initial tenability, can gain full justification.

CONCLUSION

Throughout this paper, I have argued for a new theory of justification, called Coherent Infinitism. Beginning by laying out both coherentism and infinitism

as separate theories of justification, as well as laying out important objections, I then began building towards my thesis by explaining how I see the nature of justification and its relation to beliefs. I constructed the basic framework of CI, ending by considering some possible objections to my schema. Justification is an important topic within the field of epistemology, following naturally from nearly every discussion of knowledge. For once one carefully analyzes each theory of knowledge it becomes quickly apparent that problems of justification lay at the heart of each theory. Shifting the question from "how do we know" to "how do we know that we know." The history of justification can be traced back to the time of Aristotle. I hope this paper has laid out a satisfactory possible solution to this problem.

REFERENCES

BonJour, Lawrence. 1999. "The Dialectic of Foundationalism and Coherentism. In *The Blackwell Guild to Epistemology*, edited by J. Greco & S. Ernest, 117-142. Malden, MA: BlackWell.

Davis, Wayne A, and John W. Bender. 1989. "Technical Flaws in the Coherence Theory." In *Synthese* 79, (2): 257-278. Accessed September 29, 2020. *JSTOR*, www.jstor.org/stable/20116645.

Elgin, Catherine Z. 2014. "Non-Foundationalist Epistemology: Holism, Coherence and Tenability. In *Debates in Epistemology* Second ed., edited by Steup, J. Turri, & E. Sosa, 244-255. Chichester, Sussex: John Wiley Et Sons.

Ginet, Carl,and Peter Klein. 2014. "Infinitism is Not the Solution to the Regress Problem: With Replies from Klein." In *Debates in Epistemology* Second ed., edited by Steup, J. Turri, & E. Sosa, 283-297. Chichester, Sussex: John Wiley Et Sons.

Herzberg, Frederik. 2014. "The Dialects of Infinitism and Coherentism: Inferential Justification versus Holism and Coherence." In *Synthese* 191, (4): 701-723. http://www.jstor.org/stable/24020066.

Klein, Peter. 1999. "Human Knowledge and the Infinite Regress of Reasons." In *Philosophical Perspective* 13, 297-325. Accessed October 17,2020. http://www.jstor.org/stable/2676107

Klein, Peter. 2014. "Infinitism is the Solution to the Regress Problem." In *Contemporary Debates in Epistemology* Second ed., edited by Steup, J. Turri, & E. Sosa, 274-283. Chichester, Sussex: John Wiley Et Sons.

Lehrer, Keith. 1986. "The Coherence Theory of Knowledge." In *Philosophical Topics* 14, (1): 5-25. Accessed September 29, 2020. http://www.jstor.org/stable/43153966.

Olsson, Erik. 2017. "Coherentist Theories of Epistemic Justification." In *The Stanford Encyclopedia of Philosophy*. Edited by Zalta, Edward. Accessed September 29, 2020. *https://plato.stanford.edu/archives/spr2017/entries/ justep-coherence/.*

Plantinga, Alvin. 2003. "BonJourian Coherentism." In *Warrant: The Current Debate,* 1-29. Oxford Scholarship Online, doi:10.1093/0195078624.003.0005

Van Cleve, James, and Catherine Z. Elgin. 2014. "Why Coherence is Not Enough: A Defense of Moderate Foundationalism; with reply from Elgin." In *Contemporary Debates in Epistemology* Second ed., edited by Steup, J. Turri, & E. Sosa, 255-273. Chichester, West Sussex: John Wiley Et Sons.

Defending Free Will: A Libertarian's Response to Waller and Waller's Challenge

Mitchell Baty

Kalamazoo College

ABSTRACT

Many philosophers believe that libertarian free will, or at the very least the ability to do otherwise, is required for an agent to be morally responsible for an action. The ability to do otherwise is typically understood as an agent having alternative possibilities (AP) open to them with regard to a particular decision or action. Waller and Waller (2015) challenge the notion of AP by demonstrating a difference between the intuitive motivation for AP of the Garden of Forking Paths and the way AP is formally expressed as metaphysical openness. Ultimately, Waller and Waller argue that any theory that requires alternative possibilities at the point of a free action fails. Many prominent libertarian theories, and therefore some theories of moral responsibility, fall prey to Waller and Waller's challenge. Despite this, no one has formally responded to their challenge until now. This paper examines Waller and Waller's challenge in depth and argues that libertarians can retain the ability to do otherwise by requiring that alternative possibilities be accessible to an agent at some point prior to a free action being performed.

Many philosophers believe that libertarian free will, or at the very least the ability to do otherwise, is required for an agent to be morally responsible for an action. While not uncontested, this is a commonsensical view. After all, if an agent could not have done anything else can they really be morally responsible for their action? Thought about a different way, if forces completely outside of an agent completely determined them to act in a particular way, does it make any sense to hold them morally responsible for their action? For this reason, the ability to do otherwise holds a prominent place in the literature on both moral responsibility and free will. The ability to do otherwise is typically understood as an agent having alternative possibilities (AP) open to them with regard to a particular decision or action such that they can choose between two or more different courses of action (Kane 2014, 39). Waller and Waller (2015) challenge the notion of AP by arguing that there is "an irreconcilable tension between the way in which philosophers motivate the incompatiblist ability to do otherwise and the way in which they formally express it" (1999-2000). Ultimately, Waller and Waller argue that any theory that requires alternative possibilities at the point of a basically free action fails (1211). Many prominent libertarian theories, and therefore some theories of moral responsibility, fall prey to Waller and Waller's challenge. Despite this, no one has formally responded to their challenge until now.

Waller and Waller (2015) call the formal expression of the ability to do otherwise "metaphysical openness" (1201). Metaphysical openness is the idea that an agent has the ability to do otherwise "if an agent does [action] *A* at [time] *t* in the actual world, she could have done otherwise than *A* at *t* if and only if in some possible world with the same past (state of the world) prior to *t* and the same laws of nature as the actual world, she does other than *A* at *t*" (Waller and Waller 2015, 1201). Intuitively, metaphysical openness is thought of in terms of a "garden of forking paths" (Waller and Waller 2015, 1203). The garden of forking paths is a metaphor for free will and AP used by many philosophers based on the 1948 short story by Jorge Luis Borges (Kane 2007, 6; Fischer 1994, 3-4). The idea is that our life is a garden of forking paths, when we make a basically free decision, we are at a fork in our life where we decide between mutually exclusive options that would take our life in different ways. Every time an agent makes a basically free decision, they are choosing between two or more "paths" that their life can take into the future. If an agent does not have multiple paths into the future open to them regarding a particular decision, they have no basic control over that decision and therefore it

is not a basically free decision.[1] Therefore, the garden of forking paths motivates the need for AP at the time of a basically free decision.

Waller and Waller's (2015) argument for why metaphysical openness and the garden of forking paths are incompatible is highly technical and requires some initial terminology to understand. It is important to realize that AP is a modal concept. The *ability* to do otherwise requires that it is *possible* to do otherwise. Possibility is a modal notion and requires a theory of modality to make sense of its semantic content. The most popular theory of modality is possible worlds theory which says that something is possible if and only if it is true in a possible world. For the purposes of this paper, it is unnecessary to get into the debate about what exactly a possible world is. It is enough to understand that for an agent to have AP there must be two or more possible worlds with the same past and different futures; one world in which the agent does A, call this W, and another world in which the agent does other than A, call this W_p. A point at which W and W_p are identical prior to the point and diverge after the point is called a *splitting point* (Waller and Waller 2015, 1205). A splitting point, t, can be either the *point of last coincidence* between the two possible worlds (i.e., the last point at which the worlds are identical), such as in Figure 1, or the *point of first noncoincidence* (i.e., the first point at which the worlds are no longer identical), as in Figure 2. It is helpful to think of "worlds" being identical and then diverging by analogy to a pair of mathematical lines such as in Figures 1 and 2. In Figure 1, t represents the point of last coincidence as the *final* point in the segment of the line labeled A. In contrast, Figure 2 shows t as the point of first noncoincidence as the *first* points in segments B and C. Both figures show the lines as identical in segment A, and different after the splitting point, the only difference is whether the splitting point is the point of last coincidence or the point of first noncoincidence.

Figure 1

Figure 2

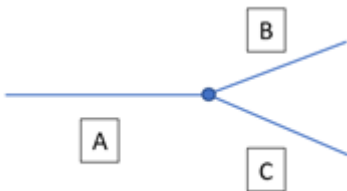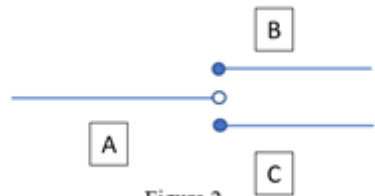1.   This is not to say that it is necessarily not a free decision. For example, using Robert Kane's (2014) variety of libertarian free will an agent's decision can be free even if they do not have AP if they had AP in the past when they performed a "will-setting action" (40). A basically free decision is one where an agent has AP directly regarding the decision at hand and not some past decision they made.

The analogy to a pair of mathematical lines is also helpful in displaying the crucial notion that time is infinitely dense. In the same way that between any two points on a line a third point can always be drawn, between any two points in time there is always another point. Waller and Waller's (2015) argument hinges on the view that time is dense because a pair of possible worlds cannot have *both* a point of last coincidence and a point of first noncoincidence if time is dense (1205). This is because there would necessarily be a point in between the point of last coincidence and the point of first noncoincidence. This makes no sense. The point of *last coincidence* is supposedly the *last* point at which the worlds are identical and the point of *first noncoincidence* is the *first* point at which the worlds are no longer identical. This means that the point of last coincidence and the point of first noncoincidence should be directly next to each other with no moment in between them. However, this is impossible if time is dense.

The final point of terminology that is important to understand is that Waller and Waller (2015) call a pair of possible worlds with a point of last coincidence *forking* and a pair of possible worlds with a point of first noncoincidence *nonforking* (1205). Waller and Waller argue that the garden of forking paths metaphor implies a pair of forking worlds, and that metaphysical openness implies a pair of nonforking worlds (1205). Already this is a problem for those that wish to hold on to both the garden of forking paths and metaphysical openness; a pair of possible worlds cannot be both forking and nonforking because they cannot have both a point of last coincidence and a point of first noncoincidence. From this point, Waller and Waller (2015) lay out possible options open to libertarians (1207).

If a libertarian wishes to retain the garden of forking paths metaphor and chooses to require a forking pair of possible worlds Waller and Waller (2015) believe they quickly run into problems (1209-1210). Recall that in forking worlds the splitting point is the point of last coincidence. A libertarian can either make the splitting point the time at which the agent makes a decision, or a moment at which the agent has yet to make the decision. If the splitting point is the time at which the agent makes a decision this would mean that the worlds diverged at the splitting point. By making a decision the agent is choosing to do one thing and not something different meaning that they worlds are no longer identical at the splitting point. However, a forking pair of worlds requires the splitting point to be the point of last coincidence, the last moment at which the worlds are identical. If the decision is made at the splitting point then the worlds are no longer identical

at the splitting point, meaning that it makes no sense to say that the splitting point is the moment at which the agent makes their decision on a forking worlds model of AP.

The splitting point must then be a point in time where the agent has not yet made the decision. In this case, at the splitting point the agent is undecided and after the splitting point the agent changes from being undecided to having made a decision. Waller and Waller (2015) make an analogy between change and movement, claiming that both are dynamic processes (1210). When something is moving at one point in time it is necessarily moving at a later point in time. Similarly, the splitting point in question must be a point of change, and any point of change must have a later point of change. This means that there must be a later point at which the agent still has not made their decision. This cannot be the case; according to the forking model of AP the splitting point is the last point of coincidence, the last point at which an agent has not made their decision. There cannot be a later moment at which the agent is still undecided.

If a libertarian is willing to give up the garden of forking paths metaphor, they presumably can say that the pair of possible worlds required for AP is a nonforking pair of possible worlds. The problem is that it seems to be impossible for alternative possibilities to be *accessible* to an agent at the splitting point on a nonforking pair of worlds (Waller and Waller 2015, 1208). By "accessible," Waller and Waller mean something akin to "available to the agent." AP seems to be inaccessible to an agent on a nonforking model because nonforking worlds have a point of first noncoincidence as the splitting point. Since the splitting point is the point of first noncoincidence, the worlds are no longer identical. The agent has already made their decision and it is no longer accessible to them to do otherwise because they cannot go back in time and undo their decision in order to do otherwise.

A potential move to make here is to say that alternative possibilities were accessible to the agent at some relevant point *prior* to the splitting point. Waller and Waller argue that this is not available to a libertarian because there is no point prior to the splitting point at which alternative possibilities are *immediately accessible* to the agent (1208). The argument is straightforward; since a nonforking pair of worlds does not have a point of last coincidence, any moment at which the worlds coincide will always have a later moment at which the worlds continue to

coincide, meaning that there will always be a later point in time at which the agent has not yet chosen between alternative possibilities.

Waller and Waller are making the assumption that alternative possibilities are only relevant when they are *immediately accessible* to the agent. While not explicitly stated in their paper, it can be inferred that what is meant by "immediately accessible" is something like "accessible to the agent in the very next moment after the splitting point." Given this understanding of "immediately accessible" and an understanding that time is dense, there is no sense to be made of "the very next moment after the splitting point" because any such moment will always have an infinite number of moments in between it and the splitting point. I am willing to concede that, if time is dense, at no point prior to the splitting point are alternative possibilities immediately accessible to an agent. However, I see no reason why alternative possibilities must be *immediately* accessible to an agent. It seems that all that is required for AP is for an agent to have alternative possibilities accessible to them at *some* point prior to the splitting point.

When asked what prevents libertarians from doing this, Robyn Waller responded with two points. First, that she would ask anyone who took this approach "to explain when prior to t is within an acceptable range to qualify as providing the alternative possibilities needed for basically free action and when prior to t is not within an acceptable range to qualify as providing the necessary access to alternatives" and that libertarians do not have this problem if they require alternative possibilities to be immediately accessible (personal communication). Second, that "the incompatibilist gets away with an illusion of metaphysical and mathematical rigor in talking about time points like 'at t' in statements of metaphysical openness" (personal communication).

Regarding the first point, it seems to me that libertarians who want to take this approach have several options. The first option is to simply choose an amount of time arbitrarily. Any amount of time prior to the splitting point would work. However, Robyn Waller is quite right to point out that choosing an amount of time arbitrarily causes a theory to lack any sort of mathematical or metaphysical rigor. In order to maintain a level of philosophical rigor there needs to be a nonarbitrary criterion for determining how far prior to the splitting point is within an acceptable range of time to qualify as providing the alternative possibilities required for a basically free action.

To find this criterion it is important to look at what makes an action free beyond simply having alternative possibilities. Libertarian theories also have another dimension of responsibility or ownership of one's actions. Robert Kane's (2014) theory of libertarian free will as Ultimate Responsibility (UR) can be used as an example of a way to find the criterion Robyn Waller is looking for (39). UR requires that an agent be "responsible for anything that is a sufficient reason (condition, cause or motive) for the action's occurring" (Kane 2014, 39). According to this theory of free will, in order for an agent to perform a free action they must be ultimately responsible for their action. Using UR, libertarians can respond to Robyn Waller by saying that the criterion for how far prior to the splitting point is within an acceptable range of time to qualify as providing the alternative possibilities required for a basically free action is any time in the agent's life where they have the capacity to be ultimately responsible for their actions.

Kane's (2014) theory has six conditions for what he calls "plural voluntary control" that must be met in order for an agent to be ultimately responsible for their action (50-51). A free decision must be:

> *brought about* by the effort of the agent, the agent had *control* over it at the time in the sense of having the power to make it be and the power to make it not be, the agent brought it about *voluntarily*, *intentionally* or *purposefully*, and for *reasons*, and could have brought about an alternative choice at the time voluntarily, intentionally, and for reasons. (Kane 2014, 51)

As expressed by Kane, there is a problem with the conditions for plural voluntary control. The second condition, that "the agent had *control* over it *at the time* in the sense of having the power to make it be and the power to make it not be" is clearly incompatible with a nonforking model of AP (Kane 2014, 51, my italics). As written, the second condition specifies that an agent have control *at the time* they perform an action, suggesting a forking model of AP. On a nonforking model this is impossible. However, there is a simple fix. Simply change the second condition to read "the agent must have had control over their action *at some time in the past* in the sense of having the power to make it be and the power to make it not be" and the condition is retained in a form that is compatible with a nonforking model of AP. All of the other conditions for plural voluntary control are compatible with a nonforking model. Using this revised version of the conditions for plural

voluntary control we are now in a position to provide a criterion for the range of time during which AP are of the type required for free will. We can say that the lower limit of this range is the time at which an agent develops the *capacity* to meet the revised conditions for plural voluntary control, while the upper limit is the splitting point. Any alternative possibilities between these two moments in time are of the sort required to provide a basis for free action.

Note that this criterion is not arbitrary; without the capacity to meet the revised conditions for plural voluntary control an agent could not possibly be ultimately responsible for their action and therefore could not possibly perform a basically free action. Despite not labeling an exact range of time, this approach is still rigorous because it points to a specific, nonarbitrary, point in the agent's life after which all alternative possibilities open to the agent are considered relevant for basically free actions. Some of the conditions for plural voluntary control may be innate to all humans from the moment they are born. However, at least the capacity to perform actions *intentionally or purposefully* and for *reasons* develops throughout childhood. The exact time at which people gain these abilities are different for everyone depending on the rate at which their brain develops. However, every agent, every person, had or will have a moment in time at which they develop the capacity to have reasons for their actions and to act with purpose or intention. Only once an agent develops the capacity to meet the conditions for plural voluntary control can the alternative possibilities open to them be of the right kind to allow for basically free actions.

Robert Kane's (2014) theory is only one example of how libertarians can create a criterion. Not every theory has the conditions for plural voluntary control. However, every theory will have a point before which agents are incapable of free action. This might be the moment a person is born, or conceived, or some later moment at which they develop the capacity for free action. The point is that every libertarian theory of free will can take this approach. As with Kane's (2014) theory, doing so might require minor revisions to the theory but it can be done.

I agree with Waller and Waller's (2015) central claim that any theory that requires "indeterminism-involving alternative possibilities at the point of a basically free action fails" (1211). However, I do not see it as a problem for libertarian theories of free will. Using Robert Kane's (2014) theory of free will as an example, I argue that the nonforking model of AP can be utilized if one is willing to give up the garden of forking paths metaphor for free will. Libertarians can say that an agent

has alternative possibilities in the relevant sense if and only if the agent has AP at some point between the emergence of the agent's capacity for free action and the splitting in a nonforking pair of possible worlds. This criterion allows libertarians to agree with Waller and Waller's conclusion but maintain that it is not a problem for libertarian free will.

REFERENCES

Borges, Jorge, L. 1948. "The Garden of Forking Paths." *Ellery Queen's mystery magazine*, August 1948. 12 no. 57. Translated by Anthony Boucher.

Kane, Robert. 2014. "Acting 'of One's Own Free Will': Modern Reflections on an Ancient Philosophical Problem." *Proceedings of the Aristotelian Society* 114: 35-55. https://doi.org/10.1111/j.1467-9264.2014.00363.x.

Waller, Robyn, R., and Waller, Russell, L. 2015. "Forking Paths and Freedom: A Challenge to Libertarian Accounts of Free Will." *Philosophia* 43: 1199-1212. https://doi.org/10.1007/s11406-015-9612-8.

Kane, Robert. 2007. "Libertarianism." In *Four Views on Free Will*, Edited by Ernest Sosa, 5-43. Oxford: Blackwell.

Fischer, John, M. 1994. *The Metaphysics of Free Will*. Cambridge: Blackwell.

Impoverished Bees: A Critical Analysis of Heidegger's Conception of Man as World-Forming and the Animal as Poor-In-World

Abigail Bergeron

Trent University

ABSTRACT

In this paper I undertake a critical analysis of Martin Heidegger's comparative understanding of man as world-forming and the animal as poor-in-world from his 1938 lectures The Fundamental Concepts of Metaphysics. I show that Heidegger imposed this framework onto his reading of Nietzsche's perspectival realism in Lecture 25 from Nietzsche, Volume One. I argue that Heidegger's understanding of an absolute metaphysical divide between the essence of humanity and the essence of animality is not only severely flawed due to its failure to deal with the problem of transposition and emergence, but that Heidegger also co-opted Jakob von Uexkull's work regarding umwelts to suit Heidegger's own narrative of human superiority, making it antithetical to Uexkull's project of humility. Heidegger claims to be contesting reductionist or mechanist conceptions of life and being, yet I argue that his understanding of animal "captivation" and "disinhibiting rings" is merely a regurgitation of these same types of theories. I conclude that Heidegger's justification for the division between man and animal is circular and groundless, and despite his protestations to the contrary, represents only an argument for anthropocentrism.

In this essay, I will be focusing solely on Heidegger's interpretation of Friedrich Nietzsche[1] in Lecture 25, and considering its relation to his much earlier writings on the topic from the 1938 lectures *The Fundamental Concepts of Metaphysics.* This paper will explore Heidegger's conception of man's being-in-the-world as compared to that of nonhuman animals, and consider how Heidegger's interpretation of man as world-forming and the nonhuman animal as poor-in-world influenced his reading of Nietzsche's position of perspectival realism and the sensuous. I will conclude with a critical analysis of Heidegger's position from a post-humanist perspective, comparing his thought to the work of Jakob von Uexkull, highlighting the inherent Anthropocentrism and latent contradictions in Heidegger's conception. I will argue that Heidegger's account is not only skewed by these flaws, but that it is no valid argument at all. Instead, Heigger presents only a circular justification for why human beings are superior to nonhuman animals as part of his wider exploration of the human Dasein.

In Part 1 I will begin this essay with an exploration of Heidegger's early thought regarding the distinction between man and nonhuman animals in his 1929/1930 lectures as drawn from *The Fundamental Concepts of Metaphysics.* Considering Heidegger's foundational premises on this topic will help me to better interpret his 1961 Lecture 25, "The New Interpretation of Sensuousness and the Raging Discordance Between Truth and Art." In Part 2, I will then analyze Heidegger's Lecture 25 "The New Interpretation of Sensuousness and the Raging Discordance Between Truth and Art" from *Nietzsche, Volume One* and I will consider HN in relation to Heidegger's own position. Lastly, in Part 3, I will provide my own thoughts on Heidegger's theory, including in relation to that of Uexkull, and Thomas Nagel, regarding the problem of transposition, anthropocentrism, and the attendant gaps in Heidegger's thought.

PART 1: HEIDEGGER'S CONCEPTION

As aforementioned, to better understand Heidegger's position in Lecture 25 concerning HN, I have decided to undertake an exploration of Heidegger's earlier

---

1.   Henceforth to be referred to as "Heidegger's Nietzsche" or "HN" to denote Heidegger's pervasive adaption of Nietzsche as separate from more objective accounts of Nietzsche's thought. While certainly a discussion on Nietzsche's perspectival realism and his influence on post-humanist thought would be relevant, the focus of this paper is on Heidegger's thought and interpretation.

work and considerations on the topic of being-in-the-world, and more specifically, his differentiation between man (as world-forming), nonhuman animals[2] (as poor-in-world), the inorganic (as wordless), and the role perspective plays with regards to reality and being-in-the-world.

What is a Man?

According to Heidegger, the being of man is being-there, also known as Dasein. What Heidegger calls *attunement* is a fundamental part of Dasein, our way of our being-there (Heidegger 1983/1995).[3] *Attunement* is the way the external world is opened up to human beings, and includes our being-with-each other. On this account *attunement* is the manner in which "Dasein is as Dasein", the way in which we as human beings exist (Kuperas 2007). *Attunement* then is not a "side-effect" but the fundamental way of being and perceiving.

Unlike other philosophers who have attempted to differentiate man in unique or special ways, Heidegger does not prescribe human beings speciality insofar as they have logos, politics, or even rationality (Kuperas 2007). Heidegger actually sees the rational conception of man as hitherto hindering the recognition and essence of *attunement.* In *The Fundamental Concepts of Metaphysics* (*FCM*) Heidegger argues that due to the narrative of rationality, the true essence of man's perspective and reality (being-there) is ignored, and his related concept of *attunement* is subordinate, man in the "first place" is a rational being (Heidegger 1983/1995). On Heidegger's conception, perception, feeling, and being, have nothing to do with innate cognitive abilities or reason. Man is not special because he thinks, man is special because he is *attuned.*

An irrevocable aspect of man's disposition for *attunement,* is man's ability to be world-forming. It is this world-forming capacity that differentiates human beings from nonhuman animals. Man is world-forming insofar as man not only can access world, and is affected by it, but that he is *attuned* to it. The world of man

2. Heidegger uses the term "animals" to denote animals as separate from human beings, however I will refrain from its usage due not only to categorical inaccuracy (human beings *are* animals), but also because the term "animals" too, is generalized to describe all kinds of living beings that can differ from each other is extreme ways. Therefore I will use more encompassing terms such as "nonhuman" and "organism" to describe living creatures that are not human beings.

3. All citations from *The Fundamental Concepts of Metaphysics* will be cited using in-text citations denoting the Part Number and Chapter Number, with a formal citation in the Works Cited List at the end of the essay.

is a "rich one", it has great range, penetrability, and it is constantly extendable and extending (Heidegger 1983/1995). In simpler terms, human beings are world-forming because only human beings have such great ability to relate to, interact with, and perceive their worlds in creative, innovative, and far-reaching ways. As per this capacity, humans can actively and attentively *comport* themselves towards others (an essential aspect of being-there is being-there-with-others) and towards their environments.

Contrary to other anthropocenterists Heidegger does not rely on transcendental narratives, or reductionist behaviourism to set human beings apart, instead, it is the human capacity for perception and relation, for *being* that quite simply, and self-evidently accounts for this speciality.

What is a (Nonhuman) Animal?

In *FCM* Heidegger outlines three distinct "realms"-that of the human as world-forming, that of the material objects (eg. a stone) as worldless, and then, that of the nonhuman animal, which according to Heidegger, falls somewhere in between these two (Heidegger 1983/995). Heidegger's project in *FCM* is to determine the metaphysical difference between the essence of humanity as world-forming, and the essence of animality. Again, unlike other anthropocentric philosophers, he is not interested in cognition or rationality, or morphological differences and the comparison of species. Instead, nonhuman animals differ from human beings because of this metaphysical understanding of their essence, that is that nonhuman animal's are poor-in-world (Heidegger 1983/1995).

Despite describing this poverty-in-the world as a deprivation, Heidegger explicitly rejects the fact he is entailing a "hierarchal assessment" between humans and nonhuman animals (Heidegger 1983/1995). The apparent counter-intuitiveness of this statement will be discussed at length in Part 3, however for now, it is important to note that Heidegger argues, as a matter of degree, that the nonhuman animal possesses *less* in their worlds than human beings do. Both *have worlds* and have accessibility to their worlds, but the animal is "confined" to its world in a way that human beings are not. There comes a point where the nonhuman animal is incapable of further expansion or contraction of their perspective worlds, and its specific domain or "environment" is limited in this potentiality for range and penetrability. Heidegger gives an example of a worker bee, which while familiar with the blossoms on the flowers it frequents, the bee

cannot *know* about the type of blossom, or the roots, or the number of leaves. The world of the bee is strictly circumscribed (Heidegger 1983/1995).

When considering the relation of nonhuman animals to the world, it follows that every animal, as an animal, has a certain set of "relationships to its sources of nourishment, its prey, its enemies, and its sexual mates" (Heidegger 1983/1995). However, to distinguish these relationships from those that humans have, Heidegger posits a pseudo-biological understanding of "capacity." Heidegger argues that when we speak of organs, we speak of capacity (in comparison to tools, which we speak of as having serviceability), therefore, something which is *capable* is something that is intrinsically regulative and self-sustaining (Heidegger 1983/1995). Nonhuman animals have *capacities* (such as the capacity for sight). The "potentiality" for sight in nonhuman animals is only a capacity, whereas for human beings-despite anatomical similarities, our potentiality for sight has a "quite different" characteristic (Heidegger 1983/1995).

As already discussed, Heidegger views human beings as being able to *comport* themselves towards things, humans *do* and *act*. Following this, nonhuman animals such as the worker bee gathering pollen, constitutes a mere *driving* and *performing* - they only have capacities for *behaviour*. The essential structure of this behaviour is grounded in what Heidegger calls "captivation", the nonhuman animal is captivated insofar as it can behave within a certain environment, but never within a world like that which human beings can (Heidegger 1983/1995). Heidegger returns to the example of the bee, citing an experiment where a bee was cut open while it was collecting honey. Heidegger highlights how the bee did not stop collecting the honey, despite the fact that the honey was escaping out its severed body just as readily as the bee was consuming it. Overlooking the gruesome nature of this experiment, Heidegger surmises that this is proof of his theory, the bee is not governed by any actual recognition of the honey as honey, but only that the "drive" (to collect honey) is merely "captivated." Whereas humans, who can *comport* themselves towards honey, would have recognized that their bodies had been severed, and would have consequently stopped collecting the honey.

Heidegger argues that this "captivation" and the totality of the organism's capacities (what it is capable of) is what determines its environmental world, what Heidegger calls its "intrinsic self-encirclement" (*Sich-Enrigen*), also described as its "disinhibiting ring" (Heidegger 1983/1995). Nothing else can penetrate

this "ring" around the nonhuman animal, including its inability for the organism to recognize itself, or other beings outside of this "ring" of capacities and behaviour. The disinhibiting ring is essentially that, a disinhibiting and delimiting insurmountable barrier. It is this "ring" that makes nonhuman animal's poor-in-world when compared to the lack of such a barrier in the human world, humans do not have "rings" because humans are world-forming.

I am not seeking to critique this position yet, however, it is worth keeping in mind when we turn to Heidegger's interpretation of Nietzsche, and we consider the way in which Heidegger has imposed this framework onto Nietzsche's thought.

PART 2: LECTURE 25

In Lecture 25 of *Nietzsche,* "The New Interpretation of Sensuousness and the Raging Discordance Between Art and Truth" Heidegger is continuing his discussion from the previous Lecture regarding Nietzsche's interpretation of reality and sensuousness.

In Lecture 25 Heidegger is exploring to what extent Nietzsche's ideas can be taken further, and he opens the lecture with a fundamental question, "to what extent is 'the sensuous' the genuine 'reality'?" (Heidegger 1961/1991, 211). This is an important question not just regarding Heidegger's interpretation of Nietzsche, but it is a question Heidegger himself cares about in *FCM,* that is, to what extent is the nonhuman animal's reality, genuine reality, especially when compared to our own.

For HN, the sensuous constitutes real reality, or rather, the perspective of the subject is that which constitutes reality (Heidegger 1961/1991, 212). The perspective, is defined as, "the angle of vision" that is incorporated and encompassed by the organism's "capacity for life" (Heidegger 1961/1991, 212). According to HN, it is this "angle of vision" that draws the "borderline" around what the organism, and how the organism encounters things in their environment. Heidegger gives the example of a lizard, which can hear a rustling in the grass, but cannot hear a gunshot fired because (on this account) a gunshot would not be relevant to its interpretation of its environment and life.[4] A gunshot would mean

---

4. Depending on the species of lizard, it is likely it *could* hear a gunshot, so Heidegger's statement is misleading. It is also likely, given the fact that many species are quite sensitive to sound and sudden movements, a gunshot *would* cause a lizard to react, likely making it flee in fear. I doubt

nothing to a lizard, it is outside its "angle of vision", while a rustling in the grass could mean a potential predator or prey. This is not incidental as perspective is according to HN, "the basic condition of all life" (Heidegger 1961/1991, 212). Heidegger continues this line of thought with further explanation of the perspectival reality for the organism, and unsurprisingly, HN takes the position that this perspective or "angle of vision" is necessarily circumscribed by a "line of horizon." It is only within this horizon that something can come into appearance for the organism in question (Heidegger 1961/1991, 212). Following this, what is "true" for the organism is what is perspectively perceived and seen as definitive, making truth then erroneous "mere appearance"- whatever the subject perceives is what it will believe is the definitive truth (Heidegger 1961/1991, 214). Truth is what appears to the organism, what is valued for the the life of organism, and it is inherently perspectival (Heidegger 1961/1991, 215).

In this short summary of the first half of Lecture 25, it is quite easy to see, after considering Heidegger's foundational position in *FCM*, how much of Heidegger's own framework is imposed onto Nietzsche's perspectival realism. The language used is seemingly identical to that which Heidegger uses to describe the nonhuman animal's poverty-in-world; *"angle of vision"* and *"line of horizon"* convey his personal interpretation of the nonhuman animal's "disinhibiting ring" which necessarily includes such "borderlines" and "horizons." Also note HN's emphasis on *drives,* and *capacities*-the lizard does not "hear" the gunshot because it cannot physically hear, rather, it does not "hear" the gunshot because it does not *need to,* it does not have the *capacity* within its ring, its life environment, its "angle of vision" to hear the gunshot for what it is, or react to it in that way.[5] For that reason, the organism is circumscribed within its environment, and it is for that same reason in *FCM* that Heidegger makes the claim that nonhuman animal's are poor-in-world when compared to human beings.

But does HN make the same comparative claims in Lecture 25? Here it is important to note that the focus in *FCM* was to discuss the metaphysical differences between humans and nonhuman animals, in Lecture 25 however, HN is concerned more with defining reality as apparent and perspectival (the sensuous) and exploring how, in that case, truth can be pursued and defined. In the first

---

a lizard would just ignore the sound of a gunshot. But as we will see with many of Heidegger's examples, they aren't quite as clear or relevant as he would have us believe.

5.   Again, according to Heidegger's pseudo-biology.

part of the Lecture, it is clear that HN is discussing the limitations of apparent reality with animals generally, he even notes that this erroneous truth of mere appearance, is the same "world" that "man resides in" (Heidegger 1961/1991, 214). So it would appear, according to HN, that man is at least to some extent, limited by this same perspectival appearance, or the "error" of perspectival realism. However, Heidegger dedicates the second-half of this Lecture to his discussion regarding the role of art as perspectival letting-shine, it illuminates, it "liberates" one's perspective in order for it to be conceived of and expanded. HN claims that because life is perspectival, and it "waxes and flourishes" with regards to its appearance. This means that truth-as it is immobilizing of this natural oscillation of life, is therefore inhibitory. Art is "worth more than truth" because it is through art that perspectival realism can realize and overcome itself (Heidegger 1961/1991, 216). It is through art that the organism can transcend the boundaries of perspective that encompass it. Art and truth thereby diverge, truth is only perspectival and thus limited, art is transcendent of perspective.

While it is not within the scope of this essay to continue a consideration of HN's view on art, it is important to note that so far it appears that HN's is offering a way out of the "disinhibiting ring" or "line of horizon" of the organism's perspectival environment. The question that must be asked then is, could this apply to nonhuman animals? On HN's account in Lecture 25, it is clear that the answer is a likely no, as while HN's does not explicitly conclude that nonhuman animals could not participate in perspectival shining through art, it is implied that art, as a shaping and form of creation, is fundamental to the essence of Dasein. Art is that which "the supreme lawfulness of Dasein becomes visible" whereas (perspectival) truth which on this account, all beings have access to, is subordinate (Heidegger 1961/1991, 217). Perspectival truth is mere appearance, it is a fixation on an apparition (recall: captivation), whereas art which belongs to Dasein, is *transfiguration (*Heidegger 1961/1991, 217). Again, while HN does not mention Heidegger's conception of world-forming, it is easy to see where this emphasis on art as enhancement of reality, conveys the same message. Man, as Dasein, is world-forming, which necessarily includes the ability to transcend the boundaries of perspective through creation and formation, and art on both Heidegger and HN's accounts, plays a fundamental role in this augmentation of reality and truth. Art and (perspectival) truth are both proper to the essence of reality, however it is art, that further enhances the extension and reach of reality and of truth,

eg. the reality and truth of Dasein. Nonhuman animals are not mentioned in the second-half of the Lecture because they do not need to be, it goes without saying for HN that nonhuman animals could never transcend their perspectival realities (otherwise known as "disinhibiting rings") let alone through such world-forming ways as creation and art. The lizard cannot even recognize a gunshot!

Sarcasm aside, this analysis leaves us with two remaining questions. If we take HN's proposition regarding perspectival truth and the transcending power of art to its fullest conclusion, *could* nonhuman animals, if they *could* demonstrate perspectival shining through art, thereby transcend their world-boundaries in the way that Dasein can? I would argue that would be an interesting consideration indeed, and perhaps, modern biology and a better understanding of nonhuman animal communication and art, could allow, on Heidegger and HN's account some nonhuman animals to be world-forming in a way akin to Dasein. For example, recent science has proven that humpback whales "sing" to each other across the spans of oceans, in songs that are consistently changing and use the very same rules that human composers use (Sagan 2010, 22). Is this not art? Or at least, according to HN's definition, a form of perspectival shining? At first, this appears to be an excellent example to defeat Heidegger's exclusion of nonhuman animals from world-forming, at least, for *some* types of nonhuman animals. Except, we must remember that Heidegger was not concerned with the particulars of different species like other anthropocentrists usually are. Heidegger's position is metaphysical, not zoological, therefore, he sees all nonhuman animals as sharing the same essence, completely ignoring any physiological similarities that certain species may have in common with human beings. For that reason, Heidegger is best described as a Anthropocentrist with a capital "A", because for Heidegger, that difference is absolute, not based on conditionals or typology. I would argue that if Heidegger were presented with the example of the humpback whale, he would reiterate that its 'singing' is not a form of art, of perspectival shining or transfiguration, but is merely a drive or a capacity-the humpback whale, regardless of its musical abilities, would remain entrapped in its giant oceanic ring, in a way that human beings just aren't.

This provides a nice segue to my second remaining question, as noted earlier, it remained unclear in *FCM* whether Heidegger argues that human beings, insofar as they are world-forming ie. capable of transcending the disinhibiting rings of nonhuman animals, or that human beings are world-forming because they are

a priori free of such disinhibition. However, in Lecture 25, it is much clearer; human beings are perspectival just like all living organisms, and as a result their truth is just as fixed and one-sided as any other organism. It is only through the unique ability of Dasein, through creation of art as perspectival transfiguration, that human beings can overcome this "truth as mere appearance, as error" (Heidegger 1961/1991, 214). Yet, what exactly constitutes this art that can inspire such transcendence? Is it art in the typical understanding-painting, dancing, singing? What of human beings that do not pursue the typical arts, what of human beings that are not creative or "shining" in this common sense, who eschew the arts and live relatively mundane and unexpressive lives? One argument that could be made is that these certain people don't transcend their perspectival truth, and that in some sense, they are not as world-forming as other people. For example, are toddlers as world-forming as adults? What about the mentally disabled? An unfortunate case can (and historically has) been made that these people are not world-forming in the superior sense, and instead remain entrapped in animal-like disinhibition. However, a second, and I think more accurate argument could be made regarding this world-forming as the *essence* of Dasein, of people, *regardless* of how individuals choose to express it. This would mean that a very broad range of human expression, creation, or formation, could be considered perspectival shining, transfiguring, and thus world-forming, perhaps everything we as humans beings *do* can even be considered in this light. Given again, that we are considering metaphysical essence here and not individual types, I would say that this answer would be the more valid one. Yet, now we are returning to our first question, if the definition for art, for Dasein and world-forming is so broad and far-reaching, how can we merely ignore the similarities and analogies between cases of creation and expression that undoubtedly overlap in cases with nonhuman animals, such as that of singing humpback whales? Like is the case with many other Anthropocentrists, these gaps are too obnoxious and prevalent to ignore. Yet, they are ignored and dismissed, because Anthropocentrists, such as Heidegger, beg the question from the start. Again, Heidegger is not concerned with particularities because his theory is meant to be absolute. The metaphysical divide that he posits in *FCM* between humans and nonhuman animals, and the one he posits in Lecture 25 regarding art as transcending perspectival truth is exactly that, a metaphysical divide about *essence,* and as such, it is insurmountable from its first premise.

Like Heidegger's foe Rene Descartes, by ignoring similarities and contingencies, Heidegger's theory is less about understanding the relation between humans and nonhuman animals, and more about constructing a metaphysical narrative about human superiority.[6] As I have shown, this is especially blatantly evident in his Lecture 25 reading of Nietzsche, where Heidegger co-opts Nietzsche's account of perspectival realism and will to power as art, to impose his own framework of Dasein as exclusively world-forming and transcendent of perspectival truth. Regardless of whether man is initially entrapped in disinhibiting rings or not, man necessarily overcomes such delimitation, in a way that other animals are a priori considered incapable to do.

PART 3: CRITICAL ANALYSIS

There are many issues with Heidegger's conception of man's being-in-the-world in comparison to that of nonhuman animals, not the least of which is its vague and pseudo-scientific quality. *Prima facie* Heidegger appears to be positing a radically original and new claim, one that he explicitly states refutes earlier rationalist, or mechanist conceptions of human superiority or the realities of nonhuman animals. However, here I argue that Heidegger's claims are nothing new, and if anything, they consist of merely convoluted interpretations of the same old rationalist and mechanist theories, including Heidegger's rampant sidestepping of Uexkull's work regarding *umwelt,* and his failure to address what forms the foundational problem of the problem of other minds, that being, transposition and emergence. I will conclude then that Heidegger's conception of humanity's relation to the world in which it dwells, when regarded in comparison to his position on the world of other beings, has very little merit. It is, to borrow from Jacques Derrida, a "violent and awkward" assertion, one that is blatantly anthropocentric, and filled with the same shortcomings as that of other absolutist Anthropocentrists.

Heidegger and Uexkull

I consider it no small irony that initially I was planning to write this essay as a comparative analysis between Heidegger and the work of Uexkull. At first, I was under the impression that Heidegger and Uexkull were radically opposed,

---

6. Except, where Descartes unashamedly admits to human superiority, yet as we will see, Heidegger continues to deny he is attempting to assert any hierarchal claim at all.

however, after reading *FCM* and Lecture 25, I realized something much more sinister. Heidegger not only adopts much of von Uexkull's theory regarding *umwelt* and capacity, and then explicitly cites him throughout his writings,[7] but that Heidegger succeeds in warping Uexkull's theories to support his own, a theory that as I will show, is glaringly antithetical to that of Uexkull's.

Uexkull and his Institute studied the perceptual worlds of humans and nonhuman animals, and were particularly interested in integrating the functionality of the organism into an understanding of its biology and life-worlds. Fundamentally, Uexkull saw the perceptions, communications, and behaviours of organisms as part of purpose and sensation that was not limited to human beings (Sagan 2010, 3). A perceptual and perspectival universe is therefore, inherently Uexkullian. Uexkull was radically opposed to the hitherto machinist and reductionist view of biology, and instead pushed for a biology that would finally and properly account for the perceiving inner-world of nonhuman animals that had been left out (Sagan 2010, 12). Through meticulous observations, illustrations, and understanding of anatomy and physiology, Uexkull posited that organisms as subjects in their life-worlds or perceived environments (*umwelt),* and recognized not only sensory inputs, but also "functional tones" of the external stimuli that they need to survive and thrive. Uexkull did not consider these as reflex or drives, but some form of phenomenal and subjective awareness, even in such humble creatures as the tick, which only recognizes three "functional tones" in its *umwelt* (Uexkull 1934/2010, 50). He argued that there was a functional reason for this subjectivity. A living organism is not a finished state but rather, a continuous process that must constantly replenish and maintain itself to prevent falling into eventual disrepair and death, and such a living organism requires at least some level of awareness of the signs (both internal and external) in order to survive (Sagan 2010, 20). In this way, Uexkull addresses the problem of emergence (how sentience and consciousness emerge) and the evolutionary contingency between species by highlighting the functionally of subjectivity which guarantees that in some form, subjectivity is ubiquitous across all living organisms from the very beginning (Sagan 2010, 27).

Uexkull not only recognized that even a lowly tick has purposeful perceptions, but that these perceptions and experiences are exceedingly diverse across species. On his account there is the human *umwelt*, and then there are different *umwelts* for different nonhuman creatures. As a perspectival realist, Uexkull's conception

---

7.   Uexkull is actually the scientist most cited by Heidegger (Sagan 2010, 6).

was so revolutionary in the sense that these *umwelts* represent such variation that an 'objective' or unified world is impossible; most living things (including us) exist exclusively in their specific *umwelts* or perceptual environments. As seen with the example of the tick, these umwelts can be constricted to a very simple degree or highly complex, but none necessarily represent the 'true' or objective external space in which we exist.

However, Uexkull was also adamant that this did not imply that human *umwelts* were exemplary or unique in any way, all organisms have unique *umwelts,* and even individual people, have their own unique *umwelts,* where tones in their environment appear to them in a way they wouldn't for other people.[8] Human beings may be unique insofar as they can employ triadic signs (written language) and other skills, but comparing human *umwelts* to that of other beings is nonsensical, there are innumerable capabilities of other species that we cannot being to imagine or envision, let alone conclude that these organisms do not exist as subjects in their environment akin to how we do.[9]

For Uexkull, this multiplicity of perspectives was what made the universe and biology so exhilarating. He did not seek to reduce or constrict our understanding of perspective to only what *we* can perceive, as he saw that humans being are just as restricted in our respective *umwelts* as other organisms are. As Uexkull writes, we can hardly grasp the true meaning of things "if we relate it only to ourselves" (Uexkull 1934/2010, 142). For Uexkull, each life-world, each *umwelt,* represented a "new world" where we must consider the living organism as a subject that affects it and perceives it as such.

It is clear that Uexkull is positing a perspectival realism akin to that which both Heidegger and HN are espousing. Truth, for the subject of such environment, or *umwelt* is inherently mere appearance. It is perspectival. The *umwelt* to Uexkull is the "ring" or the "angle of vision" to Heidegger and HN. However, unlike

---

8. He writes, "In the dog world there are only dog things, in the dragonfly world there are only dragonfly things, and in the human world there are only human things. Even more so, Mr. Shulz will only encounter Shulz things and never Mr. Meyer things, just as Mr. Meyer will not encounter Shulz things" (Uexkull 1934/2010, 64).

9. Thomas Nagel would later adopt this very same position in his foundational essay "What it is Like to Be a Bat." Personally, I find it ironic that for all of Heidegger's reference to the poverty-of-world of bees, pollinating insects actually detect flowering plants through signs otherwise invisible to those such as ourselves, as we cannot see in the ultraviolet range below 400. Who is poor-in-the-world now? (Sagan 2010, 22).

Heidegger and HN, Uexkull does not see such perspectival appearance as a negative "error" or "barrier." A simple organism is not lacking any form of truth or world by having a simple world, and a "multiform" organism is not gaining in any particular sense a "better" or "wider" world by having a more articulate one (Uexkull 1934/2010, 50).

It is this neutrality, or rather, fundamental humility that pervades Uexkull's work and that Heidegger either fails to recognize or chooses to diametrically warp and oppose. For Heidegger, humans are world-forming, our *umwelts* are richer, more extendable, and according to HN, transcendable through art. All other nonhuman animals are subsequently poor-in-world, and are constricted and circumscribed by their respective *umwelts*. Humans can hop the wall so to speak, whereas every other animal remains entrapped and impaired behind it. In this way, Heidegger explicitly co-opts Uexkull's concept of *umwelt,* the environing world, describing it as "nothing other" than his own theory of the nonhuman animal's disinhibiting ring (Heidegger 1983/1995). In *FCM* it is clear that Heidegger is attempting to take on the same project as Uexkull, to refute mechanist claims of biology and explore the relation of organisms to their environment from the perspective of that organism. However, unlike Uexkull, Heidegger does not see the nonhuman animal as a subject in its environment in the same way human beings are. Recall that nonhuman animals are "captivated" by drives and behaviours, the bee is so "captivated" in its suckling of honey, that on Heidegger's account it does not notice it has been cut in half. Notwithstanding the flaws of this alleged 'experiment', Heidegger, despite his emphasis on rejecting the mechanist view of biology, has posited merely a new flavour of the same old theory. Nonhuman animals, while not mechanical in the literal sense, are still delimited by "instinctual" and "behavioural" drives, that captivate and control them. It is these drives and capacities for captivation that prevent the organism from being world-forming in the way that human beings are. It is a mere regurgitation of the old (Cartesian) rhetoric. Nonhuman animals are reactionary, and if they are feeling subjects in any sense, their feelings and subjectivity are controlled by their physiology and instincts in such a way that human beings, due to a form of superiority, so easily evade.

Even more importantly, Heidegger also misreads Uexkull's fundamental point; Uexkull is not differentiating the *umwelt* of humans from other animals on the basis that humans are unique, or superior, in fact, in most of Uexkull's writings,

humans are rarely considered. On the contrary, Uexkull highlights how other animals stand in relation to their world differently from us because he wants to explore the potentiality for an infinite multiplicity of perspectives rather than an anthropocentric account of objectivity and reality.

The Problem of Transposition

The second problem with Heidegger and HN's position, is a problem encountered by most philosophers attempting to delineate human superiority over nonhuman animals-that is, the problem of transposition. How can a human philosopher make claims about the perspective and abilities of a nonhuman animal? This is why contemporary biologists focus on empirical data and observation, and even with that, only shady inferences are made using knowledge of anatomy, behaviour, and evolution. Uexkull worked hard to construct an understanding of how different sensory and biological capacities in animals could produce differing perspectives. This is what Uexkull described as the walk into "unknown" and "invisible" worlds, yet Uexkull also was adamant regarding the limits of this, we can watch animals and understand *how* they do things, perhaps even functionally speaking, *why* they do things, but external observation can only take scientists so far.

In the landmark essay, "What it is Like to Be A Bat", Nagel takes this problem of subjective experience to its full conclusion, subjectivity is what it is like to *be* that organism, and therefore, cannot be adequately analyzed by any material or external account. For Nagel, it is useless to attempt to evaluate or reduce this mental and subjective phenomena in a way that fails to deal with its inherent subjective qualia (Nagel 1979/2018, 167). For this reason, explanatory systems such as "functional drives" or "intentional states" or even basing it off the causal behaviour in human experiences, does not exhaust its analysis (Nagel 1979/2018, 167). Therefore, reductionist, psychological, or physicalist accounts for subjectivity do not extend far enough to ascertain the inherent qualia of an organism's perception and experience.

Nagel then gives the example of a bat, which like Uexkull's tick, has a very different physiology and sensory world than human beings. Bats use echolocation to navigate, a sense that humans do not have. So what is it like to be a bat? Nagel says this is a question that presupposes the answer; when we try to imagine what it is like to *be* a bat, that is to transpose ourselves into the body and physiology

of a bat, we are doing exactly that, transposing *ourselves* into a bat. Because our own experiences provide the "basic material for our imagination" our range for imagining is subsequently limited, imagining yourself as a bat is as far as we can go, we cannot imagine what it is like "for a bat to *be* a bat" (Nagel 1979/2018, 169). On Nagel's account, our ability to transpose ourselves into other beings (and in many cases, even other people) is thereby fundamentally restricted by the resources of our own minds.[10]

It is here that Heidegger takes a radically different position. For Heidegger, transposing oneself into another means going-along-with what it is and how it is. Heidegger does not view human subjectivity as the metaphysical barrier that Nagel and other philosophers, including myself, see it as. Heidegger asserts that the concept of even "empathy" is mistaken because we are never outside of other beings in the first place (Heidegger 1983/1995). Heidegger sees human beings as *a priori* equipped to transpose themselves into other human beings, because being-there (Dasein) means also being-with others (Heidegger 1983/1995). The capability of transposing oneself belongs to the fundamental essence of human Dasein, which means we can also transpose ourselves into other animals. The nonhuman animal, despite not having "what we call a world" has a "sphere of potential transposability" which means that we as human beings already find ourselves transposed into the nonhuman animal in such a manner (Heidegger 1983/1995).

It is this presumptive skill for transposition that gives Heidegger the authority and ability to make conclusions about the essence of animality, despite the fact that he sees it differing widely from the essence of humanity. While it is not within the scope of this essay to explore more of Heidegger's conception of Dasein, nor this role of transposability with regards to Dasein, to me, this seems like a very convenient justification for what is otherwise a theory based on groundless assumptions. I also find it ironic that Heidegger insists throughout *FCM* that we must pursue the essence of animality through the perspective of

---

10. Human beings have certainly attempted to augment our senses to make up for this lack, eg. through technological telescopes and cameras, however, as Dorion Sagan points out, we will eventually meet structural and insurmountable barriers. For example, blue whales have brains much larger than our own, and communicate across thousands of miles of ocean. This means that the *umwelt* of a blue whale may have "fabulous" and "multi-sensorous" pictures of vast mileage of ocean, yet, even if we had direct access to such imagery, it is likely that we would not be able to neurologically process it because our brains and capacities are too limited (Sagan 2010, 23).

the animal, and not misinterpret it by "crudely adopting" human psychology (Heidegger 1983/1995). Yet, he oversteps this problem by explaining how we as human beings can naturally understand the essence and perspective of other animals. Nonhuman animals cannot transpose themselves, while human beings, miraculously, can! This is regardless of the blatant fact that nonhuman animals are fundamentally *different* metaphysically as Heidegger is positing, but also different physically and anatomically. It defies the limits of common sense to accept this solution to the problem of transposition, unless you presuppose the question by accepting that human beings are superior to other nonhuman animals. However this presupposition, as I have shown, is groundless.

Like Nagel, I see transposition as one of the fundamental problems to addressing the qualia and *umwelt* of nonhuman animals, and even our own interactions with each other. While our sociality allows us the cognitive and social tools to empathize, care about, and attempt to consider other people's subjective experiences and feelings-including those of nonhuman animals, we will always find ourselves falling into the trap of referring back to ourselves and our subjective experiences as the base point of reference. With regards to nonhuman animals and other organisms, perhaps then we are doing them an injustice when we humanize them or project onto them our own emotions and interpretations, yet, this is still better than the alternative of treating them as if they are lesser.[11]

---

11. This is what I refer to as the "Problematic Dichotomy" with regards to our relations with other animals and organisms. Either, we take on a Cartesian stance which reduces animals to their mere behaviours and prescribes a lack of subjective experience (at least to some degree) thereby justifying abuse and exploitation, or we humanize them, which is not only myopic and unfair-given the capacities of many species that likely far surpass human ones, but also untenable in many cases. I would say that most people adopt both perspectives based on context and the sociozoological scale. For example, most would find it absolutely repulsive to consider stepping on the head of a mammal, such as a beloved cat, and crushing it to death-yet most do not blink to consider the same effect on a spider. A discussion of this pervasive sociozoological scale is deserving of an essay in its own right, but this is what is coined by Peter Singer as "speciesism." My solution is a proposed project of humility based on the works of Nagel and Uexkull (among others). Strangely enough, Heidegger avoids the problem of speciesism by merely positing all animals (except humans) as one metaphysical "type", he generalizes to such an absurd degree that he avoids this footfall that other anthropocentrists have gotten caught by.

Heidegger's "Abyss" as "Violent and Awkward"

In Lecture 25, HN avoids any form of comparison between the lizard first mentioned at the beginning of the lecture, and the perspectival shining of the artistic Dasein at the end. It is such a seamless segue because it is already implied that such a comparison doesn't need to be made. However, in *FCM* where Heidegger *is* positing a comparative analysis, he emphasizes repeatedly that he is not attempting to "entail" a "hierarchal assessment" (Heidegger 1983/1995). Even when discussing how the world of man, as compared to other beings, is a "richer" and "greater" one, he counter-intuitively insists that his comparison does not imply "evaluative ranking or assessment" (Heidegger 1983/1995).

