compos mentis

Undergraduate Journal of Cognition and Neuroethics
Volume 7, Issue 2

compos mentis

Student Editor
Lisa Gawel

Production Editor
Zea Miller

Publication Details
Volume 7, Issue 2 was digitally published in June of 2019 from Flint, Michigan, under ISSN: 2330-0264.

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

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

Table of Contents

compos mentis

How Responsible Are You For Your Actions?

Nadia Ayensah
Augustana College

ACKNOWLEDGMENTS
I would like to extend my profound gratitude to Dr. Heidi Storl for her immense contribution to the success of this research. I would also like to thank the Ayensah family and the Siaw family for their constant support and encouragement throughout my study.

ABSTRACT
In a world where our sense of responsibility rests solely on the existence of a notion of morality and free will, how do we make sense of responsibility when neuroscientific findings have been shown to trim morality and free will? How can a civil society held together by justice emanating from a retributive sense of responsibility keep running when the basis of retribution has been undermined? This paper examines the relationship between morality, free will, responsibility, and neuroscience so as to determine whether we can justifiably attribute moral responsibility. In this paper, I argue that moral responsibility can still be attributed, but only through a lens of free will skepticism. How would such a responsibility materialize in our contemporary society? What problems will it encounter? My research seeks to draw a comprehensive plan for acting on this new sense of moral responsibility through an examination of findings in philosophy, neuroscience and psychology.

KEYWORDS
Basic Desert Responsibility, Free Will, Free Will Skepticism, Morality, Moral Enhancement, Neuroscience, Retributive Justice, Rehabilitation System, Take-Charge Moral Responsibility

DEFINITION OF KEY TERMS

Basic desert sense of responsibility: The notion that an agent deserves punishment or reward for his actions because he acted out of free will.

Existentialism: A philosophical expression of the anxiety that there are no secure foundations for meaning and morality, no deep reasons that make sense of the human predicament.

Free will: The notion that we are in charge of our actions, and the ability to have done otherwise in a given situation. This endorses the basic desert sense of moral responsibility.

Free Will Skepticism: The idea that punishment cannot be justified by the mere possibility of free will or the feeling of freedom.

Morality: The domain of human life where we evaluate of character and action in accordance with rules that license condemnation and punishment.

Moral Sedimentation: The past patterns of proscription that shape present attitudes and guide current behavior.

Sedimentation: The phenomenon of experiencing the world and acting in it through a filter of the past without necessarily realizing it.

Take-Charge Moral Responsibility: The capacity to change one's future behavior when given the necessary means.

PART I

Introduction

For as long as humans have existed, there have been questions, both subtle and direct, on the essence of human existence and the source of the morality we hold up our actions against. This anxiety that there are no secure foundations for meaning and morality and no deep reasons that make sense of the human predicament has been termed as existentialism by various neuroscientists, philosophers and experts in relevant fields. Authors Gregg D. Caruso and Owen Flanagan have divided the concept of existentialism into three waves—the first wave dealing with the anxiety that there is no God, from which the sense of human essence was derived; the second with the anxiety that the concept of human good from which essence was deemed to come from during the European Enlightenment; and the third with anxiety the that the findings of science with regards to evolution and neuroscience nullify our concept of having the ability to conceive human essence (Caruso and Flanagan 2018).

According to Caruso and Flanagan, the first wave of existentialism was dominated by philosophers such as Kierkegaard, Dostoevsky and Nietzsche. This wave, according to Flanagan and Caruso, was "characterized as the displacement of ecclesiastical authority and a consequent anxiety over how to justify moral and personal norms without theological foundations" (Caruso and Flanagan 2018, 3). During this period, anxiety over human essence and the source of morality started to spread among people because there was a significant shift from total ecclesiasticism to visible signs of traits of atheism and nihilism. With the increase in questions as to whether or not human life and our notion of morality have any essence outside of the belief in God, there was a panic to grapple some form of meaning and associate it to humans. This then led to the second wave of existentialism in the eighteenth century, where the concept of morality and human essence being derived from God were nonexistent.