To be charitable, lets consider Heidegger's theory as he clearly intends it to be, merely descriptive of the metaphysical difference between humans and nonhuman animals. For Heidegger, the poverty-in-the-world that the nonhuman animal experiences is not a "characteristic property" but the very way in which the animal exists (Heidegger 1983/1995). Yet, this description *is* normative, it is based on other animals having *less* world or access to world, than human beings do. It denotes an essential valuation. Also, recall that Heidegger is positing a metaphysical thesis, not a zoological one, that is, he is not concerned with evaluating the differences and similarities between species, he is ascertaining an absolute and metaphysical divide. His only "proof" of that metaphysical divide is based on a pseudo-scientific understandings of nonhuman animal "capacity" and perspective, which is based on the assumption that human beings are capable of transposing themselves to actualize that knowledge in the first place. Human beings are therefore world-forming while other animals are perspectively limited. Yet this conclusion is ascertained by presupposing that we human beings are metaphysically different and superior, because only we are capable of the transposition that allows us to understand this limitation of other animals at all.

I argue that this this is quite the circular argument, and that Heidegger begs the question with his first assumption-that there is something metaphysically different (and absolute) between the essence of animals and the essence humans (who are also animals). Regardless of his protestations that he is not making an evaluative or hierarchal argument for human superiority, he nonetheless makes one from his first premise.

Even more contradictorily, Heidegger contests that his conclusion "does not mean that [animal] life represents something inferior or some kind of lower

level in comparison with the human Dasein" (Heidegger 1983/1995). Instead, he writes in suddenly humble terms, "life is a domain which possesses a wealth of openness with which the human world may have nothing to compare" (Heidegger 1983/1995). I do not need to spend time explaining how this statement not only contradicts his conclusion that anything that is not human, such as animals, are comparatively poor-in-the-world, but that it undermines his whole premise of a metaphysical division of essence and the superiority of Dasein. Even more strangely in the following paragraphs, Heidegger contradicts this apparently humble statement and all of his protestations hitherto, by again reiterating that the animal, due to the essence of its being, is separated from man by an "abyss" (Heidegger 1983/1995).

It is this contradiction and circularity which Derrida so aptly calls "violent and awkward" (Sagan 2010, 29). It is violent because it is such a blatantly absolute and Anthropocentric divide, and it is awkward because Heidegger never manages to get his initial argument for this comparison off the ground. He begs the question from the start. Human and nonhuman animals are metaphysically different, and he uses that premise to only further his exploration and consecration of human Dasein, which on his account, is at its essence, superior.

CONCLUSION

In this essay I have undertaken the task of considering Heidegger's conception of humanity in the world in which it dwells in comparison to that other beings. I have shown that this conception of man as world-forming is bulwarked by the relative inferiority of all other animals as poor-in-world. I have sought to trace the foundations of this theory from its beginning in *FCM* to explore how Heidegger uses it to interpret and co-opt Nietzsche's perspectival realism and the value of art in Lecture 25. I have also considered the merits of Heidegger's theory, and I have argued that not only does it represent a misreading of Uexkull, but that it presupposes the question with the role of transposition, completely ignoring the problem of subjectivity and emergence. By positing such an absolutist and inherently Anthropocentrist divide, Heidegger has only succeeded in his comparative account to further his explanation of the human Dasein, and by doing so, positing it as superior, regardless of his protestations to the contrary. For someone who set out to reject mechanist or reductionist conceptions of

life and being, Heidegger has managed to posit the same theory only dressed up in new terminology. I conclude that Heidegger then, has not only done an injustice to other animals by positing a circular argument premised on their deficiency, but that he has done an injustice to himself, and his philosophy of human Dasein and being. Heidegger should have contained his phenomenology to that which was accessible to him, that of the perspective and realities of human beings, and not attempted to overcome that divide by presupposing some great metaphysical difference or "abyss." Instead, his interpretation of the nonhuman animal represents an affront to the Uexkullian approach, that being one of humility towards the great and exhilarating potential for the multiplicity of perspective across the contingencies of life and even, the wider universe.

REFERENCES

Heidegger, Martin. 1961/1991. "25. The New Interpretation of Sensuousness and the Raging Discordance between Art and Truth." In *Nietzsche*, translated by David Farrell Krell, 211-20. Vol. I. New York, NY: HarperOne.

Heidegger, Martin. 1983/1995. *The Fundamental Concept of Metaphysics.* Translated by William McNeil and Nicholas Walker. Indiana University Press.

Kuperas, Gerard. 2007. "Attunement, Deprivation, and Drive: Heidegger and Animality." *University of San Francisco.*

Nagel, Thomas. 1979. "What Is It Like to Be a Bat?" In *Mortal Questions*, 165-80. Cambridge University Press.

Sagan, Dorion. 2010. "Introduction." In *A Foray Into the Worlds of Animals and Humans*, 1-34. University of Minnesota Press.

Tye, Michael. 2017. *Tense Bees and Shell-Shocked Crabs.* Oxford University Press.

Uexkull, Jakob Von. 1930/2010. *A Foray Into the Worlds of Animals and Humans.* Translated by Joseph D. O'Neil. University of Minnesota Press.

compos mentis

Issues Surrounding Attempts to Ground Political Obligation in Gratitude

Joshua Doland

University of California, Los Angeles

ABSTRACT

In the following paper I examine a theory of political obligation which seeks to ground a citizen's duty to comply with their government in debts of gratitude. After presenting a historical account of the mechanisms of this theory, I turn to its modern reception by examining works by John Simmonds, A.D.M. Walker and George Klosko. Both Simmonds and Klosko are skeptical that gratitude is an appropriate principle to ground political obligations; It seems to them that debts of gratitude aren't strong enough or necessarily appropriate for this task. In order to resist the critiques of Simmonds and Klosko, Walker seeks to adapt the theory for modern audiences. The majority of this paper will examine the dialectic between Simmonds, Klosko and Walker. After the basic dialectic has been put forth, I will conclude that although theories which rely on debts of gratitude offer a convenient and romantic way to view a citizen's political obligations to their government, these theories ultimately prove unsuccessful, as they misunderstand what gratitude fundamentally is.

One significant question within Political and Legal Philosophy deals with whether citizens have an obligation to comply with their government. In more contemporary philosophical thought there is a good deal of skepticism towards whether these political obligations actually exist. Any theory that does argue for them has the difficult task of grounding this obligation in some existing conception. One theory that attempts to do so grounds political obligations in debts of gratitude. Although the view may seem initially plausible, upon further scrutiny of its concrete formulations, it cannot adequately ground obligations citizens might have to their government.

Within the following essay I will seek to convey the central tenants of the Argument from Gratitude and interpret it in its most favorable light. Once the central view is laid out, I will ultimately claim that the debts of gratitude generated under the theory are not sufficient to ground political obligations.

In order to do so, within the first section I will offer a cursory glance at the historical interpretation of the Argument from Gratitude. In the second section I will rehearse two pressing critiques that John Simmonds brings forth in response to this historic view. In the third section I will introduce a more plausible view of the Argument from Gratitude as presented by A.D.M. Walker and attempt to respond to the critiques put forth by Simmonds. Within the fourth section I will offer critiques against Walker's revised theory of gratitude as presented by George Klosko. In the subsequent section I will then provide Walker's responses to these critiques. Finally, I will weigh into the debate and argue that Walker's revised theory of gratitude, while more plausible than others, still doesn't sufficiently ground political obligation. I argue that this is due both to the lack of independence within the theory and because of its reliance on an implausible interpretation of what gratitude is.

SECTION I: THE ORIGINAL ARGUMENT FROM GRATITUDE

The initial attempt to ground Political Obligation in debts of Gratitude comes from Plato's Crito in which he argues that the citizen-government relationship is akin to the child-parent relationship and as such, the citizen has an obligation to obey the laws of the government (Tredennik 1954, 92).

The underlying thought is often referred to as the principle of reciprocation. Simmonds characterizes the thought thusly:

> The receipt of a benefit puts one under an obligation to requite one's benefactor, to confer on him a benefit in return for the benefit one has received from him. (Simmonds 1979, 143)

Essentially, if an entity, in this case the state, confers upon a recipient, in this case a citizen, a benefit, that citizen ought to offer an equal benefit back to the state.

From this principle of reciprocation, Plato then argues towards the existence of political obligations. Walker runs the full argument as follows:

1. The Person who receives benefits from X has an obligation to requite or make a suitable return to X. (This is the principle of reciprocation)
2. Every citizen has received benefits from the state.
3. Every citizen has an obligation to make a suitable return to the state.
4. Compliance with the law is a suitable return.
5. Every citizen has an obligation to comply with the laws of his state.

As will be explored within the following section, this initial attempt at putting forth a theory of political obligations grounded in gratitude is deeply flawed for a number of reasons. It should still be noted however, that this view is attractive insofar that it offers a generally communal and friendly picture of the relationship citizens hold with their government.

SECTION II: SIMMONDS CRITIQUE OF THE ORIGINAL ARGUMENT FROM GRATITUDE

With the original argument from gratitude now laid, within this section we will explore Simmonds critiques of it. For the purposes of this essay, the focus will be placed initially on Simmonds argument against the first two premises, and subsequently on Simmonds argument against the move from the fourth to the fifth premise.

Simmonds first contests the first principle, questioning whether the mere receipt of a benefit from a benefactor is sufficient to oblige the beneficiary to reciprocate. Simmonds essentially argues that "the person who receives benefits from another does not always have an obligation to requite his benefactor, but

only in certain circumstances" (Walker 1988, 194). Simmonds isolates the following five principles that he thinks need to be present for a beneficiary to be under a debt of gratitude to their benefactor:

1. The Benefactor "must have made some special effort or sacrifice, or incurred some loss, in providing the benefit in question" (Simmonds 1979, 170).
2. In conferring the benefits to the beneficiary, the benefactor must have acted intentionally, voluntarily and without disqualifying motives, like acting out of self-interest (Simmonds 1979, 171-172).
3. The benefit can't be forced upon the beneficiary (Simmonds 1979, 175).
4. The beneficiary must want the benefit which is granted (Simmonds 1979, 177).
5. The beneficiary must not want the benefit not to be provided by the benefactor (Simmonds 1979, 178).

Simmonds think that while a government may fulfill the latter three principles, it will be unable to act under the prior two. In arguing this, Simmonds contends that a state doesn't make any special effort or sacrifice to offer benefits to its citizens, nor is it the right kind of entity, being an institution, to have motives that can properly be said to be intentional, voluntary and genuinely altruistic.

With this revision of the principle of reciprocation, the second principle in the initial argument is under threat. If the state is incapable of providing benefits in the ways necessary, then it can't be said that every citizen has received a benefit from the state.

The second main critique Simmonds offers deals with the move from principle 4 to principle 5. Simmonds argues that even if a debt of gratitude was owed by the citizens to the state, it needs to be shown that compliance with the law is in fact a suitable return. Simmonds is basically wondering why one should pay their debt of gratitude through compliance with the law instead of through some other means (Walker 1988, 194).

Simmonds then argues, that even if compliance with the law is shown to be a suitable return for the benefits conferred on the beneficiaries, it needs to be shown to be the "uniquely suitable" return (Walker 1988, 195). Otherwise, it doesn't follow that every citizen is obliged to comply with the law, as there may be other equally adequate ways of fulfilling a debt of gratitude to the government.

With Simmonds two critiques of the initial formulation of the Argument from Gratitude now put forward, within the next section we will analyze Walker's response.

SECTION III: WALKERS RESPONSE AND REVISED ARGUMENT FROM GRATITUDE

In defending the Argument from Gratitude, Walker seeks to distance himself from the historic view put forth previously and instead adapts the theory based off his understanding of what gratitude is.

For Walker, Gratitude should be understood as "a set of attitudes…towards the benefit and towards the benefactor…[showing] proper [appreciation of] the benefit and [having] goodwill and respect for [the] benefactor" (Walker 1988, 200). Walker admits that the notions of appreciation, goodwill and respect he relies on are nebulous. He next draws from this understanding of Gratitude, two obligations that follow: 1) To demonstrate to the benefactor that you have the correct attitudes towards them and 2) to not act contrary with the possession of these attitudes (Walker 1988, 200).

While this first obligation is largely symbolic, the second is to be concretely manifested when the benefactor is in need. This second obligation, unlike the first, is a continued requirement that can't definitively be met (Walker 1988, 200).

Walker offers an alternative way to distinguish between these two obligations; The first is an obligation to show that the beneficiary has goodwill towards the benefactor whereas the second obligation is an ongoing set of actions stemming from the actual goodwill the beneficiary holds towards the benefactor (Walker 1988, 200).

It is important to note that a political obligation is primarily an obligation to act in a certain way rather than to act in a certain way while also under a certain attitude or doing it for a certain proper reason.

From his characterization of Gratitude and goodwill, Walker next examines four candidate requirements stemming from goodwill that could potentially serve to ground political obligations. This list of obligations isn't meant to be exhaustive, but rather is simply a list of promising candidates. Walker argues goodwill requires the beneficiary to:

A. Help the benefactors when he is need or distress and one can do so at no great cost to oneself.
B. To comply with his reasonable requests.
C. To avoid harming him or acting contrary to his interests.
D. To respect his rights. (Walker 1988, 202)

In order to select which of these principles ought to be used to ground political obligation, Walker argues that it must be both relevant and independent.

In order for a principle to be relevant in the right way it needs to "bear on the issue of a citizen's compliance with the law" (Walker 1988, 202). In other words, political obligation must be able to be seen as a version of compliance with the principle.

In order for a principle to be independent it must not hinge upon another explanation which already supposes a political obligation (Walker 1988, 202). Essentially, if the Argument from Gratitude is to work, the principle chosen must do the explanatory work and not rely on a pre-existing foundation for political obligation. If it does so, then whatever principle it is relying upon is the grounding principle of Political Obligation rather than gratitude.

Walker argues that the first two principles, A and B, are not able to ground political obligations because they are not relevant. This is to say that Walker doesn't view a Political Obligation as akin to helping a benefactor in need or complying with a reasonable request.

Walker rejects the final principle, D, as it runs afoul of the independence principle. Principle D in invoking a respect for the benefactor's rights presupposes that the benefactor has rights to be respected. If this is the case, then some other explanation is necessary to provide the foundation on which the state can claim to have rights.

Walker takes principle C, the obligation to not harm or act contrary to the states interest, as the principle that grounds political obligation. The principle is relevant insofar that it is, generally speaking, within the interest of the state for its citizens to abide by the laws. Walker does argue that in some cases it may not be in the interest of the state for its citizens to abide by the law, thus opening an interesting space for civil disobedience, but these cases should be rare and are not the focus of the current task at hand.

Walker takes principle C to be independent as it doesn't rely upon another theory to explain the source of political obligation. Rather, political obligations are

grounded in our obligation not to harm our benefactor, which is grounded in our debt of gratitude.

Walker's revised form of the Argument from Gratitude should be taken as follows (Walker 1988, 205):

1. The person who benefits from X has an obligation of gratitude not to act contrary to X's interests.
2. Every citizen has received benefits from the state.
3. Every citizen has an obligation of gratitude not to act in ways
4. Noncompliance with the law is contrary to the state's interests.
5. Every citizen has an obligation of gratitude to comply with the law.

With Walker's alterations to the theory finally laid out, we can now turn to examine whether the theory is suited to answer the objections previously raised by Simmonds.

As fore-mentioned, Simmonds initial critique of the Argument from Gratitude questions whether debts of gratitude are formed from the government's conferral of benefits to its citizenry. Simmonds argued that while the latter three conditions for the formulation of a debt of gratitude may be fulfilled, the government is incapable of fulfilling the first two, and thus no debt is created.

Central to this critique is the issue of whether or not one can be said to feel gratitude towards an institution or only towards individuals that make up an institution. While Simmonds is sympathetic to the latter view, Walker argues that one can coherently feel grateful towards an institution.

The common objection to Walker's view is that when an individual claims to be grateful to an institution, in reality they are actually grateful to specific individuals who make up that institution. Walker dismisses this point as an objection however, noting the following two points (Walker 1988, 198):

1. Even if gratefulness to an institution is derivative for gratefulness for the people who make up the institution, it is still coherent to say that one is grateful to the institution.
2. In the case of the government, one could generate a political obligation out of gratefulness to a collective of all citizens. This doesn't interfere with the Argument from Gratitude.

By arguing the above two principles, Walker can dodge this initial objection. The second component of this critique of the principle of reciprocation as applied to

the government is whether or not it can act with the correct intention or motive. As an institution after all, it isn't a thinking being but rather a collective.

Walker accepts this critique but argues that the institution have a sufficient analogue in their stated purpose or function (Walker 1988, 199). While a government may not be said to act out of a motive to ensure the safety of its citizenry for instance, it can be said that the institutions purpose is to protect its citizenry. Walker views this as a sufficient answer to Simmonds critique.

Walker offers another defense of his initial two principles arguing that even if Institutions don't act purely out of the safety of its citizenry (or any other benefit that is conferred) a partial or mixed purpose or function is sufficient to ground a debt of gratitude (Walker 1988, 208). Though the government operates out of duty, so long as part of its function is also an intentional and deliberate effort to benefit its citizenry, that is sufficient. Walker supposes that if one was swimming and began to drown, and a lifeguard expended a minimum effort to retrieve you from the pool, though the lifeguard acted out of duty and sacrificed practically nothing to save you, a debt of gratitude can still be generated (Walker 1988, 208). Walker finds this situation analogous to the one citizens find themselves in relation with the government.

Simmonds second objection deals with whether compliance with the government's laws is the uniquely suitable return for the generate debt of gratitude.

Walker argues that insofar that a debt of gratitude to the government is a sign of goodwill directed towards the government, then not complying with laws is against the government's interested and thus not done in goodwill. One would be failing to act with goodwill and thus have violated their debt of gratitude.

Walker clarifies that it isn't necessary to actively harbor bad will when not complying with the law, but rather that an attitude of non-goodwill is sufficient to damage the interests of the government Walker 1988, 206). While citizens may have other obligations per their debt of gratitude, compliance with the law will be included in any minimal account.

Finally, Walker anticipates two objections: one dealing with whether non-compliance with laws damages the government's interests; and the other dealing with whether the state truly confers benefits on the citizenry.

The motivation underlying the first objection is roughly that individual defiance of laws doesn't damage the government in a noticeable way. Walker

mounts Parfit's theory in response, arguing that while small acts of defiance by individuals may not harm the government's interest, collective defiance would and thus each individual act should be treated with this in consideration (Walker 1988, 207).

The second objection is grounded in the idea that citizens are in a commercial relationship with the state, and since they pay taxes, the state can't be said to be conferring benefits in a way which generates debts of gratitude.

In response, Walker argues that gratitude may have a place in commercial relationships, that the citizen-government relationship is not a commercial one and that we owe debts of gratitude to other citizens and not necessarily the state. The government in this model is a redistributive agency that provides service individuals could not such as international security. From this set of arguments, Walker argues taxes don't nullify the debts of gratitude that are generated (Walker 1988, 209).

Within this section I laid out Walker's Argument from Gratitude and responded to the critiques mounted by Simmonds. Within the next section I will present Klosko's critiques of this revised theory.

SECTION IV: KLOSKO'S CRITIQUES OF WALKER'S REVISED THEORY

Klosko has two significant arguments against Walker's theory: 1) Walker's use of Parfit's theory inadequately answers the question of whether noncompliance is against the interests of the government and 2) That debts of gratitude formed without Simmonds five conditions present are too weak to ground political obligations.

Klosko regards the first critique as relatively insignificant though I will offer it additional attention in a subsequent section. Walker had previously argued that while an individual not complying with the law may not relevantly damage the interests of the government, one must consider the collective effect that would be had if a multiplicity of people didn't comply with the law. This rehearsal of Parfit's argument is reminiscent to Kant's formula of universalization.

In order to argue against this, Klosko supposes a situation in which only one person doesn't comply with the law by not paying their taxes. Since we cannot evaluate her actions collectively, as she is the only non-compliant, Klosko argues

that Walker must take her not to be obligated to pay (Klosko 1979, 354). This seems intuitively incorrect though, it seems the individual has still done something wrong in not paying her taxes. Klosko gestures towards fairness as a better explanation for this intuition then promptly drops the critique (Klosko 1979, 354).

Klosko attends closely to the subsequent objection. Klosko reasons that a political obligation "can be overridden by conflicting moral or religious beliefs, [but] in most cases…should be presumed to hold" (Klosko 1979, 354). Essentially, if Political Obligations exist they are forceful and give strong reasons to pursue certain political actions.

Klosko's concern is that Walker's model of debts of gratitude is too weak to support such Political Obligations. Under Walker's model, these obligations "would be overridden frequently, not just in unusual circumstances" (Klosko 1979, 355). This critique is nested within Walker's theory as can be seen by his example of the council member with a benefactor and an unrelated duty. Walker argued that while "[the council member] may feel an obligation not to vote for a proposal which would significantly damage [the] benefactor's interests…this obligation will almost always be outweighed by [the council member's] duty…to some wider good" (Klosko 1979, 355).

Klosko takes this as evidence of the general weakness of debts of gratitude. Such debts are thus not sufficient to ground political obligation. Though Klosko cedes, unlike Simmonds, that debts of gratitude can be formed without the complete fulfillment of all five of Simmonds principles, he ultimately argues the principles are necessary to create debts of gratitude strong enough to ground Political Obligation (Klosko 1979, 356).

In order to dispel this critique, Walker will need to be able to show that there exists a debt of gratitude that can be formed between a citizen and the government strong enough to ground Political Obligation. In the next section we will turn to see how successful Walker is in this response.

SECTION V: WALKERS REBUTTAL OF KLOSKO'S CRITIQUES

Walker, likely for the sake of brevity doesn't address Klosko's initial concern.

As for the second concern, in order to dispel it, Walker needs to show the existence of at least one obligation of gratitude that won't be typically overridden by other considerations and thus can ground Political Obligation.

Instead of providing a concrete example however, Walker merely gestures towards what Hume and Kant have to say on the importance of debts of gratitude. Walker notes that Kant "named ingratitude along with envy and malice as one of the three sins" and that Hume though that "of all crimes…the most horrid and unnatural is ingratitude" (Walker 1989, 363). While a call to authority of these two renounced philosopher rhetorically inclines one to believe gratitude is important in some sense, it does little in the way of actual convincing one of that.

To be charitable, Walker may not have engaged deeper with the objection due to the space constraints placed upon his arguments, yet the response he offers is simply unsatisfactory.

SECTION VI: REMAINING ISSUES WITH THE ARGUMENT FROM GRATITUDE

Having now considered Walkers response to Simmonds as well as the ensuing debate by Klosko, it seems prudent to weigh in and establish that Walker fails to adapt the Argument from Gratitude sufficiently in order to ground Political Obligations.

Aside from Walker's failure to produce a counterexample of a generally strong gratitude-based obligation, he also relies too heavily on Parfit's arguments. Take the case where only one individual doesn't comply with the government and the government's interests aren't damaged. It still seems that individual is doing something wrong and is obliged to reconcile their action. Walker can only account for this through its collective cost, and yet separate from the universalization of this action, the individualized instance seems wrong.

Put another way, there is a fairness concern over whether the individual in question is justified in not complying with the law. Since concerns of fairness better explain the individuals obligation instead of gratitude, it seems this might be a more productive line to follow.

Separate from this, Walker's theory reads a lot of content into the conception of gratitude. On a more classic Strawsonian picture, gratitude is an attitude individuals adopt in response to the good will that others show them. This good will is expressed by exceeding the expectations and demands put on the person. The individual who feels gratitude isn't necessarily indebted to the person who benefitted them nor are they obligated to reciprocate. If this actual expectation

was a condition upon the benefit that was conferred, the individual wouldn't feel gratitude, instead this would just be another instance of a contractual exchange of benefits. The attitude of gratitude arises when there isn't an expectation for an individual to do something, and they still do it. Separate from this one can determine whether the action is praiseworthy or whether gratitude should be expressed but this is a different question.

Although the Argument from Gratitude doesn't work, at least in its presented formulation, its friendly perspective on the relationship between the citizen and their government remains an attractive idea.

REFERENCES

Klosko, George. 1989. "Political Obligation and Gratitude." *Philosophy & Public Affairs* 18 (4).

Simmonds, John. 1979. *Moral Principles and Political Obligations*. Princeton University Press.

Tredennik, Hugh. 1954. *The Last Days of Socrates.* London: Penguin Books.

Walker, A.1988. "*Political Obligation and the Argument from Gratitude." Philosophy & Public Affairs* 17 (3): 191-211.

Walker, A. 1989. *"Obligations of Gratitude and Political Obligation." Philosophy & Public Affairs* 18(4): 359-364.

compos mentis

Everyone's a Critic: The Importance of Accountability in Artistic Engagement and Creation

Emma Fergusson
Kalamazoo College

ABSTRACT

What does it mean to properly engage with an artwork? What is the function of aesthetic perception in our daily lives? What are the dangers of passive spectatorship and how can a more meaningful understanding of various artistic mediums lead to a greater understanding of the world at large? This paper explores the importance of proper engagement with artworks—including but not limited to film, literature, and other visual mediums—and the function of art criticism as crucial to the act of viewing an aesthetic presentation. I claim that artworks actively articulate and sensuously present their meaning or significance, but only to an active participant who reflexively works through that presentation. This level of responsible viewership serves, not only to qualify the depth of one's engagement with the artistic object at hand, but also with particular worldly circumstances. Thus, one's ability to commit to an artwork serves as a gauge of one's ability to commit to their own life. While there is no necessarily right or wrong understanding of an artwork, there is a true and a false way of engaging in that work of interpretation.

An artwork embodies a complex system of aesthetic materials and techniques which actively work to present the truth, the meaning that lies within and beyond its presentation. Anyone can look at a painting, scan the words of a text, watch the screen of a movie theater without gaining much beyond the mere appearance of color, shape, character or plot. But how can we encourage an artwork to open itself up to its audience? How can we truly experience that which the artist has created, which the artwork longs to reveal, which the viewer needs to understand? We must serve as our own art critics—not in order to criticize, to place a value upon an artwork, or to ascribe it a label of "good" or "bad," but rather in order to scrutinize, to unpack, to engage more deeply, and thus to gain greater insight and understanding of that significance contained within a successful artwork. The viewer of an artwork has a duty—a duty to attribute a work with the kind of power that it wields to create change, to expand minds, to alter perspective, and to develop judgement. An artwork cannot accomplish such a task on its own—it requires equal involvement from its participants. When an individual engages only passively with an art object, they lose out not only on the aesthetic qualities of the artwork, but also on the opportunities for greater engagement with the world. Art is not only beneficial, but imperative to each individual's capacity for change, for empathy, for reasoning, and ultimately for the proper participation in one's own life. If a person cannot commit to an artwork in a meaningful way, he cannot hope to commit to his own relationships, to his own responsibilities, to his own circumstances in a meaningful way.

I. AESTHETICS OF APPEARING

In *Aesthetics of Appearing*, Seel discusses not only the functions of the artwork as an actively appearing aesthetic presentation, but also the importance of perception, and the ways in which aesthetic perception serves both the artwork and the perceiver:

> To apprehend something in the process of its appearing for the sake of its appearing is a focal point of all aesthetic perception. Of course, this perception frequently goes way beyond a mere execution-oriented sensing. In particular, the perception of artworks necessarily incorporates interpretive and epistemic attentiveness.

> However, the aim of this interpretation and knowledge is first and foremost to be with the articulating appearing of the objects…. Aesthetic attentiveness to what happens in the external world is thus an attentiveness to ourselves too: to the moment here and now. In addition, aesthetic attentiveness to the objects of art is frequently an attentiveness to situations in which we do not find ourselves and perhaps never will: to a moment now and never. (Seel 2005, 15-16)

Seel identifies the act of appearing as the primary function of an artwork: it shows what it is to convey. The act of perception—of merely registering the existence of an aesthetic presentation—thus serves as the most basic function of the viewer. However, this basic level of perception must dig deeper, beyond the initial reactions of the senses, beyond the mere acknowledgement of the presentation of an artwork. One cannot (or should not) passively consume the movie, novel, play, painting, sculpture, etc. There must be other muscles at work. There must be deeper levels of attentiveness—a churning of thoughts, ideas, connections, and ultimately an attempt at understanding (or at least digesting). This is not to say that one should multitask—consider multiple unrelated and opposing ideas which detract from the artwork, or which distract from the true task at hand. Seel warns against this sort of false engagement. Rather, these deeper levels of perception must maintain a primary focus upon the aesthetic presentation at hand. As Seel suggests, the artwork is in the process of articulating as we perceive it—this is an action, a movement, and one which the viewer must match in order to properly engage. There exists a duality to the act of engaging with the artwork: as the artwork presents and thus articulates, articulates and thus communicates, the viewer must observe and thus perceive, perceive and thus engage, engage and thus interpret. Without the adequate attention of the attentive viewer, the artwork remains stuck, trapped, frozen—silenced, and incapable of communicating its truth.

Stemming from this notion of a dual movement comes the notion of a dual result. Specifically, the idea that in properly engaging with an artwork, the viewer not only unpacks the systems at work in the piece itself, but also commits to a greater level of engagement with their own life. There is an intangible facet to this level of perception: the viewer may engage with and experience a scenario which exists outside their own life, while simultaneously achieving a stronger

grasp upon their own reality. The viewer reaches a level of understanding which exists beyond physical form, memory, and reality. They gain access to a sense of truth—not truth in the sense of reality, in the sense of events or experiences that have actually happened, but, rather, truth in the sense of honesty, of empathy, of human understanding as a result of aesthetic communication.

Seel establishes artworks—and thus the process of an individual's interaction with an artwork—as distinct from that of other objects of appearing. He labels them as objects of "different appearing"—which is to say, different from other, non-art, aesthetic objects, e.g., trees, the gears of a clock, a plastic bag blowing in the wind, etc. Seel writes:

> Objects of art are objects that are subject to an aesthetic treatment different from other kinds of aesthetic objects. Moreover, they are objects that deserve or do not deserve this treatment. In contrast to all objects of mere appearing and some objects of atmospheric appearing, the status of artworks is a *normative* status. They are objects that *merit* being experienced aesthetically, or being candidates for this recognition. It is only within the framework of such an evaluation that they can come to appearance as works of art. (Seel 2005, 110)

Approaching and interacting with an art object is necessarily distinct from interacting with objects that appear as states of affair: i.e., while the former has been actively crafted and serves to communicate through a system of highly intentional artistic techniques, the latter exists in a state of mere appearing which lacks the intentionality—and probably the complexity—of an art object. Only a different kind of understanding allows the unique elements at work within an artwork to appear. An artist has intentionally crafted the art object, has constructed a significance, which requires thoughtful interpretation. An object's status as an artwork subjects it to certain scrutiny. The art object earns the right to recognition as an artwork, and thus to in-depth attentiveness, through its creation, and through the fundamental presentation of its aesthetic qualities. An aesthetic object becomes an artwork when it is viewed as such by the attentive perceiver. Seel suggests that "Not types of sensuous presence but rather types of conception and interpretation are decisive for the art status on objects." The material which makes up the artwork—be it paint, canvas, clay, ink, film—does

not determine its status as artwork. The status of artwork exists beyond material presence, beyond simply placing these items in a certain order, or shaping them, or mixing them together. The truth that drives these aesthetic constructions, the conception behind their creation, the understanding drawn out by the viewer of the presentation, has more to do with an object's status as an artwork: "The sensuousness of the material, which an art object can share with any other object, undergoes a metamorphosis into a state that it does not share with other kinds of objects" (Seel 2005, 106). In this way, the difference between a twisted, discarded scrap of metal left at the dump—which may exist as an aesthetic object—and the twisted scrap of metal reshaped and crafted into something new, or perhaps placed alongside other discarded objects in a sort of collage by an artist—thus solidifying the scrap as an art object—lies within a kind of transformation that takes place at the hands of the artist. As Seel terms it, a "metamorphosis" takes place. Objects constructed from the same materials do not contain the same properties of significance, do not require the same level of understanding and aesthetic engagement. Art objects exists in a state distinct from that of aesthetic objects, not only because of how and why they are made, but because of the way they must be perceived by the attentive viewer.

Seel further describes this level of understanding:

> One cannot simply perceive the results of all these operations just as one can perceive stones, sounds, and colors. Rather one has to *understand* to a certain degree to what operations they are due and what functions they possess. This understanding leads not away from perception, however, but to a perception that can achieve precisely this—to grasp its objects in the organization of their material as products of a particular kind of operation. (Seel 2005, 107-108)

The greater level of understanding, of appreciation for the movements taking place within the self-articulating presentation of an artwork, the more one registers the artwork's significance, but also grasps the complexity of what the artist has done. Seel dismisses the fear of over-analyzing—of destroying the art by deconstructing it, of stripping it of its distinct, artistic properties by looking too closely into its functionality. Rather, he suggests that this process of more deeply understanding the artistic functions at work constitutes perception at its

most effective. The viewer thus comprehends the artwork not only on the level of its aesthetic appearance, but in terms of the way that appearance articulates itself, the way the sensuous materials interact, the way the artist has constructed the artwork and thus how the pieces of the puzzle have evolved and worked together to present something that exists beyond their mere appearing. The way in which Helen Frankenthaler stains her canvas in one solid plane of color, the manner in which Alfred Hitchcock moves his camera to direct the attention of the audience—these techniques benefit the overall impression of the artwork left upon the viewer, but provide an even more in-depth artistic significance when understood as aesthetic operations in and of themselves.

According to Seel, artworks are "special *kinds* of presentations.... They are made in order to be grasped as presentations of a particular kind. It is their primary functions. Their materials are organized so that they present *themselves* in such a manner that we can find *something* presented by them." He goes on to say that "What is important is this self-presentation. Artistic objects exhibit themselves, in the precise organization of their material, in order in this way to bring something to presentation" (Seel 2005, 108). The major function of the artwork—the reason for its existence—lies in its ability to present itself—to articulate its aesthetic elements to the viewer. The artwork exists as a system, containing both the truth which it desires to express, and the means to articulate that truth to an active, attentive viewer. In a novel, words present something; in a painting, the colors and the brushstrokes; in a film, the images on the screen and the way in which the filmmaker has cut them together. The twisted, discarded scrap of metal requires shaping, or framing, or coloring, or reconstructing in order to function as a self-presenting art object. The art object, thus, requires the effort of the artist to communicate, and the effort of the perceiver to decode. Without this careful consideration, artworks "are what they are in aesthetic perception, but they are not what they *reveal* to an understanding and interpretive aesthetic perception" (Seel 2005, 109). In other words, while the aesthetic qualities of the artwork exist openly, without obstruction, to any passerby, the artwork only reveals—or actively articulates—its significance to those who perceive within, and thus beyond.

What purpose is there in distinguishing artistic objects from aesthetic objects, or from the sensuous materials that make them up? What benefit does this additional level of attention provide, and why do we bother wasting our energies

at all? To get at the heart of Seel's conception of the art object, and thus to extract the essence of an artwork's function and significance, artworks,

> **Are ascribed the status of constellational presentations—the status of objects that can bring complex human conditions to light in the medium of their appearing.** They are in this respect "constructs of spirit," to use Hegel's language. "The spirit of artworks," Adorno writes in his *Aesthetic Theory*, is the spirit "that appears through the appearance." This spirit, Adorno elaborates, which "infiltrates" the sensuous appearance of a work, cannot be cognized independently of this appearance; but it must not be equated with it, just as it must not be equated with the intention of the artist. (Seel 2005, 111) (My emphasis.)

There exist few words to sufficiently describe that which artwork accomplishes, which cannot be found elsewhere, aside from the human soul itself. Seel suggests "spirit"—a notion that works on multiple levels: the artwork stems from the "spirit" of the artist; the artwork itself embodies this spirit and strives to articulate it; the viewer perceives, recognizes this spirit, and thus finds it within herself. In expressing "complex human conditions" within a constellational presentation—i.e. an art object in which the materials, conceptions, techniques and truths work to present this "spirit"—the spirit of the work remains tethered to its aesthetic appearance, but not defined by it, or contained within it. The constellational presentation does part of the work, the viewer does the rest. What the viewer interprets may differ from that which the artist has intended. But the significance of an artwork lies not within the seamless communication of ideas—not within "right" or "wrong"—but within that significant truth, that spirit, which translates to each individual in its own way. To experience an artwork in this way is to exist simultaneously beyond and within—to find greater contact with the human spirit, the human experience, as well as to see beyond that which presents itself in the form of an art object. One who takes up the work of attending to such an artwork—such a constellation— also takes up the prospect of greater understanding of one's own experience, or the experience of others, and thus opens herself up to a greater kind of existence in which the barriers between human souls begin to melt away.

II. FILMED THOUGHT

The role of an artwork's perceiver receives a deeper interrogation in Pippin's *Filmed Thought*. Pippin's position on the function of the spectator of an artwork—more specifically, the viewer of a film—centers the question of how one approaches, interacts with, and thus engages with the artwork he consumes. In using Alfred Hitchcock's 1955 *Rear Window* to illustrate his point, Pippin highlights the protagonist's activity of peering into the windows of the apartments of the people who live across the courtyard from him, secretly observing their lives from a distance:

> When in *Rear Window* Lars Thorwald enters Jeff Jeffries's apartment and asks, "What do you want from me?" if we have been noting the constant analogy between Jeff's position and movie watching, **we know the question resonates with the larger question of what we think we want from movies, and that the film has been suggesting how badly we understand the depth of that question, and its connection with non-cinematic issues,** like adopting a spectatorial position with respect to other people. (Pippin 2020, 9) (My emphasis.)

Here, Pippin uses the example of Hitchcock's *Rear Window* as a means of clearly identifying the relationship between audience and artwork—in this case, film—as perceived through the philosophical lens of a movie which presents this issue in a relatively blunt manner, exposing the nature of the average moviegoer as indulgent in the spectacle and surface-level entertainment value of the film, without truly engaging, or acknowledging the issues at work—without truly attending to the concepts and elements of presentation which the movie brings forth. As Pippin goes on to explore, Jeff Jeffries in many ways mimics the position of the viewer in relation to an artwork—taking his experiences and inadequacies allows us to take our own position as filmgoers, and perceivers of art, into greater account.

The multifaceted nature of a film, when approached as an artwork, proves imperative to the proper engagement with that film. As Pippin writes:

> The movie, one has to say in an ontological mode, *is the movie it is* only by means of this emerging, internal self-conception, a

> dimension we can miss if we too quickly apply some apparent
> formal unity, like a genre designation or a sociohistorical concept,
> or if we simply attend to the plot. (Pippin 2020, 8)

A movie exists on multiple levels: in one sense, as the product of story elements—the world inside the film, in which the characters exist, a world where, even outside of the frame, these characters continue to move about, to live their lives, to eat and sleep and discuss and think about how they will react within certain situations; on another level, the world of the movie functions as a self-conscious entity—one in which the camera decides what we ought or ought not to see. Non-diegetic elements—such as music, voice-over narration—exist within this self-conscious world, but not within the characters' world. Special effects and stylistic lighting employed as a manner of depicting the psychological status of the characters, as opposed to creating a realistic setting, influence the audience's perception of the story as well. In short: the story itself vs. the presentation of the story. If we take for granted the technical elements employed—each of the choices made by the filmmaker—and take these choices as simply suggesting genre, or adhering to plot or convention, or as presenting the story in the most entertaining way, we miss the deeper significance behind each individual choice, and the effect created by each chosen element—certainly as part of a unified system within the film itself, but also as conscious, and significant beyond mere presentation. Pippin draws from Daniel Yacavone:

> The duality between the "world **in** the movie," on the one hand,
> and the "world **of** the movie," on the other, what we see and
> attend to as depicted, and our sense of its being depicted in a
> way, or the "movie world" as a selection, highlighting, focusing,
> and, in its cinematic way, commenting. (Pippin 2020, 8) (My
> emphasis.)

In this way, movies function as self-reflective artworks: pieces fully conscious of their own function as framing a story and its themes in expectation of its perception by an audience. This duality suggests that the viewer's participation in the "world of the film" serves as equally significant to the presentation of the film—albeit from *without* the story itself—as the performances that take place on the screen and contribute directly to the story. Thus, Pippin also discusses the function of irony within the films he explores, noting that the elements at stake within the

film are often: "All not attended to as such or in an honest way by characters, but prominent for the attentive viewer" (Pippin 2020, 11-12). The filmmaker presents the artwork in such a way that it cannot do all the work on its own. A viewer who consumes the piece with a surface level of perception runs the risk of misreading and ultimately destroying the elements at work within the film.

Jeff Jeffries serves as the ideal example of the failing filmgoer. He watches the action from afar, drawn in by the spectacle and drama of the situation, but he does not take into account the larger issues of his scheme to catch Thorwald—issues which have less to do with his moral duty to capture a murderer, and more to do with his exploitation of the situation at hand as a means of personal entertainment. His fixation on Thorwald's apartment leaves him blind to the stakes of his own life. In other words, the audience member adopts a "spectatorial position" in terms of his own life, and in terms of the lives of others, rather than a position of true engagement. He attempts to involve himself in the events of the actions taking place across the courtyard, in the apartment opposite him, but only in a way that maintains his distance from the reality of the situation. For Jeff, this position shifts when Thorwald enters his apartment, and the world of the movie suddenly penetrates his reality and forces him to interact directly with the questions he has heretofore ignored. Earlier, when Jeff watches from a distance as Lisa sneaks into Thorwald's apartment, the stakes suddenly and immediately become clear: the world of the film does not exist within a bubble—it functions beyond the realm of the screen and applies to the world of the viewer in the same way that it applies to the characters. Hitchcock demonstrates this concept explicitly. Jeff only comes to realize his love for Lisa when the murder plot he has been indulging in, from a distance, puts *her* life at stake. This direct intersection between the world of the movie and world of the viewer suggests the level of interaction necessary for one to properly engage with an artwork.

As Pippin proposes, we must ask ourselves "what we think we want from movies"; the problem here lies within our inability to properly answer this question. What *do* we want from movies? Thorwald, the subject of Jeff's entertainment, serves as the personification of the film world itself—agitated by the viewer's inability to appreciate the duality of its presentation. Those who fail to consider this question beyond its most superficial answers—to be entertained, to be distracted, to be amused, to feel something—fail to grasp the nature of film and its capacity for philosophical reflection. People who ask the film to do the work

for them—to *make* them laugh, to *make* them cry, to *make* them think—ignore the work that they must do for the film, and thus do not thoroughly engage with the piece beyond its surface level, aesthetic functions. Though we engage more directly with one world of the film (the world *within* the movie) and tend to engage more subconsciously with the other world of the film (the world *of* the movie), we do not always recognize that our world, too, exists as a layer of the film world which must be engaged with. There is an active element necessary to viewership. We must *engage*, *notice*, *search*, *consider*, and *maintain* an active awareness of our position within the function of the film. The viewer ought not want the movie only to do things *to* her or *for* her. She must expect that a movie will engage in a dialogue *with* her, will ask her a question which prompts her to answer, and vice versa. A film is never complete when simply projected onto a screen—only when it begins its work in the mind of the viewer does the extent of its significance come to light. Ultimately, if one repeats to oneself while watching a movie, "it's only a film, it's only film," she is actively rejecting the film, and denying that which the artwork strives to disclose to the viewer; what we ask of the film must also apply outside of the bounds of cinema all together. The film does not speak exclusively in terms of film-related issues—it addresses the world and the state of human nature, the questions of being and the issues of existence, in the same way that philosophical text does: it deserves to be unpacked, and it deserves to be engaged with. Not only are the lives of the characters at stake, but the lives of the audience as well.

Thus, the role of the spectator requires special attention, and a greater sense of active involvement with the artwork at hand:

> It is often assumed that Hitchcock does this to show how "sinful" we are, how prurient movie watching can be, how voyeuristic it is. That is not the case, I argue. The problem Hitchcock is posing concerns not cinematic viewing as such, but an inadequate way of watching movies, Jeff's (and very likely ours). This is posed as the difference between a purely spectatorial mode of viewing, as if the film is just there for us to watch and enjoy, and a more involved mode of engagement with the film, in which what is asked of us (most attentive, interpretive *effort*) is as important as what we expect from it. (Pippin 2020, 12)

Herein lies the heart of the argument at hand: that the effort on the part of the viewer must match that of the effort on part of the artist. The viewer must not simply expect to be entertained, to sit back and allow the film to work itself out. The film begs for attention, for an amount of non-passive work in return for its presentation. Hitchcock does not present the viewer as sinful for watching, but as sinful for not watching well enough:

> We have, though, become a nation of moviegoers and, beginning at around the time of the film, 1954, of television viewers, in a way suggested by the tiny framed windows. I don't think this just means to suggest that filmed drama and comedy interest us because we like to be voyeurs, unobserved observers, but that we watch these screens *like Peeping Toms*. That is the *uninvolved spectatorial way* we watch them, as if what we see asks nothing of us, is simply there "for us"; and therein lie both the aesthetic and ethical issues.... (Pippin 2020, 31)

Pippin suggests that, beyond the voyeuristic implications of Jeff's position, and thus our position in watching a film—which therein suggests taking pleasure in the specific act of seeing but remaining unseen, of observing something that one was not meant to see—we also take pleasure in observing a scene—a dramatic or comedic dialogue, a private moment, a glimpse into the lives of other—which has no implication upon the life of the observer. A couple engaged in a heated argument is far more entertaining to the unseen observer than to the parties involved in the dispute. The viewer takes a sly satisfaction in the knowledge that she may indulge in the spectacle of conflict without facing the consequences of such a fallout. Thus, the "uninvolved spectatorial" viewer takes pleasure in consuming a movie or television show, secure in the knowledge that there remains an impenetrable fourth wall, isolating the film world from that of the real world. Herein, according to Pippin, lies the great fault of the filmgoer: to take up the role of the uninvolved spectator in one's approach to an artwork is to miss out on the greater implications of the artwork as a philosophical system. To take entertainment as a means of detachment from one's own life, or from the world at large, is to fundamentally misunderstand the function of the artwork, and to fail at one's job as filmgoer.

The audience does not literally engage with the action of the film. As Pippin puts it, "We are not *in* the action, cannot be affected by what happens, cannot intervene" (Pippin 2020, 23). As Jeff shows us, however, the failure to treat the art we witness with proper engagement can still have significant, real-life consequences:

> In one sense, yes; *I* am invisible. In another sense, this does not mean that all I must do to understand the world presented is watch, like a voyeur, a "peeping Tom," in the language of the film. In this case, here is something insufficient, deformed, about the way Jeff "watches" his little films, and that has something to do with the deformation in his relations with others, paradigmatically with Lisa. (Pippin 2020, 24)

Until the end of the film, Jeff does not register the hypocrisy of his own actions—his complete obsession with the lives of other, his absorption into the world he perceives through his window, blinds him and prevents him from critically addressing his personal relationships. When Lisa ultimately wins his attention by literally placing herself within the frame of the film he watches through his window, and an authentic and attentive relationship with her can truly begin, it is because he has felt the fear of losing her. He has realized how much he cares for her only through his proper engagement with the "movie" outside his window. He starts to appreciate her, to accept her for her differences, and welcome her affection because of a new understanding achieved through his open and active engagement with the aesthetic presentation at hand. The act of perceiving thus only has the potential to cause change, or to inspire understanding, to communicate truth and promote a greater engagement with one's own life, only if the viewer sees beyond the blinders that often prevent us from seeing another life, another world, in terms of our own.

Cinematic techniques—and within the broader context of art objects, artistic techniques in general—serve to execute that which the filmmaker wishes to capture. The tools of the artist, the materials at their disposal, must work in such a way that the aesthetic presentation properly articulates content, tone, and mood to its audience. As Seel notes the importance of understanding the functionality of the elements of the art object at work, Pippin notes that understanding the role of the film director in constructing not only what the viewer sees, but how

and why they see the images at hand, can lead to a deeper understanding and richer grasp of the art object. As he writes of the largely passive viewer: "We have learned to ignore for the most part that someone is purposefully showing us what we are seeing, has decided what we will not see, and the events are not simply magically present in front of us" (Pippin 2020, 23). In understanding the artist's role in the creation of the final product, the viewer has a greater chance of picking up on the truth that the presentation strives to communicate. Acknowledging not only *what* is presented, but the *way* it is presented, helps to develop a more acute aesthetic perception—a type of perception often lost on the passive spectator. Pippin describes the notion of a "self-conscious" film:

> Further, some directors do not want us to ignore their active involvement.... They are able also to draw our attention to the director's narrational control, and so to the presence of the camera, not just to what the camera is photographing. When we do notice, the visible narrational element is what gives the film its reflective form. Such a narrative form cannot but suggest a purposiveness, its point, and so manifests that the aesthetic object bears a conception of itself, a source of unity and ultimately interpretive meaning. It seems odd to say that filmed fictional narratives are in a sense "self-conscious," embody an awareness of themselves, but this is just an elliptical way of saying that the director is self-conscious of the point of the determinate narrative form. That point may simply be "to create funny situations" or to "scare the audience in a way they will enjoy," but it can clearly be more aesthetically ambitious; for example, to help us understand something better. This all corresponds to our own implicit awareness in experiencing an aesthetic object that that is what we are doing…. But such aesthetic attending already embodies a norm. It can be done well, or it can be done lazily, sloppily, indifferently, in a biased way, or self-righteously. (Pippin 2020, 24-25)

Most people watch a movie in the knowledge that they are watching a movie. The same self-awareness can be said of the film itself. In Hollywood films of the 1930s and 40s, studios strove to embody an "invisible" aesthetic—to craft

stories in a way that drew the least possible attention to the cinematic techniques themselves. In films that "draw our attention...to the presence of the camera," however, where we lose a certain sense of objectivity, of "realism," we gain a sense of artistic voice. We gain a sense of our role in the presentation of an artwork. In maintaining a self-conscious approach, the director not only attempts to create a purpose within the film but demonstrates to the audience that purpose as well. The audience knows they are watching a movie; the director knows she is creating a movie; yet the viewer does not feel disillusioned by this fact. This self-consciousness merely makes the process of perceptive attentiveness all the more rewarding—the aesthetic constellation at work, the layers of interior narrative as well as filmic, artistic voice that shapes the story—all of these elements, when well executed, make the viewing experience more engaging. When the artist ceases the attempts to cover up her tracks, and instead allows her movements, her ideas and her techniques to show, the viewer is allowed to see not only the final product, but the process that transforms the aesthetic into the artistic.

The powerful functional value of editing—whether subtle, unnoticeable, or distinct and jarring—can have a major impact upon the viewer of the film. Pippin writes:

> [Jeff] creates out of what he views, what is meant. Put another way, he edits together the various scenes he has seen to make his own imagined film narrative, a murder thriller.... This all of course mirrors what we see and largely what we also do when watching a film; we are always implicitly asking what it means that one sequence follows another, and we form our hypotheses the same way as Jeff, usually with more tentativeness. (Pippin 2020, 26)

The editing of a film—the order of shots between which the viewer cuts—between which the filmmaker directs our attention—suggests an implied interpretation. The viewer does not necessarily make a conscious connection between these moments, but, nonetheless, she picks up on a subliminal meaning. In *Rear Window*, for example, the scenes of Jeff observing his neighbors constantly cuts back and forth between the action taking place in a different location of the courtyard, and Jeff's reactive expressions. Sometimes he looks amused, other times shocked, delighted—and in certain key moments, terrified. Through his observations of his neighbors, Jeff pieces together the tale of a murder; through our observations of

Jeff, we likewise piece together a specific interpretation of these events, taken out of context as they are. When the film cuts from Jeff's point of view, to Jeff's face, we know that he is reacting to the scene we have just witnessed from his limited vantage point. We are only allowed to view certain moments, however, as Jeff only catches certain glimpses. We form assumptions—orchestrated through direction and editing—based upon the information we do receive. There's an element of intuition to watching a movie, regardless of the viewer's level of engagement: one must commit to these intuitions in order to make sense of shots, pasted together. The questions we ask while viewing—though not explicit—remain a crucial tool to making sense of cinema as an artform. Without the ability to make connections—to ask, What is he looking at? Why is he smiling? What is the relationship between this person and that person? What does this or that event mean?—we cannot begin to understand the film, even at a most basic level. As one's level of spectatorship digs deeper, expands beyond the level of passive viewership—if one begins to notice the patterns of editing, or the ways in with the camera reveals and conceals—one's understanding of the film reaches that state dual perception: of the world *of* the film, and that world *within* the film. This kind of active engagements reveals what lies beyond as well as within, and thus presents greater opportunity for openness and engagement outside of the artistic realm.

What makes the task of the viewer so difficult? What makes artistic engagement all the more worthy of a cause to promote? Because with great power, with great capacity for change, comes great fear:

> This is because, inevitably, as noted, we are also often invested in some way in the clarification, and that investment can be self-interest, self-deceived, biased, subject to wishful thinking, and so forth. One of the ways this can become impossible for us to avoid acknowledging is when our views intersect with the lives of others, and they respond, intervene in some way to challenge us. And, of course, one of the ways this can be *avoided* is by preventing such challenges, keeping our distance, staying inside our dollhouses or cages, psychological as well as spatial. (Pippin 2020, 28)

Pippin acknowledges a fear of the "challenge" that accompanies interacting with

an artwork—particularly an artwork which refuses to be taken at face value. The ugly moments, the scenes which too closely resemble reality—these are the artistic choices which "hit too close to home"—which make the corrupt, disinterested aim of the uninvolved spectator much more difficult to maintain. In some ways, the act of maintaining a passive stance in relation to the artwork proves more taxing that the act of actively engaging. To disengage becomes a matter of shutting oneself off from the truth of reality; this can mean taking the area of the movie theater as an isolated space—one which exists outside the tethers of reality, outside of the implications of the world. The theater thus serves as a neutral space, in which anything might happen, and after exiting the theater, the moviegoer leaves behind anything gained or lost—as if no time at all has passed. Psychologically, the artwork has little to no impact aside from small sparks of emotional engagement. Deep, troubling issues and questions remain untouched, locked within the mind of the filmgoer, and left to rest during the movie experience. This kind of fear and avoidance can serve most aptly to gauge one's inability to confront his own, personal troubles. His inability to confront the discomforts of the world. Art often looks harmless on the outside—once the real work begins, and that distance between the viewer and the aesthetic presentation has mostly collapsed, one must embrace their own vulnerabilities, their own weaknesses. There's nothing easy about true, deep aesthetic attentiveness. We refer to a film, a painting, a poem, a sculpture as an art*work* for three reasons: because the artist's *work* is creating, because of the object's *work* is presenting, and because of the viewer's *work* is perceiving. Any truth worth knowing, worth revealing, requires the kind of work that makes the outcome all the more rewarding. To avoid the challenge of a well-crafted artwork is to deny oneself, to refuse one's own chance at greater understanding—and, above all, to dismiss that which rivals even the human spirit in sheer beauty and promise.

III. CONCLUSION

What is the function of aesthetic perception in our daily lives? What does it mean to look at an artwork properly? We are, in short, lost without our ability to explore an artwork and derive not only aesthetic pleasure, but also a mode of rational thought. To look at an artwork and ignore the active nature of its self-presentation, to take it at face value, is to destroy its functionality, and to

assert one's one disengagement with life. One's ability to work as an art critic thus indicates one's ability to engage with reality—to acknowledge one's own existence within the world. Detachment, ignorance, passivity, are choices—they represent the choice of the individual to forfeit his place as a rational, thinking, feeling being, as an active part of the world. We cannot hope to lead lives of substance, lives of meaning, lives of playful engagement with the world, if we cannot partake of that engagement within the realm of aesthetic appearance. The art critic does not function to distinguish the good art from the bad art, the right art from the wrong art, or even the beautiful art from the ugly art, but rather to distinguish the active from the passive, the engaged from the disengaged, the thinking from the thoughtless, the living from the dead.

REFERENCES

Pippin, Robert B. 2020. *Filmed Thought: Cinema as Reflective Form.* Chicago, Illinois: The University of Chicago Press.

Seel, Martin. 2005. *Aesthetics of Appearing.* Translated by John Farrell. Stanford, California: Stanford University Press.

compos mentis

The Moral Distinctiveness of Torture

Ethan Galanitsas
University of Essex

ABSTRACT

Suppose there's a chemical weapon hidden somewhere in your hometown, and you have custody of the person responsible for planting it. This person is also capable of disarming the bomb; is it acceptable to torture this person in order to get the information needed? I argue that it can't be. While many could die, it seems extraordinarily unlikely that torturing this person would actually yield the intended results: one must have the right person, and they must promptly and accurately reveal the information needed. If the wrong person has been apprehended, the end result wouldn't change – the bomb in your hometown will still go off – except now an innocent person has been tortured in the last moments of their life. Furthermore, even if you do have the right person, to extract correct information from them in the required time frame necessitates someone well-practised in torture. But a well-practised torturer could never be well-practised without people to regularly practise on – which cannot be endorsed. In this essay, I use these points to illustrate that the moral distinctiveness of torture lies in its total unjustifiability.

INTRODUCTION

Torture is an abhorrent, evil act. Yet still, it appears to be a frighteningly useful tool, and good people might find themselves inclined to use it – or at least justify its usage. In this essay, I mean to discuss the moral worth of torturing, and assess whether or not it can be classed as morally distinct from all other acts. I will first elaborate on what it means for an act to be 'morally distinctive', concluding that a morally distinctive act must be something that is completely impermissible, regardless of the situation and consequences. I will then use Winfried Brugger's hypothetical scenario (Brugger 2000) to illustrate the difficulty in classifying torture as wrong a priori, before commenting on Henry Shue's writing on the problems of these kinds of hypotheticals (Shue 2006), and addressing the context that must exist for torture to be viable. The aim of this essay is to illustrate that torture *is* morally distinctive – it can never be a permissible act.

I. TO BE MORALLY DISTINCTIVE

Firstly, to address what it means for something to be morally distinct, I argue that the act would need to be impermissible in any situation. To prove this, torture – the distinctiveness of which is yet to be proved – must be temporarily removed from the discussion. Instead, I will look at other immoral acts and assert that they can all be deemed as morally permissible in a certain context.

First, killing appears to be the most wrong[1] one can do to an individual. If one loses their life, all their other rights vanish as well: if I kill someone, I necessarily stop them from accessing all of their other rights. Meanwhile, infringing upon any other right – to private property, to education, to free speech – doesn't directly infringe upon one's right to life: I can steal from someone, but they won't necessarily lose their life as a consequence. Therefore, killing can properly be deemed as the most wrong crime, because it necessarily infringes upon *all* of one's rights. Even this, however, is generally considered to be morally permissible when defending an innocent person from a guilty attacker. As Jeff McMahan writes, if a guilty party

---

1. I use 'most wrong' here instead of 'worst' because my claim is about *wrongfulness*, not about how distasteful the act is. I'd be slower to call killing the 'worst' act, since death is something everyone must experience, while no one *must* have something stolen from them for example. However, I think *wrongfulness* is most appropriately weighed by assessing damage done to one's rights.

threatened to murder a child and the only way to save the child was to kill the guilty party, "By his own voluntary action he has made it the case that either he or the child will be killed" (McMahan 2008, 94). In this case, justice demands that, if one should live, it should be the innocent party. Thus, it can rightly be said that *defending the innocent* is a powerful enough moral imperative to permit killing.

Since I've established that murder is the most wrong crime, *defending the innocent* must also be a powerful enough imperative to permit every other immoral act. If examples are considered, the truth of this seems to be clear: stealing a weapon from a madman; betraying a friend who has become murderous; and lying to a murderer about the location of his intended victim, are all examples of immoral acts which would generally be agreed to be morally permissible. Therefore, to be morally distinctive, an act would need to be impermissible even when it's needed to defend the innocent.[2]

II. THE TICKING TIME BOMB

It is largely agreed that torture ought not to be used regularly. Thus, the question which often finds the spotlight in discussions about torture is that of the *ticking time bomb*:

> Suppose the bomb is planted somewhere in the crowded heart of an American city, and you have custody of the man who planted it. He won't talk. Surely, the hypothetical suggests, we should not be too squeamish to torture the information out of him and save hundreds of lives. (Luban 2007, 252)

To posit this hypothetical in order to claim that torture should be used is to say that torture cannot be wrong a priori. This would promptly end the discussion about torture's moral distinctiveness: it has none, since it cannot always be morally impermissible. With good reason, Gross argues that torturing the terrorist responsible for the *ticking time bomb* could be justified – provided all the desired results were achieved. He proposes an *ex post, extra-legal* method of torture, where interrogators act outside the established law against torturing; they may

---

2. There is a temptation to use rape as an example of an act that is always impermissible, but it seems that torture, which can be defined as the infliction of suffering upon a victim in order to gain something, can accommodate rape in its meaning.

be legally punished if it wasn't necessary, or ratified if it saved the nation from disaster (Gross 2004, 24). This seems to support the idea that torture cannot be morally distinct: if torture can be ratified post-hoc, then it isn't totally morally impermissible.

Perhaps then, we may allow that the *ticking time bomb* case is extreme enough to allow torture, but this doesn't mean that the case can't be rejected altogether as laughably improbable – improbable enough to make the decision to torture ill-advised and wrong. I won't go as far as to say that this scenario isn't possible: Luban comments on the real 1995 example, "an al-Qaeda plot to bomb eleven US air-liners was thwarted by information tortured out of a Pakistani suspect by the Philippine police" (Luban 2007, 253). The issue is that this success isn't regular enough to allow that torture is the right thing to do – there are far too many variables that simply aren't addressed in the *ticking time bomb*. These issues are aptly presented by Henry Shue: "*The right man… Prompt and accurate disclosure… Rare, isolated case*." (Shue 2005, 233). What's laughably improbable is not the claim that the *ticking time bomb* can't happen – it can – but the claim that torture is likely to yield the desired results. It's optimistic at best to even suppose you have the right person, let alone that they will break in time; give accurate information; won't disassociate, leading to them being informatively useless; and won't die while being tortured. Gross's proposition of an *ex post* ratification is flawed largely for this reason: no one would feel confident putting their neck on the line when the likelihood of success is so vanishingly small, and the consequences of torturing an innocent are so huge. Recognising the lack of pragmatism here, a government would need to work torture into its policy in the case of emergencies, or regard torture as totally impermissible (making it morally distinctive, at least in the eyes of the law). To work torture into a nation's policy is to likely cause global catastrophe, as Shue makes clear:

> "[The] catastrophe lies on the side of undermining the taboo against torture. Then other nations will reason that if the superpower with its thousands of nuclear weapons and high-tech conventional forces cannot maintain its own security without the liberal use of secret torture, they can hardly be expected to defend their security without far more torture." (Shue 2005, 234–235)

Torture is, therefore, morally distinctive for this reason: the use of it provides a tiny chance to avoid catastrophe, thus *defending the innocent*, but it is far more likely to cause the catastrophe of normalising torture – a practice which regularly *harms* the innocent by torturing the 'wrong' person, while rarely successfully defending the innocent.

I've shown that the classic *ticking time bomb* has too many variables to allow torture to ever be morally permissible. The question becomes more interesting if we can remove some of these variables: specifically, the *right man* problem; if it's known that the person being tortured is certainly guilty, then there is no risk of breaking the *defending the innocent* imperative. Consequently, I'll comment on Winfried Brugger's more interesting hypothetical, which implies that torture can even be just when alternatives exist:

> This hypothetical takes place in your home city that is threatened by a terrorist armed with a bomb containing deadly chemical agents. He has hidden the bomb. After he has been tracked down and detained by the police, he states, credibly, that he has activated the timer of the bomb. The bomb will detonate in five hours and kill all of the inhabitants of your city and its suburbs. All will suffer a horrible death. Despite police pressure the terrorist refuses to disclose the bomb's location. Instead, the terrorist demands ten million dollars, the freeing of all death row inmates, and an airplane for his getaway. In addition, he wants ten hostages, so as to ensure a successful escape. The hostages must be ten prominent citizens of your home city. The police find that they are neither able to meet the terrorist's demands nor can they evacuate the city and the surrounding area in time. Only one solution seems to remain. They want to use physical force – torture – to compel the terrorist to divulge the location of the bomb. Are they allowed to use such methods? (Brugger 2000, 662)

Brugger admits here that giving in to the demands isn't an option. For one, "innocent people in your town would be subject to risking life and limb" (Brugger 2000, 667), which immediately counteracts the *defend the innocent* imperative. Furthermore, whether or not the terrorist will actually disarm the bomb once he's

free seems doubtful at best: "we do not know whether the terrorist would really deactivate the bomb, and what would keep him from reactivating it, once he's on the airplane?" (Brugger 2000, 667). He admits that torture doesn't guarantee the divulgence of information either, but the terrorist would certainly have more motivation to do so than if he was on an airplane to freedom (Brugger 2000, 668). Forbidding the use of torture here seems to be far too extreme. If we can agree that it is permissible to end a guilty person's life to protect an innocent person, so too should justice permit us to torture someone we *know* to be guilty to protect many innocent persons. It seems in this situation, Gross's *ex post* ratification system seems plausible here – the torturer can guarantee that there will be no repercussions for their actions, since they know for a fact that they have the *right man*.

III. A CONTEXT OF TORTURE

Brugger's hypothetical presents torture as not being morally distinctive: it is permissible at least in his hypothetical. However, it focuses far too heavily on the situation itself and not the wider context that the act of torture would need in order to exist. For torture to be successful, the torture needs to be done *well*, as Shue highlights: "successful torturers need to be 'pros', and no one becomes a 'pro' overnight. At a minimum, one must practise" (Shue 2005, 236). To imagine that the torture is a success – the torture that involves going against every basic human inclination towards empathy, without going overboard and rendering the captive informatively useless – is to imagine a good torturer. To imagine a good torturer is to imagine someone practised in torture, which imagines people that the torturer regularly practises on. As a result, any attempt to justify the use of torture in extremely isolated incidents relies on a context involving the frequent use of torture for it to be a success; otherwise, guilty individuals would be harmed without truly *defending the innocent*, since the operation is likely to fail.

One could claim that well-practised torturers are no longer necessary in a world with modern technology – surely we could create Artificial Intelligence which could torture people far more effectively than humans, and without regular practise. Setting aside the frightening and unnerving thought of AI torturers, even these would certainly need to be trialled on humans to ensure their effectiveness. Consequently, while they may need *less* practise, it would be wrong to say that

they need *no* practise, and so would still need to torture in more cases than the occasional isolated incident.

In conclusion, torture is morally distinctive because it is completely morally impermissible in the real world. Because the *right man* is so unlikely to be known in the *ticking time bomb* scenario, Brugger's hypothetical (Brugger 2000, 661–678) is the only context in which torture is acceptable *as long as the intended good results are obtained*. However, for these results to be obtained *at all* – let alone in so short a time frame – the torturer would need to exist in a context in which they can regularly practise, and therefore be effective. While imaginable, it cannot be said that Brugger's hypothetical – the only justifiable occasion to use torture – occurs regularly enough for the torturer to be well-practised. Consequently, the torturer will be inexperienced and will almost certainly be unsuccessful. Unsuccessful torture does not *defend the innocent*, so it cannot be permissible. Of course, there is a small chance the torture *could* be a success, but an action cannot be called permissible if the intended good outcome is doubtful – this is what is morally distinctive about torture.

REFERENCES

Brugger, Winfried. 2000. "May government ever use torture? Two responses from German law." *The American Journal of Comparative Law* 48, no. 4: 661–678

Gross, Oren. 2004. "The prohibition on torture and the limits of the law.*" Available at SSRN 485043*: https://ssrn.com/abstract=485043

Luban, David. 2007. "Liberalism, torture, and the ticking bomb." In *Intervention, terrorism, and torture*. 249–262. Springer, Dordrecht.

McMahan, Jeff. 2008. "Torture in principle and practice." *Public Affairs Quarterly* 22, no. 2: 91–108

Shue, Henry. 2005. "Torture in dreamland: disposing of the ticking bomb." *Case W. Res. j. Int'L.* 37: 231–9

compos mentis

Putting the "Judge" in "Prejudice:" Neutralizing Anti-discrimination Efforts Through Mischaracterizing the Motives for Prejudice

E. C. Hamilton
Concordia University

ABSTRACT

In 2018, Kate Manne argued that framing misogyny as hatred of women had the effect of neutralizing efforts to organize against it. She held that criteria for "hating women" were so rarely met that virtually no one could be said to have done so. Taking for granted that the situation against women was unfair, she argued that those who sought to correct the situation should reconceptualize what misogyny means: not as hatred, but rather understood by its perpetrators as righteous punishment for violating a perceived moral code. I argue here that every point she made against "misogyny as hatred of women" is at least applicable to "transphobia as hatred of transgender people." I say instead that this character of righteous punishment is also well-evidenced in negative responses to the civil advances of transgender people, and invite the reader to consider what this would mean from a policymaking perspective.

Published in 2018, Kate Manne's *Down Girl* was a series of essays that made several cases for why we should evaluate sexism and misogyny by what they "do" to women (Manne, Aims, 19). She contends that these systems should be understood as crude moral theories protecting male dominance, and the tactics carried out in their service are considered by their perpetrators to be righteous punishments for perceived violations of this moral code. Manne's argument stands in contrast to prior scholarship that emphasizes the role of dehumanization as an explanatory force for sexism, in which perpetrators perceive their treatment as not morally injurious at all on account of not believing women to be capable of being injured in the way they acted, i.e. that they were justified in doing so (Haslinger 102-103). Theories opposing Manne's explanation would explain sexism as an attitude akin to me using my own bike: I have committed no moral wrong in using my bike because "its nature" is to be used; the sexist thinks this to some degree about women, says the dehumanization proponent (MacKinnon 173). Manne notes, however, that the preoccupation with the sexist's attitude towards women means the disputed territory is their private knowledge of their own attitudes—philosophically speaking, all the sexist need do to defeat the charge he views women as objects or hates them is deny it (Manne, Introduction, 33). One objective of her framework then is to reopen diagnoses of misogyny by exploring the political dimensions served by this preoccupation with attitudes, namely that the implausibility of diagnosing an aggressor's attitude contrary to their stated beliefs about themselves renders it difficult (if not impossible) to organize against (Manne, Threatening Women, 4).

I'm going to argue that Manne's framework carries significant explanatory power for cissexism and transphobia as well, after reacquainting the reader with the four terms heavily referenced in this paper (sexism, cissexism, misogyny, transphobia). Thus, as I proceed to develop Manne's account as it relates to cissexism and transphobia, I will likewise be questioning the role of dehumanization ("DH") as an explanatory force for discourse and actions hostile to transgender people, instead demonstrating that Manne's framework is highly relevant in this application too. As a brief warning, I do not mean to say that dehumanization, objectification, revulsion, hatred, or other emotive responses form *no* part in prejudice—rather, I will argue that these may be possible tactics used to carry out what is felt to the aggressor as a righteous punishment, rather than "the cause" of the behaviour.

If I (and Manne) are right in this, it can provide a worry for the notion that the prejudices against women or transgender people (or people who are both) are things that can be corrected with education. We suspect the perpetrator knows full well that they inflict an injury; the actual disagreement is whether the target "deserves it." Providing salient facts is unlikely to persuade because it mis-identifies prejudice as a bottom-up logical construction with inaccurate premises rather than a top-down judgement stemming from an axiomatic commitment. If education is to be a response at all, it would likely have to be of the moral variety, and not simply an attempt to raise facts salient to their stated beliefs.

SCOPE

While this paper will endeavour to recognise when others may be making moral judgements as opposed to strictly logical arguments, it should be noted that the question of whether such judgements are justified is outside the scope of this undertaking. It is not, itself, attempting to broach an ethical question, but rather it poses a political question mediated through psychology. A handful of positions will be evaluated against my ethical stances but the core point of this undertaking can proceed even if one disagrees with my ethical positions. Likewise, I mean to defend my claim that Manne's framework appears explanatory in the cases of cissexism and transphobia while recognising that I am not in the best position to argue the same for prejudices of which I am not affected, namely racism, and I will defer to those affected by both transmisogyny and racism as to whether the framework is explanatory in their lives. Lastly, I will discuss why some people will likely be unpersuadable at various points in my argument, and what that might mean for policymakers wishing to redress the problem of discrimination in their organizations. I suggest what type of approach may be applicable in these instances, but I will not have space to elaborate on or defend those claims here.

REVISITING SEXISM AND MISOGYNY

Manne distinguishes between sexism and misogyny as follows: Sexism is the rationalization by which the belief-holder comes to understand a perceived debt to certain moral goods (e.g. caring, nurturing) owed to them from women, while misogyny might be the tactics deployed by that belief-holder to enforce those

beliefs (Manne, Discriminating Sexism, 4). Both benefit from the point that we are now focused on observable things: Stated claims in the case of sexism; actions (completed or planned) in the case of misogyny. Sexism "purports to merely be [reasonable], [while] misogyny tries to force the issue."

There are several implications which merit this choice. The first advantage is that we can discuss sexism/misogyny by evaluating specific actions and stated beliefs against those affected by it. We are no longer making any claims against an interlocutor's intentions or internal state. To explain: When we say an interlocutor "hates women," we are speaking of his true motives for his beliefs. However, our interlocutor has special access to his own thoughts. As outsiders, we can at best tease at what his true motives might be, but we are in a vulnerable position if we try to make decisive claims to the contrary of what he says about himself. Unless we happen to possess a recording where he explicitly states his motives, and is only now presenting a different story, we can't "prove" he's wrong about what he thinks. Even in the example of the recording he can just as easily claim he's had a genuine change of heart.

The effect of this "naïve conception of misogyny" as Manne phrases it has a political dimension: It makes misogyny very difficult to identify and organize against. She takes the case of the Isla Vista killings in which the murderer posted a manifesto condemning women for what he saw as a failure to express adequate sexual interest in him. Nonetheless there were several responses resisting misogyny as a diagnosis for the murderer's motives, from opponents to feminism claiming he was merely "misanthropic" (Young 2014) to DH adherents arguing that he didn't view women as meaningful agents who could be hated to begin with (Thomas 2014). Manne notes that under these stringent psychological criteria of hatred, meaningfully arguing someone or something they did is/was misogynist is rare. Since far more women are affected by misogyny than there are people who will admit to "hating" them, this conception impairs their ability to identify the sources of their threats, never mind organizing a response. It will follow that those who either wish to continue benefitting from this situation, or who at least hold it to be justified, will not be persuaded by such an observation: Indeed, it's rather the point. Asking who stands to gain from maintaining the naïve conception of misogyny is precisely one goal of Manne's framework, and it's a poignant question as we turn to cissexism and transphobia as well.

For those that do take it for granted that the reduction (or utopian elimination) of sexism and misogyny are desirable goals, Manne offers another explanation for what they are that seems compatible with what is claimed by the aggressors in gendered harassment actions. She suggests that misogyny is inspired into action by the feeling of being spurred an entitlement to a moral good, in this specific case being the deference, nurturing, procreation, care, often associated with feminine social roles (Manne, What She Has To Give, 9).

Manne's model predicts that a misogynist wouldn't hate women generally, but rather hostility would rise specifically when these goods are withheld from him (wrongly in his view). She compares it to sitting at a restaurant and receiving no service from a waiter—one probably doesn't conclude that restaurants are unsuitable as a service, merely that this particular server (or this restaurant by employing them) has failed. Further, one might anticipate confusion and maybe even indignation were the server to ask you to take *their* order. This tracks with a common objection to being called misogynist: "But I love my mother/sisters/wife." Unlike the hatred/DH model, Manne's explanation doesn't contradict him, pointing out that his love to said women may be a product of their providing the goods to which he feels he is owed, much in the same way that restaurant goers are satisfied when servers serve them. Instead of litigating our interlocutor's motives in futility, Manne's analysis lets us look at the structure of the avowed accusation: a woman of "making unreasonable demands," in response to an incident that prompts the accuser to say or imply they've been denied something owed to them, much as we might react to the server telling us to take their order at the restaurant.[1]

Sexism is then all the arguments and beliefs which lead the observer to conclude favourably in this asymmetrical debt of moral goods owed by women to men. It could include naturalizing differences where they exist or exaggerating differences where they don't, but it need only support the conclusion that this debt is justified to meet Manne's meaning of sexism.

---

1. The limitation of this comparison is that servers have at least nominally agreed to their job, and we have more freely agreed to enter the restaurant. I believe this changes the moral circumstances of the exchange described – part of the problem is that women as a whole have not agreed to, or even be consulted on whether they want to, supply this debt to men.

REVISITING CISSEXISM AND TRANSPHOBIA

Recall earlier that I said the political effect of the naïve conception of misogyny was that it neutralizes attempts to organize against it. This was an important point on my broader argument about intervention on this issue necessitating a moral education, rather than simply raising facts salient to a stated belief. I will soon show that an analogous naïve conception of transphobia exists to serve precisely the same purpose: It has the effect of casting "true antagonism" against trans people as something so formless that almost no one can be meaningfully said to have done it, rendering efforts to counter it null. But, critically, even if I succeed in doing so, this will *not persuade those for whom this disruption is the goal.* Anyone axiomatically committed to protecting what they see as the benefits of cissexism can see this sort of epistemic vandalism as, at best, the regrettable cost of enforcing their moral code.

To get an account of cissexism, I turn to Talia Mae Bettcher's work "Evil Deceivers and Make-Believers." Bettcher noted that certain responses to highly publicized violence against transgender women required more explanation than what was offered by commentators at the time (Bettcher 2007, 47). Transmisogynist murders were (and still are) handwaved away as instances of homophobia, and while Bettcher concedes this would partially explain the attitudes of murderers, it was not sufficient to explain the peculiar charge of deception that is wrongfully levelled at the victims. I will clarify on both meanings of the word "wrongfully" here—it is both the case that the murderer knew (and in modern cases, usually knows) about their victim's trans history in advance of intimacy and (I would submit) also a moral wrong that this accusation is the one that readily occurs to observers. For our purposes the question I ask is, how can one be accused of lying for never answering a question that wasn't actually asked?

The answer, I think, is our clue to cissexism, as well as its parallel relationship to sexism. The accusation that the trans victims have "lied" for "failing to disclose" can only make sense if the person holding this position believes that one's appearance is "supposed" to convey specific information about one's birth sex (Bettcher 2007, 53). In other words, this belief holds that the question *was* asked: It's asked every time you leave your house, and your appearance is supposed to be the answer. You've done something intentionally *wrong* if this is not the case--why else default to specifically *lying* in the accusation? In Manne's account of sexism,

it is made up of beliefs that lead the observer to conclude favourably in male dominance, or the situation of men being owed certain moral goods by women. We would expect cissexism to similarly be the beliefs which lead the observer to conclude favourably in cisnormativity, or the situation of people signalling their birth sex in a way intelligible to their cultural norms through their appearance.

I submit that this account of cissexism is plausible because it explains why some political commitments that would normally be at odds suddenly find themselves in agreement when problematizing transgender people. One could hold that one "has a right" to know through appearances the birth sexes of people they encounter because it's important for securing the safety of (cisgender[2]) women and another could hold that one "has a right" to know through appearances the birth sexes of people to create procreative pairings, positions held respectively by some (self-described) radical feminists and religious conservatives, and in either position one concludes that transgender people have committed a wrong in shirking the duty to provide the information to which one feels entitled. This cissexism shares a nature with sexism in that it supports belief in a moral duty, and diverges from sexism in that it lacks asymmetry, instead taken to be granted as true for all people.

A complication can be found in women who find themselves confronted for being visibly masculine, usually in gendered spaces. One may reply that it is not clear whether she is being policed in that moment for violating a perceived moral code to supply men with a desirable appearance, or whether she is being policed for inadequately conveying her birth sex. My answer is simply that it is possible to be both. A view which holds asymmetrical obligations between men and women needs a way of clearly delineating between the two to prompt the requisite obligations; sexism requires cissexism. The real complication is that the inverse does not need to be true. One can reject the asymmetrical obligations described under sexism and have different reasons for believing that knowing another's birth sex is a matter of moral duty. The "anti-sexist cissexist" is not only a theoretical possibility in this construction, but something well-evidenced by self-described women's rights advocates who campaign in favour of strict state-regulation of gender, as we will discuss soon.

---

2. Commonly defined as "non-trans," here tentatively describing one's relation to cisnormativity as described above.

Another complication to this is that another route by which interlocutors sometimes conclude disfavourably against transgender people is that it's bad to increase one's medical requirements. In other words, it could be possible that one rejects this duty to supply sex information through appearance but nonetheless believes that any biomedical transitioning increases dependence on healthcare interventions, and therefore should be at minimum not encouraged, if not actively discouraged. It is not within the scope of this paper for me to elaborate on the many weaknesses of this position, however, I will note that the motivating factor in this situation is "dependence on healthcare interventions," which would resemble Fiona Campbell's definition of abled narcissism. While such accounts are among the inventory of hostile reactions to trans people, I would say this theoretical possibility is better described as Campbell's network "that produces a particular kind of self and body... that is projected as perfect and essentially human," with disability being cast as the state of "diminished humanity" (Campbell 2012, 213). This attitude could hypothetically explain intentional biomedical gender variation as one such "diminished" state without relying on cissexism, and would be similarly vulnerable to the many objections laid against abled narcissism in Campbell.

NAÏVE CONCEPTION OF TRANSPHOBIA

Since the naïve conception of misogyny has the effect of neutralizing attempts to organize against it, we will want to see a similar effect when locating an analogous naïve conception of transphobia. With the naïve conception's preoccupation on the motives of an interlocutor, my proposal finds merit if we see discourse focused on the attitudes of those accused of transphobia. Both characteristics are present in a high-profile essay by J.K. Rowling titled "Reasons for Speaking out on Sex and Gender Issues," self-published on her blog in June 2020.

We now turn to Rowling to see which parts evidence this effect (all emphasis added):

> **'Woman' is not a costume**. 'Woman' is not an idea in a man's head. 'Woman' is not a pink brain, **a liking for Jimmy Choos** ...one of the objectives of denying the importance of sex is to erode what some seem to see as the cruelly segregationist **idea of women having their own biological realities.** ...When you

> throw open the doors of bathrooms and changing rooms to **any man who believes or feels** he's a woman – and, as I've said, gender confirmation certificates may now be granted without any need for surgery or hormones – then you open the door to any and all men who wish to come inside. (Rowling 2020)

For Rowling, bodies constitute a rallying point which "unify [women] as a cohesive political class." It is stable. It yields predictable problems to which there are political solutions. It is "real," contrasted with "liking Jimmy Choos." The body is a political locus, the appearance a frivolity. Her proclamations about what a woman isn't are framed to be a response to a trans person's argument (whose argument, one wonders?). She isn't merely stating what she believes; she's implying this is what a trans person believes: "Shoes make the woman." Frivolity, unserious, fake. Her argument is rooted in the body, the real, the concrete; her unnamed opponent is rooted in liking shoes.

She problematizes the example of women's bathrooms where she perceives the criteria for entry being reduced to "belief" and "feeling" because limited legal recognition as female might have been granted without bodily alterations. This requires elaboration: The current status-quo for most restrooms in the world is such that Mary already selects the bathroom they believe most appropriate for them. If another user in that bathroom, Sue, feels that Mary is mistaken in their selection, Sue has at her disposal several options: Asking Mary if they're sure they're in the right place, confronting Mary and stating Sue believes Mary to be in the wrong place, calling management or the police to adjudicate the dispute, or initiative violence to eject Mary herself. I delineate this because it seems to me the case that men can already choose to enter the women's bathroom if they are so inclined, and women already possess several ad-hoc avenues to address any perceived breaches. It is not often the case that bathrooms require any proof of sex to merely enter, and I doubt Rowling advocates for the inspection of ID markers at bathrooms, but it is currently true that users of a bathroom assess for breaches based on what is visible.

Thus, easing the conditions for a limited legal recognition as female would only be a problem if that legal recognition was going to be called upon after an interlocutor's appearance prompted questioning. To put it crudely: Rowling would require your appearance to convey facts about your body which are verified by a legal mechanism, which she could deploy to question you on her behalf.

So we see to start evidence of cissexism, here an argument that transgender people owe an appearance she can recognise as corresponding to their birth sex, because she wants to access public facilities used by people she knows are other cisgender women or, at best, transgender women vetted (by someone else) to meet some specific criteria. But what of the naïve conception of transphobia as hatred? Rowling again (emphasis added):

> 'They'll call us transphobic!' **'They'll say I hate trans people!' What next, they'll say you've got fleas**? …

> **None of the gender critical women I've talked to hates trans people**; on the contrary. …they're hugely sympathetic towards trans adults who simply want to live their lives …(Rowling 2020)

These are rather clear anticipatory disavowals of the motive of hatred but they are, crucially, also straightforward preoccupations with her attitudes. It is Manne's approach that allows us to examine the consequences of Rowling's rhetoric by focusing on the impact of those affected by her words. Rowling need not hate trans people to hold the views that she does. Indeed, it is much easier to explain her dismissive and derogatory tone comparing trans women to "liking Jimmy Choos" as a sort of imperial put-down directed at them for speaking out of order: Insulting, intentionally so, because they did something that warrants punishment. She benefits greatly from a lack of careful examination as to what this sort of heightened bathroom policing could promote, and she won't have to account for it if people find themselves mired in debating her hatred or lack thereof.

Note also that Manne's other predictions are also relevant here. Rowling need not be hostile to transgender people in general, for example, and she isn't. She cites "transgender adults trying to get on with their lives" as the type of trans person she *isn't* concerned with. While her ire is justified against those individuals practicing misogyny in the course of criticizing her transphobia, I note also that those making trans civil rights arguments in favour of liberalising the UK's regulation of gender make up the bulk of her targets:

> **I happen to know a self-described transsexual woman** who's older than I am and wonderful. … Being older, though, she **went through a long and rigorous process of evaluation,**

psychotherapy and staged transformation. The current **explosion of trans activism is urging a removal of almost all the robust systems through which candidates for sex reassignment were once required to pass.**

So we see that Rowling is anticipating the accusations she received by implicitly arguing they can't be true because she doesn't have any objections to trans people *in general*—just the ones interfering with her specific program—and we see that the inciting incident which prompts her reaction is the act of attempted removal or reduction of the "robust systems" responsible for regulating and surveilling deviations from cisnormativity. I hold that such surveillance would be considered valuable if one thought the core project promoting "appearance as an obligation to communicate birth sex" was a worthy pursuit while conceding that forcibly removing any option for transitioning is probably too draconian for justification. If we must tolerate breaches of this norm, holders of this position might support systems to regulate it such that the greatest threats to the project are minimized. It dovetails with the earlier support for surveillance in the sense that it seems to say "it's fine if someone, some*where*, is watching for me." If I've described this position correctly, they may even feel it is a magnanimous compromise rather than an invasive breach of transgender people's autonomy.

CONCLUSION

I believe I've made a good case for why conceiving of transphobia as a "hatred" of transgender people can, itself, be a tactic of transphobia, while also promoting cissexism by countering efforts to reduce it. Insofar as strong, emotive reactions such as disgust or rage occur when carrying out the tactics of transphobia, I believe their actual root cause to be a moral judgement, a righteously-felt indignation at having something withheld to which the aggressor feels they are owed. I look no farther than the numerous campaigns in the United Kingdom which compare transgender people to people convicted of heinous violent crimes on bases that are outright delusional (e.g.: Glinner 2020). We react emotively to the crimes described *on account of their immorality,* so these comparisons could only make sense on that basis—not accusations stemming from empirical observations, but equivalencies of moral violation.

If this diagnosis is accurate concerning a particular incident, the value for policymakers is in understanding that the correction would have to be a moral education and not simply attempting to fact-check stated beliefs. I would argue elsewhere, for example, in favour of a Rawlsian "veil of ignorance" approach in a bid to see if a transphobic interlocutor might locate injustice in framing the terms of their beliefs between Bettcher's "make-believer" and "deceiver" by pointing out that the target of such a framework is literally *trapped* by these terms. They cannot self-advocate: The ability to do so has been razed, as any objection can be written off as a product of the "delusion" or "bad faith" evidenced by the "deception." I suspect if one can properly imagine themselves subject to such an underhanded silencing tactic, they will be far more responsive than (for example) simply being told evidence of endocrinological triggers, or more bluntly that they'll just have to agree to a code of conduct or be fired. But this would require the interlocutor in question to have the skill of applying Rawls' veil to begin with, of properly imagining how they would want to solve the dilemma if they didn't know what position they would occupy in the solution. They would need, in essence, a moral education. Such a topic warrants its own treatment.

As previously stated, little of what I've outlined here will be persuasive if someone finds the "right to know" someone else's birth sex a morally worthy goal to protect. We might hope that this encourages those holding the position to be more honest and explicit that it is the position they are trying to defend, but it does not address the reasons they came to form the belief to begin with. What I do hope we recognise is that when one sees an argument that would require a person's birth sex to be made visible in some way, accusations of hatred work to terminate the discussion that would be necessary to expose that claim. I believe that the particulars of this sort of birth-sex-policing will make the stakes far clearer; I also believe that many holders of this position are aware of this, and will smother objections to that effect by avowing they "don't hate trans people."

REFERENCES

Manne, Kate. 2018. *Down Girl*. Oxford University Press. Kobo format. Downloaded February 23 2021.

Young, Cathy. 2014. "Elliot Rodger's 'War on Women' and Toxic Gender Warfare." *Reason.* Archived May 7, 2021: https://archive.is/uS1YH

Thomas, Dexter. 2014. "Elliot Rodger wasn't interested in women." *Al Jazeera.* Archived May 7, 2021: https://archive.is/NLEvl

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

MacKinnon, Catharine. 1987. *Feminism Unmodified.* Harvard University Press.

@Glinner, "Grooming." 2020. *Twitter.* Archived May 7, 2020: https://archive.fo/Xczeh

Haslanger, Sally. 1993. "On Being Objective and Being Objectified." In *A Mind of One's Own: Feminist Essays on Reason and Objectivity,* 209–253. Boulder, Colorado: Westview Press.

Rowling, J.K. 2020. "Reasons for Speaking out on Sex and Gender Issues." *jkrowling.com.* Archived November 21, 2020: https://archive.is/2cxeY

Campbell F.K. 2012. "Stalking Ableism: Using Disability to Expose 'Abled' Narcissism." In *Disability and Social Theory.* Palgrave Macmillan, London. https://doi.org/10.1057/9781137023001_13

J. Doe's Right to Moral Status: Moral Responsibility Beyond Principle of Equality

Seyeong Hanlim

Denison University

ABSTRACT

In moral discussions against animal cruelty and abortion, animals and fetuses are often considered to possess the equal moral standing as fully-functioning human adults, implying that they have the moral right to be treated as such. In his essay "All Animals Are Equal," Peter Singer takes a utilitarian approach in arguing that all living beings capable of suffering deserve a moral status. However, the strategy of determining one's moral community membership based on some generalizable standard risks excluding entities that may lack such quality. Hence, my stance is to acknowledge that animals and fetuses do not have the equal moral status as most humans, and to investigate other moral considerations that lead us to ethical obligations toward such beings that go beyond the principle of equality. Analyzing Cora Diamond's claims about the relationship between humans and animals, as well as Thompson's position in defense of abortion, I contend that our connection to and concern for co-creatures obligate us to treat those who do not possess moral community membership with decency. Since animal and abortion ethics deal with human attachments and emotions toward other entities, we may be better off focusing on the fundamental relationships and responsibilities we have with co-living beings rather than the rational approach of moral categories, rights, or considerations.

KEYWORDS

Ethical Responsibility, Animal Cruelty, Abortion, Principle of Equality, Moral Status, Human obligations

For long, animals have been a useful tool for humankind. They have various ranges of use, becoming our daily bread, decoration to hang on the wall, or life companion assisting disabled humans. However widespread such practice might be, it has not been without moral condemnation. Another historical practice that is no longer morally underground and widely debated is abortion. A common feature of moral discussions surrounding these two subjects consider animals and fetuses as beings of equal moral status, as fully-functioning human adults, and thus grant them the moral right to be treated as such. However, I believe we need a slight repositioning of focus in the moral debates regarding animal cruelty and abortion. My stance is that although animals and fetuses do not possess an equal moral standing as most humans do, there are other moral considerations that render us into ethical responsibility and obligation toward such beings beyond those merely following from the principle of equality.

In the first section, I present Singer and Bentham's utilitarian perspective granting equal moral consideration for all entities capable of suffering. I assert that utilizing sentience as a criterion for moral status is implausible because it lacks discernment in identifying immoral interests. One reason why the criterion of sentience itself is wrong, for example, is that it considers the interests of sentient yet torture-loving psychopaths as equally weighty to morally responsible adults. Second, I suggest that something I loosely identify as capability of moral agency would be a better criterion when we are trying to determine one's moral status. To the worry that making moral status exclusive to those capable of moral engagement might lead to exploitation of the morally incapable, I respond that Singer's criterion of sentience is also not as inclusive as he thinks, and thus this sort of problem exists in Singer's argument as well. Third, I argue that my criterion could be a solution to this problem as membership in the moral community obligates one to not treat entities that do not have moral status as an instrument, as doing so would be diminishing one's own moral status. I point out that as long as we continue to apply the strategy of determining one's moral community membership based on a generalizable quality, we will face the problem of having to leave out people or entities that lack such quality. Thus, my approach has the advantage of not hiding from this fact but also not implying counterintuitive moral results. In the fourth section, I integrate some arguments from Cora Diamond on the relationship of humans and animals to help provide further reason how having an exclusive moral status doesn't mean that one is allowed to do whatever one

wants to entities who lack this status. The key idea here is that of one's relationship and moral status that gives rise to moral obligations to other creatures. Fifth, I describe how this approach generalizes by applying it to debates concerning abortion. By analyzing Thompson on her defense of abortion, I supplement the principle of equality in moral status or rights. I stress that what matters more is our connection with and concern for co-creatures that renders us responsible to treat those that do not possess an equal moral status with decency. For example, even if there is no moral injustice involved in aborting a fetus that lacks moral rights and status, it remains possible for the mother to violate morality in some other way via her relationship to the fetus. I conclude with the consideration that maybe we should abandon the rational approach of moral categories, rights, or considerations when it comes to topics like animal and abortion ethics that concern moral relationships and emotions with other livings. Rather, we might do better focusing on the fundamental relationships and obligations we have toward co-living beings and the stories demonstrating them that are not strictly tangible or rational.

I. SENTIENCE AS A TICKET TO MORAL MEMBERSHIP

In "All Animals Are Equal," Singer argues that we should extend the principle of equality that applies to all members of humankind to other species, mainly animals (Singer 1989, 148). He asserts that the interests of all human beings deserve equal consideration, which means not that their interests should be treated exactly the same but should be given an equal amount of weight in our deliberation (Singer 1989, 149). To Singer, the common property trait that provides humans a moral status and thus worthy of equal consideration is the basic fact that they have interests in the first place. He then goes on to argue that an entity must be capable of sentience, or feeling pain and pleasure, to have interests. Following this line of thought, Singer argues that if sentience endows equal moral status for human beings, then should it also for animals that are capable of pain and pleasure. Therefore, it is morally inconsistent to say that it is okay to conduct scientific tests on animals and immoral to do so on humans as both deserve equal moral status. Singer's approach to animal treatment requires logical consistency when we make personal choices such as choosing a vegetarian diet or using cruelty-free cosmetics. With this perspective, it logically follows that animal eating

is a moral wrong and animal-friendly practices should be a legitimate and widely applied cause.

Utilizing sentience as an indicator of a being's interest and moral status, however, is problematic. Singer develops Bentham's idea that animals deserve moral protections as they have the capacity of sentience just like humans. This view does not deny important differences between humans and animals, but focuses on a relevant similarity of the two, the capability to experience suffering and pleasure being the common core (Beauchamp and Frey 2011, 173). However, granting equal consideration for the interests of all sentient entities is implausible as it does not discern for those that engage in morally inappropriate sentiments. It is evident that not all entities experience pain and pleasure in an identical manner, and some beings lack the moral ability to distinguish between what is right or wrong. This could lead to inconsistency of one's interest or sentience with common moral perception. For instance, consider there is a psychopathic being who derives pleasure from moral wrongs such as torturing animals and causing suffering. It does not logically follow to say that such a being deserves a moral status and thus qualifies for equal moral consideration in satisfying their interests. In fact, their immoral interests and suffering coming from not being able to afflict harm to others should not be given any moral weight. Sentience neither logically connects with one's ability to be moral, nor is a promising criterion to back up one's moral status given its goal of defining moral membership in terms of having interests at all. In the next section, I offer a better criterion for moral community membership.

II. WHAT GIVES ONE A MORAL STATUS?

Since I assert that the utilitarian model of "pain and pleasure" does not seem to be a suitable element composing moral status, it would be natural to follow that up with my own perspectives on what makes one worthy of equal moral consideration. Rather than focusing on granting equal moral status for the largest population possible, it is more plausible to acknowledge that moral status cannot be equal for all living beings as membership to the moral community should only apply to agents that are morally capable. In determining moral status, we should consider not only the ability of one to perceive or sense one's surroundings, but also to gain, think, or act morally as a result of such sentience. What matters

more in the moral sphere is not that one feels things, but whether one is able to reflect one's moral relationship and responsibilities with their surroundings and co-creatures as a result of one being able to feel certain things. I believe that moral engagement is key to moral community membership. Such qualities include but are not limited to: having a conscience, distinguishing right and wrong, utilizing moral intelligence to reflect and guide one's life, being altruistic and considerate of the marginalized, being able to break out of impulse or choices to maximize one's benefits in consideration of better good. These qualities demonstrate one's ability to engage in morality, and I believe only one to whom morality actually matters may obtain a moral status. As animals are incapable or less capable than humans to engage in this sort of complex moral activity, they do not obtain an equal moral consideration as human beings, and thus do not possess an equal moral status to humans.

An objection that follows this argument considers the treatment of humans who lack moral qualities. One might argue that if we were to give different considerations to individual interests depending on one's ability to reason morally, that would leave out humans that lack such quality from moral protection. Singer stresses this problem, and his intention behind the sentience-as-interest argument is to find a non-arbitrary criterion to make as many entities' interests worthy of equal consideration. However, Singer's criterion of sentience is also not as inclusive as he thinks, and this sort of problem also exists in Singer's argument as well. Singer argues that qualities such as intelligence, moral capacity, or physical strength should not be a basis of claim to equality (Singer 1989, 150). This is because the above-listed matters are simple factual qualities that in no way justifies us to discriminate against or be partial toward someone's interests. Nevertheless, I do not see how sentience differs from such factual abilities that cannot be a basis of the principle of equality as there are discrepancies in ability to sense that make Singer's logic prone to misinterpretation or misuse. Under Singer's logic, I do not see a plausible cause to not exploit living things that are incapable of physical or emotional pain or pleasure. For instance, human beings under paralysis, brain death, vegetative state, insensitivity to pain, or alexithymia count as not having interests under this view. Some beings that respond to external stimuli or capable of locomotion through bodily movement yet incapable of experiencing pain or pleasure, such as aquatic animals without complex nervous systems or plants (even those that are carnivorous such as flytrap or pitcher plants) also do not

possess interests. However, it would be a mistake to say that these beings do not have an interest in living life.

III. DOES MORAL INEQUALITY IMPLY DISCRIMINATION?

Under my criterion of moral engagement, the fact that animals and humans possess unequal moral status does not permit a moral agent to exploit those of lesser moral status. In other words, membership to a moral community does not endow one a right or permission to abuse others incapable of moral engagement. This is because moral community membership involves an agent reasoning and acting morally in a way that is considerate of one's surroundings and co-living beings. If one treats animals as mere instruments in achieving ends, one is thereby demonstrating one's inability of moral engagement with the community and co-creatures, which leads to the undermining of one's own moral status. Moral abilities are unequally distributed among different species and entities, which means some do not possess an equal moral status as others. Nevertheless, it is crucial to note that moral membership, once granted, obligates the member to treat other living beings with decency and respect regardless of their moral status.

If we turn back to the original problem of the criterion and principle of equality approach, the strategy of determining one's moral status based on a generalizable quality or criterion always leads to an exclusion of a group no matter how thin the criterion gets. In the first place, I believe that endowing a moral characteristic to beings based on a certain standard, in Singer's case, possession of interests or capability to feel pain and pleasure, will always encounter this problem of a limitation. Thus, my attempt is to deny equal moral status or consideration for interests for all people, animals, or entities in the first place, and imagine what a world with such inequality would look like. Moral inequality does not mean those superior in moral deliberations get to mistreat those who are lacking, as by doing so one would be undermining one's own moral status. In addition, treating other living beings or entities that lack moral community membership with dignity and kindness is not an act of charity that expects gratitude from the beneficiaries or entitles the benefactor to a right or a favor in return. Such a relationship does not involve benefits but moral obligations that are about having decency in treating and respecting one's relationship with co-creatures. It is apparent that the moral capabilities of an animal and a human being are different in their depths, and

there are discrepancies in moral intelligence among humans. To say that a dog, full-grown adult human, infant, and a mentally disabled person all deserve an equal amount of moral consideration for their interests based on their common capability of suffering or happiness would be blinding oneself from the apparent correlation between moral capacity and moral community membership and their unequal moral status in the real world. Rather, I find it more plausible to admit that not all humans, animals, or entities' interests and needs deserve equal deliberation in the moral sphere. From there, it is important to recognize that it is the obligations and responsibilities of morally advanced or capable beings coming from the moral community membership that prevents agents from mistreating the morally lacking or incapable despite the inequality.

IV. IS THERE SOMETHING BEYOND CRITERIONS?

In the previous section, I discussed that as long as we maintain the approach of using moral criterion to grant morally relevant status, we are prone to an error of excluding certain entities that do not suffice such criterion. Making such a rational, factual, and objective case through the utilitarian principle of equality seems central to Singer in discussing animal ethics (Diamond 2002, 71-72). What matters for him is the moral obligation that arises from a living's ability to suffer as shown in his response to Coetzee who argued that it is more important to consider animals in terms of our lives in relation to them. "We can't take our feelings as moral data, immune from rational criticism," writes Singer in his response to J.M. Coetzee's 1999 Tanner Lectures, The Lives of Animals (Trask 2012). However, the particular perspective that Singer demands us to see the world from renders us to the dangers of a single story. Under the logical consistency that Singer asks us to apply in making ethical decisions considering human and animal rights, all his conclusions naturally follow. Nevertheless, the demand to see the world from that perspective itself is in question. By formalizing ethical actions by attributing why we refrain or abide by certain norms solely in the principle of equality, Singer focuses too much on reasons and analytical standards, which might blur a fundamental reason or consideration of why we do certain things.

In response to Singer, Cora Diamond published "Eating Meat and Eating People," where she argues that defending animal rights involves a moral relationship between humans and animals rather than moral equality rising from

one's capacity to suffer (Diamond 2018, 723). Diamond asserts that approaching animal treatment with the capacity of sentience is not the right way of beginning the discussion. She goes on to state that:

> ...anyone who, in discussing this issue, focuses on our reasons for not killing people or our reasons for not causing them suffering quite evidently runs a risk of leaving altogether out of his discussion those fundamental features of our relationship to other human beings which are involved in our not eating them... (Diamond 2018, 724)

In other words, Diamond wants to avoid focusing too much on rationality or reasons for not eating or mistreating animals or humans because doing so blurs our fundamental relationship with those beings that could be the cause for our decency toward them. Here, the perspective diverges a little. The reason for not abusing or exploiting fellow humans or animals now becomes not merely because such actions would cause these beings suffering, but because of our relationship unattached to any normative system that makes us recognize their significance in our lives and renders us responsible for treating them as an entity with character. What matters here is not one's recognition of others' rights or moral status that acts as a preventive mechanism from moral wrongdoings, but a deeper understanding of one's relationship with surroundings and co-creatures that allows one to recognize the significance of others and approach them with care and respect.

Singer regards that denying equal moral status and consideration for all sentient beings would lead to the issue of moral anarchy where there is no moral restriction that prevents one from discriminating against those with less or no capability of moral agency. However, no matter how much we idealize equality and blind ourselves from apparent discrepancies arising from factual differences among creatures, the rights of the weaker become vulnerable when faced with the stronger voices of those in power. Therefore, it is important for moral agents to acknowledge the unequal moral status among different species, and face the problems arising from such inequality. The question now becomes why would moral agents, despite this clear discrepancy of moral abilities among species and entities, have any obligation toward beings that are outside of the moral community. This is because membership to a moral community requires one to

go further than acting upon sheer self-interest, which prevents one from abusing moral status in fulfilling one's interests. A moral agent objectively recognizes that an infant, mentally disabled person, and animals all possess unequal moral status, but an "other-concerning" component of the moral community obligates them to act with decency regardless of other being's moral status. The attitude that one needs to seek, then, is not the justification of mistreatment or exploitation of others based on factual differences, but an attempt to ensure that such inequality does not lead to discrimination. However, although it is primarily the obligation rising from an unequal moral status that renders one to seek such an attitude, there is more involved in this story, which could be answered by Diamond.

Under the analogy of equal moral consideration, it is true that some entities have a special moral status that others do not, and that can be transitioned into rights. However, this doesn't mean that we are allowed to mistreat entities incapable of moral agency due to our relationships and obligations to them. There are moral categories that are meaningful beyond what merely follows from the principle of equality that makes us meticulous and careful in making choices or conducting actions that involve our other living beings that lack moral status. Possession of moral status gives us a special status that allows us to ground relationships with others who possess the same status, and this relationship motivates us to deliberate and act in consistency with the community's shared moral beliefs and attitudes. In a moral community, it would be plausible to say that members sacrifice and cooperate to protect the weak or underprivileged, as moral agents are considerate of one's surroundings and the welfare of co-creatures. As moral agents, then, we naturally navigate what it means for us to be a member of such a community, which could be a call that connects and commits us to our relationship with co-creatures. Through such relationships, we come to see co-creatures such as animals as livings that matter just as much as our fellow community members, which commits us to treat them as entities with characteristics deserving of respect, or even alternative forms of people with moral status. If we obsess ourselves over objective duties or rationale in not abusing those without moral agency, we are unable to focus on fundamental reasons why morality requires caring for the weak and not holding others' weaknesses against them.

V. WHAT FOLLOWS FROM THIS ASSERTION FOR ABORTION?

With regards to differing perspectives in discussing moral responsibilities toward non-humans, there are other philosophers who present alternative reasoning outside of the principle of equality or rights. In "A Defense of Abortion," Thomson discusses two distinct approaches to investigate the moral permissibility of abortion. Contrasting the question of what it means to grant a fetus right to life to the matter of moral decency toward the fetus, she suggests that abortion involves another moral concern irrelevant to rights-violation. When we discuss abortion in the perspective of fetus's rights, there are some cases that one could argue abortion is an unjust killing given the fetus's status and right to life. However, Thomson deviates from this perspective and further applies matters of moral decency in explaining abortion (Thomson 1971, 59). For example, if a wealthy person refuses to return land and house documents to all debtors and share food and resources during a famine, she does not violate anyone's rights as no one but she is entitled to her property. However, turning her eyes away from a community struggling to meet daily ends would be another form of violation that involves human decency toward other beings. According to Thomson, the bar to violation of decency in the case of abortion remains high as pregnancy is physically and emotionally draining for the pregnant woman, and the decision to carry the fetus in cases when the pregnant woman is in a burdensome circumstance is supererogatory that should not be commanded by law (Thomson 1971, 60; 63-64).

Applying the previously discussed principle of equality to abortion, fetus lacks capability of both sentience (at least during certain stages of pregnancy) and moral community membership as an underdeveloped human being. The fetus is a dependent of the pregnant woman's body and an entity to whom morality does not matter, and its interests and moral status is absent or lacking than that of the pregnant woman's. In a case where the woman bears a physically or mentally challenging occupation or is not in a situation to raise an infant, she should be able to abort the fetus upon her choice and deliberation. Here, abortion is not an unjust killing of a fetus as it does not involve violating any rights, as the fetus is not a member of the moral community. Nevertheless, regardless of the woman's financial, mental, or physical incapability of childbirth and parenting, such a decision leaves the woman personal moral consequences and possible self-condemnation

for directly aborting a child. Although the woman did not commit any injustice to the fetus nor violated anyone's rights, it is her position as a moral community member considerate of co-creatures, and her relationship and attachment to the fetus that renders her to moral guilt and burden. The fact that the pregnant person did not go out of her way to bear the fetus for nine months, give birth to the baby, sacrifice for its upbringing, does not mean that she has committed a crime against humanity or moral failure. However, despite the fact a woman has made a reasonable and appropriate choice of abortion regarding whatever personal circumstance she is faced with, she would experience heavy moral self-contempt and deliberation--although the woman should not be subject to public condemnation or legal punishment--in this process. Her fundamental relationship and connection with the fetus rising from her moral community membership leaves a mark in her heart that she would possibly carry for her lifetime.

VI. FURTHER DISCUSSION

My project in this writing has followed a route that leads to considering that there are more interesting and profound moral concerns beyond what merely follows from the principle of equality in approaching abortion and animal ethics. I argue that we cannot draw the circle to moral community membership with Singer's criterion of sentience as it would grant equal moral consideration for morally counterintuitive interests. Then, I provide a what-if analysis of Singer's view based upon the principle of equality by drawing a bar to moral community membership in a way that is less inclusive yet more morally relevant than Singer. I explain that exclusive equal moral consideration for those capable of moral agency does not end up at a conclusion where moral agents can do anything they wish to entities lacking moral status. This is due to obligations rising from the agent's moral status and relationship with co-creatures of the world that refers back to Diamond's argument. Tying one's fundamental connection and relationship with humans or other non-humans to the issue of abortion, I demonstrate the importance of such personal or emotional data that redirects our discussions focusing too much on a single perspective of formal moral obligations regarding rights or moral status.