The second wave of existentialism is said to have emerged during the European Enlightenment, and moved from the idea of morality stemming from God to the idea of morality stemming from the notion of a common good. In this period, this notion of common good had to do with the fact that "we could count on human goodness and human rationality to make sense of meaning of morals" so as to give purpose to humans and guide human action as a result

(Caruso and Flanagan 2018, 3). According to Caruso and Flanagan, the works of philosophers Jean-Paul Sartre and Albert Camus lean very much towards the ideals of the second wave of existentialism. The idea in this period was that even if there was no God, we could still count on the existence of the common good to be the source of morality and human essence. Just like the view on God being the source of morality and essence couldn't stand, this notion of morality and essence stemming from the common good was also brought down. It was seen that this notion couldn't stand when human actions resulting in colonialism and holocausts came to light. Philosophers such as Sartre and Camus were horrified that humans could do this to other humans, and hence, the anxiety that essence and morality did not stem from this notion of common good arose.

The third and most recent wave of existentialism has to do with the findings of science, more specifically, neuroscience. Existential anxiety in this wave has been brought about by the findings of neuroscience providing evidence to the Darwinian claim that humans are just animals, basically nullifying the humanistic image of persons. The third wave of existentialism has been defined by Caruso and Flanagan as the "twenty-first century anxiety over how contemporary neuroscience helps secure in a partially vivid way the message of Darwin" that humans are indeed not special in any way, but rather, are merely "one kind of primate among the two hundred or so species of primates" (Caruso and Flanagan 2018, 5). This affirmation by neuroscience has led to questions surrounding the existence of free will, the relationship between the mind and the brain, the transcendence of morality and meaning and the difference between humans and animals.

For the purposes of being concise, I will only pay attention to the third wave of existentialism—neuroexistentialism. In this paper, I will examine the relationship between morality, free will, responsibility, and neuroscience so as to determine whether we can justifiably attribute moral responsibility. This paper will show that although neuroscience trims morality and free will, it affirms that we ought to be held responsible for our actions so as to preserve justice. It also shows that human behavior can be remodeled, although the thought of such remodeling can be unethical.

Morality

Due to the various existentialisms, especially the third wave, there has been a vigorous search into the source of morality. What then is this morality? In this

paper, I deem a moral person to be one who knows what is right from what is wrong, and follows rules to do the right thing, even if they do not will to do so. When we look at human behavior, the act of caring for others, even where it will be a disadvantage to the actor, seems to be tightly woven into our thinking because we see it to be the morally right thing to do. How then is this notion of caring for others woven into human existence? How does it fit in a world where survival depends on competition? Well many philosophers such as Patricia Churchland have considered these same questions and have come up with possible sources of morality as well as objections as to why morality cannot stem from some of these sources.

According to Churchland, evolutionary biologists attributed the source of morality to altruism, which is the "disinterested or selfless concern for the wellbeing of others, especially as a principle of action" per the Oxford English Dictionary. The reasoning behind this was that since morality entails performing actions for others which could be disadvantageous to the actor—it embodies traits of altruism. The main objection to this line of reasoning is that humans are wired to care about our own survival and well-being and hence, having altruistic genes from inception would have been a disadvantage to those who bore them, so, altruism could not have been a by-product of evolution. The argument here is that if some organisms had altruistic genes, they would have been killed off because organisms with non-altruistic genes would have taken advantage of the former since survival was contingent upon an organism's ability to care for itself regardless of the means. Patricia Churchland then argues that if this is the case, then morality could not have come from evolution; it must have been taught (Churchland 2018).

Churchland then talks about how religion has been thought to be the "watershed of moral values" (Churchland 2018, 27). This proposed source of morality was also received with much criticism, although it seemed to be plausible on the surface. Like Christianity, most religions have a set of rules by which they act, and from which such actions should be moral. With Christianity, one of such set of rules of conduct is the Ten Commandments found in the Holy Bible. Christians believe that if one wants to live a moral life, one ought to live by the Ten Commandments, and hence, people believe that rules of conduct such as this must be the source of morality. Churchland outlines two main objections to the notion of religion as the source of morality: the disparities between the emergence

of religion and the existence of humanity and the existence of humans whose view of God differs from that required for religion to be the source of morality. With the first objection, Churchland highlights that religion has only existed for about 10,000 years, and humans have existed for about 250,000 years, and the concept of morality is thought to have existed longer than that; hence, religion cannot be the source of morality. With the second objection, she points out that in order for religion to be the fountainhead of morality, God or whatever deity is being worshipped must be seen as a law giver and a punisher. There however exist certain groups such as the hunter-gatherer groups that are highly moral, but do not see God as a law giver and a punisher. Hence, with this objection also, morality could not have emerged from religion (Churchland 2018, 27).