In discussing animal and abortion ethics that concerns real-life human behaviors and connection with other creatures such as fetuses or cats, I cannot help but think that there is a need for somewhat intangible or informal accounts

of our experience with other humans, entities, and surrounding of this world. In this sense, exploring the fundamental relationships of the kind Diamond emphasizes in dealing with moral issues concerning human emotion with other entities becomes central. The topic of this project does not remain purely in the theoretical moral sphere, but is a matter that continues to be lived by us in real life. Therefore, we should admit the importance of both rational and emotional data in our discussions rather than attempting to separate them and discard the latter. Practically speaking, there are myriads of moral theories that make sense in considering these ethical issues, but human practices of animal cruelty or abortion is not remote from our daily lives. Since these issues are complex, worldly, and alive in their nature, we need something more than objective reason. I agree that factual descriptions of what morality requires in following a vegan diet, using products that are not cruelty-free, or getting a medical examination prior to abortion, must be given serious contemplation. However, the fact that these dilemmas and subsequent actions involve something more than moral obligations, that is, human emotions and personal life circumstances, gestures toward the inevitable need for personal stories in the debate. Without the real-life stories of the butchers, pet owners, women suffering from abortion aftereffects, moral debates on these issues will remain incomplete, lacking wide consensus and resonation among our community that is fundamental to these discussions resulting in action and change. Emotional relationships or personal accounts of those who had to face and act upon these moral dilemmas should not be undermined as irrelevant or irrational in determining morally right and wrong responses in such circumstances.

REFERENCES

Beauchamp, Tom L., and Raymond Gillespie Frey, eds. 2011. *The Oxford Handbook of Animal Ethics.* New York: Oxford University Press.

Diamond, Cora. 2018. "Eating Meat and Eating People." In *The Norton Introduction to Philosophy*, edited by Gideon A. Rosen, Alex Byrne, Joshua Cohen, and Seana Valentine Shiffrin, 723-29. New York: W.W. Norton & Company, INC.

Diamond, Cora. 2002. "What Time is It on the Sun?" In *Philosophers in Conversation: Interviews from The Harvard Review of Philosophy*, edited by S. Phineas Upham, 69–81. New York: Routledge.

Singer, Peter. 1989. "All Animals Are Equal." In *Animal Rights and Human Obligations*, edited by Tom Regan and Peter Singer, 148-62. Englewood Cliffs, NJ: Prentice Hall.

Thomson, Judith Jarvis. 1971. "A Defense of Abortion." *Philosophy & Public Affairs* 1 (1): 47-66. http://www.jstor.org/stable/2265091.

Trask, Michael. 2012. "The Ethical Animal: From Peter Singer to Patricia Highsmith." Post45. Accessed December 3, 2020. https://post45.org/2012/11/the-ethical-animal-from-peter-singer-to-patricia-highsmith/.

compos mentis

The Clarification of Philosophical Inquiry: In Defense of Metaphysics, Ethics, and Aesthetics

Michael Hyde

Central Michigan University

ABSTRACT

In this paper, I aim to demonstrate the usefulness of retaining metaphysics within the context of analytic philosophy. Utilizing the work of such thinkers as Carnap and Wittgenstein, I make the case that metaphysics (as well as aesthetics and ethics) is a legitimate component of philosophical discourse. In addition to refuting Carnap's case against metaphysics through the use of his own proposed linguistic framework, I acknowledge the reasons that Carnap wishes to reject metaphysics, and reframe the discussion in a manner that attempts to address those concerns while showing that metaphysics are useful even to an Empiricist.

The nature of this paper is twofold; firstly, I aim to briefly[1] defend the inclusion of metaphysics, ethics, and aesthetics in the study of philosophy.[2] This is in response to Carnap's (2003) *The Elimination of Metaphysics Through the Logical Analysis of Language* (*Elimination*). I will contrast it with his work "*Empiricism, Semantics and Ontology*" (*Empiricism*), working to demonstrate the tension between the two[3] (Carnap 2003). I will defend a specific interpretation of the latter and reject the former, utilizing various philosophers from both the analytic and Continental traditions to do so. Secondly, I shall then assert that the task of philosophical inquiry needs clarification. Rather than emphasizing the divide between the analytic and Continental traditions, I maintain that we must transcend the artificial divide between the two and return to the work that analytic philosophy understood itself to be doing in the 20th century; chiefly, being clear about not only what we are doing in our work, but also elucidating the motivation behind the undertaking of the work itself.

I begin with discussion about Carnap's (2003) paper *Empiricism, Semantics, and Ontology*. Here, Carnap presents the concept that we accept and utilize various linguistic frameworks based on pragmatic considerations. Questions are either internal or external to the framework, and it is only within the context of such frameworks (internal questions) that statements can be determined as either true or false. Instead of concerning ourselves with whether a framework is right, we must determine if it is useful. While this presents a surface challenge to such areas of thought as metaphysics, this is not as problematic as it may appear.

Epistemology suffers from such a view in the same way that metaphysics and ethics do. In it, statements about the nature of existence or other knowledge claims cannot be posited as absolutes, they can only be evaluated within the context of the framework that the speaker is using. One cannot claim to know anything universally, such claims are also limited to the speaker's framework. However, there is utility in the concept of pragmatic considerations. One can

---

1. The purpose of this paper is to offer a brief overview of such a defense. Some parts of this paper could easily be turned into papers of their own, and as such, I may only briefly mention some relevant considerations and wholly ignore others.

2. This paper is largely a response to Carnap's "*Elimination of Metaphysics Through Logical Analysis of Language.*" Familiarity with it and "*Empiricism, Semantics, and Ontology*" are recommended.

3. Some may consider *Empiricism* as a step back from the position outlined in the *Elimination*. While this is likely, my argument still stands regarding the inclusion of metaphysics in linguist frameworks.

imagine with relative ease a framework in which the pragmatic consideration would indicate that treating a particular claim *x* as being always and everywhere true in context of *y* would be the optimal treatment of *x*.

It is of note that the pragmatic adoption of a framework is contingent upon not only the ends with which an agent is pursuing but also the motivation behind the pursuit of such ends. That is, two agents with the same end *e* may diverge on the framework adopted because of motivation *f* compelling one and not the other. This comes into play when considering the inclusion of aesthetics or ethics into philosophical discourse. For some, the motivation of their framework would pragmatically stipulate that such things be rejected, while the others may necessitate their inclusion.

For example, an agent who was concerned with ontological commitments that mirror the aspects of the world that can be validated via sensory experience or empirical validation would have no use for any prescriptive discourse. As such, eliminating metaphysics or ethics (branches of philosophy that are excluded by the empirical stipulation) could be argued as useful or necessary. However, within a linguistic framework, if an object *x* is contained within the scope of an adopted framework, then *x* exists. As such, metaphysical objects are not barred from the philosophical discourse within the context of linguistic frameworks that include them. While Carnap believes that utilizing such frameworks will lead us away from metaphysics, ethics, and aesthetics, I aim to demonstrate the opposite.

Having demonstrated that Carnap's view outlined in *Empiricism, Semantics, and Ontology* (Carnap 2003) is not necessarily dismissive of metaphysics, aesthetics, and ethics, I next turn my attention to an earlier paper of his entitled *The Elimination of Metaphysics Through Logical Analysis of Language* (Carnap 2003). In this work, Carnap attempts to reduce "real" philosophy to epistemology and logic, discarding all other philosophical projects as either misguided art or nonsense.[4] He claims that any statements that are not able to be empirically validated are meaningless. This is not to say that Carnap believes any statements failing to meet such criteria are not of value, just that they are reduced to mere expressions. (Consider the distinction between prescriptive and descriptive statements, this is of importance and I will return to it at a later point in the paper). Metaphysics, ethics, and aesthetics are some such meaningless expressions. While

---

4.   Carnap does more in this paper than what I describe. For sake of simplicity, I ignore the work he does on meaningless statements that do not pertain to my purposes directly.

these expressions can be useful in several ways, they do not tell one anything about the world, but rather about the agent making them. Carnap writes that all metaphysics[5], "do not serve for the descriptions of states of affairs, neither existing ones... or non-existing ones...They serve for the expression of the general attitude of a person towards life" (*Elimination*) (Carnap 2003, 196).

I argue that Carnap is mistaken to reject metaphysics, ethics, and aesthetics. Furthermore, I suggest that he becomes at odds with himself in taking such a stance. In *Empiricism* (Carnap 2003), he allows the acceptance of some "meaningless" metaphysical concepts such as numbers as they are "internal" to the adopted framework. Within this context these concepts can be assigned a truth value. In *Elimination* (Carnap 2003), he insists that we reject all metaphysical, aesthetic, and ethical concepts on the grounds that they are meaningless and unable to be empirically verified. These two views are incompatible.

By eliminating metaphysics, Carnap is successful in preventing such abstract things as a God existing outside of space and time to be included in his ontology. However, he is also necessitating the elimination of frameworks that could prove useful, even to an Empiricist. While some examples include numbers or time, I aim to demonstrate larger and more nuanced instances in this paper.

A possible reason to retain metaphysics (beyond inclusion by default within a particular linguistic framework) is that an agent may conclude that their ontological framework is flawed or incomplete without it. Similarly, one may desire to utilize an old ontological or linguistic framework toward a new end, and as such must adjust or pragmatically posit new or amended metaphysical principles toward the advancement of the desired end.

These reasons demonstrate the sense in which discussion of ethics or aesthetics could be understood as attempts to build supporting structures within the context of a framework. They could also be outcomes resulting from the proper functioning of it.[6] In either case, they would be the result of pragmatic selection.

We see Carnap's process of pragmatic selection demonstrated throughout philosophy. One such area is that of logic. As Tennant (1981) points out in *From Logic to Philosophies*, how we choose our logic is contingent upon what we

---

5.   He includes aesthetics and ethics in this description as well.

6.   It is prudent to mention here that some would choose to suggest that we pursue ontological investigation separate from metaphysics, but that is something I must set aside for the time being.

wish to do with it, as well as what we take the logic itself to be doing. These considerations must be considered when choosing a framework. Am I seeking new truths about the world? Am I looking to solve a problem? Am I seeking to articulate and define a problem that I have noticed exists?

Here we must return to the prescriptive/descriptive statement distinction as previously promised. With a commitment to truth such as Carnap and others seem to possess, they are prohibited from making various prescriptive claims that one without such a commitment is free to. For many in this camp, it seems that they are okay with this. Their primary concern is that their adopted framework safeguards against various things, including what they would consider to be metaphysical nonsense. Carnap limits himself to descriptive statements only when he rejects metaphysics. This becomes problematic when considering the prescriptive nature of his work in *Empiricism* (Carnap 2003).

I sympathize with Carnap's concern about metaphysics in a particular sense. Consider something like Leibniz's famous system of monads. It is convoluted and esoteric. As Carnap and the logical positivists would point out, it is lacking any sense of verification conditions. One must either take it or leave it, with very little sense of why they should accept it. To treat monads as descriptive of the world seems far-fetched. Even embracing them as a "useful fiction" to understand the nature of reality seems questionable. How is this understanding of reality pragmatically useful? It is almost easier in such cases to take Carnap's view that such systems express a person's view of the world than to consider it a legitimate linguistic framework.

It becomes apparent that there exists some necessary sifting between what ends justify a useful fiction and what ends do not. As a believer that not all metaphysics are nonsensical, I need to produce some mechanism to differentiate those of value from those that are not, such as the monad. At this time, I have no such equipment and must fall back on Carnap's concept of utility within linguistic frameworks.

Before continuing, I must make an argument for prescriptive statements. The argument that we should select linguistic frameworks with consideration to utility is itself a prescriptive statement. It does not have a truth value and is a statement of judgment rather than a reflection of the world. As such, it seems that if Carnap and his ilk wish to take themselves seriously, it is necessary to accept some prescriptive statements that themselves are not "meaningful" by Carnap's

criteria. If we ignore Carnap's qualms and allow such statements, what criteria would he use to differentiate between acceptable ones and ones he wishes to eliminate, such as ethical ones?

I now offer two examples to demonstrate the usefulness of metaphysics. Firstly, I put forward the concept of the Tao in the Tao Te Ching. In Taoist thought, no effort is made to put forth verification conditions for the Tao. It explicitly states that anything that can be experienced, named, or demonstrated is not the Tao (Lao Tzu 1988, 1.1). And yet, the entirety of the teaching is an attempt to reveal what things can be revealed about the Tao, as well as instruction of how to live based on these revelations. The inability to verify the Tao does not seem to inhibit its ability to give meaning.

The meaning that Taoism provides does not present itself as someone's expression of a view toward life. Rather, it is the foundation for a prescriptive view of the world and a normative basis of behavior. In fact, there is a sense in which it takes itself to be describing that nature of the world, and providing prescriptive statements based on this description. This provides a tension with Carnap's view of metaphysics.

Carnap might argue that the Tao proves his view of metaphysics correct, that it is nothing more than a "personification of natural phenomena… {a} quasi-poetic expression of man's emotional relationship to his environment" (*Elimination*) (Carnap 2003, 169). Yet, Carnap himself provides the equipment to refute this. In the *Elimination*, Carnap notes that he avoids using the term worldview to describe a metaphysic because it blurs the "difference between attitude and theory", a difference that is "of decisive importance" (*Elimination*) (Carnap 2003, 170). The Tao is not an attitude, it is a theory of existence that prescribes a certain attitude toward life. As such, I maintain that it is problematic for Carnap.

Carnap may at this junction return to my admission concerning monads, asking me to demonstrate how the Tao and the world of monads differ. I would argue that they differ in their levels of verification. Neither monads nor the Tao can be experientially or logically verified. Yet, the balance and harmony that the Tao speaks can be experienced and perceived in particular ways. Beyond this, the positive results in the lives of those who follow the Tao are more evident still. What evidence do we have for monads? What reason do we have that would compel us to believe (or necessitate belief in) such an idea? There is none. I argue that

within the realm of things that are not themselves verifiable, there are levels of verifiability. These levels are what differentiates the Tao from the monad.

Even if I accept Carnap's (2003) assertion that the Tao is a mere "quasi-poetic expression" it still seems to be a fine example of my position. Accepting the existence of the Tao for a particular pragmatic purpose, such as living the best life possible, is accepting a linguistic framework for a practical end. The linguistic framework necessitates the positing of a metaphysic. If one wants to live well, and the Tao helps them do that, then their end has been achieved regardless of whether the Tao is "real."

I present my second example in a similar manner. Consider the metaphysical concepts found in existential literature such as the Absurd, Despair, or Will to Power.[7] These, I argue, are metaphysical concepts that were put forth to describe a perception of a sensation that is common among humanity. Wittgenstein takes on the question of senses and sensations in the *Philosophical Investigations* via the concept of pain (Wittgenstein 2003). He argues that no pain is not unique to an individual and that the illusion that there is a pain that is "my" pain stems from a word trick, "the substitution of 'identical' for 'the same'" (Wittgenstein 2003, PI 253-254). Pain cannot be pointed out in the world, Wittgenstein admits. We do not have a picture of pain. However, we have the image of pain portrayed by the grimace, the outcry of the individual experiencing it. And, when this is present, we have something in the world that we can experience and verify.[8] As Wittgenstein points out, "An image is not a picture, but a picture can correspond to it" (Wittgenstein 2003, PI 301).

Taking Wittgenstein's arguments about sensations (and language) seriously, we can reasonably extrapolate that argument to include sensations that are not a result of our five senses and that transcend the ability to express within linguistic confines. Some such sensations or ideas include the Absurd and Despair. These concepts demonstrate the existence of such shared non-private sensations. If this is the case, we have a picture of metaphysical concepts that do reflect the world. Additionally, we may have understood verification conditions for such sensations, that being their quality of being commonly perceived sensations among persons.

---

7.   See Camus, Kierkegaard, Sartre, and Nietzsche for more on these ideas.

8.   For more on sensation, pain, and private languages, read Wittgenstein's *Philosophical Investigations*. Particularly around 243-315. For sake of space, I must keep this discussion brief.

While the exact conditions under which one can be said to be experiencing Despair or the Absurd are not clearly laid out here, this is an undertaking for a different paper. The thinkers who put forth such concepts often dedicate their work to exploring and defining this question. For sake of this paper, I find it sufficient to argue that identifying such a state is possible in the very way that it is with pain.

Switching our attention to the next (and perhaps weaker) defense of the inclusion of metaphysics, ethics, and aesthetics in philosophical discourse, let us consider an appeal made by Grice & Strawson in "In Defense of Dogma." In their refutation of Quine, one of the remarks they make is that Quine is, "certainly at odds with a philosophical tradition that is long and not wholly disreputable" (Grice & Strawson 2003, 291). "But there is no need to appeal only to tradition; for there is also present practice," they write (Grice & Strawson 2003, 291). At risk of stealing their argument, I contend that the same argument applies to the present discussion. From the conception of Western philosophy, ethical and metaphysical reflections have been included in the discourse. This has not changed. What purpose would it serve to discard them now? While Carnap and company may feel that it would help clarify philosophical discussion, it seems that all it would do is reduce the number of possibly useful linguistic frameworks to consider.

Additionally, there is the deeper point that these topics already largely exist in the overarching philosophical linguistic frameworks. On Carnap's assertion, he and the logical positivist can choose a linguistic framework that eliminates such things, but in the larger context of philosophy they are already established. Thus, a charitable interpretation of Carnap would have us interpret him as urging us to consider a framework lacking metaphysics, aesthetics, or ethics. However, this is not consistent with what Carnap (2003) does in the *Elimination*.

Thus far, I have argued for the inclusion of metaphysics, ethics, and aesthetics in philosophy from an argument from inclusion in linguistic frameworks, from an argument from pragmatic use, from an argument differentiating prescriptive and descriptive statements, and from an argument from tradition.

I understand that Carnap believes that by removing metaphysics, ethics, and aesthetics from philosophical discourse, we will clarify what we are doing in philosophy. It will serve to remove those that rely on word games from philosophy. While I am sympathetic to this desire, I do not believe that removing those things is the way to achieve this. Instead, I argue that we must clarify what we are working towards and why we are doing our work.

How we do this is through identifying not only our linguistic framework but also our motivation for selecting it. This will address Carnap's concerns in two ways. Firstly, it will help us better identify what work we should concern ourselves with. If someone claiming to be a philosopher is utilizing a linguistic framework that is not useful to us or is at odds with our own ends, we can reject that framework and that person's work within it. This allows us to filter the positions we consider in pursuing our own work. Secondly, it will minimize the amount of interpretation that goes awry when reading other's work. Time spent formulating a rebuttal to a piece of work that one misunderstood could be put to better use undertaking some other philosophical endeavor. The number of schools of thought based on such misunderstandings that clutter the discourse will be reduced. This will be accomplished without having to sacrifice already established philosophical discourses such as metaphysics, ethics, and aesthetics for those who wish to retain them.

I wish to deviate for a moment from the discussion at hand to provide an example of the pragmatic framework selection and clarification of purpose outlined in this paper. I do this so that the reader may have a clear understanding of my view.

In Nietzsche's work *Twilight of the Idols*, Nietzsche writes that "metaphysics, theology, psychology, epistemology — or formal science, a doctrine of signs, such as logic and that applied logic which is called mathematics. In them reality is not encountered at all, not even as a problem — no more than the question of the value of such a sign-convention as logic" (Nietzsche 2020) (Reason in Philosophy, section 3). His concern is that those who commit themselves to particular logical and epistemological viewpoints in pursuit of the "real" world are missing the reality that they are pursuing. This is a concern of mine as well. In Carnap's case I am concerned that he is actively rejecting aspects of the world.

In that same work, Nietzsche also states that "One chooses logical argument only when one has no other means... Nothing is easier to nullify than a logical argument: the tedium of long speeches proves this. It is a kind of self-defense for those who no longer have other weapons" (Nietzsche 2020) (Reason in Philosophy, section 6). He previously proclaimed in that work that "I mistrust all systematizers and avoid them…" (Nietzsche 2020) (Maxims and Arrows, section 26). Nietzsche has a disdain for those who commit themselves to logical systems and arguments to understand reality. His position seems to reflect that of Wittgenstein in the PI

when he argues that philosophy should "neither explain nor deduce anything" (Wittgenstein 2003, PI 126).

Like Carnap, Nietzsche puts much weight on the verification of the senses. Carnap himself mentioned Nietzsche's work rather favorably in the *Elimination*[9] (Carnap 2003). Yet, the framework the two select seem at odds in several respects. However, this strengthens my position, demonstrating how two may have similar ends and select different or opposing frameworks towards those ends.

Many compelling arguments and nuances to consider can be drawn on this conversation from *Twilight of the Idols*, as well as much of the body of Nietzsche's work.[10] The point is that while Carnap makes a case for the removal of metaphysics, ethics, and aesthetics, there is a sense in which one could understand Nietzsche to be piecing together a framework in which logic and emphasis on reason are rejected while aesthetics is retained. Furthermore, Nietzsche illustrates well the existence of an in-between area. He neither rejects or embraces ethical judgements, but rather redefines it, or rather, places the dichotomy of life affirming and life denying in the role that conceptions of right and wrong would traditionally fulfill. He performs a similar move in exchanging the free will versus determinism for weak versus strong will.

While I am not claiming that Nietzsche was a pragmatist, I submit that one way of envisioning much of his work, including the recurring theme of transvaluation of values, is as being a pragmatic selection of values towards a desired end (life affirmation). Metaphysical ideas reflecting the world are posited in the framework constructed from these values.

Returning to the discussion around the task of clarifying our intent in philosophy, I wish to offer one more example of why I insist on keeping metaphysical, aesthetic, and ethical discourse as a part of philosophy. Consider a coherentist, Quinean web of belief epistemic framework.[11] The beliefs we hold at the center of the web

---

9. The author suspects this affirmation was largely due to misinterpretation of Nietzsche's work, but will for now charitably assume that this is not the case.

10. I do not have the space here to explore these here, let alone introduce or unpack the relevant (and often controversial) Nietzschean conceptions of truth, morality and aesthetic value. For further reading on the transvaluation of values, see *The Genealogy of Morals*, the *Gay Science*, and *Beyond Good and Evil*. For more on the will, see the *Twilight of the Idols* and the *Will to Power*.

11. I mean only to consider the general idea of a web of belief presented by Quine. The arguments I present regarding metaphysics would be repulsive to Quine, as demonstrated by such things

are the driving force behind much of our conclusions about the world. We protect these core beliefs at all costs. Not only do they inform the ways that we navigate the world, but also the way that we interpret our experiences. Our beliefs shape our experiences, and we affirm our beliefs through the interpretation of those experiences.

However, sometimes strands in our web are damaged or go missing. While learning new information and amending the conceptions that make out the outer web is necessary, damage done to the more essential beliefs near the conceptual center can be harmful to the individual. Accepted metaphysical or moral principles can act as "secondary" or removed core beliefs, allowing the deconstruction of that web to only occur up to a certain point and preventing it from reaching innermost core beliefs.

These principles that are built upon in our webs are not problematic and are even good, so long as the agent is aware of these assumptions and works to address them accordingly. It is the same approach that we use in symbolic logic; we can assume whatever we please so long as we can keep track of our assumptions and discharge them appropriately. If one is working towards a specific end, (contingent upon the agent's ability to exchange a preconception of truth for a useful fiction), incorporating such assumptions can allow them to maintain their web and build on it in ways that can help them achieve that end.

In this paper, I have demonstrated the tension that Carnap creates between two of his works; *Empiricism, Semantics, and Ontology* (Carnap 2003) and "T*he Elimination of Metaphysics Through the Logical Analysis of Language*" (Carnap 2003). I have defended the former and rejected the latter. Utilizing arguments and examples from both analytic and Continental traditions, I have sought to demonstrate the strength of the position I have advanced in this paper. Specifically, I contend that in rejecting all of metaphysics, ethics, and aesthetics, Carnap needlessly eliminates useful philosophical linguistic frameworks. I have claimed that to rectify the problem with philosophy,[12] we need to not only clarify

---

as his contempt for the idea that "there are other kinds of knowledge unfathomable by our cognition, other ways of knowing beyond the limits of our logic, which are deserving of our serious attention." He mocks this quote, calling it "incoherent." However, that is a topic for another paper.

12. This is in reference to the belief that contentions within philosophy are primarily due to misuse of language. While I can agree that some issues stem from this, I believe that many of the disputes within philosophy are caused by either one not having a clear handle on what he is doing, or

what we are doing but the motivation behind what we do. This clarification will then pragmatically dictate the linguistic frame we accept. I then demonstrated that this endeavor is compatible with assuming and accepting metaphysical and ethical principles so long as we are aware of these assumptions and account for them accordingly.

REFERENCES

Carnap, Rudolph. 2003. "Empiricism, Semantics and Ontology." In *Contemporary Analytic Philosophy: Core Readings*, second edition, edited by James Baille, 172-185. Upper Saddle River, NJ: Pearson Education, Inc.

Carnap, Rudolph. 2003. "The Elimination of Metaphysics Through the Logical Analysis of Language." In *Contemporary Analytic Philosophy: Core Readings*, second edition, edited by James Baille, 155-171. Upper Saddle River, NJ: Pearson Education, Inc.

Grice, Herbert & Strawson, Peter. 2003. "In Defense of Dogma." In *Contemporary Analytic Philosophy: Core Readings*, second edition, edited by James Baille, 290-301. Upper Saddle River, NJ: Pearson Education, Inc.

Nietzsche, Friedrich. 2020. *The Twilight of the Idols*. Translated by Walter Kaufmann and Robert Hollingdale.

Tennant, Neil. 1981. "From Logic to Philosophies." *The British Journal for the Philosophy of Science* 32 (3): 287–301. https://doi.org/10.1093/bjps/32.3.287.

Tzu, Lao. 1988. *Tao Te Ching: A New English Version*. Translated by Stephen Laozi Mitchell. New York, NY: Perennial Classics.

Wittgenstein, Ludwig. 2003. "Selections from *The Philosophical Investigations*." In *Contemporary Analytic Philosophy: Core Readings*, second edition, edited by James Baille, 110-130. Upper Saddle River, NJ: Pearson Education, Inc.

by misattributing motive or goals to the works of others. Having an incomplete view of others' motives or goals can cause similar issues as well.

compos mentis

White Privilege and Epistemic Polarity

AJ Jinkins
Portland State University

ABSTRACT

White supremacy has roots in dominant epistemic methods. In this paper, I examine knowledge production for white people as it occurs both structurally and socially, where one's social position interacts with systems of oppression to manifest a certain kind of intelligence (or lack thereof) and epistemic polarity. This examination is done in light of a person's status of dissent to show that dominant knowledge is pervasive, and despite one's best intentions can harm an individual belonging to a marginalized community as they offer testimony. I utilize the Black Lives Matter protests in Portland, Oregon in 2020 to demonstrate how white, privileged people must acknowledge their status of dissent and the ways in which their social standpoint affects what they believe to be true about power and race.

INTRODUCTION

Learning is about opening yourself up to people and experiences and seeking to understand. And when you're talking about race, it involves making horrible mistakes and getting your feelings hurt and not closing the door when that happens. White people can't content themselves with asking questions of black people. They should be prepared to face a lot of anger and not shrink from it. Too often, they've shown themselves to be unprepared for the emotional heat, too willing to check out (B.L Wilson, 2020).

I. A LOOK AT PRIVILEGED (WHITE) KNOWERS AND SYSTEMIC RACISM

I begin this paper with a discussion of systemic racism that entails reviewing the ways in which the most privileged members of society, predominantly white people, produce knowledge. I argue that knowledge production for white people occurs both structurally and socially, where one's social position interacts with the systems in place to manifest a certain kind of intelligence (or lack thereof). I refer to "privileged knowers" throughout this paper as those who operate from a dominant standpoint in their knowledge production practices. The practice I focus on is testimonial exchange, a method of gaining knowledge of the world by receiving testimony and evaluating it (with a default title to credit). The term "privileged," in this case, encompasses white people but is not limited to them specifically. Privilege may apply in any situation where a person achieves dominance over another in their epistemic endeavors (Toole, 2019). Because the context of my analysis emerges in light of the Black Lives Matter protests and civil unrest in the United States in 2020, I find it necessary to utilize the word privilege as it stands in popular discourse on this topic. And so the term "privileged" is applied to white people for the duration of this paper.

My goal is to analyze the epistemic consequences of white supremacist action that seized national attention in 2020. The bulk of this attention does not lean in favor or opposition to white supremacy; it suffices to say the nation stands at a point of exceptional polarity in terms of whether or not people believe systemic racism exists. The deaths of George Floyd and Breonna Taylor, for instance, have made it so white people are prompted to decide whether or not they believe in

racist police violence and systemic racism at large. As a result, white people have shown up in new ways and embraced an "us vs them" mentality within their own communities, families, and predecessors. The polarization among white people is in part dictated by their knowledge production practices. The two sides of white epistemic agency for discussion are the "dominant privileged knower" and the "dissenting privileged knower." They are two sides of the same coin, despite stubborn attempts to distance themselves from one another. In this paper, I argue that their polarization reveals testimonial epistemic injustices complicit in the perpetuation of white supremacy.

Dominant privileged knowers utilize dominant logic that reinforces systemic oppression. They represent a collective consciousness among white people, which Charles Mills identifies as "white ignorance," an epistemology composed of social memory and mainstream political science based on mistaken beliefs. White ignorance strongly insulates itself to"ensure that whites will live in a "racial fantasyland, [or] a 'consensual hallucination," and that the root of all this is the "cognitive and moral economy psychically required for conquest, colonization, and enslavement" (Mills 2007, 49). Dominant identity is entwined with the perceptions of privileged individuals and creates the consensual hallucination.

This fantasy occurs due to selective accounts of history revered by social memory. The resources provided in public education and media purposefully support the paradoxical illusion of white people as both racially superior and the enablers of social progress. Popular sources of information set the foundation of the collective illusion in white people regarding their place in society. Take for instance, a CNN interview with National Security Advisor Robert O'Brien on the topic of systemic racism and police brutality. On May 31st 2020, following the death of George Floyd, O'Brien said "there is no systemic racism in U.S. law enforcement, 99.9% of our law enforcement officers are great Americans, and many of them are African American, Hispanic, and Asian… there are a few bad apples giving law enforcement a terrible name." (O'Brien, 2020). O'Brien stands on a powerful media platform to commend law enforcement officers for being "great Americans." Additionally, on October 27th of this year, CNN boasted about their place in the top-five most watched news sources for the past nine months (CNN Pressroom 2020). One may assume that the average American turns on their television to find their National Security Advisor praising law enforcement in

a conversation about the murder of George Flloyd, a 46-year old, unarmed Black man by a Minneapolis police officer.

An authoritative claim portraying systemic racism in law enforcement as nonexistent is dangerous as Americans interact with information regarding this issue. Data from the Public Health Surveillance System examined circumstances of violent deaths by lethal force of law enforcement in 17 states from 2009-2012. Evidence shows victims were a majority white, but disproportionately Black with a 2.8 percent higher fatality rate. In 2015, researchers analyzed fatal police shooting data from the Washington Post for implicit bias. There were two indicators of threat perception failure they examined: "(1) whether the civilian was not attacking the officer(s) or other civilians just before being fatally shot and (2) whether the civilian was unarmed when fatally shot. The results indicated civilians from "other" minority groups were significantly more likely than Whites to have not been attacking the officer(s) or other civilians and that Black civilians were more than twice as likely as White civilians to have been unarmed." (DeGue, Fowler, and Calkins 2016). However, the researchers in this study acknowledge their access to information is limited. After Michael Brown was killed in Ferguson, Missouri in 2014, Federal Bureau of Investigation (FBI) director James Comey discovered that his agency does not collect reliable data pertaining to civilians killed by the police because it is voluntary to report police killings to the FBI. In 2014, only 224 agencies (approximately 1% of all police departments) reported a civilian killing by an officer to the FBI.

What O'Brien fails to acknowledge in his statement is the evidence that verifies the existence of racist policing. By announcing "systemic racism does not exist," he manifests a faulty understanding of the circumstances in the mind of the public. Accordingly, Mills writes on knowledge acquisition; "at all levels, interests may shape cognition, influencing what and how we see, what we and society choose to remember, whose testimony is solicited and whose is not, and which facts and frameworks are sought out and accepted." (Mills 2007, 24). O'Brien's statement lacked factual information regarding police activity to support his claim; his authority nonetheless granted him excessive credibility and shaped the national opinion on the issue. The idea that law enforcement as a system exists to do anything other than "serve and protect" its citizens runs counter to mainstream American narratives that honor the morality of police officers. A kind of epistemic inertia then emerges, fostering an individual's disinterest in confronting the

oppressive policies evident in police work. Privileged Americans have no need to investigate the evidence because it is easier, comfortable, and more familiar to accept the predominant narrative.

There are two ways in which white privileged knowers arrive at their illusions of society: dominant logic and structural ignorance. I borrow from Linda Martín Alcoff, who articulates how a person's situatedness affects their position to know; "some situations are in positions of ignorance, even though the knowers in those situations may have identical access to the relevant facts." (Alcoff 2007, 39). The perception of the world held by an individual is altered by their heuristic meaning-making processes informed by the cognitive norms of assessment privy to maintaining systems of oppression. Dominant logic thrives in this space. Social memory solidifies the illusions among white knowers that actively distort inaccurate accounts of the world. Mills argues social memory is then inscribed in textbooks, generated and regenerated in ceremonies and official holidays, concretized in statues, parks, and monuments. When I grew up, I was taught in public school that racism was an issue of the past. We learned the bare minimum in regard to oppression of people of color. Subjects included slavery, Abraham Lincoln, segregation, and Dr. Martin Luther King's "I Have a Dream" speech. The only required reading other than this speech for high school students was the Narrative of the Life of Frederick Douglass, written over one hundred and fifty years earlier. Our state-approved education overlooked the history of systemic oppression and thus perpetuated it. The standard American education led us to believe racism is finished due to the evolution of our collective moral consciousness with the moving words of Dr. Martin Luther King.

Misinformation paves the way to white ignorance, straightforwardly for a racist cognizer, but also indirectly for a nonracist cognizer who may form mistaken beliefs (e.g., that after the abolition of slavery in the United States, blacks generally had opportunities equal to whites) because of the social suppression of the pertinent knowledge (Mills 2007, 21). Mills identifies the two pathways as the non-racist cognizer and the racist cognizer. This is a point I expand upon in the next sections. There is plenty more to say about the production of white ignorance; for the sake of my argument, I move on to an examination of the other epistemic agent at hand: the dissenting privileged knower.

The Dissenting Privileged Knower

Moments arise when privileged epistemic narratives are pierced. There are white people, privileged knowers, who utilize sources of information offered by (or more likely, demanded from) individuals from marginalized communities when producing knowledge. However the information enters this knower's circle, be it university studies, social media, pop-culture— the privileged knower's dominant position is challenged by counterevidence and taken into consideration. Consideration does not equate to integration of such evidence, therefore the privileged knower is not absolved of their dominant positioning and the consequences that flow from it. I call this privileged knower the dissenting privileged knower, because this person allows counterevidence to oscillate them away from the ideology that reinforces domination. "Dissenting" is preferable to "dissented" because this position requires continual awareness of non-dominant informative sources and self-reflection.

Dissenting knowers take the time and effort to learn about realities outside their own within the comfort level their privileged positioning allows. For instance, this knower may take part in a meeting where HR representatives mediate a conversation about sexism in the workplace. In a constructed arena, he is entitled to participate in a discussion as to how it exists and what to do about it. This dissenting knower pays attention and listens to this new information, whereas a dominant privileged knower might roll his eyes and watch the clock for the duration, or opt to play devil's advocate to agitate the room.

On the subject of systemic racism, the dissenting knower has a myriad of evidence at their disposal. And it seems to be making a difference. During 2020, Americans engaged in what New York Times crowd data analysts call the biggest movement in U.S. history. It is important to note who exactly protested: "More than 40 percent of counties in the United States — at least 1,360 — have had a protest. Unlike past Black Lives Matter protests, nearly 95 percent of counties that had a protest recently are majority white, and nearly three-quarters of the counties are more than 75 percent white" (Buchanan, Bui, and Patel, 2020). Perhaps it is the video footage of police murdering Black people, like George Flloyd in Minneapolis or Breonna Taylor in Louisville, that has awakened a distrust in the status quo, something that is lost to the privileged knower, whose conceptual barriers lead her to believe it's just a few bad apples. Nonetheless, a movement of

dissenting knowers, integrating non-dominant information in their consciousness, has emerged.

Despite his intention, this knower still operates from a privileged standpoint, therefore seeking to operate with themselves at the center of their understanding. With this behavior, the dissenting knower's focus is off-balance. On standpoint epistemology, Brianna Toole argues "although newly developed conceptual resources may become intercommunally shared, dominantly situated knowers may be initially reluctant to adopt the resources developed by marginalized communities" (Toole 2019). Conceptual resources of the non-dominant knower develop so individuals may better understand their experiences in virtue of belonging to a marginalized group. These resources are not included in the meaning-making processes of dominant groups, and so even as the dissenting privileged knower attempts to form an understanding of non-dominant logic, they inevitably error because their conceptual resources are not suited to make sense of the experiences of marginalized people.

All this is to say the dissenting privileged knower is not exempt from perpetuating epistemic injustice. While the latest information may influence one's intent for the better, their privileged standpoint prevents them from complete awareness of how dominant logic affects their epistemic processes.

II. EPISTEMIC INJUSTICE IN TWO FORMS

The Dominant Privileged Knower and Testimonial Injustice

On June 6 2020, filmographer Mandy Rosen recorded footage of an altercation between a Black woman (P1) and a white man (P2) at a Black Lives Matter (BLM) protest. This took place in Huntington Beach, California.

> P1: "I have walked down the streets in my own community of Huntington Beach, where I'm from, and people have told me they were gonna shoot my n***** head off."
>
> P2: "No they didn't."
>
> P1: "Yes, they did! Yes, they did! The police were there!"

P2: "No, they didn't. Don't make s*** up. No one feels that way. Nobody hates you guys." (Rosen 2020)

Yelling surrounds them with the phrase "white privilege! white privilege!" The man asks the crowd of protestors what they are doing, protesting for Black Lives Matter, to which the woman responds on their behalf that they are hurting. He demands that she quit being the victim and the crowd shouts back at his rebuttal while preventing a physical altercation from unfolding. What I find most compelling is that the protestors are using the words "white privilege" at the time of this altercation to describe the wrongdoing on the man's behalf.

This man's privilege prevents him from being a decent participant in the testimonial exchange. Credibility excess inflated the sense of entitlement in his perception of the event, hindering his ability to engage responsibly with the speaker. Recall that during testimonial exchange, there is a title of credit inherent in the speaker's statement necessary for this type of exchange to be epistemically justified. Miranda Fricker's Acceptance Principle acknowledges this; "in the absence of cues for doubt, we surely accept most of what we are told without going in for any active critical assessment, and so our experience as hearers can seem to be that we are trustful unless and until some prompt for doubt is picked up on" (Fricker 2007, chap. 3). The Acceptance Principle entails a reasonable default credibility in one's interlocutor. Kristie Dotson extends this theory; "a transaction is a social act for which the self-conscious awareness by both agent and patient of their respective and complementary roles that comprise the act's structure is part of the very idea of the act itself" (Dotson 2012). If this man were to follow the typical terms of testimonial exchange in this transaction, one could assume he would have behaved differently. Rather, he deviated from the respective structure of testimonial exchange, his reason being identity prejudice held against his interlocutor.

As we know, testimonial injustice "occurs when a speaker is given less credibility than deserved (suffering a credibility deficit) because of an identity prejudice held by the hearer" (Fricker 2007, chap. 2). Often identity prejudice is "sufficient enough to cross the threshold for belief or acceptance so that the hearer's prejudice causes him to miss out on a piece of knowledge." The dominant privileged knower that this man represents interacts with testimony from a person within a marginalized community to the extent his privilege allows. The power of his identity prevents him from looking further than the bounds of

his own perspective, because dominant logic suggests that what he would find outside of matches his preconceptions. Tirelessly the non-dominant knower is denied credibility despite a collection of conceptual resources obtained in their lifetime; when the dominant member of society determines what is and is not epistemically verifiable, the conceptual resources made available, from which a dominant knower could learn, are lost.

Both identity prejudice (along with the credibility deficit it beckons) and systemic ignorance are what drive testimonial injustice in this event. Identity prejudice is the criterion inseparably tied to structural prejudice that materializes in testimonial exchange. Dotson argues "to understand this criterion for testimonial injustice, one need only imagine all of the ethically suspicious reasons one could have to hold the judgment that African Americans are liars or paranoid, especially when testifying to continued racial dis- crimination" (Dotson 2012). Using identity prejudice as his foundation in the exchange, the man in the video attacks the information he receives according to stereotypes of his interlocutor. He asks the crowd why they protest, to which the woman replies that they are hurting. He says no you are not, quit being the victim. He utilizes the emotion driving her testimony in his attempt to deny its truth value. The dominant privileged knower in this video sees himself as omniscient and the protesters as overly-emotional people "playing the victim." His identity prejudice reinforces his illusion despite the fact that he is the odd one out at a protest that blatantly identifies the contrary.

The dominant privileged knower does not seek to understand but finds grounds to reinforce what he believes to be true about the world. Knowledge production continues to be an insulated, circular process that leaves the agent stagnant and brutally unaware of reality that exists outside of his dangerous solipsistic perspective. His situatedness makes it so he does not have to do the epistemic work necessary to find truth, because finding truth provides him with no benefit in a world that seeks to justify the experience of the oppressor. Toole uncovers the nature of epistemic oppression; she argues:

> The situatedness of the dominant knower will not make salient those features of the world that the marginalized knower's conceptual resources attend to. As a result, the dominant knower can use this fact to preemptively dismiss the knowledge claims of a marginalized knower, as well as to dismiss the conceptual

> resources required to understand those knowledge claims. (Toole 2019)

Without access to the non-epistemic social facts that support what an epistemic agent is in a position to know, he is unable to claim any stake in the truth of the world outside of his own experience. His social perceptions of his interlocutor go unchecked, and so he is epistemically culpable. This injustice is unethical because it is an obstacle to truth and denies the speaker her capacity as a knower. Denying credibility to her testimony erases the danger of the racial prejudice she has experienced. If this man carries on with the preconception that he lives in a post-racial society, he will not notice racist behavior or policy when it is in his midst, therefore allowing it to continue. On top of this, denying truth to her experience gaslights her, which is a form of emotional abuse.

One could argue that in an absence of sufficient evidence, it is necessary to critically observe testimony as it is received. This is the inferentialist approach defined by Fricker, where the obligation of a hearer—if she is to gain knowledge through testimony— is to go in for a piece of reasoning that justifies accepting what others tell us" (Fricker 2007, chap.3). The hearer must make an inference based on criteria such as "the reliability of people like that about things like this" (2007). This varies from the non-inferentialist, who enjoys a default uncritical receptivity to what she is told, therefore behaving in accordance with the typical mechanisms of testimonial exchange. Of course, default uncritical acceptance of what one is told poses issues of justification. Placing critical capacity on the back burner, according to some, is epistemically irresponsible. In the case of my presented example, it would then be reasonable for the man questioning the young woman to be critical of her experience because it is his duty as an inferential hearer to seek reason in accepting what he is told. She has no evidence outside of her testimony to verify the truth of it, therefore the man has grounds to discount what she tells him. This is especially true because their exchange reared on the emotional side.

I respond to this idea with the notion that there were more reasons than not to act responsibly in this exchange. For one, the exchange took place at a Black Lives Matter protest where hundreds of other people gathered in solidarity with the topic of her testimony—systemic racism. Their conversation occurred on territory dedicated to fighting against oppression, so it would have been the epistemically responsible choice to entertain information that sheds light on the reality of existing as a marginalized person in a predominantly white city like Huntington

Beach, California. Second, the objectivity inherent in Fricker's inferential approach is problematic given the non-epistemic social factors that drive this interaction. Dominant logic seeks to absolve non-dominant logics. The racially-motivated circumstances change the circumstances of testimonial responsibility on behalf of the hearer. It is not useful to analyze using purely epistemic factors when evidence of racial prejudice often does not exist in the conceptual awareness of the agent attempting justification. In a sense, "this dynamic is produced by the fact that dominantly situated persons frequently take their own misunderstandings to be substantive objections," and so a lack of quantifiable proof in testimony is taken to be a shortcoming of the interlocutor (Berenstain 2016, 3.3).

The Dissenting Privileged Knower and Conditional Testimonial Injustice

The dissenting privileged knower is unique in his willingness to have non-dominant sources of information—epistemic and the non-epistemic conceptual resources belonging to individuals within marginalized communities— influence his knowledge production process. His perception of the world has been altered in some amount with the awareness of opposing standpoints from his own. In other words, the dissenting privileged knower has acknowledged that his privilege affects his efforts in understanding the world. Non-dominant sources of information originate from those who experience oppression. The dissenting privileged knower demonstrates a breakage in dominant logic when a variety of source information is explored, creating space for epistemic responsibility beneficial to the agent as well as those around him.

Taking a more objective view of the world than their counter-knower (the dominant privileged knower), the dissenting privileged knower does not subscribe to a particular perspective and aims to maintain rationality in their cognitive approach. The dissenting knower is more objective because their epistemic processes require acknowledging the opposing realities of others that fulfill their reformed sense of understanding. Traditionally, this is an acceptable epistemic move as rationality is a reliable pathway in finding truth. However, is there truth about the world that is universally accessible? Quill Kukla tackles this question as it assumes a tie between ontology and aperspectivity. Assuming the existence of universally accessible truth denies any epistemic advantages particular to social groups (in the scope of standpoint epistemology already established). My point is that objectivity in pursuits of truth is necessary, but there is a limit to what a

person is in a position to know—constructing the epistemic injustices a dissenting privileged knower is capable of inflicting. A dissenting privileged knower inflictings an epistemic injustice the moment one centers herself in discussion regarding her privilege. The information a dissenting privilege knower allows to alter their perspective of the world is filtered through what the comfort of white privilege permits. Consequently, the information that passes the barrier centers around their person; how they can do better, how privilege affects them, and how they can benefit with the humbling effect ontological objectivity allows. What remains impermeable are the discomforting feelings necessary to truly transform their perspective. This is a process of internalization that the structure of privilege enforces.

When the barrier causes harm to an epistemic agent bestowing information upon a privileged individual, I call it "conditional epistemic injustice." Conditional epistemic injustice is specific to the dissenting privileged knower when confronted with information regarding his role in oppressive behavior. It occurs when the hearer's objective approach to absorbing new information filters testimony through the privilege barrier and prioritizes their experience (and comfort) as a dominant knower instead of the speaker's. The injustice itself is rather insidious because the privileged knower may remain unaware of the barrier's existence. In testimonial exchange, the barrier reflects the individual's original status of dissent, and it creates conditions for testimonial exchange that will ultimately affect what information is allowed passage. The hearer yields to their privilege and stops listening when they feel attacked, typically when uncomfortable emotions arise. Their original status of dissent determines the limits to which they are willing to actively engage in testimony that challenges their privilege.

Conditional epistemic injustice is isolated in the terrain I have mentioned previously. That is to say there is a possibility that later on, the hearer may have a change of heart once they process the emotions that prevented them from responsibly engaging in testimony. However, regardless of the long-term effects for the dissenting privileged knower, the injustice has immediate consequences for the speaker who is willing to share information regarding their experience with oppression at that moment in time.

The terms of engagement set by the privilege barrier during testimonial exchange are honored so long as the speaker prioritizes the comfort of the hearer. The hearer may be fully prepared to listen to testimony that criticizes them, as their

reformed sense of understanding allows them to see that listening is a crucial step in changing behavior. The hearer may even be inclined to believe the testimony as all norms of testimonial exchange are followed up until the terms set by the hearer are broken, so the speaker feels relatively safe when beginning the exchange. Unfortunately, the speaker has no way of knowing beforehand how her testimony will be received by the hearer. But she will become aware when her testimony is discredited as the hearer retracts to the confines of their privilege.

Consider an example of a woman telling her boss that she was assaulted by a coworker in the workplace. She may approach him and tell him she was assaulted by a male coworker and is adamant that he needs to be taken off of the schedule. She felt safe doing so because this manager boasted about being a feminist and belivier of women in a staff meeting regarding this topic. So the manager responds as she expected, with concern, and asks her what happened. She says it happened after their shift; she was finishing up side work when he grabbed her thigh, and it was not the first time he touched her without consent. He reminds her that he does not tolerate harassment and will handle it, because she has a right to be safe at work.

The manager later follows up saying that she must have had a misunderstanding. She says no, there's no misunderstanding, and asks why he would think so. He responds by saying that this coworker claimed it was an accident and showed remorse. Furthermore, he admits he has worked with this server for years and knows he is a good guy. He tells her she is overreacting but can adjust her shifts if she is that concerned. She responds with frustration, saying "I thought you said you believed women." He tells her he does, to which she replies "if you believed me, you would not have to ask for his side of the story. Also, it says a lot that you trust a man's word more than mine." He responds aggressively with "hey, believe me, I'm a feminist. I grew up with sisters. I'm just doing the right thing and getting all of the information before making a move." He then tells her she's being too emotional for work and sends her home.

In this example, immediately upon hearing that there was an event of harassment, the manager was inclined to believe the testimony provided to him by the server who experienced it. He has undergone the proper training to understand the in-and-outs of workplace harassment and pledged to be an ally. However, when engaging with the other person involved, a man like himself, he switched gears and believed that it was just an accident. The terms

of the testimonial exchange—unbeknownst to the speaker—were that she had credibility in her testimony, pending further investigation from the man involved. The manager's pronounced identity as a feminist gave her a false sense of safety in coming forward with such a difficult topic, only for the epistemic processes of the manager (namely, verification and conditional listening) to gaslight her and cause emotional distress on top of the trauma of sexual assault.

This example demonstrates the role of the dissenting privileged knower observing testimony from a knower of a marginalized group (in this case, a sexist male superior and a female employee). He was open to her testimony initially but his privilege filter rendered her credibility conditional. It caused her emotional harm when she criticized his behavior; no longer was he supportive, and he sent her home rather than allowing her to finish out a shift. My goal with this example is to identify how privilege is still inherent in the dissenting knower, despite their status of dissent. The manager in this example knew the proper moves to take when handling sexual assault in the workplace, signifying his understanding of the issue at hand. However, when his own behavior was challenged, the conditions of his epistemic approach became clear and he ensured that his comfortability was at the center of the experience.

Conditional epistemic injustice is harmful because it gives the speaker a false sense of security during testimonial exchange. The dissenting privileged knower centers themself when partaking in testimonial exchange with a person belonging to a marginalized community. Conditional epistemic injustice is specific to an interaction between a dissenting privileged knower (who has taken steps to dismantle their oppressive behavior) and an epistemically oppressed knower in times when testimonial exchange covers issues of privilege. An exchange of this kind covers oppression— be it racial discrimination, police brutality, workplace harassment, and the like— and identify it for the sake of fixing it. This transformative discussion territory is imperative for dominant knowers who wish to join marginalized persons in solidarity with their own liberation efforts. It is imperative in the elimination of systemic oppression. The insidious nature of conditional epistemic injustice does the dirty work of centering privileged people in conversations meant to break them from their oppressive perspectival tendencies.

III. PRIVILEGED EPISTEMIC AGENTS AND THE PERPETUATION OF WHITE SUPREMACY

Thus far I have explained two versions of the privileged knower as they exist apart from one another. I homed in on privileged white knowers in my claims. In this section, I will bring these two knowers together in terms of a similar issue they enable, that being the perpetuation of white supremacy. I plan to utilize the aforementioned behaviors described as substance to support this claim.

The dissenting knower has entertained information from sources outside of their epistemic community. This person obtains awareness of their privilege and begins to perceive the world having a grasp of the role it takes in their perception. The point I have made is that despite their status of dissent, the dissenting privileged knower is still operating from a position of dominance and that comes with its own epistemic problems. Sarah Hoagland writes "dominant logic not only works to obscure interdependent relation, it is a practice of conceptual coercion; in significant ways it forecloses the possibility of a destabilizing critical response, recognizing only those responses that reinforce its own status" (Hoagland 2007, 102). The dissenting privileged knower has a unique status wherein they maintain epistemic patterns of domination while pursuing an engagement with other realities that challenge the privileged illusion they have come to know. The process of being an ally requires a breakdown of this illusion, and it takes tremendous work. Dominant logic affects privileged knowers when they receive new information. What they allow to pass through their privilege barrier is contingent upon how it affects them and how they perceive themselves.

This behavior is almost identical to that of the dominant privileged knower, whose status is maintained by insular practices of information-sourcing to actively discount conflicting views of the world. The difference between the two is their intent and the degree of severity in their behavior, but the impact is the same: the privileged knower sources information that centers themself and their privilege when making an effort to understand issues of oppression. They discount information crucial to gaining ontological objectivity regarding their privilege and uphold white supremacist ideology. Because of this, the status of dissent held by a privileged knower matters greatly. The on-going work of self-reflection is mandatory for the knower to be labeled as "dissenting." Epistemic humility is at the heart of this process.

Dissenting privileged knowers and their status of dissent resemble the phenomena of white redemption. The epistemic centering of oneself—in the case of white redemption— reinforces white supremacist systems with dangerous consequences for those who continually suffer from it. Sullivan writes "white redemption is a troubling goal for racial justice struggles; it has very little connection with ending white domination of people of color" (Sullivan 2014, 70). It centers the white person on issues of racial injustice. The epistemic agent then seeks to understand their privilege to stand in solidarity, but more importantly, showcase themselves as an ally. In the case of 2020 BLM protests, for instance, filmmaker AJ Lovelace spoke about protestors to the Los Angeles Times: 'It was obvious to me that people were out there to say they were out there," Lovelace said. "White girls would agitate the police and then cry when they responded. This isn't how a protest works" (Logan, 2020). In July of 2020, I witnessed at a Portland protest white protestors stopping at various points of a march to take photos of themselves and graffiti property in iconic areas. Not once did they join the chanting of the names of those who had been murdered. Behaviors like this distract from the purpose of civil disobedience. If they were not there to support the movement and protect Black protestors, they were there for the wrong reasons. Insincere solidarity, a symptom of self-centeredness, is damaging to the integrity and impact of the BLM as well the lives of those it's meant to protect.

White redemptive behavior among dissenting privileged knowers prevents necessary discussion of systemic racism with dominant privileged knowers. Dissenting privileged knowers include in their process of radicalization cutting ties with other white people in their communities. They may consider themselves "good white people" in the sense that they are not-racist, placing them on a moral high ground in comparison to "bad white people." Sullivan argues "distancing oneself from their white forebears can appear as the only viable way for contemporary white people to (not) deal with their racist history. It can seem to be the best way to demonstrate that they disapprove of slavery and that they are not racist like previous white generations" (Sullivan 2014, 60). The dissenting knower investigates their relationship with oppression as an individual rather than a piece of an expansive system; the individual therefore denies their proximity and interdependency to racist white people, generating a lack of accountability that perpetuates white supremacy.

Dominant privileged knowers have no reason to look beyond information that satisfies their illusions of the world; consequently, "if one insists on the public and counterpublic transcript as the only sense, one promotes both an epistemology and ethics of ignorance within the logic of oppression. For dominant logic must erase resistant logics, render them invisible, render them nonsense, to maintain its own legitimacy" (Hoagland 2007, 109). Dissenting privileged knowers have a proximity to dominant privileged knowers that creates space for sharing conflicting sources of information. They may not have the ability to radicalize their dominant privileged knower companions, but they have the responsibility to disrupt the epistemic processes that enforce their logic. It is necessary to uphold the obligation of dispersing counter information if the dissenting knower truly attempts to dismantle systems of oppression.

An objection to this responsibility is that it is in poor taste for a privileged person to invest themselves so deeply in liberation efforts and assume the white savior complex, a role that materializes when a white person attempts to help people of color in a way that is self-serving. The white savior complex is another instance of a privileged person centering themselves in their understanding of oppression and racial injustice. If a dissenting privileged knower takes it upon themselves to educate their peers on issues of oppression, they are actively silencing marginalized voices.

There is truth to this objection. It is inappropriate for a white person to speak over a person of color regarding their lived experiences. However, the responsibility to disperse information a privileged person receives from marginalized folks is contingent upon their proximity to both dominant privileged knowers and those who are epistemically oppressed. Their positioning provides them with the means to activate the information they have gained and inspire change without expecting this to be accomplished by marginalized persons. For too long, this sort of educational work has been demanded from marginalized persons, an occurrence Nora Berenstain calls epistemic exploitation:

> Epistemic exploitation occurs when privileged persons compel marginalized persons to produce an education or explanation about the nature of the oppression they face. Epistemic exploitation is a variety of epistemic oppression marked by unrecognized, uncompensated, emotionally taxing, coerced epistemic labor. It maintains structures of oppression by centering the needs and

> desires of dominant groups and exploiting the emotional and cognitive labor of members of marginalized groups who are required to do the unpaid and often unacknowledged work of providing information, resources, and evidence of oppression to privileged persons who demand it—and who benefit from those very oppressive systems about which they demand to be educated. (Berenstain 2016, 2)

The polarization of privileged knowers as an epistemic event inhibits social progress. The matriculation of information into one's conceptual practice has no use unless it is a continual practice that prompts self-reflection and changed behavior. In order to dismantle oppressive systems at work, a dissenting privileged knower must uphold a lifelong commitment to adjusting their perceptual capacity. Neglecting to do so places the burden on marginalized persons entirely, which not only allows the dominant logic supporting white supremacy to go unchecked, but also negates the cognitive labor of the agent. If the dissenting privileged knower fails to activate the information they have been given by marginalized persons and disperse it within their communities, there was no purpose in receiving it in the first place. The dissenting privileged knower has epistemically exploited a marginalized person's labor and made no effort towards eliminating the systems of oppression they sought to understand.

REFERENCES

Adedoyin, A. Christson, Sharon E. Moore, Michael A. Robinson, Dewey M. Clayton, Daniel A. Boamah, and Dana K. Harmon. 2019. "The Dehumanization of Black Males by Police: Teaching Social Justice—Black Life Really Does Matter!" *Journal of Teaching in Social Work* 39 (2): 111–31. https://doi.org/10.1080/08841233.2019.1586807.

Bandyopadhyay, Ranjan. 2017. "Volunteer tourism and 'The White Man's Burden': globalization of suffering, white savior complex, religion and modernity." *Journal of Sustainable Tourism*, 27 (3): 327-343, doi: 10.1080/09669582.2019.1578361.

Berenstain, Nora. 2016. "Epistemic Exploitation." *Ergo, an Open Access Journal of Philosophy* 3 (22). https://doi.org/10.3998/ergo.12405314.0003.022.

Buchanan, Larry, Quoctrung Bui, and Jugal K. Patel. July 3, 2020. "Black Lives Matter May Be the Largest Movement in U.S. History - The New York Times."https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html.

CNN Pressroom. October 27, 2020."CNN Has Most-Watched Prime Time Ever in Total Viewers." https://cnnpressroom.blogs.cnn.com/2020/10/27/cnn-has-most-watched-prime-time-ever-in-total-viewers/.

Collins, Patricia Hill, and Patricia Hill Collins. 2012. *On Intellectual Activism*. Philadelphia, PA: Temple University Press. ProQuest Ebook Central.

DeGue, Sarah, Katherine A. Fowler, and Cynthia Calkins. 2016. "Deaths Due to Use of Lethal Force by Law Enforcement: Findings From the National Violent Death Reporting System, 17 U.S. States, 2009–2012." *American Journal of Preventive Medicine*, National Violent Death Reporting System: Analyses and Commentary, 51 (5) Supplement 3: S173–87. https://doi.org/10.1016/j.amepre.2016.08.027.

Dotson, Kristie. 2012. "A Cautionary Tale: On Limiting Epistemic Oppression." *Frontiers* 33 (1): 24-47,106-107. http://stats.lib.pdx.edu/proxy.php?url=http://search.proquest.com.proxy.lib.pdx.edu/scholarly-journals/cautionary-tale-on-limiting-epistemic-oppression/docview/1000731292/se-2?accountid=13265.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press. Oxford Scholarship Online. doi: 10.1093/acprof:oso/9780198237907.001.0001.

Grasswick, Heidi. 2018. "Feminist Social Epistemology." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University.

Kukla, Rebecca. 2006. "Objectivity and Perspective in Empirical Knowledge." *Episteme: A Journal of Social Epistemology* 3 (1): 80–95. doi:10.1353/epi.0.0005.

"Kyle Heckler," Mandy Rosen Photography, June 6, 2020. Instagram direct message to Author, September 7, 2020. https://www.dropbox.com/s/z4aj700xu5r2d0s/Kyle%20heckler.mp4?dl=0.

Logan, Erin B. September 8, 2020. "How White People Gentrified Black Lives Matter." *Los Angeles Times.*" https://www.latimes.com/opinion/story/2020-09-04/black-lives-matter-white-people-portland-protests-nfl.

Pohlhaus, Gaile. 2017. "Varieties of Epistemic Injustice 1." In *The Routledge Handbook of Epistemic Injustice*. London: Routledge (1) . https://doi.org/10.4324/9781315212043.

State of the Union. May 21, 2020. "'No, I Don't Think There's Systemic Racism,' White House National Security Adviser Robert O'Brien Says about US Law Enforcement Agencies. 'There's a Few Bad Apples That Are Giving Law Enforcement a Terrible Name' #CNNSOTU Https://T.Co/Tdwo9XfTQ7." Tweet. *@CNNSotu* (blog). https://twitter.com/CNNSotu/status/1267091151556947968.

Sullivan, Shannon. 2014. *Good White People: The Problem with Middle-Class White Anti-Racism*. Albany, United States: State University of New York Press. ProQuest Ebook Central.

Sullivan, Shannon, and Nancy Tuana. 2007. *Race and Epistemologies of Ignorance*. Ithaca, United States: State University of New York Press. ProQuest Ebook Central.

Toole, Briana. 2019. "From Standpoint Epistemology to Epistemic Oppression." *Hypatia* 34 (4): 598–618. https://doi.org/10.1111/hypa.12496.

Wilson, B. L. June 8, 2020. "I'm Your Black Friend, but I Won't Educate You about Racism. That's on You." *Washington Post*. https://www.washingtonpost.com/outlook/2020/06/08/black-friends-educate-racism/.

Wilson, Marguerite Anne Fillion. 2019. "Teaching Whiteness in the Neoliberal University: Positionality, Privilege, Resistance, and Transformation." In *Teaching with Tension: Race, Resistance, and Reality in the Classroom*, edited by Bolton Philathia, Smith Cassander L., and Bebout Lee, 217-36. Evanston, Illinois: Northwestern University Press, 2019. doi:10.2307/j.ctv8d5sf6.16.

compos mentis

Degrees of Naturalism in Epistemology and Philosophy of Science

Matthew Kern
Kent State University

ABSTRACT

Naturalism in philosophy of science explicitly came on the scene in 1969 with the publication of W.V. Quine's "Epistemology Naturalized." Quine's view can be seen as a backlash to prior foundationalist and positivist views that understood epistemology and philosophy of science as independent of and prior to science itself. Here Quine argued that it is not fruitful to try to justify scientific practice by means of an a priori philosophical theory or method. Instead, we ought to take the results of the sciences for granted, and not worry about trying to answer Hume on skeptical worries about induction, according to Quine. This results in a "naturalistic" project in which the normative discipline of epistemology is replaced by the descriptive discipline of human psychology. First, I will examine Quine's paper in depth. Second, I will examine the responses of Jaegwon Kim and Alan Berger, who both advance, in their own ways, the objection from normativity. Third, I will examine a naturalistic response from Ronald N. Giere. I will ultimately argue for an intermediate version of naturalism, in which the data of the sciences are taken as relevant to philosophy without standing in need of justification by philosophy, and yet philosophy cannot be replaced by science as Quine would wish.

I. INTRODUCTION

Naturalism in philosophy of science explicitly came on the scene in 1969 with the publication of W.V. Quine's "Epistemology Naturalized." Quine's view can be seen as a backlash to prior foundationalist and positivist views that understood epistemology and philosophy of science as independent of and prior to science itself. Foundationalism (in this context) is the view that scientific knowledge ought to be deduced from indubitable first principles and that the sciences need to be justified from an independent perspective by means of a purely rational methodology (i.e., something like an inductive logic). The positivists took a foundationalist approach to philosophy of science in several ways: (1) their reduction of the meaning of scientific statements/theoretical language to observation statements/observational language, (2) their use of the analytic-synthetic distinction to create a sharp divide between the empirical subject matter of the sciences and the conceptual subject matter of philosophy, and (3) their insistence on the need for formulating an inductive logic which would secure the epistemic privilege of science in relation to all other forms of human inquiry.

By the second half of the twentieth century, many philosophers of science had grown deeply skeptical of this program. Reductions of meaning to observation were notoriously fraught with problems, since one would have to state every possible piece of potential experience that could confirm or disconfirm a statement for the reduction to be legitimate. This proved to be a mightily difficult task, though theorists such as Rudolf Carnap (in his work *Aufbau*) continued to attempt to reduce the meaning of a statement to a "sense-datum language." However, the biggest opponent of the positivists at this time was Quine, who attacked both elements (1) and (2) of the positivist program in his influential "Two Dogmas of Empiricism." Quine attacked (1) on the grounds that there is no such thing as the meaning of a statement, since the same statement can mean different things (i.e., have different confirmation and disconfirmation conditions in light of possible experience) depending on how those statements are holistically related in each individual's web of belief. Quine attacked (2) on the grounds that the term 'analyticity' did not have a clear meaning, since it was closely tied to other confusing notions such as 'synonymy,' etc.

It is the third part of the positivist program – the search for an inductive logic – that Quine more explicitly attacked in his later 1969 article "Epistemology

Naturalized." Here Quine argued that it is not fruitful to try to justify scientific practice by means of an a priori philosophical theory or method. Instead, we ought to take the results of the sciences for granted, and not worry about trying to answer Hume on skeptical worries about induction. This results in a "naturalistic" project in which the normative discipline of epistemology is replaced by the descriptive discipline of human psychology. Science is not in need of philosophical justification, but provides tools for understanding how human knowledge acquisition works. Therefore, there is no need for anything like epistemology as traditionally conceived (as an a priori search for normative principles that can provide a firm foundation for knowledge and tell us what statements are justified, etc.) There were many responses to Quine. Some were sympathetic to traditional epistemology, while others thought that Quine was on the right track but went too far in proposing the replacement of philosophy by science.

First, I will examine Quine's paper in depth. Second, I will examine the responses of Jaegwon Kim and Alan Berger, who both advance, in their own ways, the objection from normativity. Third, I will examine a naturalistic response from Ronald N. Giere. I will ultimately argue for an intermediate version of naturalism, in which the data of the sciences are taken as relevant to philosophy without standing in need of justification by philosophy, and yet philosophy cannot be replaced by science as Quine would wish.

II. QUINE'S "EPISTEMOLOGY NATURALIZED"

I will begin with Quine's original 1969 paper, "Epistemology Naturalized." Quine begins by tracing the origin of the foundationalist views that he is fighting against. Quine identifies the distinction between a theory of concepts/meanings and a theory of truths in the attempted reduction of mathematics to logic (Quine 1969, 70-71). He further traces back the history of this distinction to Hume, who identified objects with sense impressions (a form of reductionism about meaning), and who denied that we can justifiably make statements about the future based on past experience (the denial of the rationality of induction) (Quine 1969, 72). Later, meaning was identified primarily with sentences and set theory was employed for the sake of categorizing sense impressions (Quine 1969, 73). Quine cites Carnap's *Aufbau* as an example of this kind of work (Quine 1969, 74). Meanwhile, progress in answering Hume on the side of a "theory of truths" was still unsuccessful;

this was considered a scandal for philosophy in that it eliminated any hope for certainty (Quine 1969, 75). Carnap had sought to find a "rational reconstruction" of how we make inductive inferences, but Quine challenges this task and suggests that we take a simpler approach: "The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology?" (Quine 1969, 75).

Quine immediately responds to the objection that such a strategy would be flagrantly circular. We cannot deduce science from observations, so why should we keep trying? In order to understand how we come to know the world, we ought to use anything at our disposal, even the results of the sciences that pertain to human perception (Quine 1969, 76). The project of translating science into observation terms and logic has failed and will always continue to fail, so why should we keep trying to do that (Quine 1969, 76)? At the very best, these translational efforts are stilted and fabricated, argues Quine. We cannot formulate the "difference that the truth of a given statement might make to experience" in a single sentence; rather, whole theories have experiential import (Quine 1969, 79). Here Quine links his confirmation holism with the hopelessness of performing the translation that the positivists were after, and more importantly, with the impossibility of offering an inductive method that could justify scientific practice from a perspective outside of science (Quine 1969, 79-80). In other words, the indeterminacy of translation removes the possibility of translational schema that could justify scientific inferences by means of a purely sensorial or experiential language.

At this point, Quine provocatively suggests that the positivists had in a sense not gone far enough: not only is metaphysics dead— epistemology (as traditionally conceived) is also dead (Quine 1969, 82). With the failures of the positivists, we are only left with epistemology as a chapter of the natural sciences (Quine 1969, 82). The project now is to understand how a human subject, after being exposed to certain patterns of sensory information, etc., comes to have a description of the world (Quine 1969, 83). Quine makes some ambiguous statements at this point. He says that "something like the old rational reconstruction" can still go on, but "we can now make free use of empirical psychology" (Quine 1969, 83). Epistemology and natural science are "in" each other as component parts (Quine 1969, 83-84). For Quine, awareness ceases to be a necessary condition for

justification; instead, we will talk about an organism that is being stimulated by an environment (rather than internal states). Instead of observation statements, we will have statements for which "all the speakers of the language give the same verdict when given the same concurrent stimulation" (Quine 1969, 87). The end of traditional epistemology should not lead to the "epistemological nihilism" of Kuhn, Polanyi, and others; instead, we can finally make progress by means of psychology, linguistics, and evolutionary biology (Quine 1969, 87-90). Quine presents a provocative and strong critique of traditional epistemology, but what does his project really amount to, and can it possibly serve as a replacement for traditional epistemology?

My central contention at this point is that Quine's pitch for naturalized epistemology is rather vague, though the general thrust of his argument is strong. First, Quine is quite right that there is no straightforward way to answer Hume on his own terms. There really is no way to non-circularly justify the very practice of making inductive inferences. Furthermore, it really is impossible to reduce scientific theories to observational terms. Many scientific theories appeal to unobservable entities such as genes, quarks, electrons, etc., and we cannot cash out the meaning of statements with those terms by speaking of the readings of an electrometer, or any similar strategy. Our scientific theories make statements about the structure of reality by quantifying over these entities. In fact, one might think that science is essentially defined by the positing of unobservable entities to explain observable phenomena. Quine is also quite right to cast doubt on the possibility of an inductive logic or a rational reconstruction of the inferential steps that an ideal scientist would make to justify a theory (a theory of scientific justification). Furthermore, there is ultimately no way to justify science from outside of science. The claim that science does not need a justification is more questionable, but it may be the only way for a scientific philosopher to allow scientific theory into philosophy. But naturalized epistemology is not circular, since it does not even try to justify science.

Nevertheless, it seems that Quine makes too much of a leap in suggesting that we replace epistemology with psychology/linguistics. It does not follow from the fact that we cannot provide an a priori metaphysical/epistemological foundation for science that we must eschew the very notion of justification. To be fair to Quine, it is not clear whether he does want to remove the concept of justification, since he does say that "something like traditional epistemology" will

still go on. But the large thrust of his argument is against the idea of any normative epistemology, and I fail to see how 'justification' could not be a normative notion. Even if knowledge could be naturalized (which seems more plausible, though this would not be knowledge in the sense of "justified true belief"), how could justification be? Finally, it is not clear how naturalized epistemology could allow us to answer questions about the foundation of morality, what a just society would look like, and other normative questions. But it seems that a comprehensive epistemology should tell us how we can know the answers to these questions.

Now one might be an anti-realist about all these notions, but that would require separate argumentation from the arguments for naturalized epistemology. And even if we could talk about how the world impacts our receptors and influences our perception (we can, since we have psychology), it is unclear what that would tell us about what beliefs we should have beyond trivial beliefs that are based on immediate sensory experience. How can naturalized epistemology tell us what to believe about who will win the next presidential election? The defender of Quine will certainly have an answer related to statistics and history, but it seems that this would go beyond the strictures that Quine places on naturalized epistemology here, since those parts of the web of belief are very far removed from sensory irritations. So while Quine's position accounts for the difficulties of prior foundationalist approaches, it seems to leave open other, essentially normative questions (essentially prescriptive questions, to be exact) about knowledge as such unanswered, and it is not clear that Quine has given us a good reason to reject these questions as ill-formed. I will now turn to two philosophers, Kim and Berger, who express some of my criticism of Quine in more detail.

III. THE NORMATIVE OBJECTION TO NATURALIZED EPISTEMOLOGY

My main concern with naturalized epistemology is its seeming elimination of normativity from the concept of knowledge. Jaegwon Kim expresses this worry in his paper "What is 'Naturalized Epistemology?'" Kim begins by recounting the Cartesian epistemic project. Descartes' goal was to find, from indubitable foundations, propositions that were worthy of belief (Kim 1988, 229). This project involves two components: (1) formulating the standards by which propositions are evaluated for belief or rejection and (2) determining what beliefs one ought

to hold by applying the criteria of (1) (Kim 1988, 229). Modern epistemology has inherited these two goals and reformulated them to an extent. The goals are now (1) to find the conditions for justifiably taking a belief to be true (these conditions must be stated without the use of epistemic terms for them to be useful) and (2) to discover what beliefs we are justified in holding (Kim 1988, 230). Kim especially stresses the parenthetical clause of project (1): "the criteria of justified belief must be formulated on the basis of descriptive or naturalistic terms alone" (Kim 1988, 230). Justification thus has a crucial role to the epistemic project as traditionally conceived, and it provides the normative dimension to the concept of knowledge (Kim 1988, 230). The foundationalist strategy is to begin by finding a set of "directly" justified beliefs and then justify all other true beliefs by relating them to these initial "basic" beliefs (Kim 1988, 231).

At this point, Kim presents Quine's arguments for naturalized epistemology. He says that Quine's claims about reductionism (of meaning) and inductive logic are justified, but claims that no one would quibble with Quine on these matters (Kim 1988, 232-3). The controversy comes about when looking at Quine's conclusions. For Quine is not only arguing against the usefulness of the project of justifying science, but also wants us to replace traditional epistemology with a psychology of cognitive processes (Kim 1988, 233). Kim pinpoints the removal of justification and thus normativity from epistemology as the heart of Quine's project (Kim 1988, 233-4).

I think Kim is fair in representing Quine's position here. Kim's objection is that a nomological theory of belief fixation would only provide a causal description of the relation between sensory input and cognitive output, and this nomological relation could not fulfill the same role in our epistemic efforts as an evidential relation would in a traditional epistemology (Kim 1988, 234). Nomological patterns depend on the psychology of the organism, while evidential factors abstract away from psychology (Kim 1988, 234). The point is not that psychology is irrelevant to epistemology; rather, it is that naturalized epistemology cannot be epistemology at all since it removes the normativity of key terms such as 'evidence' or 'justification' (Kim 1988, 235). In other words, naturalized epistemology cannot be a replacement for traditional epistemology because it is not even trying to do the same thing. A nomological description of belief fixation is just not the same thing as a normative prescription that tells us what beliefs we should hold. Kim concludes that epistemic values must be "consistent with" the facts, but there

must be value for the project to even be considered an epistemic one (Kim 1988, 235-6).

Kim lays out my worries about normativity quite well and advances the objection to Quine in a very straightforward and blunt manner while accurately representing Quine's project. Alan Berger nuances this objection in a way that may be further helpful.

In his paper "The Quinean Quandary and the Indispensability of Nonnaturalized Epistemology," Berger covers much of the same territory and argues for basically the same position as Kim. However, he contributes to the dialectic further with a few more arguments. Once again, Berger recounts the traditional Cartesian project of justification, in which one searches for justified basic beliefs and derives chains of justification from those beliefs to all other justified beliefs (Berger 2003, 367-8). Berger insightfully notes that the first part of the Cartesian project – the formulation of criteria of justification – drops out of Quine's project, since Quine adopts, from the beginning, the criteria of science for this task (Berger 2003, 368). All of this is familiar ground from Kim.

Berger contributes something new when he attributes a view that he calls "epistemic if-thenism" to Quine. This is a strategy for reducing normative language to descriptive language. It offers us conditionals with antecedents that specify the epistemic goals that one is seeking and with consequents that tell one how one can achieve those goals (Berger 2003, 372). Berger says that Quine appears to be committed to the conditional "If you seek the best way to discover truths and make predictions, then use science" (Berger 2003, 372). But for Quine, the only way we have of identifying truths is by taking them to be the beliefs that we are rationally justified in holding now, and for Quine, those are of course the beliefs justified by science (Berger 2003, 372-3). So the conditional becomes "if you seek to make statements that are scientifically acceptable, then use science," but this is rather unhelpful (Berger 2003, 373). Furthermore, Berger objects that Quine's definition of an observation sentence as "one which all speakers of the language give the same verdict when given [during] the same concurrent stimulation" leaves Quine open to cultural relativism, since this would result in different communities having different observation sentences (Berger 2003, 373-4). The only way that Quine can avoid relativism is to take on board "epistemic authoritarianism," the view that the truth should not be judged independently of the standards of our community (Berger 2003, 374).

Berger's next objection is that there is a circularity within Quine's argument for holism. Quine's thesis for epistemic acceptability is "A theory is epistemically acceptable iff it is compatible with the data" (Berger 2003, 375). As a matter of implication (a matter of logical relations), we cannot deduce something that is logically contradictory with premises that are based on the data, but Quine seems to mean that we should not make such an inference (a matter of psychological reasoning) (Berger 2003, 377). But then Quine is saying that we ought to follow this epistemic principle, and he is doing traditional normative epistemology (Berger 2003, 377). In addition, Quine's arguments for confirmation holism are not based on scientific evidence, but hold as a matter of philosophical principle (Berger 2003, 378-9). Thus, Quine does not seem to take a consistent naturalist stance. Berger goes on to argue that Quine's argument for holism is self-refuting, since it relies on deduction and the notions of validity/invalidity, which Quine sees as contingent truths (Berger 2003, 380-1). But these principles are indispensable to science, and yet cannot be part of science, so naturalized epistemology fails (Berger 2003, 381).

Kim and Berger advance the normative objection against Quine on much the same terms. It just seems that a causal description of human sensory faculties cannot replace a prescriptive account of how we should evaluate propositions. In addition, Berger points out that Quine seems to be inconsistent, since his argument for holism is not based on scientific principles. It is at least difficult to see how it could be. Perhaps one could argue that Quine observed patterns of belief fixation (which need to be understood in behavioristic terms for Quine) in many individuals and came to formulate holism as an empirical hypothesis in this way. However, it seems that confirmation holism is a normative matter for Quine, since any adjustment "can" be made to one's web of belief in light of contradicting evidence. There is of course an ambiguity in this "can," but there is at least some reason to think that Quine is inconsistent here.

However, Berger's last point about Quine's argument for holism being self-defeating is misguided. All that Quine needs to say is that he accepts deductive and implicative notions such as validity/invalidity as central parts of his web of belief. Because they are central to his web of belief, he can use them in arguing for holism, even though he will have to recognize that he "could" have to drop them in light of future experience. But in the end, I take Kim's formulation of the problem with Quine's naturalism to be the main sticking point. Quine is right

about many things, as I noted above, but he seems to go too far in his talk of replacement. But to counter these concerns, I will now examine the work of another naturalist, someone who is sympathetic to Quine. This is the work of Ronald N. Giere.

IV. A NATURALISTIC RESPONSE TO THE OBJECTION FROM NORMATIVITY & WHERE I PLACE MYSELF

In his paper "Philosophy of Science Naturalized," Ronald N. Giere defends naturalism in philosophy of science against two foundationalist approaches, "methodological foundationalism" and "metamethodology" (331). Giere's naturalism incorporates Kuhn's proposal for a "role for history" in philosophy of science and incorporates a decision theory model to explain theory choice in the sciences (Giere 1985, 331). Giere begins by defending the claim that Kuhn was a naturalist about philosophy of science. Kuhn's denial of a distinction between a context of discovery and a context of justification as well as his talk of "gestalt switches" (as opposed to "rational evaluation of theory") and "persuasion" (instead of "rationality") seem to mix philosophy of science with psychology (Giere 1985, 332). However, even many theorists who were sympathetic to Kuhn in terms of providing a role for history did not accept his naturalism; philosophers such as Lakatos and Laudan still sought to show how scientific inquiry was rational in the course of history (Giere 1985, 332). Giere's goal is to show that Kuhn was quite right in adopting naturalism and refusing to construct a "rational methodology" of this sort (Giere 1985, 332).

Giere recounts three objections against naturalism: (1) the circle argument– it is circular to use science to investigate scientific methods, (2) the argument from norms– naturalism can only describe methods used by scientists, but the point of philosophy of science should be to prescribe what methods scientists should employ, and (3) the argument from relativism– naturalism would be unable to say that evolutionary theory is better than "creation science" (Giere 1985, 333-4). The circle argument is usually used to justify some form of foundationalism. It is in the background of the positivists' claim that logic was at the foundation of the scientific method (Giere 1985, 335). Carnap kept trying to make this project work, but logic did not help us evaluate actual scientific theories, and there was no technical way to cash out the notion of an "initial probability of all hypotheses" (Giere 1985, 335).

Giere notes, however, that this project was also tied up with a circular problem—this inductive logic was supposed to be an explication of our pre-reflective notion of evidential support, and it was supposed to provide evidence in turn for that same notion (Giere 1985, 336). The problem is that "The logic is at best descriptive of our intuitions. It does not insure us that our intuitions themselves are correct" (Giere 1985, 336). Giere also charges the "metamethodology" of Lakatos and Laudan as falling prey to a similar circular problem— this program also only ends up reflecting our intuitions about rationality (Giere 1985, 337).

Giere here finds a crucial insight about the normativity involved in assessing programs in philosophy of science/epistemology. The reality is that all programs involve normative assumptions from the start, and naturalism is not different from any other program on this score. But at least the naturalist is upfront about these assumptions. So, I do not take the circle argument to be a serious threat to naturalism. However, Giere also seems to think that he has answered the argument from norms here. He notes that the foundationalist or metamethodologist can only provide a description of our pre-existing intuitions, and not actually prescribe those intuitions as representative of rationality as such (Giere 1985, 338). However, this response does not seem satisfactory to me. He has not shown that we should give up our search for criteria of justification and simply replace that quest with a search for the causal principles underlying human perception. He has only shown that the normative output of an epistemological theory/theory in philosophy of science will "reflect" the normative input baked into that theory from the beginning. He is correct to point out that this is the case with non-naturalist approaches.

However, as Berger noted, there are normative assumptions baked into Quine's project as well, and it is unclear how this normativity would not be reflected on the output side of analysis too. Quine may very well deny this (and he must for epistemology to be "naturalized" in his sense), but it seems that he implicitly wants us to move from an understanding of psychology to an acceptance of certain human cognitive and behavioral procedures as the best way to know the world, at some point in the move from input to output. If he does not want us to do that, then he is not offering an epistemology as all in the first place. So, we are back to the worries of Kim and Berger again at this point. The Quinean will likely argue in reply that I am assuming a normative conception of epistemology and trying to impose it on Quine, and that Quinean may be right. I do not know if there is a way to avoid question begging at this point, but it still does not seem

like Quine or Giere have done enough to convince a traditional epistemologist to give up his/her discipline.

Giere presents his own version of naturalism at this point, which he calls "an evolutionary perspective." Much of it seems to mirror Quine's own project, though there is more of an emphasis on the role of evolutionary history. The goal is to "explain how creatures with our natural endowments manage to learn so much about the detailed structure of the world...this problem calls for a scientific explanation" (Giere 1985, 339-340). Again, I believe that this is a worthy enterprise, and it is certainly important and relevant for how we should understand knowledge and knowledge acquisition, but I do not see how it can fulfill the role of traditional epistemology in the intellectual landscape. Also, to continue a point that I had made earlier, I do not see how this project would tell us how we form beliefs about phenomena that are far removed from basic-level sensation. The relationship between sensation, perception and beliefs is also one fraught with difficulties. I am not even sure if 'belief' is a scientific notion. It seems that we can have a nomological account of sensation/perception, a psychological account of belief fixation, and an epistemological account of what beliefs we should have at the same time. These projects are of course interrelated, but distinct in goals. However, Giere is quite right to criticize both traditional rationalists and traditional empiricists for being unable to get beyond subjective experience and find the actual mechanisms that govern our conceptual and linguistic abilities (Giere 1985, 340). And he is right to ignore the objection that invoking evolutionary theory to explain how we know about the world would be circular (Giere 1985, 340). Nor does a physicist need a philosopher to tell him/her that what he/she is doing is justified. In this sense, I support some version of naturalism.

Finally, Giere argues that an evolutionary perspective can deal with the problem of norms and relativism. According to this perspective, norms are a product of a stage in human evolution in which society became so complex that enforced patterns of social behavior were needed to avoid social chaos (Giere 1985, 341). This leads Giere into questions about whether his evolutionary perspective can avoid cultural relativism about epistemic norms. He does not think that he needs to be committed to metaphysical realism, the view that "there is exactly one true and complete description of 'the way the world is'" (Giere 1985, 342). I too find a notion such as this unhelpful for the purposes of metaphysics and epistemology,

and Giere is right that other animals may have equally "good" representations of the world that will differ from our own.

Giere then rebuts Putnam's contention that a naturalist would provide a vacuous definition of rationality. The lesson of naturalistic epistemology is that there is no sharp boundary between animals and humans, and thus between the irrational and the rational (Giere 1985, 342). I again agree with Giere here. The failures of providing an inductive logic or a rational reconstruction have shown that we cannot maintain such a sharp boundary. However, I think that a "replacement" version of naturalism would have to drop any notion of rationality at all, since this is a normative notion. But I think that the notion of rationality is still useful—we need only drop an essentialism about rationality, as Giere claims (Giere 1985, 343). Giere then goes into an extended discussion about the human need to find some essential difference between animals and humans (Giere 1985, 343). I think that this is mostly a red herring, though I recognize that this discussion is in the background when discussing these issues. I do not see how it is directly relevant to the question of whether some form of naturalized epistemology is the most fruitful philosophical program.

Next, Giere provides a naturalistic account of theory choice based on his evolutionary program. From an examination of the way actual scientists are trained and the pedagogical strategy of scientific textbooks, Giere argues for "constructive realism" (Giere 1985, 344-6). This is a realism which "understands hypotheses as asserting a genuine similarity of structure between models and real systems without imposing any distinction between 'theoretical' and 'observational' aspects of reality" (Giere 1985, 346). This realism is not metaphysical in that it is not committed to there being "a single complete description of reality" (Giere 1985, 346).

Then Giere advocates for a "bottom-up" approach to theory choice, rather than the traditional "top-down" approaches (Giere 1985, 347). The point is that a choice between theories is of the same kind as any other choice that an individual makes during his/her life (Giere 1985, 347). Giere advances a decision-theoretic model of theory choice, which is centered around a decision problem defined by a set of possible options and a set of possible states of the world (Giere 1985, 347). The agent's desires are ranked in relation to these states of the world (Giere 1985, 348). Agents adopt a strategy that will fulfill their minimum satisfaction level— there is no need to pick *the* rational choice (Giere 1985, 348).

Giere explains how this model would work by presenting the example of a major revolution in geology related to plate tectonics (Giere 1985, 348-353). One important thing to note is that he allows professional interests to play a role in the process of individual theory choice; this stresses the fact that he wants to talk about human agents with normal desires (Giere 1985, 352). Giere ends his paper by advocating a middle way between a universal naturalistic philosophy of science and one which can only apply to local historical periods (Giere 1985, 353-55). While all these later parts of Giere's article are secondary to his main points, they do provide an interesting look at what an "evolutionary epistemology" could look like. I do not have any quibbles with this part of the paper, though I do not want to necessarily be committed to these notions. I will now conclude by summarizing the core dialectic as I see it, restating my position, and considering an old objection one more time.

V. CONCLUSION: A MODERATE NATURALISM

I will now summarize the dialectic and explicitly situate myself within it. Quine is quite justified in looking for an alternative way to do epistemology, because the foundationalist and positivist strategies that he had inherited had failed. The reduction of scientific meaning to observational language simply cannot be done; scientific statements about unobservable entities do have meaning and are at the heart of many powerful scientific theories today. The distinction between conceptual truths/empirical truths (or analytic truths/synthetic truths) rests on vacuous circularities, does not explain the linguistic phenomena that it is supposed to account for, and is not even needed to do the epistemic work that it was introduced to address. Any attempt to create a purely formalized inductive logic or method of rational reconstruction to secure science as a rational enterprise is doomed to fail. We cannot answer Hume on his own terms. All of this is quite true. And it seems that a good starting point for a scientific philosopher such as Quine is science itself. Those scientific results cannot be justified by an a priori philosophical argument that secures the rationality of the scientific enterprise. It does seem that we should use the results of the sciences in our philosophical endeavors, given the powerful theoretical and technological fruits that they have given us. But I take it that this is not too controversial and results in only a minimal naturalism. The question is whether Quine is right that we should remove

normativity from epistemology and settle for a causal-nomological account of human perception in place of traditional epistemology. It is this claim, if it is to be attributed to Quine (and it is not entirely clear, since Quine leaves some ambiguity in his view), that I just cannot make sense of. I do not know what a non-normative epistemology would even be. Perhaps one could charge me with holding some notion of analyticity in making this claim. However, I am not saying that Quine is logically incoherent in talking about naturalized epistemology. I can only say that his conceptual scheme is incommensurable with my own on this point. And that is not a knockdown argument against Quine; it is just a personal confession.

On a more purely philosophical note, I do think that the objection from normativity is something that any promoter of naturalized epistemology must face. This objection has much more weight behind it than the "circularity" objection or the "relativism" objection. For it seems that in making scientific determinations about how the human perceptual apparatus works, we are not giving the inquirer any guidelines about what beliefs to accept and what to reject. We are only describing the processes that, in an extended and somewhat unclear way, "result" in belief fixation. And it is at the level of beliefs that normativity comes onto the scene, since a set of beliefs can be coherent, incoherent, rational, irrational, etc. When we are just talking about how axons travel down certain nerve endings and provide information to the brain, we are operating in a different space, a space of pure description. This space is only indirectly related to the space of reasons where we can understand beliefs as rational or irrational, etc. This space does not disappear with the death of foundationalism. We still need to use terms such as 'rational' to indicate our commitments in the space of reasons, and just because we are no longer attempting to find a theory of rationality that could justify science, we are not free to abandon a theory of rationality as such.

Giere's use of decision theory even hints at the naturalistic possibilities for a theory of rationality. This theory would talk about the beliefs and desires of a human being as a social organism, but it would still be a theory of rationality, and 'rationality' is normative notion. It is still the domain of philosophy to work with these normative notions, and there is never going to be a science (in the sense of controlled experimentation) that tells you about these notions. Yet they are ineliminable parts of our conceptual scheme, and we continue to use them as we navigate the space of reasons. Kim and Berger, despite advancing many concerns that do not put a dent in Quine's project, both express this in their own ways.

Giere shows the promising directions that naturalism could go in, but does not quell this fundamental concern with regard to the more extreme "replacement naturalism"/naturalized epistemology of Quine. The possibilities for a naturalist epistemology are limitless, but a naturalized epistemology in Quine's sense is not an epistemology at all.

REFERENCES

Berger, Alan. 2003. "The Quinean Quandary and the Indispensability of Nonnaturalized Epistemology." The Philosophical Forum 34 (3 & 4): 367-381.

Giere, Ronald N. 1985. "Philosophy of Science Naturalized." Philosophy of Science 52 (3): 331-356.

Kim, Jaegwon. 1988. "What is Naturalized Epistemology?" In Philosophical Perspectives. Vol 2: Epistemology, edited by James Tomberlin, 229-238. Atascadero: Ridgeview Publishing.

Quine, W. V.. 1969. "Epistemology Naturalized." In Ontological Relativity and Other Essays, edited by W.V. Quine, 79-90. New York: Columbia University Press.

compos mentis

Understanding the Other: Spaulding and the Inclusion of Being-in-the-World in Folk Psychology

Omar Khali

Eastern Michigan University

ABSTRACT

In her book, How we Understand Others, Shannan Spaulding attempts to expand the orthodox theory of mindreading, which claims that we mostly attribute beliefs and desires to explain the behavior of others. She argues that mindreading tends to be much more complex and messy than what is assumed, for there are additional folk psychological dimensions to mindreading which must be considered. The various goals we have when mindreading, the stereotypes that are implicit in our interpretation of others' behavior, and the way in which situational context affect our explanatory ability are a few facets of mindreading that Spaulding considers. Despite Spaulding's pursuit to expand mindreading having merit, her position is nonetheless limited in the sense that she does not properly account for a fundamental way in which we understand others; namely, as being wrapped up in the world. I argue in the following essay that any discussion of mindreading, or folk psychology, must account for how we see the other as a subject engaged in the external world. I use Spaulding's book How we Understand Others as a way to illustrate an account of mindreading that falls short in considering how the other's worldly involvement affects how we explain and predict the behavior of others. The first part of the essay elucidates on Spaulding's position on mind reading. In the attempt to provide a more comprehensive account of mindreading, I explore the phenomenological concept of being-in-the-world as it relates to explaining the behaviors of others. The last section demonstrates how empathy can be further developed when emphasizing the other's being-in-the-world in explaining behavior.

KEYWORDS

Philosophy of Mind, Folk Psychology, Phenomenology, Social Cognition

Our social interactions with others vary from walking past someone on the street to arguing with a loved one. These are merely two of the many ways in which we interact with others. Essential to the system of interaction is explaining and predicting the behavior of others; for this is what folk psychology primarily concerns itself with. The orthodox view contends that folk psychology mostly consists of mindreading, "the capacity to make sense of intentional behavior in terms of mental states" (Spaulding 2018, 1). Shannan Spaulding, in her book How We Understand Others, argues that while she agrees with the orthodox view, mindreading turns out to be much more complex and messy than what is assumed. Despite Spaulding's basic pursuit to expand mindreading having merit, her position is nonetheless confining in the sense that she does not properly account for a fundamental way in which we understand others; namely, as being wrapped up in the world. I argue in the following essay that any discussion of mindreading, or folk psychology, must account for how we see the other as a subject engaged in the external world. I use Spaulding's book How We Understand Others as a way to illustrate an account of mindreading that falls short in considering how the other's worldly involvement affects how we explain and predict the behavior of others. The first part of the essay elucidates on Spaulding's position on mindreading. In the attempt to provide a more comprehensive account of mindreading, I explore the phenomenological concept of being-in-the-world as it relates to explaining the behaviors of others. The last section demonstrates how empathy can be further developed when emphasizing the other's being-in-the-world in explaining behavior.

MINDREADING AND EXPLAINING BEHAVIOR

To understand my focal thesis in this essay, it is important that I elucidate Spaulding's exposition of mindreading and how she responds to the various perspectives. She begins her argument by first stating that while she still defends the standard mindreading story (i.e. the orthodox view), she concedes that mindreading is not as "simple, uniform or accurate as the philosophical literature on mindreading suggests" (Spaulding 2018, 3). In other words, we mostly attribute beliefs, desires, and other mental states to a person in order to explain and predict their behavior; however, there are additional moving parts to mindreading that must be explored.

The "broad scope of mindreading claim" —the claim that mindreading is essential for social interactions—which Spaulding supports is not without its challenges. Proponents of embodied and enactive cognition contend that most of our folk psychology is not based on mental state attribution, but rather a mixture of non-mentalistic, embodied understandings predicated on perception. For example, we perceive in the "other person's bodily movements, facial gestures, eye direction, and so on, what they intend and what they feel" (Spaulding 2018, 9). A more sophisticated social interaction occurs when a child is capable of engaging in shared attention behaviors. Abilities such as "following gazes, pointing, and communicating with others about objects of shared attention" are elements of advanced social interactions which do not involve attributing mental states (Ibid). The former embodied practice is what embodied cognition calls "primary intersubjectivity" and the latter "secondary intersubjectivity." Attempting to substantiate their objection to the broad scope of mindreading claim, embodied cognitivists like Shaun Gallagher assert that our phenomenological accounts reveal that our ordinary interactions with others do not involve explicit attempts to understand the mental states of others. Spaulding takes issue with this position by pointing out that the evidence yielded by the phenomenological method is not novel, reliable, and relevant, thus rendering his objection to the broad scope of mindreading claim flawed. The second challenger to the insistence on mental state attribution being the primary social navigator comes from pluralist folk psychology. Similar to the embodied cognition perspective, pluralists (Kristin Andrews, Victoria McGeer, etc.) maintain that there is more to folk psychology than merely mindreading for the purpose of explaining and predicting the behavior of others. Pluralists note that sometimes our goal in a social interaction is not to ascertain the other's mental state, but rather to shape their mental state in accordance to our desires.

Unlike embodied and enactive cognition, Spaulding is more sympathetic towards the pluralist perspective; for she dedicates the third and fourth chapter of her book considering how the pluralists' contributions to folk psychology can expand our understanding of mindreading. Although, this is not to say that she agrees with everything pluralists espouse. Spaulding disagrees with the tendency pluralists have to separate non-mindreading social practices (like personality trait inferences or stereotypes) from mental state ascriptions. She argues that they cannot be separated from mindreading by virtue of the fact that they "interact

in messy ways" (Spaulding 2018, 17). The third chapter of How we Understand Others draws the reader's attention to the ways in which mindreading is affected by aspects of social interaction such as stereotypes and situational context; for these are features that the orthodox mindreading theories often neglect. Situational context refers to how a certain environment—whether that be in an airport, in school, on a bus, etc—contains implicit rules, norms, and behavior scripts which shape how we explain and predict the behavior of others. Those who live in populated urban settings like New York, will typically associate someone walking up to them with the desire to sell them some product. This explanatory behavior (explaining why the man is walking towards them) becomes shaped by being familiar with the context of walking in the streets of New York.

Stereotyping describes the tendency to associate a particular group with a set of characteristics or attributes. This can take on rather neutral form, as in associating children with being innocent, dependent, and vulnerable. Stereotyping or social bias can also be negative, as in the categorization of black Americans with unpleasant words like "terrible" and "horrible" (Spaulding 2018, 28). Spaulding notes that these developed associations prime or filter how we explain the behavior of others; her example of the man attempting to explain the behavior of his female colleague illustrates this. When a male faculty member explains that his colleague is less productive or is teaching a fewer number of courses due to her being a mother now, "he is explaining her behavior by categorizing it as an instance of a familiar pattern of behavior, namely, women putting their careers on the back burner to focus on their families'' (Spaulding 2018, 53). The association of being a mother and prioritizing family over a career is implicitly deployed in his attempt to make sense of her behavior. Moreover, the goals inherent in the man's attempt to make sense of his colleague is another dimension of mindreading which Spaulding highlights in this example. Explaining the colleague's behavior with the goal of accuracy may not be the focal aim in his mindreading, it is likely that his explanation "serve[s] to protect his ego as he regards his colleague as a competitor in terms of research productivity…" (Spaulding 2018, 55). Since I am mostly concerned with accuracy as the main goal of mindreading, there is no need to explore the other goals. The principal takeaway from the fourth chapter is that sometimes our main goal is not to accurately determine the beliefs and desires of others, but instead to boost our self-worth (as evidenced by the previous example) or to manipulate others.

Khali

BEING-IN-THE-WORLD AND THE OTHER

While Spaulding's efforts to expand mindreading appear to be successful, there is a feature of explaining the behavior of others that is missing in her exposition; namely, how the other is engaged in the world. This concern can be made manifest when observing the following example of "how mindreading works," found in the second chapter of the book (Spaulding 2018, 7). If we suppose that Lawrence (a colleague of yours) agrees to meet you at your office at 11 a.m., however, ten minutes pass and he has yet to show up. A possible explanation for his behavior might be that he desires to meet with you but believes that the meeting is at 12 p.m. Or perhaps Lawrence knows the meeting time but does not desire to meet with you. Spaulding also mentions that it may be the case that he "desires to meet with you, has a true belief about the scheduled meeting time and location, but something urgent has come up and he cannot come to your office"(Ibid). Notice how the final plausible explanation involves more than merely beliefs and desires about the meeting; for it involves an alteration in Lawrence's relation to the world. Attributing beliefs and desires to explain and predict the behavior of others, might not be helpful in many social interactions. Exploring the phenomenological concept of "being-in-the-world" and how it relates to our social interactions will capture the importance of considering the other's engagement with the world when explaining behavior.

The term "being-in-the-world" was first introduced by 20th century philosopher Martin Heidegger to emphasize the continual interaction we have with the world—whether that be engaging with others in conversation, being occupied with a task, or any of the various interactions we have with the world. Essential to this notion is recognizing that the world is not some neutral and static space; oftentimes the world complicates our desires to achieve some task. Heidegger remarks that our "concernful absorption in the work-world which lies closest to us" is disturbed once a piece of equipment breaks (Heidegger 1962, 103). For example, the carpenter faces resistance from the world when her power drill breaks; for now her preoccupation with building a chair must be deferred. Another primary element of being-in-the-world involves our encounters with others. When we interact with others, we typically do not consider them as objects with some semblance of consciousness; rather, we implicitly recognize them as having a similar conscious experience. Not only do we implicitly assume that they

have perceptual and rational faculties akin to ours, but there is an assumption that they are meaningfully involved in the world in their own unique way. This is to say that we understand others as having their own obligations, facing difficulties in the world, having particular desires and beliefs, and so on. Susan Bredlau, in her book The Other in Perception, says something similar in a succinct way: "[W]e discover that we encounter these bodies as perceiving subjects engaged with the natural and cultural world that we, too, can perceive rather than as thinking subjects engaged with ideas that are not immediately accessible to us" (Bredlau 2018, 29). In sum, not only do we find ourselves being-in-the-world, but we also take it as a fact that others are immersed in that same world. This has a few informative implications for discussions about mindreading and explaining the behavior of others.

Let us return to the example of Lawrence being late to the scheduled meeting. You take into consideration Lawrence's being-in-the-world precisely when the possibility that something urgent came up is subsumed with the array of plausible explanations for why he is late to the meeting. Another potential conjecture you might consider is the possibility that Lawrence is stuck in traffic, thereby preventing him from arriving on time. And, assuming that he desires to meet with you and has a true belief of the meeting time, explaining his behavior in terms of his mental states would not be helpful in this aforementioned possibility; it is simply the case that traffic is preventing him from being punctual. One might suggest that how the other is engaged with the external world falls into the broader domain of folk psychology, which makes it distinct from mindreading (making sense of behavior in terms of mental states). I however, contend that the other's being-in-the-world and their mental states cannot be separated as wholly distinct properties of a human; or as Spaulding would remark, they "interact in messy ways" (Spaulding 2018, 17). This implies that it is common for us to also consider how their interaction with the world contributes to explaining their behavior. Moreover, the necessity to not divide these properties into distinct categories is demonstrated in the consideration that Lawrence's behavior can be explained by him attending some urgent matter. Spaulding or other defenders of orthodox mindreading might claim that we are still attributing mental states to Lawrence when we entertain the possibility that something urgent came up, namely, the belief that the urgent matter was more important than the meeting. I concede that this mental state ascription might be occurring despite us not being

aware of it; however, this ascription follows from the consideration that changes in Lawrence's external world are interfering with his plan to meet with you. In other words, attributing mental states (i.e. the belief that the urgent matter was more important than the meeting) to explain Lawrence's inexplicable behavior would not have been possible had you not entertained the possibility that he experienced some significant alteration in his being-in-the-world. Another messy interaction can be found in the case where a faculty member is explaining why his colleague is less productive in her work. Assuming the faculty member is not stereotyping nor motivated to protect his ego (as described in the previous section), he attributes her reduction in productivity to having a child precisely because he knows how having a child (with all the effort and time required) typically interferes with other dimensions of one's being in the world—like teaching or conducting research, for example. We can grant that inherent in the explanation is an implicit recognition that the colleague believes that childcare is currently more important than her profession, however, this only follows from conjecturing how external circumstances affected her being-in-the-world (namely, having a child). The foregoing examples involve explaining the behavior of others from a distance, that is, the other's behavior is not being examined in a perceptual and direct way. A woman hammering a nail into a chair is an example of another's actions directly conveying their desires and intentions. However, another's mental state is not always provided to us just by merely looking at them; many occasions require us to inspect how that person is engaged in the world in order to make sense of their behavior. Imagine you are walking along a street in an area that has many stores in close proximity to each other, and suddenly see a man across the street dressed in a batman costume. A common method deployed in understanding his behavior would be to look at the stores he is walking towards. Once you do that, you will find a comic book store holding a small costume convention. Observing the other's relation to their environment provided the necessary information in explaining their intentions and desires; this is related to the embodied cognition's commitment to not viewing the "mind as something to be studied independently of the body and its environment" (Spaulding 2018, 9)

These three cases of explaining the behavior of others illustrate the messy interaction between mental state ascriptions and accounting for the other's engagement with the world. This messy interaction impels me to conclude that any theory or discussion of mindreading must account for the way we interpret

others' behavior through the understanding that they are subjects being-in-the-world. If we draw our attention back to Spaulding's criticism of the pluralists, we find her critical of the "idea sometimes implicit in pluralist folk psychology that we can separate mindreading from these other social practices (other folk psychological tools typically not subsumed into mindreading)" (Spaulding 2018, 18). And Spaulding takes great efforts in the third, fourth, and fifth chapters of her text expanding the theory of mindreading to include these non-mindreading social practices. However, if the goal is to include those folk psychological tools which interact with mindreading in "messy ways," in order to devise a ..".more comprehensive, cohesive, and plausible account of mindreading," it would be remiss to not consider how the other's relation to the world (their being-in-the-world) affects explaining and predicting their behavior (Spaulding 2018, 72).

EMPATHY AND BEING-IN-THE-WORLD

Besides increasing the accuracy of explaining the behavior of others, deliberately considering the other's being-in-the-world can also engender greater degrees of empathy in our folk psychological practices. As mentioned in the fourth chapter of How We Understand Others, standard theories of mindreading presuppose that accuracy is our primary goal when attempting to understand the behavior of others. This, however, is not always the case; for "we have various goals in our social interactions, and accuracy is just one of them" (Spaulding 2018, 43). Moreover, we can have more than one motivation influencing our interactions, where one is primary and others are secondary. Even our goals in understanding our own behavior can be skewed by self-serving motivations, which is evidenced by the self-serving attributional bias. This self-serving cognitive bias describes the tendency to attribute our successes to internal factors (e.g., hard work or intelligence) and attribute our shortcomings to external factors (e.g. other people or bad luck). The basic mechanism structuring this bias can be found in our social interactions with others. Like when explaining the behavior of individuals we are fond of or consider to be part of our "in-group" (family members, close friends, etc), it is likely that being empathetic and charitable are part of our goals; Spaulding remarks something similar when she says, "[w]e tend to have more favorable attitudes and empathize more with in-group members" (Spaulding 2018, 32). This means that a shortcoming of an in-group member is

likely to be explained by external factors interfering with their engagement in the world; for this is a basic feature of group-serving attribution bias. On the other hand, being uncharitable or confirming our unpleasant beliefs about the other are goals affecting the social interaction we have with those we deem despicable, which usually results in focusing on internal attributes (like their beliefs, desires, personality, etc.). For example, if you think highly of Lawrence, it is more plausible that you would attribute his unpunctuality to external factors, like being stuck in traffic or some urgent matter arising. Conversely, if you did not think highly of Lawrence, unpleasant internal factors such as being disrespectful or not valuing the meeting are plausible explanations you would consider. What is important to notice is that empathy or being considerate seems to shift our focus to the other's being-in-the-world—how they are interacting with the world— when explaining their behavior. As Shuan Gallager and Dan Zahavi note in The Phenomenological Mind, "[e]mpathy is defined as a form of intentionality in which one is directed towards the other's lived experiences" (Gallagher and Zahavi 2008, 183). This phenomenological definition of empathy is meant to convey that empathy attempts to consider the totality of the other's experience, which includes accounting for their external circumstances as well as their mental states. If one is attempting to become more empathetic (especially towards those who they deem as an out-group) in their folk psychological practices, a possible step in that direction would be to focus on how the external world might be influencing their behavior.

CONCLUSION

As mentioned in the introduction, I am sympathetic to Spaulding's efforts to expand mindreading in order to make it a comprehensive and cohesive domain of cognitive science. The orthodox accounts of mindreading fail to capture the complexity that goes into explaining the behavior of others, which in turn negatively affects how one might try to change their maladaptive mindreading practices. While I concur with Spaulding's expansion of mindreading, there are nonetheless aspects of our folk psychology needed to be accounted for, given they are inextricably tied to mindreading. I argued in this essay that any discussion of mindreading must account for how we see the other as a subject engaged in the world. I use Spaulding's book How We Understand Others as a way to illustrate an account of mindreading that falls short in considering how the other's

worldly involvement affects how we explain and predict the behavior of others. The first part of the essay elucidated on Spaulding's position on mindreading. The next part explored the phenomenological concept of being-in-the-world as it relates to explaining the behaviors of others. Finally, I considered how the other's being-in-the-world is relevant to empathy in our folk psychological practices.

REFERENCES

Bredlau, Susan. 2018. *The Other in Perception: A Phenomenological Account of Our Experience of Other Persons*. Albany, NY: SUNY Press.

Gallagher, Shaun and Dan Zahavi. 2008. *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. New York: Routledge.

Heidegger, Martin. (1927) 1962. *Being and Time*, trans. By John Macquarrie and Edward Robinson. New York, NY: Harper & Row Publishers.

Spaulding, Shannon. 2018. *How We Understand Others: Philosophy and Social Cognition*. London: Routledge.

compos mentis

Freedom or Determinism: In Light of Recent Neuroscientific Discoveries

Corine LaFrenier
Hope College

ABSTRACT

Recent advances in neuroscience and cognitive psychology have led researchers and philosophers alike to grapple with questions concerning humanity's free will. The findings of Libet, Treisman, and more recently, Barrett may initially seem to point confidently towards a deterministic stance on human free will; however, after a careful review of their findings it becomes increasingly clear that their findings suggest the opposite: humans have the intrinsic freedom to decide, to act, and to feel. In this essay, I assert that the dissonance between science and free will is fabricated from a misunderstanding of what the scientific literature has suggested in light of determinism. Furthermore, I would contend that free will is not a fallacious concept that, by nature, cannot coexist with the deterministic-like hypothesis that science has only somewhat suggested. By examining the findings of Libet on readiness potentials, the results of modern studies in cognitive psychology on attention, and recent research on the constructive nature of emotions, I propose an intermediate perspective, one that encompasses the triumphs of modern-day science without discrediting our freedom. In this way, I assert that there is no discord between modern scientific findings and the basic tenets of free will.