If neither altruism nor religion could have given rise to morality, what then is the source of morality? Churchland answers this question in her consideration of neural connections in the brain as a possible source of morality. Here, Churchland considers the state of brains as at the dawn of existence: from the time period where warm-blooded organisms had to compete with cold-blooded organisms for survival. During this period, warm-blooded organisms already had some sort of advantage over the others because they could still search for food when the sun had gone down. They were however also at a disadvantage because they needed much more food than the cold-blooded animals did to survive. Hence, there must have been a way in which warm-blooded organisms could have survived in the competition. Churchland suggests that these organisms could have survived "by ramping up their postnatal learning abilities", which included rigging the brain of mature animals to care for the infants until they were mature enough to survive on their own. With this then, the mature animals were made parental by changing their sense of self-survival that had to do with a "care of me" to a "care of me-and-mine" (Churchland 2018, 30). According to Churchland, the ability to rig an animal's brain for it to start caring for other animals than itself on its own demonstrates clearly that the bonds we form and the love we feel are embodied in the neural circuitry. Even from this point in Churchland's argument, we can see that neuroscience seems to support the notion of morality of some sort (Churchland 2018, 30).

What exactly in the brain then is responsible for this development of parental capabilities? Patricia Churchland answers that oxytocin and vasopressin receptors in certain crucial parts of the brain are what adjust the circuitry in the brain to

facilitate parental behavior, which entails bonding. Churchland argues that this bonding pattern, regulated by oxytocin and a palette of other neurochemical and neurohormones working in their proprietary circuitry, is the basic platform for morality, where morality has to do with for caring for others. If that is the case then, circuitry supporting this cluster of behaviors is the neural platform for morality.

Relationship between Neuroscience and Morality

How are neuroscience and morality connected? If they are connected, does neuroscience affirm the concept of morality or does it deny the concept? These are pressing questions that have been asked and taken on by philosophers, neuroscientists and psychologists alike. In this section, I will look at the work of philosophers Paul Henne and Walter Sinnott-Armstrong. The first established question to be addressed in this sub-section is how neuroscientific findings and the concept of morality as applies to human actions are connected. In "Does Neuroscience Undermine Morality?" by Henne and Sinnott-Armstrong, the philosophers talk about how neuroscience does not necessarily undermine all moral judgments. In this work, the authors start by considering the main reason why people think that neuroscience undermines all moral judgments: that all moral beliefs have supposedly been shown to stem from an unreliable source. Here, Henne and Sinnott-Armstrong affirm that yes, some moral judgments have been shown to come from unreliable processes. With this affirmation, the main argument now takes this form: if moral judgments are shown to come from an unreliable process, then the agent of the moral judgments does not know whether or not the judgments are correct, and if the agent does not know that the judgments are correct, they cannot be assumed to be so. They further argue that we cannot generalize neuroscientific findings with all moral judgments since there are different kinds of moral judgments which result in activities in different parts of the brain. Personal moral judgments, for instance, give rise to activity in that part of the brain responsible for social and emotional processing while impersonal moral judgments activate the part responsible for working memory (Henne and Sinnott-Armstrong 2018, 58). Greene et al., in an experiment on the emotional engagement in moral actions, come up with two scenarios—the trolley and the footbridge scenarios—both in which one person is sacrificed to save five people. What the research showed was that people were more willing to redirect a trolley headed for five people in the direction of a single person than pushing one person

off a footbridge to prevent five people from dying. The explanation given by Greene et al. for this discrepancy is that the footbridge scenario produces much more emotion, while the trolley scenario is more distant and feels void of emotion, thereby resembling a nonmoral judgment more than a moral judgment (Greene et al. 2001). Just from the observation of this experiment, we can see the reasoning behind the argument that the sample of moral judgments researched on by neuroscience cannot be representative of all moral judgments. The philosophers argue that these findings about different parts of the brain being activated for both actions that seem to be similar bring more understanding to our conception of pain that it is not a blanket emotion to which we should have the same response in each case. Rather, this shows that even within our conception of pain, there is a heavy variance, and hence, decisions in every situation should be catered to the specific kind of pain corresponded in the brain.