The recent advances in biochemistry, molecular biology, and cognitive neuroscience have led many to question how constructs like freedom—and other constructs that are so closely tied to the human experience—can coexist with the expanding body of literature suggesting the presence of a strong deterministic predisposition to the human condition. For, if we are the product of things that, for the most part, are vastly out of our control, like our genes, our environment, and our neurology, how do we reconcile our unique experiences as human beings? How do we explain the tenets of free will and personal autonomy in the face of scientific findings that point towards determinism? These questions have become stumbling blocks in the minds of many people, scientists and philosophers alike, leading them to believe that free will simply cannot coexist in the realm of modern-day science that seems to point heavily towards the role deterministic predispositions play on the human experience. Proponents of the belief that modern science provides ample evidence that the human condition is largely deterministic are not apt to believe that we are autonomous creatures, free to make our own decisions, thereby posing some serious questions as to how we think, act, and behave. Likewise, those who champion the belief that we are wholly free and autonomous individuals are criticized for their ignorance in light of the scientific evidence. This divide in perspective, while very real, is grossly unnecessary. I would argue that the dissonance between science and free will is fabricated from a misunderstanding of what the scientific literature has suggested in light of determinism. Furthermore, I would contend that free will is not a fallacious concept that, by nature, cannot coexist with the deterministic-like hypothesis that science has only somewhat suggested. By examining the findings of Libet on readiness potentials, the results of modern studies in cognitive psychology on attention, and recent research on the constructive nature of emotions, I propose an intermediate perspective, one that encompasses the triumphs of modern-day science without discrediting our freedom. In this way, I assert that there is no discord between modern scientific findings and the basic tenets of free will.

Modern scientific brain imaging techniques have allowed researchers to delve into the depths of the mind and quantify the initiation of a voluntary act. The neural delay for conscious intention, described by what neuroscientists call the readiness potential, is a neural marker that is "recordable on the head starting up to a second or more prior to the actual muscle activity" (Libet 1992, 262). This electrical charge has been thought of as "an indicator of brain activity

that is involved in the onset of a volitional act" thereby leading researchers to question whether "the conscious wish or intention to perform that act precede[s] or coincide[s] with the onset of the preparatory brain processes, or does the conscious intention follow that cerebral onset?" (Libet 1992, 262). Researchers performed a comprehensive study to determine which comes first: the conscious intention to perform an act or the readiness potential. The results clearly showed that onset of the readiness potential precedes the subjects' conscious wish to perform an action by about 350 msec (Libet et al., 1982). These findings suggest that our actions are predetermined by neural activity prior to our conscious decision to perform those actions, therefore posing some serious questions as to whether we are free to perform those actions in the first place. To describe these findings, Libet and colleagues developed what is called the "cerebral time-on theory" which states that "the transition, from an unconscious mental event to one that reaches awareness and is consciously experienced, can be a function of a sufficient increase in the duration" of the neural activities such as the readiness potential (Libet 1989). In this way, "neural activities whose duration is below some minimum substantial duration could mediate a mental function that remains unconscious" but, "when such activities persist for longer than a minimum time of up to about 500 msec, subjective awareness of the mental function can appear" (Libet 1992, 265). While the explanation proposed by Libet and colleagues seems to accurately describe their findings and offer a strong argument for determinism, it is imperative to note, as Libet does, that "the changes in durations of the appropriate neural activities may be affected, for example, by changes in the attention process" (Libet 1992, 265). Therefore, slight alterations in our attention, can contribute to whether or not these readiness potentials essentially become volitional acts. And because attention is a vastly conscious cognitive construct, this leaves room for free will to play a role.

Cognitive psychologists define attention as "focusing on specific features, objects, or locations or on certain thoughts or activities" (Goldstein & Mackewn, 2005). In most, but not all cases, when we focus our attention, we are doing so via conscious volition. According to Treisman's widely accepted model of attention, the attended message (the message that we focus on) gets through stronger than the weaker, unattended message and that the strength of the attended message has a direct relationship to our conscious ability to attend to that message

(Goldstein & Mackewn 2005; Treisman 1964). But what implications does this have for free will in the context of the readiness potential?

Combining these findings in neuroscience with those demonstrated in cognitive psychology leaves ample room for free will to coexist in the realm of readiness potentials. Drawing from Libet's "cerebral time-on" theory, the longer readiness potentials persist in our unconscious, the more likely they are to become a part of our subjective awareness (Libet 1992, 265). In this way, readiness potentials can be likened to the unattended messages that pass through the attenuator at a weaker strength—they do not enter into consciousness until we consciously attend to them. Once we consciously attend to them, only then are action potentials fired resulting in the initiation of a voluntary act. Therefore, it is our conscious attentive processes that give rise to the action potentials that characterize our volitional acts. In this way, by adjusting our perceptual load to attend to these readiness potentials, we are able to control our actions. To put it another way, readiness potentials don't evoke voluntary action, only action potentials can do that, and so by directing our attention to a readiness potential, it enters into our conscious experience and allows us to be free to perform voluntary actions. So the question isn't so much about which comes first, "the conscious intention to perform a voluntary act" or the readiness potential; but rather, the question becomes this: At what point do we decide to attend to this background neural activity? As Libet points out, the fact that we can adjust the rate at which these readiness potentials enter into consciousness via our attention (or lack thereof), points directly to the notion that we, as human beings, are mostly free to attend to messages and stimuli in our environment, or at the very least, have the capacity to freely attend to these stimuli.

So once we attend to these messages, are we essentially "stuck" in such a way that we have to perform these actions because the background neural activity entered into consciousness?

To put it simply, no. Even Libet recognizes that, once the wish to act enters consciousness (as a result of prolonged readiness potentials, at least according to Libet) "the subject can consciously control the outcome of that volitional process, for example, by vetoing, i.e. blocking its final expression as a motor act" (Libet 1992, 270). For, "if the conscious control function itself is initiated by unconscious cerebral processes, one might argue there is no role at all for conscious free will even as a controlling agent" (Libet 1992, 270). Essentially, if consciousness merely

describes our perception of our own existence, and does not lend itself in any way to control our actions or give us the liberty to freely choose our actions, then what is the purpose of consciousness? There would be no purpose. In this way, Libet distinguishes between the "conscious control of an event" and the phenomenon of "becoming aware of the volitional intent" to perform said action/event, arguing that these are separate phenomena (Libet 1992, 270). By making this distinction, Libet recognizes the limitations of his findings and, "given the difference between a control and an awareness phenomenon… we should recognize that there is presently no specific experimental test of the possibility that conscious control may require an unconscious cerebral process to produce it" (Libet 1992, 270).

Thus far we have seen how Libet's discovery of readiness potentials, when coupled with more recent research in cognitive psychology, leaves room for human free will. However, many still question the validity of free will in regard to other facets of the human experience, such as our emotions. Proponents of determinism would claim we are largely the byproduct of our neurophysiology and neurochemistry—our thoughts, feelings, and emotions are largely predetermined and wholly out of our conscious control, and thus, any actions performed while in negative emotional states cannot be fully attributed to the person experiencing the emotion. The deterministic perspective on emotional free will is based largely in the classical view of emotions, which posits that emotions are analogous to fingerprints in that they are composed of "a distinct pattern of physical changes" that result in our experiences of them (Barrett 2017, 3). In this way, "the 'fingerprint' of emotion is likewise assumed to be similar enough from one instance to the next, and in one person to the next, regardless of age, sex, personality, or culture" (Barrett 2017, 3). Evidence for this perspective comes from rather primitive studies in which researchers flash an exaggerated facial expression on a screen and ask participants what emotion they are viewing, and "from this evidence, scientists concluded that emotion recognition is universal" (Barrett 2017, 7). These researchers then reasoned that "the only way expressions could be universally recognized, [was] if they [were] universally produced" leading them to the drastically bold and wide-sweeping conclusion that "facial expressions must be reliable, diagnostic fingerprints of emotion" (Barrett 2017, 17). This explanation is satisfactory for the proponents of determinism; for it is intuitive that we all share common emotional expressions that serve as "fingerprints" that ultimately render us enslaved to our biological systems. However, with the surge in technological

advances came the ability to look at emotions as they are experienced in real-time. Facial electromyography (EMG) and Facial Action Coding (FACS) allowed researchers to observe minute changes in facial expressions that were previously undetectable. Furthermore, the evidence from these studies "present a serious challenge to the classical view of emotion" in that "in study after study, the muscle movements do not reliably indicate when someone is angry, sad, or fearful; they don't form predictable fingerprints for each emotion" (Barrett 2017, 7). This finding captured the attention of psychologists and researchers alike, highlighting the need for a more comprehensive hypothesis for the nature of emotions, and leading many to believe that there is no "one size fits all" facial expression for each emotion we experience.

But what implications do these findings have on not only the theory of emotion, but also on the basic tenets of free will? As aforementioned, the deterministic viewpoint describes emotions according to the classical view, in that emotions are these innate, natural responses that are biologically ingrained and have universal features. If emotions are intrinsically rooted in our biological system, then when we act in accordance with our emotions—when we react out of anger or frustration—we cannot be held responsible for those actions as they are the result of our biology and not our free will. But modern-day science offers little evidence for the classical view of emotions. Instead, an overwhelmingly large amount of research points to what Barrett calls the constructive nature of emotion. According to Barrett, the brain synthesizes sensory inputs to create an emotion in real time, and then adjusts these responses accordingly. Furthermore, Barrett advocates for emotional free will, citing countless studies that show that humans are largely in control of their emotions, and that with this heightened awareness of our control over our own emotions comes a great deal of responsibility for our actions.

Perhaps this is best illustrated in the United States legal system, which distinguishes between crimes committed in the "heat of passion" and premeditated crimes. The U.S. legal system "assumes that emotions are part of our supposed animal nature and cause us to perform foolish and even violent acts" thereby offering a sort of cop-out explanation that "partially mitigates a person's responsibility for his actions" (Barrett 2017, 221). This explanation, termed the heat-of-passion defense, "depends on assumptions from the classical view of emotion" (Barrett 2017, 221). Barrett cites the example of the Boston Marathon

bomber, Dzhokhar Tsarnaev who's trial weighed heavily on his expression of remorse for his actions, for if he showed remorse in the courtroom and pleaded that he was caught up in emotions over which he had no control, he would be less likely to receive the death penalty for his crimes. In this way, the court system has fallen into the deterministic-like classical view of emotions that render us no longer responsible for our actions that were performed out of intense emotion. However, as growing amounts of scientific literature indicate, this deterministic viewpoint of emotions is unfounded in modern neuroscience and psychology. To put it simply: emotions are not biologically ingrained responses that are out of our control; but rather, science has shown that humans have nearly full control over their emotions. And with emotional control comes emotional free will. And, as with any form of free will, comes responsibility for one's actions.

In conclusion, the discovery of readiness potentials and the findings and theories of Libet, while they may initially seem to point to a wholly deterministic theory of human behavior, do in fact, leave ample room for free will, especially when these findings are taken in light of recent research in cognitive psychology on human attentional paradigms. The distinction between awareness and consciousness is necessary when interpreting these findings. Moreover, other facets of the human condition also point to freedom over determinism. While the classical view of emotions was the leading theory behind emotional regulation for a substantial amount of time, recent advances in research have shown that emotions are not ingrained into our biological systems. With this paradigm shift from the classical view of emotions to the constructive nature of emotions came an inherent shift from emotional determinism to emotional free will. The recent findings of the constructive nature of emotions coupled with growing evidence for emotional free will have created the need for humanity to take increased responsibility for their emotions and their actions. For increases in freedom undoubtedly come with increases in responsibility. In total, the surge of findings in neuroscience and cognitive psychology seem to point towards freedom over determinism and, as a result, there is no inherent dissonance between recent scientific discoveries and free will.

REFERENCES

Feldman Barrett, Lisa. 2017. *How Emotions Are Made: The Secret Life of the Brain*. Mariner Books.

Goldstein, E. Bruce, and Angie Mackewn. 2005. Essay. In *Cognitive Psychology Connecting Mind, Research, and Everyday Experience*: 89–94.

Libet, Benjamin. 1982. "Brain Stimulation in the Study of Neuronal Functions for Conscious Sensory Experiences." *Human Neurobiology*: 235–42.

Libet, Benjamin. 1989. "Conscious subjective experience vs. unconscious mental functions: A theory of the cerebral processes involved in models of brain function." Cambridge: Cambridge University Press: 35–49.

Libet, Benjamin. 1992. "The Neural Time - Factor in Perception, Volition and Free Will." *Revue De Métaphysique Et De Morale 97* (2): 255-272.

Treisman, A. M. 1964. Selective attention in man. *British Medical Bulletin* 20: 12-16.

compos mentis

Is Beauty Objective?

Zhixuan Li
The University of Hong Kong

ABSTRACT

In discussions of "beauty" as an aesthetic and evaluative term assessing and describing people's looks, very often people use the term as if there is a shared objective standard of beauty. This mistaken conception of beauty as objective can be understood in three layers: firstly, the term "beauty" is used as if it means that a certain set of objective standards has been met; secondly, it is presupposed that these standards have existed throughout human history and will continue to exist as such even despite what appear to be significant changes to these standards; thirdly, it is often presupposed that we are all equally capable of achieving these standards and that we should all aim to meet these standards. However, this conception of beauty as objective is mistaken. I seek to make clear that: the aesthetics of people's looks is very often shaped by racism, patriarchy, heteronormativity and other systematic oppressions in our society. The term "beauty" is much more of an oppressive tool than an innocent realist appraisal or aesthetic judgment.

INTRODUCTION

In discussions of "beauty" as an aesthetic and evaluative term assessing and describing people's looks, very often people use the term as if there is a shared objective standard of beauty. To justify this, people appeal to a variety of mechanisms and instincts that are taken to underlie the objectivity of beauty. People argue that attributes such as "small waists", "big eyes", "pale skin" are just "naturally beautiful." The word "naturally" here is used with an implication that beauty is part of nature and therefore, objective.

This mistaken conception of beauty as objective can be understood in three layers: firstly, the term "beauty" is used as if it means that a certain set of objective standards has been met; secondly, it is presupposed that these standards have existed throughout human history and will continue to exist as such even despite what appear to be significant changes to these standards; thirdly, it is often presupposed that we are all equally capable of achieving these standards and that we should all aim to meet these standards. However, this conception of beauty as objective is mistaken.

In what follows, I will present a variety of reasons for understanding beauty as socially constructed, rather than objective. In particular, there are several constructive forces that shape what beauty is and its vast significance within society. I seek to make clear that: the aesthetics of people's looks is very often shaped by racism, patriarchy, heteronormativity and other systematic oppressions in our society. Discrimination based on looks falls into a bigger picture on the societal and historical level. The term "beauty" is much more of an oppressive tool than an innocent realist appraisal or aesthetic judgment.

In the First Part of this paper, I will spell out the example of the intellectual construction of "aesthetics." I will draw examples from Chinese literary traditions, contemporary evolutionary science as well as Western philosophy of "taste" to illustrate the biased nature of the concept as it is conceived of in different intellectual traditions.

In the Second Part of this paper, I will showcase examples of the social force "aesthetics" carries in contemporary societies. I will draw on examples such as Ugly Betty and cosmetic surgeries in China to illustrate how beauty is corrupted: it is a tool of oppression. Most importantly, we are not afforded equal status in the accumulation the social capital endowed by beauty. This section will develop the

point that beauty should not be treated as a common natural or good goal that everyone should be striving for.

In the Third Part of this paper, I will spell out the racialized nature of the aestheticization of women of color. With Jari Jones and "Mulan" as my illustrations, I will showcase how their races affect people's aesthetic judgment. This section will fully develop an argument against the perceived consistency of "natural" beauty throughout human history.

In the Fourth Part of this paper, I propose to think beyond the existing frame of beauty. I will incorporate the example of Radical Body Positivity to challenge the presupposed progressiveness of the Body Positive Movement. I intend to reveal the neoliberal nature of the Body Positive Movement, as well as the new pedagogy of beauty in this era. In the end, I propose that we should aim to abolish the authority we grant beauty, given that very little good has come from the pursuit of beauty throughout human history.

THE FIRST PART: THE INTELLECTUAL CONSTRUCTION OF "AESTHETICS"

In this section, I will lay the groundwork for the basic concept of "aesthetics" as a form of intellectual oppression. I will do so by way of three illustrative examples, which I do not take to be unusual or special, in regard to their underlying conceptions of aesthetics as objective. In each case, I will argue that these conceptions are incorrect: they fail to recognize the social forces at play in their own intellectual accounts of the objectivity of aesthetics. In the first illustration, I will reveal the sexism in the concept "aesthetics" itself by showing how women have been treated as aesthetic objects within the Chinese literary tradition of "sickly beauties." In the second illustration, I will further elaborate the "objectivity" of aesthetics as perceived by people and find reason to object to a common argument based on evolutionary theory (such as explaining appreciation of small waists and big buttocks as attraction of fertile women). In the end, I will illustrate the authority of "taste" in the philosophical tradition. From this section, it will be very clear that aesthetics is constructed by a male-centered intellectual world. Literature of "sickly beauties", science of "evolutionary theories", philosophy of "taste", they are all but part of the general patriarchal oppression in terms of human knowledge, only in the name of "aesthetics."

"Aesthetic" as projecting females as objects

A common appreciation of beauty comes from appreciation of femininities as weaker, inferior and fragile. Immanuel Kant openly appreciated "his fair sex" as agreeable to "human nature", or indeed, men's nature (Korsmeyer and Weiser 2021). He even differentiated "sublimity" from "beauty" with a gendered glance: how ordinary "prettiness and elegance" is inferior to, for example, Beethoven's music (Zangwill 2021). The subjectivity of male viewers presupposes the objectification of female bodies, and what's more, the "feminine beauty" is wedded with lesser traits –sublimity could only come from men's creation, no matter how "enjoyable" or "agreeable" Kant's "fair sex" are (Korsmeyer and Weiser 2021). And such a cultural tradition is not unique to the West. In this section, I will incorporate the aesthetic of "sickly beauties" in late imperial China to illustrate how "aesthetic", besides casting women in a light of objectification because it was dominated by a male-centered intellectual world, also manifests as an appreciation of female weakness, disease and distress. The unhealthy fetishism of "sickly beauties" in late imperial China illustrates the innate sexism in aesthetics as a discipline as well as the cross-cultural nature of intellectual oppression in the name of "aesthetics."

Gender is crucial to perception of illness in Chinese cultural tradition. In Charlotte Furth's A Flourishing Yin, she discusses how gender and illness presuppose each other in the Chinese medical tradition. The feminine element Blood is inferior to the masculine element qi in Chinese medicine: "Blood was feminized as secondary and dependent in the hierarchy of bodily energies, unable to stand alone as a primal body vitality as qi did." (Furth 1999, 74) The social gender is wedded very well with the biological difference as well as the cultural interpretation of such difference. Menstruation, understood as the most feminine issue in medicine, is "identified with ideal normality" (Furth 1999, 77). Menstruation is supposed to be on a regular basis, presupposing fertility and good health. Deviance from such a norm, though, may be understood in many ways: it could be a sign of failure to fulfill various social roles as a woman. Cheng Congzhou, a famous Chinese medical practitioner in the 17th Century, wrote about a case in which he was frustrated about a woman's irregular menstruation: "Why can't she be regulated?" (Furth 1999, 256) Upon observation of the woman's social interactions, he concluded that the woman's disease came from her reluctance to take the medicine provided to her by her sister-in-law, and it was a "negative

emotion" as well as social dysfunctionality that caused her disease (Furth 1999, 256). Medicine as a male-centered intellectual discipline interprets diseases already in a gendered way, and such medical knowledge laid a foundation for literary tradition of "sickly beauties."

With the notion of "sickly beauties", illness in women is commonly fetishized as an aesthetic object in the Chinese cultural tradition. In A Dream of The Red Mansions, Lin Daiyu, a female leading character, is constantly ill with her sensitivity and passion. As Zhang Xinjun and Liu Qizhi's wrote, "her illness originates from her own physical and psychological states and constitutes her existential condition. Lin embodies the kind of Morbid Beauty composed by her illness, grace, talent, perfectly to the taste of feudal Chinese scholars" (Zhang and Liu 2015, 60). The patriarchal control implied by her disease could not be more obvious: her weak body could not physically escape masculine power, and on top of that, her constant romantic imagination and passion toward a male lover created an emotional chain. As Janice Radway wrote in Reading the Romance, the emotional needs fulfilled by merely reading romantic stories may consolidate patriarchy because it jeopardizes the future "that might otherwise be formulated as demands in the real world and lead to the potential restructuring of sexual relations" (Radway 2009, 213). Lin, and many other passionate female characters, are perfectly described by Radway: their emotional dependence on men is reinforced by their engagement with literature, which constitutes a major aspect of the image of "sickly beauties."

What's more, given that A Dream of The Red Mansions was written by two male authors in late Imperial China, it is safe to conclude that Lin Daiyu as a sickly beauty was a cultural creation. Female characters such as her – graceful, lonely, talented, loving, ill – are but a mirror of the mental state of male scholars and intellectuals. Just as how Louise Edwards noted: "Young women are imprisoned in both realms by discourses of feminine purity, whereas men can exploit or transcend moral/immoral and mythic/mimetic boundaries." (Edwards 1994, 58) Aestheticization of her and her disease reveals the nature of human knowledge, cultural traditions and the intellectual world being dominated and written by men. Just as how Cornelia Klinger has argued on the overall masculine production of art: "The idea of the artist as genius contains a marked intensification of the belief in the subject's sovereignty, autonomy, and creativity." (Klinger 2017, 344) Such belief of male artists as well as intellectuals just manages to conceal the deep deprivation of sovereignty, autonomy and creativity of women as artistic objects.

"Aesthetic Objectivity" as defended by evolutionary theory

Besides the Chinese medical and literary tradition that describes "gendered illness", Western science has also been a major force in terms of intellectual oppression in the name of "aesthetics." One "scientific" justification for appreciating women with small waists and big hips is that: the physical attractiveness of such female bodies suits the need for evolution because men appreciate more fertile women. In this case, not only is a woman with a smaller waist and a bigger hip "objectively beautiful or attractive", such attraction is justified by the natural mechanism studied by science, and thus, should hold true throughout human history and in the future. In the Introduction, I outlined three layers of the conception of the objectivity of beauty: firstly, "beauty" is used as if it means a certain set of specific standards; secondly, these standards exist throughout human history as constant and will remain constant in future; thirdly, we are equally capable of achieving these standards and should aim to achieve them. Here we see the first two layers of objectivity: appreciation of small waists and big hips is "constant scientific truth." I will take issue with this claim using a critical-feminist perspective on science to show a fundamental male-centeredness in the discussion of female fertility and mate value.

The common appreciation of small waists and big hips is theorized as a small Waist-to-Hip Ratio (WHR), which is believed to be aesthetically preferred because it purportedly implies the good health and fertility of women. In Jeanne Bovet's study of Evolutionary Theory and Men's Preferences for Women's Waist-to-Hip Ratio, she went through the entire literature of evolutionary biology that has attached "mate value" to women with smaller WHR. It was commonly believed by scientists that "a woman with a high mate value will increase the reproductive success of her mate(s)." (Bovet 2019, 2) If such logic works, then it is concluded that women who have smaller waists and bigger hips "increase the number and quality of descendants a man will have (including the ones he has with other women)." (Bovet 2019, 2) She further analyzed a variety of claims such as that in the discipline of evolutionary biology. In the end, Bovet appraised the use of WHR in many analyses, if not all.

I argue that in the discussion of WHR, "mate value" as a result of low WHR, as a scientific term, is fundamentally erroneous. The definition of a woman's mate value is supposed to simply means her own place in the sexual marketplace. However, it becomes at the hand of the reproductive success of her mate, even

when the mate does not have children with her. WHR is supposed to be relevant to fertility and health in pure biological terms. However, it becomes a tool to measure her "ability to reproduce for men" –fertility is conceived as "useful to men" in this sense.

Not only is the term embedded with a patriarchal assumption of women's reproductive behaviors, the term itself is not that useful to explain intimate partner violence in a more gender-equal society. In Khandis Blake and Robert Brooks' study of men with high mate value, they surprisingly find out that: "When mate value was high, gender equality increased men's support for male-to-female IPV (intimate partner violence)" (Blake and Brooks 2018, 7). Namely, most wanted men in most gender-equal societies could be potential abusers because of the loss of privilege he could have had in a gender-inequal society, and yet, they are considered "men with high mate value." If we take the meaning of "high mate value" as indicator of reproductive success of one's mate, then men with high possibilities to enhance the reproductive success of his mates simultaneously endorse domestic violence. Given a general patriarchal structure in our society, the children his mates bear should be more likely his than not. It is argued that: "Both the high mate value partners of paired high-value men and their greater resource investment means that female defection and infidelity are far more costly" (Blake & Brooks 2018, 3). That is to say: "mate value", a seemingly gender-neutral term in evolutionary science, does not measure the potential good a man could bring to his female spouse. "High mate value" male is perceived with higher IPV endorsement and more offspring, but "high mate value" female is perceived with more offspring her male partner could potentially have.

In Sharyn Clough's Beyond Epistemology: A Pragmatist Approach to Feminist Science Studies, she acknowledged the long-rooted male-centeredness in evolutionary biology by tracing back to Charles Darwin. Sharyn Clough summarized that: "Expanding his theory of natural selection to include sexual selection, Darwin hypothesized that secondary sexual characteristics do indeed have a biological function because they better enable the individual to attract a mate, fend off competitors for that mate, and/or provide for the care of offspring." (Clough 2003, 47) Incorporating a Victorian mindset, Darwin attempted to explain away the fact that in nature, males tend to have "secondary sexual characteristics" in order to attract females, whereas in human society women are more trained to attract men. Darwin wrote extensively and somehow contradictorily on the

subject, arguing for the superiority of men with their "superior secondary sexual characteristics" (Clough 2003, 50) because the masculine characteristics symbolize courage and intelligence. He would argue that a civilized white male of his time is of great physical attraction to women, ignoring the fact that female appearances are far more emphasized among his contemporaries (resulting in some horrible inventions such as corsets).

Such a biased approach to nature underpins evolutionary theory. Scientists such as Darwin turned a blind eye to female oppression in human society in contrast to female superiority in nature. Moreover, Darwin and his precedents attempted to justify superiority of human males with coined terms such as "superior secondary characteristics" and "mate value", presenting male as the center of sexual selection. Aestheticization of low WHR becomes part of the patriarchal judgment on women. There is no gender-neutral discussion of reproduction, attraction and bodily traits. The seemingly objective tone in describing women's body shapes does not neutralize the fact that it is a one-way judgment from men to women and it comes from a male-centered scientific tradition. "Aesthetics", on top of such construction and justification, could not be argued as objective.

Authority of "taste" in philosophy

In this section, I will primarily take issue with David Hume's "Of the Standard of Taste", given the authority this essay carries in the argument for a fixated standard of taste constituted by judgments of well-cultivated viewers. I argue that his idea of "better or worse aesthetic judgment" is prevalent in the Western tradition of philosophy and thus, contributed to a general intellectual oppression.

Hume asserted the authority of "Standard of Taste" very early on in his piece. He argued that: "It is natural for us to seek a Standard of Taste; a rule, by which the various sentiments of men may be reconciled; at least, a decision, afforded, confirming one sentiment, and condemning another." (Hume 2006) As a philosopher of subjectivism, Hume did not abolish the natural differences of people's aesthetic sentiments towards objects. However, sentiment and judgment are perceived very differently in Hume's philosophy. Hume defended his idea of a normative judgment of taste by incorporating some of his rivals that are against the standard of taste: "There is a species of philosophy, which cuts off all hopes of success in such an attempt and represents the impossibility of ever attaining any standard of taste. The difference, it is said, is very wide between judgment and

sentiment. All sentiment is right; because sentiment has a reference to nothing beyond itself, and is always real, wherever a man is conscious of it." (Hume 2006) Appreciating each and every one's sentiments as "real", Hume did argue for a "rule" or "decision" which is normative enough so that we could judge people's sentiments with it. People's ideas of pulp fictions being nicer than Shakespeare is not denied, but rather judged and ignored by the mainstream: "Though there may be found persons, who give the preference to the former authors; no one pays attention to such a taste; and we pronounce without scruple the sentiment of these pretended critics to be absurd and ridiculous. The principle of the natural equality of tastes is then totally forgot, and while we admit it on some occasions, where the objects seem near an equality, it appears an extravagant paradox, or rather a palpable absurdity, where objects so disproportioned are compared together." (Hume 2006)

Following the absolute standard of taste, Hume also argued for a good critic. "It is impossible to continue in the practice of contemplating any order of beauty, without being frequently obliged to form comparisons between the several species and degrees of excellence and estimating their proportion to each other. A man, who has had no opportunity of comparing the different kinds of beauty, is indeed totally unqualified to pronounce an opinion with regard to any object presented to him." (Hume 2006) In Hume's opinion, only a man with sufficient opportunities of "comparing the different kinds of beauty" can be trusted to remain "in the practice of contemplating any order of beauty." A person who is immersed in pulp fictions and popular culture is thus not at a place to appreciate or judge Shakespeare. A good critic, after being exposed to a variety forms of sentiments and thus well cultivated, form a sentiment that is "right" or that adheres to the "Standard of Taste."

Hume's argument left many ambiguities. "So what are the rules of taste, over and above any rules or principles involved in sound judgment about the object of taste? Surprisingly, Hume never offers a clear case of one." (Gracyk 2020) The first ambiguity lies in his inability to analyze or describe the absolute standard of taste. According to Timothy Costelloe: "The standard, rather, as a general rule in its second influence, is an abstraction from actual practice that articulates how one ought to judge if one is to judge correctly in matters of beauty." (Costelloe 2007, 14) Interesting enough, such an explanation fails to address anything concrete as to "how one ought to judge." One may use as many examples as possible,

comparing Harry Potter to Shakespeare, comparing The Lord of the Rings to Divine Comedy, but the line between "popular culture" and "classics" remains debatable, because classics to our times were sometimes considered inferior to older texts in their own times.

The second ambiguity lies in the arrogant tone of his argumentation. Unable to describe or analyze fully what defines an aesthetic sentiment, Hume seemed to attach the absolute standard of taste to his own standard of taste as "how one ought to judge." When he argued that: "The same Homer, who pleased at Athens and Rome two thousand years ago, is still admired at Paris and at London." (Hume 2006) Hume seemed to attach absolute aesthetic value to Homer given that he is widely accepted in the Western societies. Similarly, when he compared Ogilby and Milton, or Bunyan and Addison, he also gave credit to the latter authors without any hesitation and argued that "Though there may be found persons, who give the preference to the former authors; no one pays attention to such a taste" (Hume 2006). One may have to notice his abruptness. Nothing has been developed as the criteria between a good author and a bad author, only because he took for granted that his social group, as cultural elites, have better tastes; the former authors are denounced. What Hume manifested here is the arrogance of an educated man: "No one pays attention to such a taste" –a taste of working class, a taste of uneducated public, a taste of other cultures. It is compatible that Hume argued that aesthetic judgment should be made by "good critics", a person with enough cultivation and thus, belongs to Hume's own social group. Monique Roelofs argues in her book The Cultural Promise of the Aesthetics that Hume has an exclusive nature in his "Standard of Taste" by "contrasting genuine taste with a range of inferior ones (those of Indians, women and blacks)" (Roelofs 2014, 152), and I fully agree to such a comment.

In the First Part of my thesis, I have used three useful illustrations from literature, science and philosophy to prove the biased nature of "aesthetics" from the very root of it –the intellectual construction of "aesthetics" was at hand of multiple social forces, and most importantly, patriarchy. "Sickly beauties" are fetishized with a deep fetishism of power over a weak female body; aestheticization of women with low WHR is justified by a problematic tradition that exaggerates male power in sexual selection in human societies; Hume's "Standard of Taste" was wedded with his own social power to issue and perpetuate the taste of his own social group. I have also showed that the first layer of the objectivity of beauty

incorrect – namely, beauty is meant to meet certain objective standards, but since the standards themselves are corrupted in such intellectual traditions, it is incorrect to characterize beauty as objective. My argument provides a groundwork for the revelation of the social force of aesthetics in contemporary societies.

THE SECOND PART: CONSTRUCTIVE FORCES IN CURRENT SOCIETY THAT CONTRIBUTE TO THE CONSTRUCTION OF BEAUTY AND THE SIGNIFICANCE OF BEAUTY

In this section, I will present analyses and examples of the different constructive social forces involved in aesthetics. The goal of these analyses and examples is to make clear that society consists of a variety of systems that contribute to the construction of beauty itself, and its mistaken status as objective. Beautiful looks are significant in our society, not because of an innocent manifestation and pursuit of beauty, but because of the hierarchies it creates and societal systems it contributes to. The current beauty standard is constructed in order to, not only oppress some people with less preferred looks, but also to justify existing oppressions of other forms. I will give three examples that best spell out the systematic oppression and privilege.

The Detested Ugliness

In this section, I will draw on examples from popular culture. In these examples, popular culture reinforces the detestation of ugly people and it attempts to evict ugliness from the public sphere. These examples make clear that our society is systematically detesting and subordinating ugliness in order to consolidate the significance of beauty.

"Ugly people are unwanted in the public space": one way to consolidate the doctrine of spatiality of ugliness is to justify mockery and annihilation of ugly people in our daily life. "Ugly laws" are historically significant, but no longer exists to ban people with "offensive looks" from public places (Przybylo & Rodrigues 2018, 5). However, our popular culture does so nowadays by making fun of people's looks. In Yeidy Rivero's analysis of telenovela Yo soy Betty la fea, a Latin-American adapted version of the U.S. comedy Ugly Betty, the narrative "re-articulated colonial, gendered, class, and Eurocentric dominant discourses of female aesthetics" (Rivero 2003, 65). Why is Ugly Betty ugly after all? Is it because

she is non-white? Is it because she is a working-class woman? Is it because she is fat? One cannot give a clear-cut answer to the abovementioned questions, but it is worth noting that: the complex fabrication underlying a seemingly objective judgment of her looks is laid out to audiences almost instantly when they compare her with "beautiful women." People are nudged into believing that if you're like Ugly Betty, you should really change yourself.

Despite open mockery of women's looks, popular media nowadays tends to annihilate fat bodies. In Dina Giovanelli and Stephen Ostertag's analysis of "primetime programming on the major television networks" in the U.S. (Giovanelli and Ostertag 2009, 291), they shockingly found a lack of fat women on TV. There were only two fat women on television primetime programming in March and October 2005, and they were very often "devoid of any sexual and romantic desirability or interest" (Giovanelli and Ostertag 2009, 291). By being fat, these women are taken to be not feminine enough. By being fat, these women are taken to be not wanted by men. By being fat, a woman's only mission in life should be to lose weight. It is safe to say that our media is consolidating the significance of beauty and heterosexual attraction almost every day by means of deeming fat, ugly, non-white, poor people as unworthy of public attention or an object of mockery and judgment by a generalized public.

The Mythologized Beautification

"To be beautiful is to be successful", such a doctrine is internalized by modern people no matter how we define "success." Women in China have been keen on plastic surgeries in order to add value to their "sexual capitals." Beautification and building up one's body are seen as equivalent to climbing up the social ladder – a myth that is perpetuated by popular culture and media. The value added on beautification in our society is not natural: it is meant to justify the privilege that we give to beautiful people. After all, if it "takes effort" to be beautiful, then beauty can be naturalized as a form of capital that everyone is free to accumulate, despite unequal conditions to do so.

Beautification is a gendered practice that constrains femininity. We live in a society that commodifies and spreads people's anxiety over their bodies and looks, especially women's. In Wen Hua's studies of Chinese women who received plastic surgeries, she was astonished to see the increasing number of students as recipients of plastic surgeries, whose parents actually supported such decisions.

In her account, parents believed that: "A pretty face is a worthwhile long-term investment for my daughter's future" (Wen 2013, 77). It is worth noting that: to those parents, they didn't want their children to "live on their looks", but still recognized a great deal of value to the beautification of their children. Sexual capital is ranked inferior to other forms of capitals. The young girls described here in no way intended to commodify their bodies to the extent of a prostitute or mistress. However, they do believe that a beautiful face "eases things" for women in society. Good looks get you good jobs, good husbands, good social rankings, etc. And plastic surgeries, as if magical, would suddenly beautify a face to such extent. It is not only a blind belief in the technologies, but also a huge trust in the value of beautification. They invested so much in their daughters' looks because they are nearly certain that beautification would bring more positive outcomes in the future. And the truth is such a belief is justified: in China, job descriptions sometimes specify "'above-average looking', 'good-looking', with 'an elegant demeaner', 'height over 1.65m'" as their requirements (Wen 2013, 87-88). The outcome of beautification, though mythologized and exaggerated, is still largely positive.

This example illustrates to us how beautification has been made into a wholesome industry that conforms to patriarchy. In a prevalent advertisement slogan: "There is no ugly women. There are only lazy women." Women are encouraged to dislike their unaltered looks, treat themselves as an underdeveloped project, and beautify themselves through a variety of means. They are convinced that beauty is a useful tool and capital. They internalize the effect of beautification as possibilities of social mobility and pay for it. Such acts are acts of submission. Though we are thrown into a complex social fabrication and hierarchies of people that we do not and cannot choose, active beautification should be treated as an act of submission and conformity. Conformity on an individual basis is consolidation of significance of beauty in our society, because faced with discrimination based on looks, people choose to look better to diminish the negative effect instead of showing defiance or protests. This is exactly how social aesthetics gains its power over people.

Even though women seem to buy the narrative of magical beautification with their investments on their looks, men are more anxious with the connotations of social status of their bodies. Beautification to men is more or less related to building up an ideal, muscular body in order to claim a place in society. In Susan

Bordo's analysis, "well-muscled body has become a cultural icon; 'working out' is a glamorized and sexualized yuppie activity" (Bordo 1993, 195). If one fails to build up his body, then he is immediately judged to be less assertive and lack of discipline. This is a moral judgment triggered by aesthetic assessment.

In some ways, the aesthetic judgment of looks seeks to explain one's failure or success by means of laying emphasis on beautification. It is as if saying: "If you don't beautify yourself, then you are lesser. Just because you are lesser, you become such a loser at your job." We often overlook the socio-economic context of a person's success or failure, but in turn appeals to aesthetics to explain a person's situation. It is treated as if beauty is a capital that everyone can freely accumulate, but indeed we do not have the same conditions to accumulate it. Not only are we born differently with natural looks and genes, we are also born into different families and different situations. We are socially conditioned to have more or less leisure time and freedom to care for our bodies. However, in Susan Bordo's analysis of popular media portraying beautification: "The ability of the (working-class) heroine and hero to pare, prune, tighten, and master the body, operates as a clear symbol of successful upward aspiration, of the penetrability of class boundaries" (Bordo 1993, 195). It is treated as if an over-focus on body, despite the lack of conditions to do so, is a showcase of character which preconditions a person's future success.

I have explained how individual choices to beautify themselves contribute to the significance of social aesthetics in our society. From the examples, it should be clear that the beauty of looks is not the ultimate goal for many people. Beauty becomes a "means" or "signifier" of success, capital and social status. The hierarchy based on looks is meanwhile the hierarchy already present in society in general. We can easily conclude from here that beauty is no longer an intuition. Or even if it is something a priori, it is as a priori as other forms of subordination. Beauty, though claimed to be natural and objective, merely serves the interests of privileged minorities that benefit from its social capital.

"Beauty" as a Tool of Oppression

As I have explained, "beauty" has become the "means" or "signifier" of success, and therefore serves to perpetuate oppressions in our society. Following this line of thinking, I will explore the privilege people claim because of their "beautiful looks", and the discrimination based on "ugliness." From

such examples we can easily see the nature of "beauty" as an oppressing tool instead of an objective, harmless judgment. To illustrate the situation better, I want to use the term "interlocking oppression" (a term that was inspired by Robin James who used it as "broader interlocking systems of patriarchy, whiteness, and heteronormativity" 2013, 102), namely that the significance of beauty we described before as interwoven with racism, sexism and heteronormativity, colonialism, etc.

I will raise an example from Sharlene Nagy Hesse-Biber's The Cult of Thinness to explain the interlocking nature of oppression in the name of beauty. Hesse-Biber, in order to illustrate the "cult" nature of keeping thin in modern America, interviewed Delia, a 90-pound college cheerleader for her ideas on body control. Delia said to Hesse-Biber:

> Going in, I know I weighed like 93 or 94 pounds, which to me was this enormous hang-up, because I'd never weighed more than 90 pounds in my entire life…I knew people were going to be looking at me…So I would just not eat, work out all the time…and when we would do lifts he'd (my squad partner) always say, 'Delia, go run. Go run, you're too heavy.' I hadn't been eating that day. I had already run 7 or 8 miles and he told me to run again. And I was surrounded by girls who were all so concerned about their weight, and it was just really this horrible situation…When I was eight I wanted to be president of the United States. As I grew older and got to college I was like, 'Wow, it's hard for women.'… As I just figured, God, how much easier would it be for me to get married to somebody I know is going to make a lot of money and just be taken care of… (Hesse-Biber 2007, 13-14)

In Delia's case, the interlocking nature of beauty as an oppressing tool is quite clear. Delia, an already underweight college girl, fearing what people might say on her 3-4 pounds extra weight, "would just not eat, work out all the time" (as cited in Hesse-Biber 2007, 13). As Hesse-Biber has noticed: "Her greatest stress at college had nothing to do with academics." (Hesse-Biber 2007, 13) It is fair to judge from existing information that Delia would put all her energy in keeping thin and attractive and thus let go of her academics. And Delia openly said struggling in a job market is too hard and she would love to "get married to somebody I know is going to make a lot of money and just be taken care of" (as cited in Hesse-Biber

2007, 14). Having put almost all her energy in body control, Delia very likely loses her competitiveness as a worker but increases her sexual capital. Delia's beauty, having cost her energy and time, becomes an oppressing tool on her, because she has thus lost ability to earn her livings independently. Meanwhile, one may notice such a "choice" on body control could not be interpreted as her own freedom of choice: when Delia's squad partner could so easily pass judgment on Delia's body to cause her anxiety. It is safe to say: Delia is trapped by a patriarchal judgment on her body and, in order to fulfill the judgment, her choice of beautification costs her opportunities to economic independence. Just as Hesse-Biber has argued: "Women may choose to spend thousands of dollars on body-work instead of investing in a purchase with a higher economic, educational, or intellectual return. In fact, by obsessively perfecting their bodies, they also are buying into a set of patriarchal values that may make them ever more dependent on men for approval and success." (Hesse-Biber 2007, 22) This is the oppressing nature of beauty on women in a patriarchal society.

This is what I want to apply to the oppression related to looks. We can hardly argue that by virtue of being ugly, someone is denied basic rights openly and causally (nowadays, we don't have Ugly Laws). But given that ugliness is very often unwanted, and that beautification is a symbol of good character and success, contemporary people can use the word "beauty" or "aesthetics" as justification for their prejudice. After all, if ugliness is detested and objective, and beautification is mythologized as objective, what is the reason for not looking good or at least trying so? In the past, when the richest guy in a TV series managed to date a girl, we think it was because of the attraction of his wealth, but nowadays, if the same thing happens, we would be convinced that it is because of his personality or looks. Seldom do we reflect on the fact that: without a privileged socio-economic status, how could he work out so often? He would have had to work. Without a privileged family, how could he remain his innocence and kindness? He would have had to survive. His privilege is transformed and expressed fully by his looks, or the "aesthetics" we use innocently. When we grant him favor, we seldom notice that such a favor is not for our own appeal, but rather for his social status. And such privilege consolidates his advantage in society and transforms back as a form of social or economic capital. In this way, oppression based on looks is perpetuated.

We can view the interlocking oppression from another angle: the actual manifestation of fat oppression in the name of "aesthetic instincts." In Anne Eaton's

studies of "fat oppression" in contemporary society, she draws our attention to "the role of aesthetics in instituting and maintaining oppression" instead of the other way around (Eaton 2016, 38). It is very often heard that people who dislike and detest fatness or overweightness out of its "unhealthiness", as if we first make judgment on people's health status and then make aesthetic judgment that they are ugly because "we are displeased by the state of unhealth and its causes" (Eaton 2016, 44). But Anne Eaton would like us to think: given that there are a variety of beautification activities that imply health risks (e.g., over exposure to sun light, over-dieting, wearing corsets), why would people choose not to "diminish the aesthetic value of the outcome" (Eaton 2016, 46)? It is as if we already have two tracks of thinking when evaluating these projects. We are astonished by its beauty (such as a tanned body, a slim body, a small waist), and then after our reason comes back, we begin to question ourselves: is it worth it? Is it worth taking so many efforts to achieve beauty? Even at this stage, we seldom object to the notion of "beauty": we only question the means but not the ends. But when it comes to fat people, we immediately think: "Losing weight is the best thing to do. It is for the health and for our eyes." Seldom do we reflect on which is more primary and fundamental. Such implicit judgment changes our perception of fat people. We refuse to make adjustments on facilities to cater for their needs but request them to fit in the facilities we have. We refuse to present more plus-size models but react with anger to anyone overweight on screen and call him/her "promoting overweightness." Just as Eaton has argued: "This notion of collective taste is meant to acknowledge the important fact that some ethnic, racial, and sexual minorities, as well as some individuals, do not adhere to the dominant aesthetic, and that this is sometimes on purpose as a result of having cultivated strategies of resistance" (Eaton 2016, 41).

In the Second Part of my thesis, I have clearly illustrated the social forces behind "aesthetics" in contemporary societies. The distaste of fat bodies, the mockery of non-white Ugly Betty as well as the justified privilege we give to beautiful people are all interwoven with each other. "Beauty" is no longer an innocent aesthetic term in such contexts –it becomes a tool that oppresses the ugly and justifies the privilege that the beautiful enjoy. I have successfully developed that: we are not equally capable of being beautiful, and I thus refuted the third layer of objectivity of beauty.

THE THIRD PART: THE ARBITRARINESS OF BEAUTY WITH WIDE SCOPE: HISTORICAL, RACIAL AND POLITICAL FACTORS

After spelling out the nature of "beauty" or "aesthetics" of people's bodies and looks as a tool of oppression, I would like to draw from some real-world examples to show the outcomes of these mechanisms. The arbitrariness of aesthetics is apparent by noting the constantly changing standards of beauty. The standard of beauty changes within the scope of history, cultures and values. I explain the racialized aesthetics of a hypothetical Asian woman Jenny and of Jari Jonnes during the Black Lives Matter movement. Such examples help illustrate the fact that "beauty" is not a stable set of standards throughout history and among cultures. The nature of aesthetics as a tool of oppression could not be more obvious with the Asian woman Jenny and Jari Jones here: when people's race is interwoven with their status in the hierarchy of beauty, as well as the preferred traits of their looks, one can hardly justify aesthetics as objective: aesthetics is in itself racialized or even racist.

Aesthetics as Racialized Identities: a hypothetical case of "Mulan"

Aesthetics of bodies vary from country to country, but such divisions are never superficially a "cultural difference." In Monique Roelofs's account, "racialized aesthetic nationalism" is "a position comprising modes of address and embodiment that recruit aesthetic forms, meanings, and experiences in the service of racialized, nationalist conceptions and attitudes and in turn, bear the imprint of these understandings" (Roelofs 2014, 151). In a sense, a lot of racialized misconceptions and prejudices are perpetuated by a specific taste from certain cultures because "racialized aesthetic nationalism" "concerns a privileging of the culture that is ascribed to a given nation or ethnic group over cultures that are attributed to other nations or ethnic groups" (Roelofs 2014, 152). As the example of Ugly Betty shows, aesthetics could work as a mechanism to ensure the superiority of whites in post-colonial Latin-America. However, the subjectivity is more of a consideration to Roelofs: "Racialized aesthetic nationalism is a matter of embodied, relational existence." (Roelofs 2014, 152)

I want to raise an example as a hypothetical woman named Jenny as an example of racialization of aesthetics. She travelled to America from China.

Her name is Jenny. She comes from China and suffers hugely because of her refined jawlines, small eyes, thick lips and tanned skin.

People in China denounce the looks of Disney Mulan as "countryside women" and Jenny resembles greatly Disney's Mulan. This is a nationalistic aspect of Chinese aesthetics of women's looks: the detest of Mulan is a way to counter Hollywood's mode of understanding Chinese beauty.

However, after she travelled to the U.S., people appreciated her "countryside looks" and called her "Asian Princess" or "Mulan." People did that of course out of pure kindness and as compliments, but however, there were non-Asian men with an unhealthy fetishism of Asian women among the group.

There is a particular racialized aspect to their taste in the way Jenny looks, because Americans never appreciate the same traits on white people. Just as how Roelofs spells out: "…the disposition to seek out preferred qualities such as purity, order, or formal novelty in artworks by white Europeans and Anglo-Americans, coupled with the tendency to read for devalued or ambivalently valorized features such as embodied rhythms, sexualized passions, and fantastic contents in the works of black diasporic, Latin American, and Latino/a artists" (Roelofs 2014, 152). Similar to the art world in which people look for "devalued" traits of white artists from black artists, the usually detested traits of a face, after transferred to an Asian face, seem ideal. The otherwise unfavored small eyes are appreciated by white viewers on an Asian face. The racialization of her looks is deeply embedded in a post-colonial, racist society in which her less preferred looks contribute to the manifestation of white superiority.

Jenny liked being favored in such a society. With huge pride, she began posting selfies on social media, signifying her boosted confidence. It all went well until she saw Beth, a chubby Chinese girl, promoting "Body Positive."

> She began to body shame Beth because she was certain that
> herself, as an "Asian Princess", was preferred, whereas Beth
> would not be. She was so proud of being deemed "beautiful"
> that she called other people unlike her, "ugly."

Here we see how racialized aesthetics becomes part of the identity of Jenny. After benefiting from a certain hierarchy of aesthetics, Jenny chose to perpetuate the system in which she benefited most and contributed to subordinating other people. That is because: "Racialized aesthetic nationalism represents powerful currents of aesthetic desire and passion at the heart of individual's daily aesthetic life worlds." (Roelofs 2014, 153) Jenny would not support "Body Positive" even though she benefited from a tolerant, diverse society that appreciate her, only because she was proud of her own body shape. Her aesthetic life world is shaped by such applause that she could not see things beyond it. After witnessing a gap between different systems of aesthetics, Jenny did not choose to disbelieve or deconstruct them but rather chose to consolidate one of the systems. The racialized taste of her looks becomes part of her identity: she saw herself merely as an "Asian Princess" from a white American man's point of view, instead of from the point of view of a fellow Chinese immigrant woman in relation to Beth. The aesthetics system of her surroundings changes her self-positioning in the society.

From the example, we can see clearly how aesthetics manifests itself arbitrarily in a racist society. The different standards of beauty across cultures (Chinese culture and American culture) could be explained by cultural relativism or pluralism, but different standards of beauty on white people and Asian people within a single culture could not be explained easily. The racialized aspect of aesthetics manifests in a racist way: the oppressed racial group in society are assigned with less favored traits. Even while foregrounding Asians and appreciating Jenny's looks, the overall aesthetic environment is still deeply discriminatory to Asians. Social aesthetics as such contributes to the perpetuation of racism, because the beauty of a racial group is defined and appreciated by a presupposed white audience with a "top-down" attitude. Americans' appreciation of an ugly woman by Chinese standards naturalizes racial hierarchy: after all, if "beautiful Asians" look similar to "ugly whites", then it is totally justifiable that "whites are naturally superior."

Meanwhile, the arbitrariness of beauty manifests subjectively and contributes to the oppression of beauty. "Racialized aesthetic nationalism is grounded in everyday aesthetic patterns of meaning making and experience we enact as we

conduct relationships with one another and the environment." (Roelofs 2014, 156) Jenny's own privilege gained by a "Mulan look" is justified by herself in her own meaning making process, and then she turned a blind eye to the fetishism and embedded discrimination of Asian women who are her fellows. The so-called beauty we gained in a racist, sexist, heteronormative society corrupts our subjectivity with the abovementioned values, just as how Jenny distancing herself from her fellow Chinese women is mainly because of her higher position in a racist patriarchy.

From this illustration it is clear that beauty is not a fixed set of standards among cultures. The predominant social groups tend to attach higher values to the aesthetic features of their own culture, and lower values to other cultures. In terms of social aesthetics, racialization contributes to a major part of aestheticization of Asian women. The "lesser" traits such as small eyes are attributed to the "lesser" social groups. Besides the innate hierarchy, the "racialized aesthetic nationalism" in this process becomes an existential situation of people involved.

Is Jari Jones Ugly or Beautiful? a case of "politicized aesthetics"

During the Black Lives Matter movement in 2020, Calvin Klein chose Jari Jones, a black transgender plus-sized model to appear on a massive billboard in Manhattan. Such an act is by all means political, but the act went from "a politicized commercial activity" to "politicized aesthetics." This section will draw from the history of imagery representation of African Americans to explain the importance of images to minorities. Meanwhile, I will incorporate debates about the aesthetic judgments of Jari Jones in Chinese social media and illustrate the nature of aesthetics as politicized.

Throughout the history of African Americans, the image of them being a fellow human occurred relatively late. In Celeste-Marie Bernier's analysis of Frederick Douglass, a famous African American abolitionist living in the 19th century, she found out that the representation of runaway slaves up until 1863 was still circulating "an object of bleeding, suffering humanity" (Bernier 2012, 288). Douglass, being an African American himself, challenged such image with his own appearances as a cultural elite and intellectual. Meanwhile, he appraised photography to a great extent because "pictures enable us to see ourselves as if from the outside and, from this more distanced view, to contemplate and access ourselves, drawing up plans for improvement" (Wallace and Smith 2012, 7). Such

objectification of photography to Douglass is not a bad thing, because instead of white viewers, he "imagined a much more autonomous African American viewer… African Americans are the primary and most important viewers of their own images" (Wallace and Smith 2012, 8).

Jari Jones' occupation of a Manhattan billboard was a heated topic in China, because it brought up the debate of how beauty is "sacrificed" to politics. Given the history of such representation of black people, it made perfect sense for people to be surprised when Jari Jones occupied the massive billboard in Manhattan: after many long years of seeing representations of bleeding objects, now a black model can righteously claim a place in the middle of New York. It makes full sense when we see such pictures as progress. However, with Douglass' presupposition of a self-reflective black viewer of their own images, he somehow neglected the polysemic nature of photography. When Jari Jones was met with Chinese viewers, what we saw was commentary such as the below: "'Beautiful is beautiful, weird is weird,' a male user commented on the Twitter-like Weibo. 'I am not going to buy CK anymore. This kind of model makes me want to vomit.'" (Zhou 2020) Such outspoken misogynistic views do not dominate Chinese media. The debate about whether "political correctness" went too far was also a voice: a debate about whether Calvin Klein "sacrificed" the standard of beauty for political purposes ensued. However, it is worth noting that, the entire debate hinges on one big assumption: "Beautiful is beautiful, ugly is ugly" and, "Jari Jones is ugly."

Just as I've argued previously, aesthetics is almost always subjected to political factors. Hiring Jari Jones certainly amounts to a strong level of outspoken support for the Black Lives Matter and LGBTQ Pride movements, but the silent endorsement of underweight and white models almost always serves the interest of dominant sexist, racist ideologies. Very few people with a firm aesthetic judgment on Jari Jones' body have displayed enough self-reflective thought towards their own default judgments of beauty. And just as I argued about fat oppression: such judgments of "beauty" contribute to oppression.