It is with Greene's analogy that Henne and Sinnott-Armstrong argue that neuroscience trims morality by reshaping our understanding of the concept since it shows which parts of the brain are activated during the formation of certain moral judgments. They argue that neuroscience trims our judgments in the sense that if it has been shown that judgments on inequity and those on homosexuality activate the same parts of the brain, then we cannot say judgments on inequity come from a reliable source while those on homosexuality do not. Neuroscientific findings basically help us to categorize the reliable and unreliable moral judgments (Henne and Sinnott-Armstrong 2018, 64). Henne and Sinnott-Armstrong then propose a solution to show which moral judgments are true and which are not—the use of higher order principles such as Order Effects Undermine Reliability (OEUR) and higher order inclinations. With OEUR, the philosophers explain that those judgments that are altered when the line-up of evidence leading up to an action are presented in different orders cannot be true, and those that remain the same regardless of the order the evidence is provided are true. With the higher order inclinations, they suggest that these can also be used to tell the reliability of a moral judgment. Before doing so however, Henne and Sinnott-Armstrong suggest that the difference between conservatives and revisionists be drawn. With the conservatives, the philosophers argue that such people believe that moral judgments are for the most part right, and hence should be followed even if there have been cases where those judgments have led us astray. Contrary to the conservative, the revisionist argues that moral judgments "often deviate from

what theories give us reason to believe is correct, [so] moral judgments should be revised to bring them in line with theory" (Henne and Sinnott-Armstrong 2018, 62). With this then, the reliability of a moral judgment will depend on whether conservatives or revisionists are in play. Henne and Sinnott-Armstrong however argue that we do not know enough about higher order beliefs to know whether or not neuroscience undermines moral judgments, hence, we can only deduce from our findings that neuroscience merely trims and categorizes moral theory.

Free Will

A question that often accompanies existentialism is whether or not we are really free. Do we freely choose to act in certain ways, or do we simply carry out what is dictated to us? If we do, what impact should our will have on our owning of those actions? Although the relevance of this question to the topic might not be evident at the stage, we will see how morality, free will and responsibility work hand-in-hand to construct the notion of justice we hold. There are speculations as to what controls our actions—we ourselves, some neurons in the brain over which we have no control, and even some supreme deity of some sort. However, for concision sake, I will only address those arguments regarding neuroscientific findings.

Relationship between Neuroscience and Free Will

How much free will do humans actually have in their actions? Philosopher Jesse Prinz argues in his work titled "Moral Sedimentation" that humans do not really have free will in moral decisions because we do not formulate the moral values according to which our actions are carried out (Prinz 2018). Prinz relies on the understanding of sedimentation as put out by Edmund Husserl and Maurice Merleau-Ponty to finally come up with his notion of moral sedimentation. The main claim in Prinz's work here is that morality is sedimented in that it is socially conditioned. With this, he explains that none of the values according to which we judge the morality of an action are actually freely formed by us because all our views and traditions are sedimented. With the sedimentation here, Prinz means that whatever knowledge and views we have are influenced by prior knowledge—"prior knowledge informs present encounters with the world, shaping how we interpret things, and gives us the impression of a pregiven order" (Prinz 2018, 88). If one is applying previous knowledge onto a present encounter, wouldn't one be

aware of this? Prinz answers that with sedimented values, they are so hammered into our daily lives that they almost feel as though they are innate so one will not be aware of the influence of sedimented values on current encounters. As to how this occurs, Prinz explains that "sedimented traditions extend enduringly through time since all new acquisitions are in turn sedimented and become working materials" (Prinz 2018, 88) by being enshrined in language and getting accepted passively through enculturation. On enculturation, Prinz argues that it has gone so far that "we do not just inhabit a natural world; we also inhabit a cultural world", as the natural world comes with the cultural view on how to relate to it. At the end of his analysis, Prinz comes to the conclusion that we only feel like we are free agents, but we are actually not free since all the values according to which we act are subtly imposed on us by societal constructions. Apart from this imposition jeopardizing the notion of free will we have, Prinz also points out that moral judgments have been shown to prompt emotional activity in the brain through neuroimaging studies, and these emotions we use to process moral judgments are imbibed through social constructs (Prinz 2018, 95-96). Hence, whether we look at psychology or neuroscience for the existence of free will, human beings do not actually have free will.