In this section, I have examined the aestheticization of women of color in order to prove the racialized nature of aesthetics. It is quite clear that "beauty", though being conceived of as "objective" when Jari Jones is denounced as ugly, is to the contrary and to a great extent subject to historical, racial and cultural factors. The understanding of the objectivity of beauty as a fixed set of natural standards throughout history and across cultures is thus totally misguided.

THE FOURTH PART: DOES BEAUTY EXIST "ELSEWHERE"?

The abovementioned examples illustrate a clear picture of the arbitrariness of beauty as constituting and being constituted by racism and sexism in society. Beauty has been corrupted as a tool of oppression and is interwoven with other forms of oppression. I am whole-heartedly against the use of the word "objective" when it comes to discussion of beauty. As I have said, by "objective" is implied a natural, stable set of beauty standards that is constant despite social change. Beauty is clearly not constant through social change as I hope the examples above have illustrated. The concept of "objectivity" neutralizes oppression based on looks, which I hope was revealed by the examples I considered in Part 2. However, even despite these examples, an objection could be raised based on a realist view of aesthetics: the so-called beauty we have gone after throughout human history, though may not reveal beauty to us now, may show us the way to "real", "essential" or "true" beauty later. The disputes of the content of "beauty" does not make it less objective. Beauty still exists but exists "elsewhere" as a mind-independent entity waiting for us to pursue. In this section, I will argue that: aesthetic realists, by overly focusing on what "beauty" is, overlook what it does and has done in human societies.

Aesthetic Realism and its objections

In this section, I will take issue with aesthetic realism in the following way: firstly, I want to refute the idea of abolishing the significance of viewers in aesthetic evaluations; secondly, I will argue against the authority of "beauty" as a concept granted by aesthetic realists, even without a clear definition of it.

Aesthetic realists would argue that beauty as a mind-independent entity is irrelevant to disputes of people's opinions and about how it is that people apply the term or concept. To a realist, beauty exists despite my viewing in a "mind-independent manner" (Joyce 2016). For example, Jenny seems to uphold a realist ideal of beauty in which she stands high up in the hierarchy. To her, her own look as approved by her surroundings brutally equals beauty, while Beth's does not. It is by all means the audiences, as in herself and her surroundings, that creates this aesthetic hierarchy, but her belief would be that such hierarchy is the "right one", "a priori" or "a fact." As to the different standards of beauty in China and America, Jenny may agree to how Inês Morais has argued: "aesthetic disagreements need

not imply an error or misapprehension on the part of any of the disputants…
but only a difference in sensibility towards the same features" (Morais 2019, 87).
To Morais, "the same features" matter more than "different sensibilities" of the
spectators. She builds her aesthetic realism on a premise that what we talk about
beauty is solely about "features" instead of "subjective preferences" (Morais
2019, 88). However, one may notice how the same traits, such as small eyes, are
perceived differently on an Asian face and a white face as I pointed out before.
The seemingly objective discussion of "features" does intertwine with subjective
racism. The external audience and judgment that constitutes the arbitrariness of
beauty in our society to her is just ignored. The implicit racist aesthetic judgment
to her becomes a mere "dispute of opinions." In general, realists seem to believe
in "taste", "a capacity that permits good judgments about art and the beauties of
nature" (Korsmeyer and Weiser 2021).

Secondly, aesthetic realists argue that beauty does exist, only "elsewhere" as
beyond human knowledge. Inês Morais argues that: "(the disputes of what beauty
is) …need not count against there being a genuine, non-mental, reality that we
can know even though only in part… In effect, the limits in our knowledge of an
area should not lead us to claim that there is not an independent reality." (Morais
2019, 88) One cannot ignore the similarity of this account to a religious one,
when believers argue that their pursuit to God or supreme beings though not yet
succeed, certainly will, because God as a mind-independent entity is waiting for
their discovery, even after thousands of years of religious disputes and warfare. I
will not take issue with the absolute existence of beauty beyond human intellectual
reach here, just as I will not take issue with disputes about whether God exists.
The only question here is the authority of such concepts and what they have done
to human society, as a matter of fact. I have illustrated the historical change of
aesthetics of African Americans based on Fredrick Douglass' and Jari Jones' avant-
garde representation of their group. Just think of their contemporaries, think of the
hegemony of corset-wearing European ladies and underweight Victoria's Secret
models, and think of the disputes Douglass and Jari Jones underwent when they
subverted the "standards" of beauty. I do agree that their acts enrich the notion
of "beauty" and expand our knowledge of aesthetics, but is it worth it? Do we
have to be dominated by a racist, sexist, heteronormative standard of beauty for
so long until we rebel against it only to "expand our knowledge"? Do we really
want to call such historical change, together with oppressions in the history (such

as feet-binding and corsets) part of a pursuit of "true, real or essential beauty" as "a genuine, non-mental, reality that we can know even though in part" (Morais 2019, 88)? The oppressing nature of beauty in our society comes from an implicit notion of its objectivity, just as how churches do wrong things in the name of God. No matter whether God or beauty exists out of human reach, there is full reason why we should be cautious against the claim of their authority as implied by its "mind-independent objectivity."

All in all, I have pointed out a fundamental error of aesthetic realists: even with an absolute concept of "beauty" in mind, people may form erroneous judgments because of their own racism, sexism and other negative social factors. Following such an argument, it is very erroneous to grant authority to the "absolute beauty" because human activities in this pursuit may cause many harm (such as feet-binding, wearing corsets, over-dieting). The aesthetic realists have undermined the effect of aesthetic disputes, and thus should be taken issue with.

Counterforce or new hegemony? Beauty in the Body Positive Movement

"Beauty exists elsewhere", such a statement could be understood as a realist appeal to unreachable beauty or an intention to create new hegemony of beauty. One example I will raise is the appreciation of plus-sized models during Body Positive Movement. If the traditional white, underweight Victoria Secret's models are an embodiment of racism, sexism and heteronormativity, then will someone just the opposite come to the rescue? If our subjectivity is corrupted by sexism, racism and heteronormativity, then will more aesthetic experience with underprivileged groups such as those represented by the black, transgender, plus-sized model Jari Jones help? By presenting Jari Jones in Manhattan, is Calvin Klein trying to create a new hegemony of beauty? Will the counterforce of such media representation actually help deconstruct the presumed objectivity of beauty or vice versa?

By casting Jari Jones' image as "promoting overweightness", many people interpret her possession of attention as a pedagogy of beauty in a Body Positive era. Helana Darwin and Amara Miller categorize the Body Positive movement as further divided into "Mainstream Body Positivity, Fat Positivity = Body Positivity, Radical Body Positivity, and Body Neutrality" (Darwin and Miller 2020, 1). While Mainstream Body Positivity activates "self-love" in a generalized postfeminist language (Darwin and Miller 2020, 7) and Body Neutrality activates "body

acceptance" as a mental state (Darwin and Miller 2020, 11), Fat Positivity activists "advocate for a focus on the systematic discrimination that fat women experience instead of the body image issues that women experience more generally" (Darwin and Miller 2020, 8). Such attention to overweightness is very easily understood as "promoting fatness" and overlooking women with BMI under 30. It is safe to say that Jari Jones' positive stare is a footnote to this movement: a movement with an exclusive highlight on plus-sized women in a positive, confident manner. One of the most famous mainstream Fat Positive models is Ashley Graham. But the difference between Jari Jones and Graham is obvious: Jari's appearance is also deeply wedded with Radical Body Positivity, a movement in which all remaining judgments on bodies are removed. Hairless, white, cisgender plus size models with hourglass bodies such as Graham were replaced by people such as Jari Jones: black, transgender, "totally out of shape." Even among body positive activists, she is the "ugly one." As I have said, such representation upsets the objectivity of beauty from the root: her daring look symbolizes a returned gaze by what has been perceived as aesthetic objects. But the issue still remains: the judgment of her body will not cease because of her subversion of aesthetic hierarchy. She was considered "brutally ugly" by some, "an expanded version of beauty" by others, but her media presence is nevertheless judged in terms of some "objective" notion of beauty. The pedagogy of her existence does not come from presenting her as beauty itself, but rather comes from the basic fact that she is an aesthetic object: how could an object subvert the aesthetic system it lives in? Her look is, thus, just perceived as "pure ugliness", or "sacrificing true beauty to political correctness."

Jari Jones' media presentation is not only objectification, but also conformity to a new hegemony in the Body Positive era. As I've said earlier, the issue of body image is very prevalent in modern societies, but Fat Positive advocates seem not to correctly consider the roots of body hatred and alienation. By presenting overweight models in a confident manner in contrast to a presupposed status of low self-esteem, people are generally convinced that the real issue of body acceptance lies in the gap between mainstream hegemony and bodily realities. If the unhappiness of fat women comes from their deviance from Victoria's Secret, then what do thin women feel bad for? If people like Jari Jones, whom is totally unfavored by our society, could embrace herself, then why couldn't I? Whereas in reality, there are endless cases of woman with BMI between 18 to

30 cast themselves in a negative light. The hegemony of such images does not teach people "what beauty is", it rather teaches people "an ideal attitude toward mainstream aesthetics." In a traditional narrative, the central change of an obese body to a normal body lies in not only the change of physicality, but also curation of emotional pains, whereas in the Body Positive Movement, by presenting a pedagogy of the right attitude toward one's body, it effectively "position[s] the central lack as body positivity itself…the body must undergo the ritualized public discourse and subsequent catharsis that body positive websites facilitate" (Sastre 2014, 938). Such emphasis on personal change and agency inevitably falls into the bigger frame of neoliberalism's "rhetoric of choice-driven, bodily-oriented self-improvement" (Sastre 2014, 932). Only this time, the "choice" and "improvement" is not to adhere to the hegemonic body, but to adhere to the hegemonic attitude toward one's body.

From this example we can clearly see how Body Positivity, a seeming challenge to mainstream social aesthetics, is nonetheless problematic. The true issue still lies in the authority and objectivity we grant on the notion of "beauty", despite a proposal to change its content.

CONCLUSION

From my argument, it is made clear three things: firstly, perceiving beauty as a set of objective standards is erroneous, because these standards are constructed by a racist, sexist, patriarchal society (such as WHR in science, appreciation of small eyes on Asian faces); secondly, the standards of beauty have changed so much throughout human history and across cultures that we should instead of pursuing it, question the authority and objectivity we grant to claims and judgments of beauty; thirdly, people are not in equal situations to adhere to beauty standards in their own societies, given for example, different socio-economic status. From such arguments, we can conclude that the idea of beauty as objective is deeply questionable. As a way to think beyond the frame, I also raise the case of the Body Positive movement, to illustrate that many efforts to abolish aesthetic standards have fallen into the trap of neoliberalism in contemporary societies. It is left unknown what should be embraced: a society in which the traditional ugly is considered beautiful, or a society in which ugliness and beauty does not exist or matter?

REFERENCES

Zangwill, Nick, "Aesthetic Judgment", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/aesthetic-judgment/>.

Bernier, Celeste-Marie. 2012. "A 'Typical Negro' or a 'Work of Art'? The 'Inner' via the 'Outer Man' in Frederick Douglass's Manuscripts and Daguerreotypes." *Slavery & Abolition* 33, no. 2: 287–303.

Bordo, Susan. 1993. *Unbearable Weight: Feminism, Western Culture and the Body*. Berkeley: University of California Press.

Bovet, Jeanne. 2019. "Evolutionary Theories and Men's Preferences for Women's Waist-to-Hip Ratio: Which Hypotheses Remain? A Systematic Review." *Frontiers in Psychology* 10: 1221.

Blake, Khandis R, and Brooks, Robert C. 2018. "High Mate Value Men Become More Accepting of Intimate Partner Abuse When Primed With Gender Equality." *Frontiers in Sociology* 3: 28.

Costelloe, Timothy M. 2007. Aesthetics and Morals in the Philosophy of David Hume. Routledge.

Clough, Sharyn. 2003. *Beyond Epistemology: a Pragmatist Approach to Feminist Science Studies.* Rowman & Littlefield Publishers.

Darwin, Helana, and Miller, Amara. 2020. "Factions, Frames, and Postfeminism(s) in the Body Positive Movement." *Feminist Media Studies*: 1–18. DOI: 10.1080/14680777.2020.1736118

Eaton, A. W. 2016. "Taste in Bodies and Fat Oppression." In *Body Aesthetics*, edited by Sherri Irvin, 37-59, Oxford: Oxford University Press.

Edwards, Louise P. 1994. *Men and Women in Qing China: Gender in the Red Chamber Dream*. E.J. Brill.

Furth, Charlotte. 1999. *A Flourishing Yin [Electronic Resource]: Gender in China's Medical History, 960-1665.* EBSCO Publishing: eBook Collection.

Korsmeyer, Carolyn and Peg Brand Weiser, "Feminist Aesthetics", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.),

URL = <https://plato.stanford.edu/archives/spr2021/entries/feminism-aesthetics/>.

Giovanelli, Dina and Ostertag, Stephen. 2009. "Controlling the Body: Media Representations, Body Size, and Self-Discipline." *The Fat Studies Reader*, edited by Esther Rothblum and Sondra Solovay, 289-296. New York: NYU Press.

Gracyk, Theodore, "Hume's Aesthetics", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/hume-aesthetics/>.

Hesse-Biber, Sharlene Nagy. 2007. *The Cult of Thinness*. 2nd ed., New York, Oxford: Oxford University Press.

Hume, David. 2006. "Of the Standard of Taste." California State University, Long Beach. https://web.csulb.edu/~jvancamp/361r15.html

James, Robin. 2013. "Oppression, Privilege, & Aesthetics: The Use of the Aesthetic in Theories of Race, Gender, and Sexuality, and the Role of Race, Gender, and Sexuality in Philosophical Aesthetics." *Philosophy Compass* 8, no. 2: 101–116.

Klinger, Cornelia. 2017. "Aesthetics." *A Companion to Feminist Philosophy*, edited by Alison M. Jagger and Iris Marion Young: 341–352. Oxford: Blackwell Publishing Ltd.

Joyce, Richard, "Moral Anti-Realism", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/moral-anti-realism/>.

Morais, Inês. 2019. "Arguments for Aesthetic Realism." In *Aesthetic Realism*, 85-103. Cham: Springer International Publishing.

Przybylo, Ela, and Rodrigues, Sara. 2018. *On the Politics of Ugliness*. Palgrave Macmillan.

Rivero, Yeidy M. 2003. "The Performance and Reception of Televisual 'Ugliness' in Yo Soy Betty La Fea." *Feminist Media Studies* 3, no. 1: 65–81.

Roelofs, Monique. 2014. *The Cultural Promise of the Aesthetic*. Great Britain: Boomsbury.

Radway, Janice. 2009. *Reading the Romance.* 2nd ed., The University of North Carolina Press.

Sastre, Alexandra. 2014. "Towards a Radical Body Positive." *Feminist Media Studies* 14, no. 6: 929–943.

Wen, Hua. 2013. *Buying Beauty: Cosmetic Surgery in China.* Hong Kong: Hong Kong University Press.

Wallace, Maurice O, and Smith, Shawn Michelle. 2012. *Pictures and Progress: Early Photography and the Making of African American Identity.* Duke University Press.

Zhou, Viola. 2020. "China Trends: Black Trans Model Triggers Debate and a Student's Offensive Tweets." *Inkstone*, 30 June 2020, www.inkstonenews. com/society/china-trends-black-trans-model-triggers-debate-and-students-offensive-tweets/article/3091009.

ZHANG, Xinjun, and LIU, Qizhi. 2015. "Aestheticization of Illness in A Dream of The Red Mansions." *Comparative Literature--East & West* 23, no. 1: 57–70.

compos mentis

Confirmation Biases and Emotional Probabilities

Derek Minton

Central Michigan University

ABSTRACT

This paper focuses on the idea of belief forming, and how it relates to subjective interpretations of evidence regarding everyday theories. Hypotheses with an emotional attachment may undermine our ability to recognize confirmation biases within closely held beliefs. Should we allow ourselves to do so, the subjective origin of our beliefs can be acknowledged to recognize confirmation biases within certain settings. To point this out is to acknowledge we possess the casual explanation for believing what we believe to be true based upon what is best available to us. To adjust this, we must understand the root causes of our own subjective biases. We will work around this issue by adhering to other probable theories first, before asserting that our beliefs are the best available. To do this, I will first introduce Kahneman's (2011) distinction between system types, followed by a psychological experiment involving teaching methods (Allen and Coole 2012), and then discuss the issue between Feldman and Cohen (2001) regarding context sensitivity. The conclusion of this paper will be in favor of searching for more probable theories before we assert our beliefs, thus, eliminating the need of a casual explanation for the availability of belief forming. Thereby recognizing that emotional attachments have a direct impact on confirmation biases within our subjective origin. Being made aware of this will provide more truthlike theories to be available than what we previously had access to.

I. INTRODUCTION

When one finds themselves deciding between rival theories regarding beliefs, we assume that emotion played a role in the acceptance of that belief. In most cases, strongly held beliefs are closely related and justified under our emotional attachments. Although our beliefs are not always accurate, they are reliable enough to confirm a subjective opinion to hold. Scientific researchers and a person that has, we will say, "everyday knowledge," comparatively reason in similar modes of theory testing. They differ in the resources and methods available to know if their theory is best fit to hold. It is unlikely a person with everyday knowledge will have the necessary means to test their theories in the same methodical way a scientific researcher would. In most cases, a person with everyday knowledge does not have access to the resources or knowledge available to claim their theory is the best available. As one with everyday knowledge would easily admit they are not an expert in a certain field, and pass this off as not knowing something. In a scientific environment, whether the hypothesis is of a subjective or objective origin, we believe the scientist who created the hypothesis did their best to not use their own bias to formulate the theory.[1] It may be argued acceptance and formation of a hypothesis both adhere to some type of bias, but a well formulated scientific hypothesis must face the challenge of crucial tests to prove its fitness. Everyday beliefs do not rest upon being critically analyzed by a scientific researcher but can still play a role in being socially accepted within a group of peers. Since scientific research is best left up to those in the field of theory testing, I will be focusing on everyday knowledge, opposed to scientific theory in this paper, as the framework concerning confirmation biases.

II. CONFIRMATION BIASES

Confirmation bias is the tendency to interpret new evidence as confirmation of one's currently existing beliefs. If one is not previously aware of a better available theory, then one is more likely to believe their theory is the best available. Personal beliefs can disrupt the ability to be open to other possible solutions should new evidence become available. As it can be understood, proof of new evidence that

---

1. We should assume that scientists are able to look at all the best available theories, because they are more readily available and knowledgeable when one goes into a scientific field.

forces one to alter their prior beliefs can be disorienting. Reacting towards this new information may be met with criticism and strong feelings of emotion, as opinions can react before adhering to other reasonable possibilities. For everyday problems, our beliefs are typically strongly self-biased. This can lead one into forming a confirmation bias towards certain kinds of evidence. Thereby limiting the types of opinion one can hold for that matter.

Confirmation bias may interfere with reasoning by not giving another person merit to their point of view, should one feel their beliefs are being threatened. These biases can range from age, sex, race, political party, religion, or any previously held stereotypes towards a group of people. We can acknowledge that stereotypes exist, as knowledge of them existing does not make a person who knows this one who stereotypes, because one understands that it is not always the case that a belief is biased due to a stereotype being present. This simply provides an additional level of understanding.[2] There is a tendency in this self-reliant mode of reasoning to internalize the hypothesis that our beliefs are the best we have available. While this can produce a boost of confidence in our assertions, if we instead consider more probable theories first, or at least entertain the idea of there being more, the dependency to rely upon our self-reliant bias decreases, thus, increasing a potential for more truthlike beliefs. I want to show that by searching for more probable theories first, i.e. not counting on your theory to be the best available, we can eliminate this type of self-reliant confirmation bias. Based on the grounds that we know there are more truthlike theories available than what we have access to in our limited everyday situations.

Our own beliefs may be true in some cases, along with the further belief of how often we believe we get our beliefs correct, but this is the same type of confirmation bias that is reliant upon the tendency to accept that our beliefs are the best available (Fitelson 2012, 85-88). Thereby affirming that for any belief to be truthlike it must first go through the process of our own subjective origin.[3] This notion falls under the additional tendency to rely upon the belief that we possess the ability to spot the best available theory. To claim this would be to accept that our own beliefs are the best we have available, because we are not

---

2. To even mention the word "stereotype" makes it near impossible not to talk about the kinds of stereotyping, but this issue of simply mentioning a word will be looked at more in depth later.

3. It is hard to determine how each person does this, as it is likely we all possess different limits and criteria for our subjective origin.

in the position of another point of view. There is no way to disprove a hypothesis such as this that takes the position of affirming its own superiority. The problem we now face is finding supporting evidence that a belief is solely dependent on misguiding our emotions to push off any rational explanations anyone else may have. I will show throughout the rest of this paper how confirmation biases that have a strong attachment to well-guarded beliefs only need to be emotionally triggered to confirm a hypothesis. Thereby overriding any previous subjective possibilities, one may have.

III. THE TWO SYSTEMS

In this next section I will focus on the different systems of thinking Kahneman (2011) provides concerning the formation and acceptance of a belief. If we can show that objective probabilities can be confirmed without the reliance upon a strongly held belief, then confirmation biases are simply the product of our subjective emotional responses.[4] If we adhere to other probable beliefs first, while understanding a hypothesis may be emotionally charged, we can better align our beliefs towards more truthlike hypotheses. Which should be the goal we are trying to achieve if one is a proponent of believing more truthlike theories. The contrary appeal to beliefs before subjective probabilities allows a hypothesis with a strong emotional attachment to overpower any rational theory. The question we should now ask is this: Why do we have this tendency to affirm an emotionally superior hypothesis even when we know it is self-reliant upon its emotional attachment? It is not uncommon to say "human nature" provides an adequate explanation, but I believe Kahneman (2011) provides us with a better way of explaining why this occurs. He separates the mind as two different systems of thinking: as system 1 and system 2. He describes system 1 as being responsible for quicker, more emotional responses, while system 2 is slower and more reliant upon logical interpretations. We can use this as a base to show that reasons to hold certain beliefs can be doubted, as Descartes may add, even "my own mind can be doubted." However, I would argue this kind of theory Descartes is known for is prone to inflict a certain emotional response on the individual. If the idea is provoked under the guidance of system 1, then a belief (for example) such as nihilism carries with it an emotional

---

4. It may be the case that an objective probability like winning the lottery has a subjective pull on the individual's tendency to gamble, but this does not change the odds of the lottery.

attachment that insinuates one must give up all beliefs. A proponent of system 2 may be more skeptical of the origins surrounding its appeal and reasons for accepting a belief such as nihilism. This is a problem, because when we have a well-formed hypothesis, once we accept it, the willingness to change theories decreases under both systems. If the willingness to change theories decreases enough into a firmly held belief, this may fall into the trap of confirmation biases. If there is any correlation to be found between accepting a hypothesis and strongly elicited emotions, the potential for an ulterior motive increases in these cases, as any new evidence will support both rival theories to be more truthlike under the scope of their confirmation bias. Simply hearing the word, "nihilism" may inflict a quicker response time for a person dependent on system 1 as guiding their responses, because emotion plays an important role in this mode of belief forming. If an emotionally charged theory succeeds in its emotional drive, its effect then becomes a matter of reasoning in the following way: by confirming the most likely theory to be truthlike, it acts as a certainty that the most probable theories will turn out to be true. By acknowledging the implications certain beliefs entail, we can use system 2 as a conductor to these types of theories before adding them to our well-founded beliefs (Kahneman 2011, 13-30).

IV. REWARD AND PUNISHMENT RELATING TO BELIEF

It now seems all we would have to do is look specifically at beliefs eliciting stronger emotions, under the guidance of system 1 or 2, and conclude them to be biased. When we are under the influence of either system, in another part of his book, Kahneman provides a reasonable explanation for our tendency to go along with the stronger emotional belief (Kahneman 2011, 176-183). In this segment, he provides an example from his own personal experience, relating this to a time when he was giving a speech in front of a group of flight instructors at an Air Force base. Kahneman claimed rewards for good performance work better than punishments for bad behavior. This was argued by one of the instructors at the base, contesting the opposite of what he was saying was true. For the instructor, punishment seemed to work better than reward, citing his example of screaming into his cadet's ear made them perform better. The instructor's evidence was that most good performances were typically followed by a lackluster job, which made him further believe rewards to be less effective than punishment. Kahneman

pointed out to the flight instructor that his observation was correct, but his inference was off. The point being made here by Kahneman was that punishment or reward did not have an effect determining the outcome of the result. He demonstrated his explanation to the other cadets using the following method. He drew a target on the floor, and then asked the group of cadets to throw a coin at it. Once this was done, they wrote down their names in rank order on a chalkboard of who was closest to the target. When performing a second experiment, those who were far off the mark the first time tended to do better, and those who were closest to the target tended to do worse. Kahneman explained, the names on the board did not cause the outcome of the result, it simply was an observation of regression towards the mean. The instructor may have held a bias that his cadets were lazy or lacked the ability to perform without the use of punishment or reward hanging over them. As the instructor believed that without punishment or reward, the cadets would not perform to their best ability. This belief the instructor held is more prone to be self-reliant upon itself and to adhere to beliefs before any other subjective possibilities. In this example, punishment typically has a stronger effect on the individual's psyche because it acts as a weapon against the commonly held belief that nobody likes being humiliated. Reward can give a positive influence, but failure to receive reward, I assume is not punishment in all instances. To avoid punishment in the situation the cadets were in was to some degree a reward. One does not know if a cadet is performing to the best of their ability unless they were the cadet themselves, which cannot be proven by observation alone. This reliance upon observational evidence supports any belief one may hold should it be fitting. As seen in this example, it is much easier to identify the mistakes of others than oneself.

Within the human psyche, it may be a physical impossibility to determine what exactly causes these biases, besides acknowledging them as our opinions and the actions we consider to be our beliefs. We can infer certain biases if we see them, but it is difficult to trace these origins by observational evidence alone. For example, a doctor typically believes their patient's word as true without question. Both the patient and the doctor's goals align with maintaining good health of the subject. A doctor is not encouraged to lie to their patient about medical practices because they understand the consequences of doing so may lead to their job termination. In terms of job loss, it would not be beneficial for a doctor to assume that most patients behave in this manner. A patient could lie to their

doctor for different reasons, but the consequences the doctor faces are usually far greater than for the patient to lie about their medical condition. If we do not assign a prior probability that the doctor or patient, "could be lying to us," there is no subjective probability to uncover. But we could also say, that if we do not assign a prior probability that the doctor or patient is telling the truth, there is still no subjective probability to uncover. If the setting provides an opportunity where lying is acceptable, i.e. a poker tournament, then it would not always be in your favor to tell the truth. It is an issue of moral obligation if one believes instances like this apply to all situations. We usually do not have to physically tell a person to do this, but as complicated as emotions play a factor into everyday life, the melding between emotional responsibility and actuality have made it difficult to tell how often we should listen to ourselves instead of other people. We can infer to the best of our ability what somebody meant by their actions, but this again depends upon what we are adhering to first, belief or subjective probability.

I believe there is another variable to take into consideration here; that is the immeasurability of emotion. If we backtrack a bit to our original inquiry, it seems that beliefs only need to be emotionally triggered to go along with certain conformational biases because they are the best theories available to us. To combat this, if we appeal to other subjective possibilities before asserting our beliefs, we can better equip ourselves to understand the emotional pull a belief may have. Different kinds of beliefs allow for various degrees of emotion to have a role in accepting a theory, as these all pass through our subjective origin. The pronouncement of a belief should not carry with it an overbearing of emotion, as this would increase the chances of there being a confirmation bias. I now suggest that by looking at the methods in which we are taught as adolescents, we should encounter an explanation as to why emotionally charged beliefs are more likely to be considered truthlike by others, rather than beliefs not needing to say within their pronouncements, "I think I am right, but I could be wrong." Thus, the crux of what we are trying to prove (Sher 2001).

V. METHODS OF TEACHING

To avoid confusion, this next section examines a phycological experiment involving the education system and student learning. It was believed this experiment would show that the more engagement there was in the classroom

from students, the more this led to advanced learning. "Advanced learning" in this instance referred to higher test scores, and a student's ability to recall scientific misconceptions sometime later that year (Allen and Coole 2012, 387). The study was used to compare the effectiveness between two different teaching methods employed on students when learning about scientific misconceptions. This was done to determine if an emotional response played a role in advanced learning. The researchers in this study were focused on the connection between student emotion and remembering lessons about scientific misconceptions. The methods on how the students learned is what we would like to pull out from this study. In the experiment, this was done by allowing students to first predict an outcome of an activity, opposed to the other teaching method that did not allow students to make predictions potentially addressing their confirmation biases (Allen and Coole 2012, 388-90).

The teaching methods used in the study were conducted on randomized students split into control and treatment groups. The control group used a traditional approach of teaching primarily found in English and American schools, while the treatment group relied on reductio-ad-absurdum techniques. Both groups were given a scientific misconception to debunk, and teachers presented the material in their orderly fashion. The researchers conducting the test referred to these as Misconception Intervention Lessons, or (MI) throughout the paper. They were looking to answer three primary questions. These were:

1. Would students who recall (MI) be able to identify correct science?
2. Would one lesson incite a greater frequency or depth of emotion in students?
3. Is there a causal link between learning and student emotions?

The control and treatment lessons focused on the Newtonian principles of gravity, and the heavy/fast misconception. In the controlled lesson, the speed of falling was introduced as the topic to debunk, while the treatment group was first asked to predict if a larger or smaller object would hit the ground first if dropped. The researchers predicted the treatment group would demonstrate a significantly higher learning gain than the control group. The structure of the control lesson went as follows:

1. A pre-test to gauge the student's prior knowledge on the subject.
2. Four mini experiments around the theme of gravity.[5]

---

5. The four mini experiments were carried out by students rotating from station to station conducting

3. The teacher reveals the answer of each mini experiment.
4. A post-test to measure what the student learned.

The treatment lesson structure went as follows:

1. A pre-test to gauge the student's prior knowledge on the subject.
2. Students make predictions.
3. Five activities interrogating the student's predictions.
4. Gathering the data and making a final prediction.
5. Teacher reveals the correct answer, and a post-test to measure what the student learned.

Between these two methods, the main difference is the treatment group was able to make a prediction while the control group did not. Instead, the control group was first introduced to a theme, rather than a question being posed. The teacher also showed how to perform the experiments in the control group, while the activities in the treatment group focused on interrogating student predictions.

The experimenters assumed students in the treatment group would perform better on the post-tests because of the emotional response one would feel from getting their prediction correct. They reasoned this is what drove the students to benefiting on the post-tests because they remembered the results of their predictions more vividly based on how they felt. At the end of both post-tests, all the students were administered a different emotional responsive test. This test inquired the students to answer how they felt when the teacher revealed to them the correct answer. This test asked a range of questions concerning how strongly the student felt after the answer was given. Excited, surprised, pride, relief, embarrassment, where some of the possible answers, with each emotion correlating to how strongly they felt at the time via numerical assignment.

After the emotional responsive tests concluded, the experimenters compared results and found there was no emotional correlation to be tied in with higher test scores of the misconception lesson. Admittingly, the treatment group did perform better than the control group immediately after the lesson, but these benefits dissipated six weeks after the initial post-tests were conducted. As this was later proven when the experimenters gave the same post-test again to all the students in the study. The tests conducted six weeks later where the same post-tests administered directly after the misconception lesson, which showed neither

---

the experiments set up by the teacher, who demonstrated how to do each experiment prior to the children doing it themselves.

group performed better at long term retainment of the scientific misconception. In other words, there was no well-founded evidence supporting the idea that emotional stimulation led to advanced learning. However, they argued this study does not provide the evidence of an entire conceptual change potentially happening within the student's cognitive development. This also does not favor one teaching method over another. Although within the traditional method, it may not debunk a specific confirmation bias a student may still have. These results are to show how different teaching methods created a greater short-term growth for students when scientific misconceptions addressed their confirmation biases (Allen and Coole 2012, 393-403).

VI. GROUP MEDIATED TRUTH

While at first this seems counterproductive to the problem we are trying to solve (why our strongly held beliefs react before adhering to other reasonable subjective possibilities), I will point out the difficultly attempting to accurately gauge a young student's emotions based on a paper test. Perhaps the students did go through an exhaustive list of alternative answers on the emotional response test, only to put down what they believed to be what they thought the experimenters wanted to hear. This is considered by the experimenters as a "group mediated truth," which I will mention briefly. A group mediated truth would be if the student took into consideration what other classmates may have also answered, a nod towards appealing to subjective probabilities first instead of strongly held beliefs. Allen and Coole (2012) mention a different study conducted by different researchers that ties into this. In this study, a teacher asked one student to compare the readings of thermometers left in beakers of cold water. One of the beakers had a piece of wool wrapped around it while the other beaker remained unaltered. The teacher then asked the student to report back what the readings on the thermometers were. While the thermometers showed the same temperature on both beakers, the student reported back the beaker with the wool around it was five degrees warmer than the beaker without the wool around it. When the teacher asked the student to check again very carefully, the student then acknowledged the temperatures were the same. This may have been an actual mistake by the student, or they may have just wanted to get the correct answer. The most likely explanation is the student somehow believed

the wool warmed the water, since it was placed there. Incidentally believing the thermometer's reading to be inaccurate. However, the experimenters pointed out the student's ability was called into question, and the misconception played right into their confirmation bias. The confirmation bias in this case was that the wool was somehow able to "warm up" the water, since it was placed there, thereby inhibiting the student's ability to give the correct answer.

In this experiment, we can confidently guess the student probably held a prior bias, but we simply do not know for sure whether they appealed to their own beliefs, or a group mediated truth, since this would need to be asked. As previously stated, the follow up (MI) tests administered six weeks after the initial post-tests found no benefits retained by either the treatment or control group. The wool experiment was independent of this study, but similarities on debunking scientific misconceptions are apparent. By this I mean, if one is given the opportunity to make a prediction, especially when the consequences of being wrong makes one feel embarrassed, it is likely one will fall trap to a popular confirmation bias. (Allen and Coole 2012, 393-95).

VII. ADDRESSING CONTEXT SENSITIVITY

The methods on how we learn about a misconception, whether it be scientific or social interaction, should help us better understand this next segment concerning context. We should now ask: Where do these misconceptions come from if they need to be addressed in the first place? A problem we run into, is if one believes the methods forming the hypothesis contain a certain bias. Having a good reason to doubt the entire theory itself and related theories based on the methods used is an adequate reason to justify doubt. However, the source of these claims must be understood as trending towards a type of confirmation bias. In any case, there does not seem to be any adequate reason to accept a less likely hypothesis unless one were skeptical of the methods used under which it was formed. To address this issue, I propose that we need to look at different logical relations and how these misconceptions can form under contextual errors.

A common problem we face is confusion on the context of which we are given, or lack thereof. For example, beliefs can be held towards questions that are context sensitive, such as: "Is Greg tall?" Compared to: "What is the probability that Greg is tall?" These are not exactly questions that have a direct answer without

prior context. Any evidence supports or denies the hypothesis by observing how tall Greg is compared to your own subjective opinion on what qualifies as "tall." If anyone is to agree or disagree with your belief that, "Greg is tall," we expect most people act from their best intentions to judge for themselves what is truthlike, because "being wrong" in this instance is strictly subjective interpretation. Thus, leading back to your criteria of tallness within your subjective origin. It should not provide evidence towards purposefully misguiding an opinion. As these intentions would be less than truthlike and be a cause for creating misconceptions.

In most cases, subjective theories are usually confirmed by a strongly held belief. There are no objective measures to determine whether, "Jim is a nice person," without resorting to the biases we tend to have on the actions observed by "Jim." It can then be a challenge to prove that the actions Jim took do not provide evidence that, "Jim is a nice person." To call this out would seem to conclude that any theories surrounding "Jim" are biased because you took action into account. It may not be true that "Jim is nice," and in fact, "Jim is not a nice person," or perhaps, "Jim is a mean person." But this belief still depends on passing through our own subjective origin. Where these originate from, whether a person is used to system 1 or 2, provides an easier opportunity to prove the opposite is true. To explain, if one has a prior belief that "Jim is a nice person," under the emotional influence of system 1, a single "not nice action" is more likely to provide stronger evidence for the counterhypothesis. Again, depending on where a person assigns their prior assumptions to begin with, in contrast, a person dependent on system 2 may be less inclined to assume the actions by Jim go against the hypothesis that, "Jim is a nice person." A person under the guidance of system 2 may notice this because they already know the action of Jim is only inferred from their own subjective origin. To tell somebody, "Jim is not a nice person," may be just as emotionally strong as, "Jim is a nice person." Tone will provide an accurate gauge for this exclamation, as we now start to see how difficult it is for one to form or change a subjective belief. Either way, an emotionally charged hypothesis will overpower any subjective probability the individual may have previously held. This teeters on the notion of, "what one considers to be strongly elicited emotions." However, I insist we cannot recognize biases without resorting to subjective opinion, since this would diminish any type of reliable measurement. The attack on a strongly held belief relays us back to agree with the stronger emotional hypothesis, thus, creating the tendency to go

along with confirmation biases. If the context is misunderstood in these cases, then there must be an issue of convenient contexts switches within our language that seem to go unnoticed until they are brought up, that is, within the right context. I will attempt to show how this occurs throughout the remainder of this paper.

VIII. CONTEXT SWITCHES IN LANGUAGE

A hypothesis should extend to all available contexts if it adheres to a standard usage of language by a society. By "standard usage," what I determine to be is that it is understood within a community of same language speakers that interpretations can vary depending on the everyday background knowledge of different people. If a person knows that a word can be used to describe two different things (i.e., homonyms), then they possess adequate knowledge about the language, and therefore obtain a standard usage of language. To merely mention a word without knowing how it is properly used to describe a fact or thing within a system of language is next to speaking nonsense. This I believe is how evidence can be construed to support either side of an argument, each in a coherent manner, but different regarding its emotional pull. We will next look at an example of how context switches can occur without our awareness of it even happening.

A tactic used via context switches to convey misleading information within a standard usage of language is called priming. Priming insinuates that it must be true that, "Greg is tall," if the priming information is new, and in the memory of someone thinking about the tallness of Greg. If we are given the priming evidence that, "Everyone in the town Greg is from is tall," it would seem very unlikely that Greg is not tall. It would make us question the origin of the priming information if it did not follow that Greg was tall. If the priming information were instead, "Everyone in the town that Greg is from is below five feet." It would make us more skeptical that the conclusion, "Greg is tall," is false. This again brings up the distinction Kahneman makes regarding system 1 and 2. If we already understand what is meant by when we say, "Greg is tall," under the guidance of system 1 or 2, then there is no need to adhere to context sensitivity. If we are being primed to believe something else, misconceptions can occur from our inability to recognize these context switches. For example, if one strongly believes the death

penalty should be abolished, as death is an emotionally charged occurrence, any new evidence[6] will support the hypothesis that, "The death penalty should be abolished," if one holds this belief. A different individual may also hold the belief that the death penalty should only be used in extreme circumstances, if the evidence is the convicted criminal was in fact a mass murderer. It is safe to assume that in general, we believe the acceptance of one hypothesis denies the counterhypothesis of the same kind. In this example, you cannot be a supporter of abolishing the death penalty while also supporting the case that it should only be used in extreme cases, this does not get rid of the problem. This makes it so that you subjectively decide upon what you consider to be "extreme instances," rather than addressing the original problem; in this case "the death penalty." These context switches conveniently change the attention, and appeal instead to the emotional dependency of the proposition itself. There is a failure to acknowledge context shifts because we are stuck on our previously held beliefs.

IX. PLACEMENT OF NEGATION IN LANGUAGE

Feldman is a proponent against holding onto our knowledge ascriptions, claiming that we should only appeal to context sensitivity when a contextualist solution is not plausible.[7] "Sometimes," he claims, "We just have good arguments for opposing views (Cohen 2001, 87)." But I believe that sometimes bad arguments come about from reasoning in such a manner. The issue of context sensitivity leads us back to the problem of not knowing what to do in everyday situations. Claiming that we already know what we mean by something does not entail that a hypothesis has been cleared of confirmation biases. I will argue this point, on the grounds that being made aware of multiple different contexts can only increase our understanding of different background relations. Being aware of these context shifts will provide us with more truthlike theories, as this should only have a minimal effect on our ability to self-analyze. For example, if one is speaking in front of group of people, we understand the phrase, "All men are

---

6.   This could be a story of how devastating it was to find out an accused criminal was falsely convicted to the death penalty.

7.   Feldman (2001) provides examples of moral disputes that can be argued without an appeal to context sensitivity such as abortion or the origins of dinosaurs.

created equal," applies to everybody.[8] It is the concern of some that there are examples of it not being the case that, "All men are created equal." Since, if one is to speak in front of a group of people who hold the same interpretations of the words being spoken, it is less likely that a different interpretation will come about. If one provides a counterexample to "All men are created equal," two different interpretations are unintentionally produced. These are, "Not all men are created equal," and "All men are not created equal." According to Feldman (2001), in cases like this it is believed the majority will understand that "men" in this sentence still means "everyone," so there is no need to adjust this statement. When speaking to a mixed group of men and women the context still adheres to the firmest held belief in the room, but we must understand that different variables are present. These variables include different interests the group may have as pertaining to certain ideologies. If one is speaking to a group of women, and the speaker is a man, then the issue would be shifted more on the phrase itself, that, "All men are created equal," applies to no person in the room who holds the same belief, because the counterexample is right in front of them.[9] It should be safe to agree everyone has their own method of interpreting truthlike statements because we all come from different unique backgrounds. Learning about these methods has a different appeal to everyone, as context sensitivity will call to attention different background relations within a group. The issue of conflicting evidence occurs when a hypothesis depends upon the transparency of the group understanding the words used in the statement are referential to an existential or universal claim. If universal and existential statements are not properly understood within a group, this limits the number of interpretations that can be made upon context awareness. Our dependency of using existential statements that apply to universal theories inhibits our ability to confirm evidence. As it is the case, the number of counterexamples available does not diminish the hypothesis that, "All men are created equal." Nor does it provide evidence that the context is misinterpreted. What I believe to be the cause of our misconceptions is where one positions the negation, "not" in our cognitive sentencing. If we apply this to our example, "All men are created equal," we can see a bit more clearly where negation relies upon our ability to properly categorize counterexamples. Creating both, "All men are not created equal," and "Not all men are created equal." Once

8.   Should one hold the belief that, "All men are created equal."

9.   That the entire room is of the opposite gender.

this is called into question the stronger the belief is, the less likely it is that one will appeal to counterhypotheses, or, the stronger the belief is, the more likely it is that one will not appeal to counterhypotheses. It is possible that all subjective theories are prone to be confirmed by a stronger emotional theory, but these emotional eliciting theories typically include universal terms to exaggerate the point being made.

Relating back to our dependencies of system 1 and 2, it is safe to assume the majority believes, "All men are created equal," applies to everybody because of its dependency of being proven false. The problem does not rest upon whether we should or should not ignore our everyday knowledge, it is an issue of understanding why it is easier to use this method of language in everyday situations. This tendency to drive out any negation related hypothesis settles on the belief that our best available theory is whatever we have available at the time. It may be fitting to envision a well confirmed belief within a group that a claim is untrustworthy, since confirmation of universal and existential statements remain difficult for us (Sprenger 2013, 737). However, retaining this information does not raise or lower any subjective probability one may hold. If the group is convincing enough, it can overpower just about any subjective probability one may hold.[10] Should it be the case that a hypothesis calls attention to conflicting context resolutions, the probability of the hypothesis put into question has a higher chance of conforming to a confirmation bias.

X. CONCLUSION

I have shown that a tendency in this self-reliant mode of reasoning is to internalize the theory that our beliefs are the best we have available. If one puts forward their closely held beliefs first before adhering to any subjective probabilities, they may be prone to being overridden by an emotionally driven hypothesis. While adhering to beliefs before subjective probabilities can produce a boost of confidence in our assertions, if we instead consider more possibilities before asserting our beliefs, the dependency to rely upon our self-reliant beliefs decreases, thus, decreasing the potential of falling into confirmation biases. To

---

10. The Asch conformity experiments were a group of studies ran by Solomon Asch examining how individuals (in this case students) defied a majority group and the effects of these influences and beliefs.

internalize the theory that our beliefs are not the best we have, believing this prior to asserting our beliefs, is an appeal to searching for more subjective probabilities. Under this, if by accepting any hypothesis to be truthlike, you simultaneously confirm other closely related hypotheses to be truthlike as well. On the condition it is well known that any rival theories are not formed based on subjective opinion as we saw with the "tallness" example. This applies to our everyday knowledge assertions, and comparatively, we should note that scientific hypotheses are not based around emotional input so we should accept these as rational. The formation of a hypothesis is typically founded upon the basis that any well-prepared theorist is not going to include their own subjective bias surrounding the theory. It may be argued that any hypothesis put forward will have some subjective bias, but as we saw with the scientific misconception lessons, if one is first given the opportunity to predict a different hypothesis as better fitting, it is the methods used that gain the advantage over an emotionally driven hypothesis. We can eliminate this type of self-reliant confirmation bias, on the grounds there are more truthlike hypotheses than what we initially believe to be available, before we apply a probability to our own as being the best available theory.

REFERENCES

Allen, Michael, and Hilary Coole. 2012. "Experimenter Confirmation Bias and the Correction of Science Misconceptions." *Journal of Science Teacher Education* 23 (4): 387–405.

Cohen, Stewart. 2001. "Contextualism Defended: 'Comments on Richard Feldman's' Skeptical Problems, Contextualist Solutions." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 103(1): 87–98.

Feldman, Richard. 2001. "Skeptical Problems, Contextualist Solutions." *Philosophical Studies* 103: 61–85.

Fitelson, Branden. 2012. "Evidence of evidence is not (necessarily) evidence." *Analysis* 72 (1): 85-88.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus And Giroux.

Sher, George. 2001. "But I Could Be Wrong." *Social Philosophy and Policy 18*(2): 64-78.

Sprenger, Jan. 2013. "A Synthesis of Hempelian and Hypothetico-Deductive Confirmation." *Erkenntnis* 78 (4): 727–738.

compos mentis

Semantics is not Just Semantics: On the Status of Formal Semantic Theory

Juan Ignacio Murillo Vargas

University of Toronto

ABSTRACT

What does it take to understand a sentence? To most linguists and many philosophers of language, understanding a sentence is a two-part process. First, we must understand a sentence's semantic content, as captured by a semantic algorithm that shows how the truth-conditions of a sentence are determined by the meaning of its parts. Second, we must understand a sentence's pragmatic content, as deduced from its semantic content and how that content would be cooperative in the context of the conversation. If we know both of these, we have understood the sentence. However, this picture says little about language understanding as a cognitive activity. Our phenomenal experience of our linguistic competence does not involve "truth-conditions" or the kinds of pragmatic derivations linguistics and philosophers of language have assumed. There is a gap, here, between what traditional theorizing says, and what linguistic competence is like.

In this paper, I argue that this gap is problematic: semantics should capture, at some level of abstraction and detail, the cognitive activity involved in understanding a sentence. I also argue that this is not accomplished by assuming semantic algorithms are part of our cognitive architecture, as scholars like Chomsky have proposed. Instead, I propose that we ought to take formal semantic theories to merely be reifications of the underlying cognitive activity that language understanding involves. I conclude by outlining how developing this proposal, both philosophically and empirically, leads to the conclusion that language understanding is possible because we, as a linguistic community, choose to keep our language sufficiently uniform so that our language processing system can be epistemically reliable.

KEYWORDS

Semantics, Semantic Competence, Language Understanding, Cognitive Activity

INTRODUCTION

What is involved in understanding the following sentences?

  (1) Snow is white.

  (2) The llama ate the kale.

  (3) Juan should go outside more often.

One prominent answer in linguistics and some philosophy of language is as follows. First, you need understand the sentence's semantic content. For declarative sentences like (1) – (3), this involves knowing the sentence's truth-conditions: what the world would need to be like for the sentence to be true (cf. Heim and Kratzer 1998). These truth-conditions are determined by the meanings of sentence's parts, and how they are put together (cf. Frege 1997a, Frege 1997b, Frege 1997c, Frege 1997d). The task of a semantic theory, per this traditional view, is to capture and generate an algorithm that describes how the parts generate the truth-conditions of the whole sentence. (For examples of this approach, see Frege 1997a, Frege 1997b, Frege 1997c, Frege 1997d, Russell 2010, Lewis 1970, Kaplan 1989, Heim & Kratzer 1998, etc.)

The traditional story then tends to appeal to something like Grice 1993. He thought that what a sentence conveys in a conversation is determined by two parts: the sentence's semantic content, as given by a semantic theory; and the sentence's pragmatic content, wherein a competent hearer derives how the sentence's semantic content fits within the context of a conversation (cf. Grice 1993, chapter 2). The rules for this derivation are Grice's maxims: the rules all rational interlocutors must follow, and assume each other to follow, for their utterances to be cooperative in the context of the conversation (again, cf. Grice 1993, chapter 2). Take the conversation in (4), for instance:

  (4) A: Hey, do you know where I could find some gas?

       B: There's a gas station around the corner.

The semantic content of B's utterance in (4) is, roughly, "there is a gas station in our vicinity." However, that does not by itself answer A's question. Instead, A assumes that B is trying to be cooperative, and is thus genuinely trying to answer the question. So, A deduces that B must mean something of the form "there is a gas station in our vicinity, and that gas station has gas." This is a case of what Grice calls a "conversational implicature:" the additional pragmatic content that a speaker infers from a sentence's semantic content and how that content would be cooperative in the context of a given conversation.

The traditional view on how we understand a sentence, then, is that we i) know its semantic content, in virtue of knowing how its parts determine its truth-conditions, and ii) we know its pragmatic content, in virtue of knowing how the semantic content fits into the context of a conversation. If we know both things, we have understood the sentence.

Notice, though, that this view says little about language understanding as a cognitive activity—something we, as subjects, do.[1] Phenomenally, we do not experience ourselves using a complex algorithm or Gricean derivation to understand each other. We do not take ourselves to be interested in "truth-conditions," or any other formal machinery, when we speak to each other. There is a gap, here, between what traditional theorizing about language understanding says, and what the ordinary language understanding experience is like.

This paper has two goals. First, I argue that this gap is a problem for traditional theorizing: any successful account of language understanding should appropriately capture the fact that language understanding is a cognitive activity. Second, I try to satisfy this desideratum, by arguing that traditional semantic theories are reifications of semantic competence. In §1, I discuss the first point. In §2, I discuss the second.

---

1. For the purposes of this paper, I will be assuming that language understanding is a cognitive activity. While the kinds of semantics I am discussing come from a legacy of behaviorism—Quine was among those who popularized the Tarski-inspired semantics we use today—it would be problematic if being a semanticist required being a behaviorist.

1: SEMANTICS SHOULD CAPTURE COGNITION

This section argues for the following desideratum:

(5) SEMANTICS-AND-COGNITION: a good semantic theory appropriately captures the fact that semantic competence is a cognitive activity, at some level of abstraction and detail.

Where "semantic competence" is defined as:

(6) Semantic competence $=_{df}$ an ordinary speaker's ability to understand a language. This involves being able to understand a sentence's semantic content, and being able to apprehend its pragmatic content.[2]

The argument for this desideratum proceeds as follows:

1. A semantic theory is an algorithm that calculates a sentence's truth-conditions based on the meanings of its parts [definition from the introduction].
2. A good semantic theory either appropriately captures that semantic competence is a cognitive activity, or it does not [tautology].
3. A good semantic theory does not appropriately capture that semantic competence is a cognitive activity [assumption for *reductio*].
4. A good theory captures what a phenomenon is like at some (discipline-specific) level of abstraction and detail.
5. Either semantic competence is not a cognitive activity, or a good semantic theory does not concern semantic competence at some level of abstraction and detail [From 3, 4].
6. Semantic competence is a cognitive activity [assumption].

---

2. There is some debate as to whether there is such a distinction between semantics and pragmatics, cf. Récanati 2010. Since this paper is partially focused on what linguists say language understanding involves, and since contextualism is not a live option for them, I will proceed as if there is a distinction between semantic content and pragmatic content.

7. So, a good semantic theory does not concern semantic competence at some level of abstraction [from 5, 6].

8. If a good semantic theory does not concern semantic competence at some level of abstraction and detail, it is unclear what it tells us about language.

9. It is unclear what a good semantic theory tells us about language [from 7, 8].

10. The result in 9 is absurd and must be rejected.

11. Both disjuncts in 5 are false [from 6, 10].

12. The assumption at 3 must be false [from 11].

13. THEREFORE: a good semantic theory appropriately captures that semantic competence is a cognitive activity, at some level of abstraction and detail [definition of SEMANTICS-AND-COGNITION; from 2, 12].

The key premises at play here are 4, 8, and 10. Let us defend each in turn.

The claim that "a good theory captures what a phenomenon is like at some level of abstraction and detail" should be uncontroversial for most, but let us cover our bases. For one thing, this is clearly our intuition. I suspect most people would rebuke a physics that could not handle how ordinary, marcroscopic objects move. Furthermore, the whole purpose of developing a theory, we might think, is to understand what something is like and how it works to some capacity. If so, a theory that does not capture a phenomenon at a certain standard of abstraction and detail fails to be a theory at all—it runs against the very point of theory-crafting.

What about "if a good semantic theory does not concern semantic competence at some level of abstraction and detail, it is unclear what it tells us about language?" Well, if a good semantic theory does not concern semantic competence, then it either concerns some other dimension of language (say, phonological or syntactic competence) or it does not concern language at all. The former disjunct is clearly absurd: if semantics captures phonological or syntactic competence, it is entirely redundant. There are other fields, with other formal models, that study those things. So, lest we want to say that semantics is redundant, we would have to say that it is unclear what it tells us about language.

Finally, why should we think that semantics not concerning language is an absurd result? For one thing, this would require semanticists to revise what we take

ourselves to be doing, since we take ourselves to be studying natural language (e.g., Heim and Kratzer 1998, 2). Furthermore, and again, if semantics did not concern natural language, the discipline would be at high risk of being redundant, since it is simply unclear what else semantics could concern. Unless someone has an account for what else semantics could be about, we should think it absurd that it is not about natural language.

Since we have defended premises 4, 8, and 10, we can safely conclude—at least for now—that Semantics-And-Cognition is a desideratum for any good semantic theory. However, notice that it leaves indeterminate the level of abstraction and detail that a semantic theory can or should operate in to be consistent with Semantics-And-Cognition. I take this issue up in the next section.

2: SEMANTICS AS A REIFICATION

2.1 Why Not Hard Internalism?

If we are coming from a linguistics background, we might think that there is already a view of semantics consistent with Semantics-And-Cognition: Chomsky 2000 proposes a view that I will be calling "hard internalism."[3] He thinks that, insofar as we want to study language through naturalistic inquiry (e.g., the kind of inquiry in physics, chemistry, neuroscience, etc.), we should only care about "I-languages;" an individual speaker's internal, intensional associations between an expression in the lexicon, with phonological and semantic features, and how it fits into the syntax (Chomsky 2000, 26). Why does he think this? Because most of the other terms we might use to describe a public language—e.g., "communicative intent," "mutually held beliefs," etc.—are folk concepts, not natural kinds, and so they simply cannot be part of a naturalistic inquiry (cf. Chomsky 2000, chapter 3).