Taking a more generous stance than Jesse Prinz is philosopher Walter Glannon in his work titled "Behavior Control, Meaning and Neuroscience". In his work, Glannon specifically narrows down on an experiment conducted by Libet on the decision-making process in the brain. Libet's experiment features the use of electroencephalography to prove that a subject's awareness of the intention to perform an action is preceded by neural activity by hundreds of milliseconds. That is, before a subject is aware of a decision he is about to take, this decision is formulated by the neurons. Libet stands on this and argues that we have no causal role in our actions and decisions, and consequently, no free will (Glannon 2018, 148). Glannon grants that yes, Libet is right about the fact that neural activity precedes awareness in some actions, but he disagrees that we do not have any causal role in our actions—that we merely carry out what is dictated to us by neural activity. Glannon distinguishes between what Alfred Mele calls the proximal and the distal intentions. He argues that the difference between these two and the components of the latter are what we need to observe when evaluating the role we play in determining our actions, because the actions used in Libet's experiment do not have to do with the everyday moral decisions we

take. According to Glannon, "distal intentions are long-range conscious plans that may precede the performance of an action by days, weeks, months, or even years. Actions performed at a particular time may have a physical and psychological history that extends into the past" (Glannon 2018, 149). With this then, we can see that our decisions may be swayed one way or the other in response to the historical and social connotation we associate with it. It is based on this foundation that Glannon argues that although our actions may be initiated by neural activity, they are ultimately determined by us based on the meaning we attribute to certain things. Therefore, this is how neuroscience trims free will—it shows us to what extent we play a role in our actions, and in which situations we do not have a say in the actions we take.

PART II

Implication of Relationship between Neuroscience, Free Will and Morality

Findings from the preceding sections show that the relationship between neuroscience, free will and morality is not fundamental to the pursuit of justice. On the surface, it seems as though free will is a necessary condition for determining responsibility, and the fact that neuroscience seems to deny the fact that such free will exists will seal the deal on responsibility. That is, it will make all things permissible as the excuse-extensionist model advocates. But this is far from the case because morality, which has been shown to be trimmed by neuroscience, conveys a sense of responsibility that can be independent of guilt. What then does this mean for the various justice systems adopted by mankind?

On what basis do we determine whether or not a person is responsible for their actions? What implications do our judgments from these bases gave rise to? There are so many factors that contribute to our current belief in retribution and our concept of guilt and justice. What we do not do is sit down to carefully analyze these factors that go into our systemic practices. There have been several speculations and theories as to how human behavior can be explained: reductionists claim that human interaction can be explained with scientific findings, while existentialists who tend to be more philosophical claim that human behavior should be explained in accordance with philosophy. With our current system of justice, there is an emphasis on responsibility where an agent is supposed to be blamed and punished for actions that are deemed to be morally wrong and praised

and rewarded for those that are deemed to be morally right. In such a system, we see justice to go hand-in-hand with free will, responsibility and retribution.

Focquaert et al., in their work titled "Free Will Skepticism, Freedom, and Criminal Behavior", argue against the basic desert sense of responsibility which advocates that blame and punishment and praise and reward is deserved. Being free will skeptics themselves, Focquaert et al. argue that although the notion of guilt is fully embedded in our societal functions, there are strong moral and scientific reasons to abandon the basic desert sense of moral responsibility and adopt a sense of responsibility which pursues justice without retributivism (Focquaert et al. 2018). Retributive punishment and free will skepticism are heavily opposed because retributivism has to do with the punishment being justified on the grounds that the person deserves to be harmed because he knowingly did the wrong thing, and free will skepticism says that we do not in fact have the choice to decide the course of our actions. Rather than relying on the flawed basic desert sense of responsibility, these free will skeptics have come up with the take-charge responsibility to determine whether or not particular agents are responsible for certain actions and what measure will be employed to remedy the source of the action. With this form of responsibility, we see that having or lacking human agency and a capacity for take-charge responsibility implies having or lacking the freedom to change one's future behavior if given the means to do so. Although this take-charge responsibility rejects the idea that free will should play a role in the attribution of responsibility, it places a rather heavy emphasis on human agency—the capacity for an agent to do otherwise in the future given the necessary means to do so—as this is the only way to determine whether the supposedly responsible person's actions were influenced by factors outside of his control. Hence, rather than advocating for retribution because a person deserves punishment in a desert sense, this method vouches for rehabilitation and leading a crime-free life. As opposed to retribution which merely punishes agents and possibly causes more harm to them than good, this take-charge responsibility addresses structural impediments, encourages reformation and offers better solutions to the problems arising from structural impediments.

With our current justice system, the only motive behind holding people responsible for their actions is to punish them because they supposedly deserve such punishment by virtue of acting out of their free will—this is otherwise known as retribution. This way of going about the justice system has been argued to be

flawed by specialists such as Focquaert et al. because studies have shown that we do not have the free will needed to act and be responsible in the sense. In fact, it has been observed that the genetic brain structure of criminals is substantially different from that of non-criminals. With this then, how can you deem a person "guilty" and deserving of punishment when he has no control over the cause of his actions? The level of absurdity of thinking a person deserves punishment for performing an action is the same as that of blaming an epileptic patient for having a seizure and thinking they deserve a certain consequence as a result. It has also been shown that environmental factors contribute immensely to criminal behavior. With this then, how can we say a person deserves punishment for something that was caused by factors beyond their control in the first place?

Since we cannot deem people responsible based on a basic desert sense of moral responsibility, and hence cannot allow retribution to dominate the justice system, do we then just let people who perform immoral actions go scot-free? No. For the purposes of preserving justice, Focquaert et al. have come up with systems that recognize responsibility while promoting justice by attending to the emotional needs of victims (Focquaert et al. 2018). These systems, deemed as psychological and behavioral interventions, can actually help restructure the brain. Under the psychological and behavioral interventions, we can have mindfulness training and attention training whose practice have been shown to increase amygdala functioning, especially after love and compassion meditations. Moving to more scientific solutions, we have what has been termed as moral enhancement. This system, as has been adopted by the Defense Advanced Research Projects Agency (DARPA), has to do with physically tweaking parts of the brain to produce desired actions from an agent. With these systems, the agent is not being deemed as deserving some sort of punishment, but rather, he is seen as having or not having the capacity to change his actions given the necessary conditions, and is worked on from there. This is a system that will promote the general welfare of society as it looks into even the smallest causes undesirable actions and eliminates those causes so as permanently rid the society of vices in the long run.

With the attention training, it is designed to rectify psychological disorders initiated by the Cognitive Attention Syndrome (CAS). According to the MCT Institute, CAS is linked to internal metacognitions that "control thinking and attention which is biased in psychological disorder and lock the individual into persistent patterns of negative thinking and attention that are difficult to control

and contribute to anxiety and depression" (MCT Institute 2018). Attention training then aims at helping the individuals in question to focus on negative thinking so as to redirect their thoughts to more positive things. This redirection of thoughts in turn reduces the individual's anxiety and depression which plays a key role in criminal indulgences. Mindfulness meditation takes a similar route in that it seeks to enable the individual in question catch himself in his thoughts and control those thoughts, steering away from the negative ones and engaging all thoughts in an impartial way.

Other than the ethical question behind retributive justice, why would we want to adopt a new system of justice when the system in effect now seems to serve its purpose? Well when we take a closer look at the workings of our current justice system, we see that it does not actually cater to any problems with issues on justice. The first flaw with the status quo is that it does not even cater to justice at all. Justice is seen to be the egalitarian treatment of all actors in a specific situation. From this definition, we can deduce that there are three groups of agents whatever justice system that is in place has to cater to—the accused, the accusers and the general public. Our current retributive justice system only seems to cater to the emotional needs of the accusers, and ignores the justice supposed to be catered to the accused and the general society. According to the United Nations Office on Drugs and Crime (UNODC), people who go into prisons mostly come out with health issues such as "Psychiatric disorders, HIV infection, tuberculosis, hepatitis B and C, sexually transmitted diseases, skin diseases, malaria, malnutrition, diarrhea and injuries including self-mutilation" due to the poor conditions in the prison environment (UNODC 2019). Apart from health implications, the current prison system also leads to social implications. UNDOC reports that family structures are disrupted as a result of the time spent locked up. Ex-convicts also face social implications in the form of stigmatization. It is no surprise that those who have gone through the prison system are stigmatized in the sense that it is even difficult for them to land a decent job upon their release. This in turn does not motivate them to lead a crime-free life, as they tend to fall back to their old ways to survive. With this then, we can see that the current justice system does not cater to the needs of the accused.