Of course, Chomsky owes us a full account of what "naturalistic inquiry" is, or why we might want to study natural language under its lens. Nonetheless, we might be tempted to accept his view because it easily satisfies Semantics-And-Cognition. If what semantics captures are an individual brain's intensional associations, then semantics just spells out the cognitive architecture associated with semantic competence.

Such a radically internalist view of semantics, however, infamously leads Chomsky to conclude that semantics does not concern "reference" or "extension:"

---

3. Pietroski 2018 is a more recent example of this type of view.

> Contemporary philosophy of language follows a different course. It asks to what a word refers, giving various answers. But the question has no clear meaning. The example of "book" is typical. It makes little sense to ask to what thing the expression "Tolstoy's War and Peace" refers, when Peter and John take identical copies out of the library. The answer depends on how the semantic features are used when we think and talk, one way or another. In general, a word, even of the simplest kind, does not pick out an entity of the world, or of our "belief space." Conventional assumptions about these matters seem to me very dubious. (Chomsky 2000, 17)

The argument looks weak, but it follows from the rest of Chomsky's framework. If we are to be studying natural language in general—and semantics in particular—under a naturalistic lens, then "reference," which is not a natural kind, and is not solely part of an individual speaker's intensional associations, simply cannot be the object of study. "Extension," for the same reason, falls by the wayside too.

I find Chomsky's view implausible for three reasons. For one, there might be a layer of normativity fundamental to language (e.g., rational patterns of use, norms of communication, etc.), and a naturalistic lens—at least as Chomsky seems to understand it—would miss out on that entirely. Moreover, there is something prima facie implausible about adding highly complex semantic algorithms to our cognitive architecture, especially considering they are not part of our phenomenal experience as language users. Finally, most semantic theory appeals to "extension," if not "reference" (cf. Heim and Kratzer 1998); so losing out on those notions, just to satisfy Semantics-And-Cognition, is simply not desirable.

What we should want, I think, is a view of semantics that keeps "extension" and "reference," but satisfies Semantics-And-Cognition without adding semantic algorithms to our cognitive architecture. (We should also leave some room for normativity in language use while we are at it.) In the next section, I will be proposing and developing my preferred way of accomplishing this: by treating formal semantic theories as a reification of the cognitive aspect of semantic competence.

2.2 Soft Internalism

Since I have lost all pretense of humility by this point, let us call my view "soft internalism:"

> (7) Soft internalism $=_{df}$ the view that semantic theories are just *reifications* of the cognitive activity involved in semantic competence.

This proposal, again, straightforwardly satisfies Semantics-And-Cognition. If semantic theories are just reifications of the cognitive activity involved in semantic competence, then, by definition, they capture what semantic competence involves cognitively soft internalism and hard internalism, then, is two-fold.One, soft internalism does not propose nor entail that the algorithms semanticists study are literally part of human cognitive architecture. Two, and consequently, soft internalism takes semantics to be operating at a higher level of abstraction than what hard internalists, like Chomsky, typically suppose.

Soft internalism comes with a few upsides. For one thing, it is more parsimonious: unlike Chomsky, I think we can and should keep the explanatory power of semantic theory without adding it to our metaphysics of the mind. Moreover, it also paves the way for normativity, extension, and reference to play a fundamental role in semantic competence, precisely because soft internalism views formal semantic algorithms as mere reifications for whatever is metaphysically at bottom in semantic competence.

For this proposal to work, however, we need to give a derivation from the cognitive activity involved in semantic competence, to what semantic theories actually look like. Otherwise, we would have no account of how it is that semantics, as a reification, can capture the cognitive aspect of semantic competence. If so, we might be tempted to accept hard internalism instead, since the story would be much more straightforward: semantic algorithms capture the cognition involved in semantic competence because they are what is at bottom in semantic competence.

One derivation to solve this is challenge is as follows:

> 1. Semanticists present native speakers with written utterances of a language, and develop formal models that i) capture our intuitions on what that utterance means, and ii) predict the intuitive meaning of similarly-structured utterances.

2. Native speakers process written and natural language incrementally: as an utterance occurs, their brains make forecasts about what the rest of the structure will look like.
3. Our intuitions about what an utterance means are generated by what our forecasting system delivers [from 1, 2].
4. Linguistic structure is relatively uniform.
5. If linguistic structure is relatively uniform, then the forecasting process will work the same way in both the written language case and the natural language case.
6. The forecasting process will work the same way in both the written language case and the spoken language case [from 4, 5].
7. The intuitions that semanticists model match the spoken language case [from 3, 6].
8. The spoken language case involves, among other things, an ordinary speaker's semantic competence.
9. THEREFORE: The intuitions that semanticists model match, among other things, an ordinary speaker's semantic competence [from 7,8].

Let us defend premise 1, 5, and 8 first, since they are easier. 1 is a basic fact about how semantics research is conducted (e.g., Heim and Kratzer 1998, Copley 2006, Yalcin 2007, von Fintel and Gillies 2010, Silk 2018, etc.). Premise 5 is a relatively plausible empirical claim: if linguistic structure is uniform, there would be no prima facie reason for the human processing system to differentiate between written and spoken language—the underlying structure would be the same. (There is some empirical backing for this idea; see Kuperberg and Jaeger 2016 for an overview.) Finally, premise 8 is straightforward; when we communicate with each other using natural language, that involves us producing sentences others can understand and understanding the sentences others produce. The second clause of that sentence is simply the definition for "semantic competence" I gave above.

This leaves us with premise 2 and premise 4 to defend. Premise 2 is an important empirical claim, which is often called "incrementality" in the psycholinguistics literature:

(8) INCREMENTALITY =$_{df}$ the way the brain processes language is not by "waiting" for the utterance to finish. Instead, as we hear parts of the utterance, the brain generates and processes forecasts[4] as to what the rest of the structure will look like.

Incrementality has been found across many different experiments, for both written and spoken language, and across different experimental paradigms (Rubio-Fernandez and Jara-Ettinger 2020, 1; again, for an instructive overview, see Kuperberg and Jaeger 2016). An intuitive motivation for this claim, however, comes from so-called "garden path" sentences. Take (9):

(9) The old man the boat.

At first glance, (9) might seem ungrammatical. It is not though: "the old" is the subject, "man" is the verb, and "the boat" is the object. The reason it appears ungrammatical is because "the old man" tends to occur as a single noun phrase. So, our brain forecasts the structure in (9) to treat it as such. When this forecast turns out to be wrong, we initially process the sentence to be ungrammatical.

Now, only premise 4 remains to be justified. At first, we might be tempted to say "it is an empirical claim" and leave it at that. However, a lot of research (especially in the Generativist tradition) implicitly presupposes that language is uniform. And relying on the empirical work, to support the presupposition underlying the empirical work, would create a nasty vicious circle. Instead, we need to give a philosophical argument here. While it will take us on an odyssey of sorts, it will also let us say something interesting about the epistemology of language understanding in the process. So, off we go.

2.3 Why Language is Relatively Uniform

To give the argument for the claim "linguistic structure is relatively uniform," let us start with the following first principles:

(10)   ARBITRARINESS =$_{df}$ there is no intrinsic relationship between a natural language expression and the content it conveys.

---

4.   The word "forecast" is used intentionally: there is an interesting question as to whether these cognitive processes are representational, perceptual, or neither. I will not take that question up here.

Instead, the relationship is *arbitrary*: members of a linguistic community choose what the association is.

(11)   STRUCTURAL-RATIONALITY $=_{df}$ humans are, or at least aim to be, structurally rational. We do not want our beliefs to be inconsistent, and we do not want our actions to be inconsistent with our beliefs, needs, or goals.

(12)   COMMUNICATIVE-NEED $=_{df}$ the mind has a basic need to communicate.

Arbitrariness dates back to Ferdinand de Saussure (Saussure 2003, 79. Holdcroft 1991, 52 explicitly attributes the idea to Saussure; and Chomsky 2000, 9 also mentions it in passing, though he does not cite Saussure). In some sense, though, it just has to be right. If it were false—that is, if there were some intrinsic relationship between expressions in natural language and the content they convey—we would be unable to explain how different languages could have different words to convey the same content. Considering this is obviously the case (e.g., "dog" and "perro" refer to the same kind of cute canine), Arbitrariness must be true.

Of course, these arbitrary associations are not the result of some communal deliberative process. There is no "original position" where English speakers decided "dog" would refer to a type of cute canine. Instead, our collective choices on the relationship between expressions and their content are much more subtle.[5] For instance, we often use expressions deferentially or parasitically: we simply use them to mean whatever other speakers, usually experts, take the expression to mean (e.g., non-physicists like myself use "quantum physics" to refer to whatever it is that physicists understand by "quantum physics.")

Structural-Rationality, meanwhile, is standardly accepted amongst philosophers of language (e.g., Williams 2020, 14). Moreover, insofar as semantics is also concerned with rational patterns of inference (as Kaplan 1989 seems to think, cf. chapter 17 and 18), something like Structural-Rationality must be right. For argument's sake, I am also supposing that, since this kind rationality is minimal

---

5.   I will not try to develop a full list or account of these choices here. In general, though, I take the choices to be more akin to the "invisible hand" economics sometimes appeals to, rather than any kind of deliberate, simultaneous choice-making activity.

and merely formal, most individuals and collectives are structurally rational in the way Structural-Rationality prescribes. While most readers might (rightfully) be skeptical, remember: Structural-Rationality makes no prescriptions on the content of the relevant beliefs and wants. So, most subjects' beliefs or wants—no matter how prima facie absurd they are—can easily meet it.

Communicative-Need is decidedly trickier to defend, and I simply do not have the space to do so here. So, let us interpret it as an empirical claim: as part of the evolutionary process, humans developed a need to communicate with each other to increase their chances of survival. This claim still needs empirical substantiation, but it is plausible enough for now.[6]

With these three principles at hand, all we are missing is Incrementality, which we already saw earlier in §2.2. Now, we can build the argument for the claim that linguistic structure is relatively uniform:

1. Linguistic structure is either relatively uniform or not [tautology].
2. Linguistic structure is not relatively uniform [assumption for *reductio*].
3. If linguistic structure is not relatively uniform, we would not know when or whether we understood someone's utterance, since our forecasting system would often be wrong [presupposes INCREMENTALITY].
4. We do not know when or whether we understood someone's utterance, since our forecasting system is often wrong [from 2, 3].
5. Our decisions surrounding linguistic structure will be consistent with our need to communicate [from ARBITRARINESS, STRUCTURAL-RATIONALITY, and COMMUNICATIVE-NEED].
6. The need to communicate entails that we need to know when or whether we understand each other.
7. Our decisions surrounding linguistic structure will be consistent with our need to know when or whether we understand each other [from 5, 6].

---

6. This is not to say there is no philosophical justification to go along with it; I think there is and should be.

8. 4 contradicts 7, since we would not make a choice that would prevent us from knowing when or whether we understand each other [from 4, 7].

9. Since 4 follows from 2 and 3, and 3 is true, the assumption at 2 is false [from 8].

10. THEREFORE: Linguistic structure is relatively uniform [from 1, 9].

For this argument to work, we need to justify premises 3 and 6. The latter is straightforward: if we did fail to understand each other, it is unclear how we would count as "communicating" in the first place. The more philosophically interesting premise, however, is 3. Notice that it is a linguistic version of what David Hume said when he first posed the problem of induction. ("Induction," for Hume, is an inference about the future based on present evidence; see Hume 2011, 78.)) For induction to be justified, nature needs to be uniform (or, at least, relatively uniform), since only then could experiences in the present reliably tell us about the future (Hume 2011, 79). Except, the same constraint applies to our brain's forecasting system: if language were not relatively uniform, then the forecasting system would constantly misfire, as it did when it first encountered the garden-path sentence in (9).

However, language, unlike nature, is entirely up to us. Arbitrariness entails that we, as a linguistic community, choose what linguistic structure looks like. And we choose to keep that structure relatively uniform because we have a need to communicate, and that need requires us to keep things such that our semantic competence is epistemically reliable. This is a deeply interesting result for the epistemology of language understanding: we reliably understand each other because we choose to make linguistic structure relatively uniform, such that reliable understanding is possible.

Moreover, we have also given an argument for the claim that "language is relatively uniform," which was the remaining premise to justify in §2.2. So, we have also completed our derivation for how semantics, seen as a reification, can capture the cognitive activity involved in semantic competence.

3: SUMMARY

This paper asked two questions: 1) should semantic theories relate to the cognitive activity involved in semantic competence, and 2) if so, how can and should they do so?

§1 answered 1) in the affirmative: if semantics fails to capture what semantic competence involves cognitively, its explanatory role in a broader account of language is unclear.

§2 answered 2) in three parts. §2.1 argued against a Chomskyan solution, since his view is implausible for a variety of reasons. §2.2. put forth my own proposal: semantics is best seen as a reification of the cognitive activity involved in semantic competence. I also gave and defended a derivation for how semantics captures this cognitive activity: if language is relatively uniform, and if written and spoken language are subject to the same forecasting system, then the intuitions semanticists model will match what ordinary speakers understand in virtue of their semantic competence. Finally, §2.3. defended the claim that language is relatively uniform. As language users, we choose to keep linguistic structure uniform so we can meet our cognitive need to communicate.

I think what I have said here raises at least three further questions. One, we need to fully defend the claim that the mind has a need to communicate, either empirically or philosophically (ideally both). Two, we can ask further questions about the status of the other theories in linguistics, i.e., what phonology, syntax, and morphology are. Finally, if soft internalism is correct, we should also find out what cognitive realities underlie our essential semantic tools (e.g., "propositions" and "possible worlds," in the way Stalnaker 1970 and Stalnaker 1999 use them). If they are merely reifications for underlying cognitive activity, we should be able to find what these tools serve as reifications for.

REFERENCES

Almog, Joseph, John Perry, Howard K. Wettstein, and David Kaplan, eds. 1989. *Themes from Kaplan*. New York: Oxford University Press.

Chomsky, Noam, and Neil Smith. 2000. *New Horizons in the Study of Language and Mind*. 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511811937.

Copley, Bridget. 2006. "What Should '"should"' Mean?" *Language Under Uncertainty Workshop, Kyoto University*, 22.

de Saussure, Ferdinand. 2003. "The Linguistic Sign." In *Classic Readings in Semiotics: For Introductory Courses*, 69-87. Ottawa: LEGAS.

von Fintel, Kai, and Anthony S. Gillies. 2010. "Must... Stay... Strong!" *Natural Language Semantics* 18 (4): 351–83.

Frege, Gottlob. 1997a. "Begriffsschift: A Formula Language of Pure Thought Modelled on That of Arithmetic." In *The Frege Reader*, 47–78. Blackwell Publishing.

Frege, Gottlob. 1997b. "Function and Concept." In *The Frege Reader*, 130–48. Blackwell Publishing.

Frege, Gottlob. 1997c. "On Concept and Object." In *The Frege Reader*, edited by Michael Beaney, 181–93. Blackwell Publishing.

Frege, Gottlob. 1997d. "On Sinn and Bedeutung." In *The Frege Reader*, edited by Michael Beaney, 151–71. Blackwell Publishing.

Grice, Paul. 1993. *Studies in the Way of Words*. 3. print. Cambridge, Mass.: Harvard Univ. Press.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell Textbooks in Linguistics 13. Malden, MA: Blackwell.

Holdcroft, David. 1991. *Saussure: Signs, System and Arbitrariness*. 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511624599.

Hume, David, and Lorne Falkenstein. 2011. *An Enquiry Concerning Human Understanding*. Broadview Editions. Peterborough, Ont. ; Buffalo, NY: Broadview Press.

Kuperberg, Gina R., and T. Florian Jaeger. 2016. "What Do We Mean by Prediction in Language Comprehension?" *Language, Cognition and Neuroscience* 31 (1): 32–59. https://doi.org/10.1080/23273798.2015.1102299.

Lewis, David. 1970. "General Semantics." *Synthese* 22 (1–2): 18–67. https://doi. org/10.1007/BF00413598.

Pietroski, Paul M. 2018. *Conjoining Meanings: Semantics without Truth Values*. First edition. Context and Content. Oxford, United Kingdom: Oxford University Press.

Récanati, François. 2010. *Truth-Conditional Pragmatics*. Oxford ; New York: Clarendon Press.

Rubio-Fernandez, Paula, and Julian Jara-Ettinger. 2020. "Incrementality and Efficiency Shape Pragmatics across Languages." *Proceedings of the National Academy of Sciences* 117 (24): 13399–404. https://doi.org/10.1073/ pnas.1922067117.

Russell, Bertrand, and David F Pears. 2010. *The Philosophy of Logical Atomism*. London: Routledge.

Stalnaker, Robert. 1970. "Pragmatics." *Synthese* 22 (1/2): 272–89.

Stalnaker, Robert. 1999. "Assertion." In *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford Cognitive Science Series. Oxford ; New York: Oxford University Press.

Williams, J. Robert G. 2020. *The Metaphysics of Representation*. 1st ed. Oxford University Press. https://doi.org/10.1093/oso/9780198850205.001.0001.

Yalcin, Seth. 2007. "Epistemic Modals." *Mind* 116 (464): 983–1026. https://doi. org/10.1093/mind/fzm983.

compos mentis

The Liberation of Philosophy: The Principle of Sufficient Reason as an Anti-Racist Principle

Ember Reed

University of Puget Sound

ABSTRACT

The method of intuition, the view that the best philosophical perspective will maintain as many of our intuitions as possible, is one of the pillars of analytic philosophy. Unfortunately, the reliance on intuition by analytic philosophers has created conditions such that the biases of those who do philosophy, predominantly those of hegemonic identities, are accepted as a basis for philosophical knowledge. This problem can be solved by rejecting intuition as a basis for philosophical knowledge and instead relying on a methodology derived from the Principle of Sufficient Reason.

In Chapter 7 of the Parmenidean Ascent, Michael Della Rocca discusses how the "Method of Intuition," or "the MI" (2020, 184), has become one of the "pillars of analytic philosophy" (2020, 183). The MI relies upon the idea that the best philosophical picture will maintain as many of our intuitions as possible. However, as Della Rocca points out, this is an intellectually conservative philosophical methodology, as it fails to ask whose intuitions we are counting. In the United States and the rest of the English-speaking world—in other words, the domain of analytic philosophy—academic philosophy is largely done by members of hegemonic identities, specifically those who are wealthy, white, heterosexual, and/or cisgender male. Worse still, those who lack hegemonic identities are often unable to challenge hegemonic assumptions without putting themselves at risk of ostracization. Within such a context, philosophy that relies on the MI is an anti-intellectual pursuit serving only to reinforce hegemony rather than to challenge it. So what's a wanna-be philosopher who wishes to advance liberatory politics, i.e., anti-hegemonic politics, to do? To my mind, there are two seemingly contradictory ways in which philosophy must be liberated, one relating to the material circumstances of philosophy and the other relating to the methodology of philosophy. In the context of philosophical work, we must seek out those with anti-hegemonic intuitions and identities to take part in our philosophical project. This will free philosophical work to be done in opposition to hegemony and, hopefully, allow the direct support of ideological ends that can be justified. In terms of methodology, we must reject the MI and, in doing so, free philosophy from an overreliance on intuition. After all, if intuitions cannot be trusted to provide reliable answers about the nature of reality, then why would we involve them in our philosophy? In rejecting the MI as a philosophical pillar, I will defend replacing it with the Principle of Sufficient Reason (henceforth referred to as "the PSR"). For the purposes of this essay, I will assume that we are generally in agreement that diversifying the field of philosophy is a worthwhile endeavor, and, accordingly, I will focus on my more controversial claim: that the PSR should replace the MI as a central philosophical pillar.

Reliance on intuition has always existed within philosophy, insofar as all forms of thought rely on intuition to some degree or another. What individuals choose to study, what ideas they choose to entertain, and what ideas they reject before fully considering them can generally be traced back to intuitions. However, just because intuition is at the root of what we choose to think about does not mean

that we ought to adopt it as a formal philosophical principle. Intuitions are often wrong. If you ask a person who lacks knowledge of physics what will happen if you drop two objects of identical shape but of different weights, they will likely respond that the heavier object will fall faster. Yet study of the physical world has shown us that this intuition, like so many others, is blatantly false. This should be of great concern to a defender of the MI. Analytic philosophers seem to want intuitions to be at the base of philosophy—to guide philosophy and perhaps even to constitute the subject matter of philosophy. Kit Fine, for example, is

> of the opinion that real progress in philosophy can only come from taking common sense seriously. A departure from common sense is usually an indication that a mistake has been made. If you like, common sense is the data of philosophy and a philosopher should no more ignore common sense than a scientist should ignore the results of observation. (2012)

Yet common sense or intuition tells us that the heavier object falls faster. If our intuitions are consistently wrong, and philosophy is merely based on our intuition, why are we doing philosophy at all? It seems that philosophy grounded in intuitions must commit itself to attacking physics, and ultimately empirical knowledge in general, in order to ensure that the heavier object falls more quickly.

The analytic philosopher is likely to concede that intuition is often imperfect. Della Rocca notes that, "as is agreed on all sides nowadays, appeals to common sense or intuitions are fallible" (2020, 186). However, "[d]espite this apparent fallibility, such appeals to intuition or common sense are, as it were, the gold standard in analytical philosophy" (Della Rocca 2020, 186). This position seems profoundly unpalatable if not, in fact, unsustainable. If we all agree that intuitions are fallible, perhaps we should stop making arguments from intuition. At the very least, we should refrain from making intuitions fundamental to our philosophical practice. In his Meditations on First Philosophy, Descartes says, "All that up to the present time I have accepted as most true and certain I have learned either from the senses or through the senses; but it is sometimes proved to me that these senses are deceptive, and it is wiser not to trust entirely to anything by which we have once been deceived" (1989, 74). Similarly, since we have often been deceived by our intuitions, it is unwise to trust them entirely.

Yet while intuitions are potentially untrustworthy in any context, in the specific context of analytic philosophy they may actively undermine our search for truth. Philosophy is not a field done by a representative sample of the population. According to the American Academy of Arts and Sciences, "traditionally underrepresented racial/ethnic minorities received 17.0% of all bachelor's degrees in philosophy" in 2014, while, "[a]t the master's level, traditionally underrepresented racial/ethnic minorities earned 10.2% of philosophy degrees awarded" and "completions of philosophy doctorates by traditionally underrepresented minorities" stood at only "7.9%" (2021). This data illustrates a trend that also holds true for marginalized identities outside of the context of race. In this context, when we talk about philosophical intuitions, we are talking about the intuitions of a specific subset of society that holds the biases of hegemonic identity markers. A defender of the MI will likely argue that the intuitions relied upon by analytic philosophers have little to do with the category of race. For example, the Mary's room thought experiment (first introduced in Frank Jackson's article "Epiphenomenal Qualia") posits a color scientist, Mary, who has perfect scientific knowledge of color but has never seen colors other than black and white. The experiment then asks us if Mary would learn something new if she were to see red for the first time. This is used to draw out the intuition that, in such a circumstance, Mary would learn something new. This intuition, my opponents will insist, has nothing to do with race, nor any other hegemonic identity. In reply, I happily concede that there may well be particular cases in which appeals to intuitions are not linked to race or affected by philosophers' racial identities, and the Mary's room thought experiment may constitute such a case. My claim is not that intuitions always and everywhere lead us astray because of underlying biases, but that reliance on intuition is sometimes biased and, thus, does not constitute a sound foundation for doing philosophy in general. Perhaps the MI does give the right result in the Mary's room case, but it may provide very distorted and biased results in other instances.

Ethical intuitionism is an ethical theory that posits that the ethical choice in any given circumstance is not based on any a priori circumstances (be they deontological rules, consequalist calculus, or a form of personal virtue ethics), but rather can be derived from one's intuitions about the circumstances at play. Ethical intuitionists often use a variety of trolley problem thought experiments and judge people's emotional responses to various courses of action to determine how ethical these courses of action are. One example of a rather troubling thought

experiment is the "fat man" version of the trolley problem, a variant which was proposed by Judith Jarvis Thomson, one of the most outspoken defenders of the MI. Thomson makes her acceptance of the MI clear when she insists that

> the main, central problems [of philosophy] consist in efforts to explain what makes certain pre-philosophical, or nonphilosophical, beliefs true. Which beliefs? Philosophers differ in their interests, but the ones that have interested philosophers, generally… are those that we rely on in ordinary life…. I still think that way of understanding what philosophy is roughly right. (2013, 54)

In accordance with the MI, Thomson's "fat man" variant asks us if we would be willing to push a fat man in front of a trolley to save five other people. She emphasizes the man's weight, describing him as "fat, so fat that his body will by itself stop the trolley, and the trolley will therefore not reach the five" (1985, 1403). While most people are unwilling to push the fat man in front of the trolley (most people will pull a lever), the inclusion of his weight, and his maleness for that matter, is highly troubling. After all, overweight people constitute a non-hegemonic identity that our culture is biased against, and, if we used the same thought experiment but with different identity markers, would people's answers change? Would we feel comfortable changing our ethical model to incorporate social intuitions positing that it was more ethical to push a Black man in front of a trolley than a white woman? Thought experiments of this kind remain troubling if we remove identity markers altogether, as people project identities onto hypothetical individuals who lack them, normally prescribing them identity markers that match their own, or hegemonic, identity markers. In this manner, an intuitionist moral framework, and ultimately any ethical framework that relies on intuition, only upholds identities already believed to be of higher worth by those utilizing it.

As concerning as ethical intuitionism is when practiced by individuals, it becomes even more troubling when expanded to the political sphere. After all, it is in this arena that the vast majority of its consequences will play out. To ask those in a favorable position within a social structure to intuitively judge the moral worth of such a structure is clearly preposterous. For instance, it would be hard to deny that the intuitions of a person who holds other people as property will be biased with respect to the subject of slavery. Yet such will always be the intuitions most trusted by philosophy done under hegemony. To see a contemporary example

of political intuitionism, we must examine the philosophy of the infamous Dr. Jordan Peterson. Peterson came to prominence in 2017 and gained great support through spreading misinformation regarding the legal consequences of Bill C-16, an expansion of the Canadian Human Rights Act. While this law merely expands the Canadian Human Rights Act to include gender identity and expression to the list of prohibited grounds of discimination under the Canadian Human Rights Act (Department of Justice 2018), Peterson claimed, "These laws are the first laws that I've seen that require people under the threat of legal punishment to employ certain words, to speak a certain way, instead of merely limiting what they're allowed to say." This is of course absurd, as Bill C-16 is a standard amendment and directly paralleled protections that already existed for other marginalized identities. How could Peterson, a professional academic, make such a mistake? As it turns out, holding views directly out of line with empirical reality is perfectly in line with Peterson's epistemological beliefs. In one of his lectures, Peterson says:

> You already know everything in some sense, you've got a map that covers the whole world. Which is sort of why you can function, and, so as long as everything's going fine, you don't really have to adjust your map and you don't have to think. But then, if you come across something that makes you think, then what [that] means is that part of the way you were thinking was wrong, and so when you think something—when you're forced to—then some little part of what you were, your map, the way you represent the world, it has to die because it was wrong.

In the context of his overall discussion, Peterson is suggesting that reconsidering the ways in which we engage with the world is negative, a form of death. For Peterson, we already know how the world works; we have knowledge in the form of reliable intuitions. To reconsider how the world works, and our place in it, is, for Peterson, not worth risking our very being. When viewed through this lens, Peterson's ability to deny reality regarding the legal consequences of Bill C-16 makes quite a bit of sense. For him to challenge his intuitions about what it would mean to give trans people human rights would be, in his eyes, the equivalent of a little part of himself dying. As demonstrated by Peterson's political advocacy, the consequence of such a view is the continued persecution of trans people or any other group that those in power intuitively believe to be lesser than themselves.

Many within the analytic tradition may find themselves aghast that I dare compare them to Peterson. Yet as much as such a comparison feels below the belt, I can't, for the life of me, find a meaningful distinction between their treatment of intuition, in the last analysis, and Peterson's. Indeed, many titans of the analytic tradition have made claims reminiscent of Peterson's regarding the challenging of intuitions. David Lewis, for example, said that "[o]ne comes to philosophy already endowed with a stock of opinions. It is not the business of philosophy either to undermine or justify these pre-existing opinions to any great extent, but only to try to discover ways of expanding them into an orderly system" (1983, 88). This view—that philosophy should not challenge, but merely systematize, our intuitions—amounts to what Della Rocca calls "the taming of philosophy" (2020, 260): preventing philosophy from raising any real challenges to the way our society operates, preventing it from being able to raise claims that may be seen as "dangerous," and thus, ultimately, ensuring that philosophy always happens to find itself aligned with the interests of hegemony.

The defender of the MI will likely argue that we can distinguish between intuitions that are fundamentally rooted in hegemony and those that are not. Yet it seems difficult to draw a principled line between intuitions that rely on racist, sexist, or other hegemonic assumptions in some way and those that do not. Ideally, intuitionists could commit themselves to rejecting any intuition that directly upholds hegemony while accepting non-hegemonic intuitions. Yet this is not feasible for at least three related reasons. Firstly, it might be impossible to draw such a line at all, as the way hegemony affects how we think about the world is highly complex and many intuitions likely spring from it in unconscious ways. We may not consciously recognize that a given intuition involves a shade of subtle racism, for example, or that it is tacitly homophobic. For this reason, even intuitions that at first appear to be non-hegemonic might uphold hegemony in subtle ways. For example, it is possible that even the Mary's room thought experiment tacitly relies upon particular cultural biases among those doing philosophy that relate to how they categorize color, and other communities might construe or classify colors differently. Secondly, it seems impossible to draw such a line in a non-circular or non-question-begging fashion, as ultimately the matter of whether or not a given intuition is rooted in hegemonic assumptions will itself generally be determined intuitively. Finally, drawing such a line prior to doing philosophy and as a starting point for philosophy makes determining which intuitions are

hegemonic and which are not a pre-philosophical matter that would, thus, itself seem to be outside of the bounds of philosophical investigation. As a result, such a move may make philosophy unable to challenge hegemony. And again, even if it were feasible to draw such a line, there would still be no good reason to assume that non-hegemonic intuitions are any more accurate than hegemonic intuitions are. Ultimately, totally rejecting the MI has a dialectical advantage over merely rejecting particular intuitions, as the defender of the MI has no coherent way to distinguish between those intuitions that are fundamentally rooted in hegemony and those that are not or those that are reliable and those that are not.

There is only one solution. If philosophy is to be liberated from hegemony, philosophers must reject the MI. As an alternative starting point, I will defend a method based on the Principle of Sufficient Reason, or PSR—i.e., the view that, for each fact F, F has a cause that fully explains F. This method will reject arguments from intuition in favor of arguments from explicability, i.e., arguments that rest on the assertion that there are no inexplicable facts. Beginning with the PSR will make philosophy as a whole less vulnerable to the biases of its practitioners. Accordingly, the remainder of this essay will be devoted to defending the PSR as a possible basis for philosophy moving forward.

In the same way our imagined intuitionist is placed at a dialectical disadvantage by merely rejecting certain intuitions while retaining others, our imagined non-rationalist is placed at a similar disadvantage insofar as they must draw an arbitrary line between those explicability arguments they accept and those they reject. Della Rocca makes this clear in his paper titled "PSR," where he argues that, because non-rationalists (including most analytic philosophers) already accept many explicability arguments, i.e., arguments that begin from the assumption that a particular event is explicable or has an explanation, they must either draw a principled line between those explicability arguments that they accept and those that they reject—or accept the PSR in its entirety. Leibniz provides us with a simple example of an explicability argument in the form of a discussion of Archimedes' scales:

> [Archimedes] takes it for granted that if there is a balance in which everything is alike on both sides, and if equal weights are hung on the two ends of that balance, the whole will be at rest. That is because no reason can be given why one side should weigh down rather than the other. (Della Rocca 2010, 2)

This is, of course, because, if one of the weights went down and the other went up, and both weights were identical (and all other considerations were equivalent), such an event would be inexplicable. Yet most opponents of the PSR would not wish to suggest that such a thing is possible. At the same time, however, to accept that no things are inexplicable would simply be to accept the PSR. Therefore, the denier of the PSR must draw a principled line between those explicability arguments that work and those that do not, and doing so has proven to be an extraordinarily difficult task. Della Rocca thus says:

> One can see my argument as highlighting an important dialectical advantage that our imagined rationalist (i.e., me) has over our imagined non-rationalist (i.e., you). I have been arguing that the non-rationalist who accepts some explicability arguments has no non-question begging way to avoid the rationalist position, i.e., no non-question begging way to avoid the PSR. By contrast, the rationalist who accepts the necessitarian implication of the PSR is not under any pressure, as far as I can see, to accept the non-rationalist position, i.e., to deny the PSR. In this way, the rationalist position is internally coherent in a way that the position of the non-rationalist who accepts some explicability arguments is not. (2010, 13)

Della Rocca's argument shows that accepting the PSR (all explicability arguments) is a far stronger position than doing the contrary.

However, it is not immediately evident how an acceptance of the PSR helps us with philosophy's general subservience to hegemonic identities, even after we have accepted that the PSR is true. After all, the mere fact that a philosophical principle is true does not necessarily make its adoption useful in the fight against racism, sexism, homophobia, or any other bigoted superstition. One might ask, "How does reasoning based on the PSR differ from reasoning based on the MI such that it is able to avoid the biases currently ingrained in analytic philosophy?" The answer is that philosophical reasoning moving forward from the PSR can be made far more objectively than reasoning derived from the MI. Rather than relying on arbitrary intuitions, in accordance with the MI, one who starts from the PSR proceeds from explicability arguments based on well-defined axioms. This does not mean that philosophy can ever exist outside of cultural and social

constraints. While the PSR liberates philosophy from such constraints as much as is methodologically possible, the practitioners of philosophy are, in general, going to be both seeking employment and avoiding putting forth views that could lead to their assassination. Rationalists, i.e., defenders of the PSR, have never stood outside of history. Spinoza, for example, held reactionary views regarding both women and pagans. However, these parts of his overall picture can be shown to lack justification in accordance with the PSR, and can be rejected, while the vast majority of the Ethics has stood the test of time as an incredible work of rationalist philosophy. Unlike the MI the PSR does not place intuitions beyond the scope of philosophy, it places all our assumptions about the world on the table and holds them up to scrutiny. While rationalists themselves may be racist, sexist, homophobic, etc., arguments made from the PSR, precisely because and insofar as they cleave to this criterion of thoroughgoing rationality, will never endorse such biases. This is parallel to how the racism of mathematicians is of little consequence regarding the accuracy of mathematical proofs per se. On the other hand, the MI, as we have seen, often does sanction such superstitions.

However, Spinoza did in fact have reactionary views about women and pagans. So while rationalism will not ultimately justify the biases of hegemonic identity, it might not directly challenge them in the way those of us committed to liberatory politics would like. It is entirely possible for hegemonic biases to maintain themselves within a philosophical discipline that accepts the PSR, and such biases, while no longer able to damage the arguments themselves, will still damage what questions are asked by philosophers and the influence that academic philosophy exerts over society as a whole. Tragically, it seems that a shift in the philosophical paradigm in favor of the PSR over the MI will not single-handedly solve racism within the field; efforts both to diversify the field and to fight the various bigotries within it remain necessary. Yet while embracing the PSR will not single-handedly solve bias, the choice to accept the PSR over the MI is fundamentally an anti-racist one. While hegemonic biases can still exist within a philosophical paradigm dominated by the PSR, they will not be able to go unquestioned. In conclusion, it is necessary to replace the method of intuition with the Principle of Sufficient Reason as a basis for future philosophical endeavors, both because the MI is deeply flawed and the PSR is true and because the MI legitimizes bigoted superstitions while the PSR does not.

REFERENCES

American Academy of Arts & Sciences. 2021. "Racial/Ethnic Distributions of Degrees in Philosophy." https://www.amacad.org/humanities-indicators/ higher-education/racialethnic-distribution-degrees-philosophy.

Della Rocca, Michael. 2020. *The Parmenidean Ascent*. Oxford: Oxford University Press.

Della Rocca, Michael. 2010. "PSR." *Philosophers' Imprint* 10 (7): 1–13.

Department of Justice, Government of Canada. 2018. "Bill C-16: An Act to amend the Canadian Human Rights Act and the Criminal Code."

https://www.justice.gc.ca/eng/csj-sjc/pl/identity-identite/statement-enonce. html.

Descartes, Rene. 1989. *Discourse on Method and the Meditations*. Translated by John Veitch. Buffalo, NY: Prometheus Books.

Fine, Kit. 2012. Interview with 3am Magazine. https://316am.site123.me/articles/ metaphysical-kit?c=end-times-archive.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127–36.

Lewis, David. 1983. "Introduction." In *Philosophical Papers*, Volume 1, edited by David Lewis, ix-xii. New York: Oxford University Press, 1983.

Peterson, Jordan B. 2020. "The Brilliant Arrogance of Nietzsche." https://www. youtube.com/watch?v=h2n756iAgX0.

Thomson, Judith Jarvis. 2013. "How It Was." In *Portraits of American Philosophy*, edited by Steven M. Cahn, 47–62. Lanham, MD: Rowman and Littlefield.

Thomson, Judith Jarvis. 1985. "The Trolley Problem." *The Yale Law Journal* 94 (6): 1395–1415.

compos mentis

Concepts, String Theory, and Separate Mental Reality

Candice Wiesner

Eastern Michigan University

ABSTRACT

Concepts are an enigma. How is it we are able to interact with abstractions such as mathematics, or, concepts? This perplexity has proven ripe fruit for philosophers. For many philosophers, concepts have a definite structure and exist in a separate realm. Conversely, for psychologists, concepts exist in the head and as such do not have a definite structure; concepts take on a probabilistic structure for psychologists. These two views are conventionally seen as at-odds, however, this paper presents arguments supporting both views. Both views are accommodated when we allow for separate realms, separate realities, to exist. String Theory is consistent with there being such realms. In the following, a number of concepts are combined, each in essence consistent with accepted ideas, resulting in a qualitative explanation for concepts, with the promise of an eventual clear cut basis for understanding consciousness in general.

If there were no mathematicians, would there be prime numbers, and if so, where would they be? We can certainly think about prime numbers, therefore, they exist in our consciousness; but if we do not think of them, they would not be anywhere. Mathematical concepts do not exist as anything more than representations in our conventional world. This puzzle has long been proven ripe fruit for philosophers; concepts are an enigma. How is it we are able to interact with abstractions such as mathematics? One suggestion is that perhaps mathematical concepts are somehow located 'in the physics'. However, such a drastic solution is not necessary to explain our ready access to mathematical ideas. Instead, this paper puts forth neural networks as providing an adequate explanation to abstract concepts. In shifting the locus of concepts (or mathematical thinking) to another realm what begins to emerge is an argument for the existence of some kind of separate realm. Perhaps we have only replaced one mystery by another; why should such a separate realm exist at all? String theory, involving as it does spaces having more dimensions than the usual three, and also a non-unique vacuum state, is consistent with there being such a 'separate realm'. This paper puts forth the idea that some aspects of mentality involve a realm of reality largely, but not completely, disconnected from the phenomena manifested in conventional physics. I will support this analysis by first explaining the difference between philosopher's idea of concepts and psychologist's idea of concepts. While these two views are conventionally seen as at odds, the analysis put forth accommodates both positions. Next, I argue that it is possible to accommodate both views through the acceptance of separate realities. Explaining string theory, we see it is consistent with there being such a 'separate realm'. Following, I bring the analysis together with explaining how one might access such a separate realm. It is natural to postulate a shared mental 'bubble', realized by neural networks, whose contents are available to life-forms. I will also address critics concerns such as the empirical testability of such an analysis. In the following, a number of concepts are combined, each in essence consistent with accepted ideas, resulting in a qualitative explanation for concepts, with the promise of an eventual clear cut basis for understanding consciousness in general.

Concepts are a contested issue. To philosophers, many believe that concepts are abstract objects. Since abstract objects do not have a spatial location, it follows that concepts are not located anywhere. In contrast, psychologists insist concepts are located "in the head." They characterize concepts as mental representations.

According to Shapiro, "whether concepts are abstract objects, or whether they are physically realized mental representations, is an issue about the ontology of concepts - about what they are" (Shapiro 2019, 81). This raises questions about the nature of their structure. For many philosophers, concepts are like definitions: they consist of a collection of other concepts that constitute individually necessary and jointly sufficient conditions (Shapiro 2019, 82). Take, for instance, the concept of 'triangle'. The conditions necessary for triangularity are three-sided plane figure. If we see a four-sided figure, we know this does not meet the conditions necessary for a triangle. For psychologists, according to Shapiro, "a dominant view treats concepts as something like packages of information about a topic" (Shapiro 2019, 82). Take, for example, the concept 'dog' consists of a potentially open-ended list of features such as four-legged, paw-footed, fur, snout, and tail. Some of these features may be more representative than others. Whether the object falls into the extension of a concept, that is, an instance of the concept, depends, then, not on whether the object satisfies some set of necessary and sufficient conditions, but on whether the object possess enough of the features that comprise the concept (Shapiro 2019, 82). To the philosopher, a definition is either satisfied or it is not. It makes little sense to think about a three-sided plane figure as being only 45% triangle. For the philosopher, it either is or is not a triangle. To the psychologist, concepts allow for graded application or judgements. This is similar to the judgement a whale is not a mammal. Owing to the fact whales exhibit less mammal-like characteristics: they do not have fur and live in the water, for instance. For the psychologist, the ability to judge helps us categorize things that may not be representative of all the features of that kind. We can notice that the ontological and structural questions are independent of one another. Shapiro writes,

> One might hold that concepts are abstract objects with a probabilistic structure. Or, one might conceive of concepts as mental representations with a definitional structure. It just so happens that many of the philosophers who have defended the idea that concepts are abstract objects are also committed to their having a definitional structure; and psychologists who assume that concepts are located in the brain have abandoned the idea of definitional structure. (Shapiro 2019, 82)

The two views need not be at odds, necessarily. Surprising as it may seem, separate mental realities can allow for both views to hold. This is the idea that some aspects of mentality involve a realm of reality largely, but not completely, disconnected from the phenomena manifested in conventional physics. In a separate reality abstract concepts exist. When we interact with these abstract concepts in conventional reality, they are mental representations only because the true concept exists in another reality. Concepts existing in a separate mental reality have a definitional structure, while the mental representations we construct have a probabilistic structure. These mental representations are based on the definitional structure, but, because they are merely representations, they take on a probabilistic structure. In fact, concepts in conventional reality need to be probabilistic because it is how we identify kinds that assist in the classification of particulars into kinds. Take, for example, the concept 'bird'. If the definitional structure includes "critter that flies," birds that do not fly could not be classified as birds. We use the flexibility of concepts in conventional reality to make easier the identification of kinds. Biologically speaking, it makes sense to lessen the cognitive load to allow for flexibility in identifying; think of how much more information one would need to remember if we sterilized the process of identifying. Psychologists are right to explain how we interact with concepts in conventional reality, while philosophers are correct in identifying a separate mental reality.

According to Brian Josephson of the Department of Physics at the University of Cambridge, "the idea of a disconnected realm does have precedents, for example in the way two of the fundamental forces (the strong and weak forces) play no role in large areas of physics and chemistry, whilst in other contexts they have a very important part to play" (Josephson 2003). The fundamental forces are the gravitational force, the electromagnetic force, the weak nuclear force, and the strong nuclear force. What Josephson is referring to is that since the weak and strong nuclear forces act over an extremely short range, the size of a nucleus or less, we do not experience them directly, although, they are crucial to the very structure of matter. Nuclear forces determine the relative abundance of elements in nature because they indirectly determine the chemistry of the atom. Everything we experience, on the tiniest level, owes itself to the strong and weak forces. Note that string theory, involving as it does spaces having more dimensions than the usual three, and also a non-unique vacuum state (in fact, a very large number of such states), is consistent with there being such a 'separate realm', in a way

that the Standard Model, with its unique vacuum state contained within a limited number of spatial dimensions, did not (Josephson 2003).

The idea of a separate reality is a strange one; we were raised to believe the word 'universe' meant everything. Speculative though string theory is, there is reason to take it seriously. I am going to describe the possibly of a separate reality in three parts. In part one I will highlight a Noble Prize winning idea and a profound mystery that arises from the results of these findings. In part two, a solution to this mystery will be provided. It is based on an approach called string theory; this is where the idea of a multiverse, or separate reality, will come into the story. Finally, in part three I will describe a cosmological theory called Inflation that will pull all the pieces of the multiverse theory together.

Part one starts in 1929 with the great astronomer Edwin Hubble establishing that space itself is stretching, or expanding. This was revolutionary as the prevailing wisdom was that on the largest of scales the universe was static. In 2011, scientists Saul Perlmutter, Brian Schmidt, and Adam Reiss were awarded the Nobel Prize in physics after making pain-staking observations of numerous distant galaxies to find that the universe's expansion was not slowing down, but was, in fact, speeding up (Greene 2012). This raises the question: what is causing the universe to expand at an ever-increasing speed? Einstein's theory of gravity provides a compelling answer. According to Einstein, gravity not only attracts but also pushes things apart. Einstein believed the universe to be full of a sort of cosmic mist and the gravity generated by this mist would be repulsive (Greene 2012). Repulsive energy explains why the universe is not only expanding, but also speeding up in its expansion. The mystery (as promised) is that when astronomers worked out how much dark energy must be infused in space to account for the cosmic speed up they found a spectacularly small number: 1.38×10^{-123} (Greene 2012). The mystery is in explaining this peculiar number. Why is it we live in a universe with this particular amount of dark energy instead of any other? The only approach that has made any headway to explain it invokes the possibility of multiple universes.

Moving on to part two. According to physicist Brian Greene,

> The central idea of string theory is if you examine any piece of matter, ever more finely, at first you will find molecules. Then, you find atoms and subatomic particles. The theory says if you could probe smaller, much smaller than we can with existing technology,

> you would find something else inside these particles: a little, tiny, vibrating filament of energy. A vibrating string. (Greene 2012)

Just like the strings of a violin, they can vibrate in different patterns. The different kinds of vibrations produce different kinds of particles: electrons, quarks, photons, etc. They all arise from vibrating strings, a kind of cosmic symphony. All of the richness of the world around us emerges from the music of these little, tiny strings. The mathematics of string theory has internal inconsistencies, however, unless you allow for extra dimensions of space. We are all used to the three dimensions of space: height, width, and depth. String theory says that on fantastically small scales there are additional dimensions, crumpled to a tiny size, so small we have not detected them (Greene 2012). Even though the dimensions are hidden, they would have an impact on things we can observe because the shape of the extra dimensions constrains how the strings can vibrate. In string theory, vibration determines everything: the strengths of forces, particle masses, and most importantly, the amount of dark energy, would be determined by the shape of the extra dimensions (Greene 2012). If we knew the shape of the extra dimensions, we should be able to calculate these features. However, we do not know the shape of the extra dimensions. As the list for potential candidate shapes soared, some scientists dismissed the idea of string theory as never being able to empirically testable. Other scientists have turned this perplexity on its head taking us to the idea of the multiverse. According to Greene, "each of these shapes are on an equal footing with every other shape. Each is as real as every other in the sense that there are many universes each with a different shape for the extra dimensions" (Greene 2012). This proposal has a radical effect on the amount of dark energy reveled by the Nobel Prize winning results.

If there are other universes, each with its own shape for the extra dimensions, then the physical features of each universe will be different (Greene 2012). In particular, the amount of dark energy in each universe would be unique. The mystery of the number 1.38×10^{-123} takes on a wholly different character. In this context, the laws of physics would not explain one number because there is not only one number; there are many numbers. We find ourselves in a universe with the particular amount of dark energy we have measured because our universe has conditions hospitable to our form of life. It may seem like we are allowing for a lot of speculation, that these ideas are mere conjecture. However, we may be able to account for multiverse theory. This brings us to part three.

When we speak of the Big Bang, we often have an image of a cosmic explosion that sent space rushing outward. However, the Big Bang leaves out something important: the bang. The theory tells us about what happened after the Big Bang, but gives no insight to what powered the bang itself. As stated by Greene, "this gap was filled by an enhanced theory of the Big Bang called Inflationary Cosmology. It identifies a particular fuel that would naturally generate an outward rush of space" (Greene 2012). The important detail for us is that this fuel is so efficient, it is virtually impossible to use it all up. It is likely that the Big Bang was not a one time event. The fuel not only generated our big bang, but countless others as well, each giving rise to its own separate universe. Our universe becomes one in a cosmic bubble bath of universes. If we meld this with string theory, each of these universes has extra dimensions. The extra dimensions take on a wide variety of shapes. The different shapes yield different physical features and we find ourselves in this universe because our universe has the right amount of dark energy for our form of life (Greene 2012). A big question remains: could we ever confirm the existence of other universes?

There is a way that in the future it might happen. The inflationary theory holds that when the bang happened, it generated a kind of cosmic finger print. A pattern of slightly hotter and slightly cooler spots throughout space that modern telescopes have observed. If there are other universes, the theory predicts that every so often universes collide. If our universe got hit with another, that collision would generate subtle temperature variances across space that we might one day be able to detect (Greene 2012). As exotic as this theory is, it may one day be grounded in observation.

So far we have established the possibility and precedent of separate realms. The logical question that arises next is how do we interact with these possible realms? Consider all of what the brain does in visual perception. Here, the primary information from the visual receptors goes through various levels of processing until it ends up as a high-level representation of the content of the visual field. It is not unreasonable to identify mathematics as a similar process, except that higher levels of abstraction are involved in this case. With the visual case, the mechanics are straightforward: the visual field typically contains, for example, edges, for which abstraction a dedicated neural system has evolved, related to our ability to perceive edges. It is hard to see why we should have such ready access to higher mathematical abstractions having little connection with experience (Penrose

1994). One resolution of the problem would be for mathematical concepts to be in some way 'in the physics', rather than being emergent properties of brains. Such a drastic solution is not necessary to explain our ready access to mathematical concepts; neural networks can provide an adequate explanation. The explanation is of a biological character, taking account of the fact that information processing is an essential component of biological functioning, but with only very specific informational processes having a life-supportive character. According to Josephson,

> While it is commonly taken that the informational processes involved are mediated by ordinary physical means, it is not a logical necessity that this should be the case. Some informational processes in an organism are specialized to the nature and circumstances of the organism concerned, however, some have a more abstract and universal character, and so could be mediated by a quite different system with which individual organisms would interact. (Josephson 2003)

Next, we observe that a form of proto-life, defined as the first life, or fluctuation patterns surviving longer than typical patterns do, can be hypothesized as occurring at the Planck scale. Briefly, the Planck scale is a tiny number; it is hard to wrap ones mind around how small it is. An analogy might help: if you were to take an atom and magnify that atom to be as big as the observable universe, then the Planck scale would be roughly the size of the average tree. The tree is to the observable universe, as the Planck scale is to an atom (World Science U 2015). Evolution of such proto-life is expected to involve evolution of the accompanying informational systems also. We get to the proposed model by supposing that the ordinary physical component and the informational component can evolve separately and that the informational component can even survive the creation and destruction of individual universes, remaining as an ever-present background with which new universes, Planck scale fluctuations, and more developed life forms can all beneficially interact. Assuming an indefinitely extended time scale, the most persistent part of the informational background can evolve indefinitely, so that its dynamics might come to include features corresponding to mathematical concepts and operations as well.

We need to add another piece of detail to our model. In order that it can model individual thought, we suppose that individual life forms can perturb the background state so as to create a localized 'thought bubble'. The concepts exist in a separate realm and our neural networks have evolved with the capacities to access such a realm. We create mental representations to make use of the information we access from these separate realms. String theory is consistent with there being separate realms.

The problem such an analysis has to face is that of explaining that, if such a mechanism for concepts and mathematical thinking exists, why do we not have ready access to these separate realms in a more concrete way? In other words, why does consciousness mostly interact with conventional physics? This is not an insuperable objection. We must bear in mind that in the biological realm the phenomena that manifest are governed not only by what is physically possible, but also by which of those physically permitted possibilities are likely to be of overall benefit to the organism concerned (Josphenson 2003). An undifferentiated sensitivity to all the realms would result in chaos, and tend to be disadvantageous rather than of benefit, leading to the individual being overwhelmed by the resulting chaos. This would be distracting and interfere with constructive activity. The right way to think about concepts is that we can access them in favorable conditions, such as learning or memory tasks, but we are not overwhelmed by our abilities to access other realms. True concepts are not located "in the head," but, after accessing these separate realms where concepts exist, the mental representations we create are accessed in our minds. We use the flexibility of these models to understand better and interact with our environment. This lessens the cognitive burden placed on us by allowing for flexibility in our models to account for the classification of particulars into kinds.

The implications of this speculative theory offer interesting insights into the understanding of consciousness in general. The picture that emerges is one of an everlasting informational system that has existed through many, many creations and destructions of universes. This is a shared system, among all beings. In essence, there is a part of us that is shared by all beings. While our biological organism appears separate from all other living creatures, this is not the full story. It is a story of compassion; all beings are connected through their shared participation of the informational system. It also implies a reduced role of our biological system and the importance we place on it. The Buddhist concept of

'we are not these bodies' applies here. Conventionally, yes, we are. However, we when exit the mundane world and begin to see ourselves outside the ego construct, that is made to protect our very biological system, we can play with the notion that we are not these bodies only. We are connected to a much deeper, much more mysterious force. An understanding such as this, if proven correct, would alter the very way we conceive consciousness.

We set out to show that philosopher's idea of concepts and psychologist's idea of concepts do not need to be at odds. Both hypothesis can be accepted and accommodated for if we allow for separate realities. String theory is consistent with there being such separate realities. The evolution of our neural networks includes mechanisms designed for abstractions, such as edges. It is not illogical to reason our neural networks have evolved to perceive other abstractions such as concepts (also referred to as mathematical thinking), or having connections with separate realms. As we can penetrate the background informational system, we can formulate what is referred to in this paper as a 'thought bubble'. This is the basic interaction between us and the separate realm. We use the thought bubble to create mental representations in the mind of concepts. We can then use these models in conventional reality by applying them to the classification of kinds to particular kinds. The flexibility of these models reduces the cognitive load; we do not need to categorize in finer and finer detail. Both the psychologist and philosopher's view correctly identify, they are simply identifying different aspects of the same thing. Mainly, concepts exist in a separate realm with a definitional structure and when we interact with them we create mental representations of concepts with a probabilistic structure. This exhilarating picture is the crossroads of modern science with ancient philosophical ideas. To accommodate something that is as foreign to our current physical picture as is the phenomenon of concepts, we must expect a profound change- one that alters the very underpinnings of our philosophical viewpoint as to the nature of reality.

REFERENCES

Josephson, Brian. 2003. "String Theory, Universal Mind, and the Paranormal." *Publisher not listed*.

Penrose, R. 2005. "Shadows of the Mind." *Oxford University Press*.

Shapiro, Lawrence. 2019. "Embodied Cognition."

Sheehy, Noel, Chapman, Antony J, & Conroy, Wenday. 2016. "Biographical dictionary of psychology."

WorldScienceU. 2015. "Can you put the Planck scale into PERSPECTIVE? How small is it?" *Youtube.com*.

cognethic.org