Data from the Bureau of Justice Statistics also shows that people who have served time in prison have an eighty-three percent (83%) chance of being rearrested (Alper et al. 2018). Of course, these figures might be altered when

we consider the severity of the crimes, the role of environmental factors in the crime, as well as the role psychological factors play. However, when examined without respect to these differences, the 2018 update of the Bureau of Justice Statistics found, by keeping tabs on 401,288 prisoners over a course of nine years, that forty-four percent (44%) of ex-convicts were rearrested within a year after being released, sixty-eight percent (68%) within three years, seventy-nine percent (79%) within six years and eighty-three percent (83%) within nine years (Alper et al. 2018). With this data then, we see that the so-call "prison reform" we have in effect now does not actually reform criminals, but just ends up holding convicts captive for a given period of time, making them worse than they came in in most cases, and releasing them back into the public. With these facts then, we can see why there is an urgent need to replace our current justice system with one that actually caters to justice and reforms convicts.

Possible Counterarguments against the Findings

The first counterargument that will arise is the supposed restructuring of the brain in the name of moral enhancement as a preventative measure. If this restructuring of the brain is even possible, how ethical will such a practice be? Who will be in charge of this restructuring? All these questions amount to substantial counterarguments that could weaken the very foundations on which the implications of the relationship between neuroscience, free will and morality lie. In the article titled "The Pentagon's Push to Program Soldiers' Brains: The Military Wants Super-Soldiers to Control Roots with Their Thoughts" by Michael Joseph Gross, the author goes more into detail about how DARPA projects started and where they are now. Gross points out that the DARPA projects started with the purpose of healing injury and curing sickness (Gross 2018). Of course, healing injury and curing sickness, just like moral enhancement, look more like they will benefit the society than hurt it. The problem here is that the actions of the agency are not impeded by bureaucratic oversight and scientific preview, as any other activity that has this high of a risk will be. Hence, there is no guarantee that the agency only works on those projects they tell the general public. In fact, Gross points out in the article that public support is drawn for DARPA projects by hiding the true projects from the public and showcasing those that the public genuinely needs. For instance, the agency draws support by advertising bionic arms and hammering on their importance, but they do not tell the public about

their intention to make super humans such as the proposed 24/7 soldier who could go for a week without sleep. DARPA also has The Restoring Active Memory Program where neuroprosthetics are developed to alter memory formation so as to counteract traumatic brain injury (Gross 2018). This program seems beneficial, even to our moral enhancement such that criminals with traumatic experiences that dictate their actions can be rehabilitated and reformed. But how far is too far? This is almost like wiping out a part of an agent and fitting that with new memories. At what point, will this modified agent stop being a human being? These projects such as those carried out by DARPA and moral enhancement all aim at making the perfect human. But doesn't this perfect human resemble a robot more than a human? Does that mean that robots can also be considered as human?

Who will be in charge of this power-wielding process? The findings of Michael Joseph Gross on the DARPA projects have clearly demonstrated that if this process is left in the hands of the government, the military or any such body that will have an interest in mooching off agents' superhuman tendencies, then the experiments and actions could spiral out of control. In fact, Gross also revealed that according to a Silicon Valley recruit, DARPA is not only interested in damaged bodies, but also in healthy bodies, which questions their purpose to merely cure illness and heal injuries. He points out that the driving goal for DARPA has now become "to make human beings something other than what we are, with powers beyond what we are born with" (Gross 2018). By our desire for moral enhancement, are we then giving the go-ahead for the government to create robots and pose them off as humans?

Another objection has to do with the slippery-slope that this restructuring can give rise to, and how difficult it might be to put an end to it. Such an instance is vividly depicted in the movie Gattaca directed by Andrew Niccol. This movie has to do with the use of genetic engineering to modify zygotes to make "perfect" human beings called "valids" whose entire life stories are known before birth. This genetic engineering gained so much popularity which led to the unmodified humans, known as "invalids", to be seen as inferior and hence not have access to certain opportunities. In this movie, the "valids" were always preferred to the "invalids" since they were seen as more efficient and more suited for all respectable roles (Gattaca 1997). The situation depicted in the movie seems to be where humanity is headed now with its development and research, especially in trying to eliminate imperfections in human beings. How sure are we that this

moral enhancement will not create a ripple effect that will end up making the superhuman dominate the actual human, since humans are known very well for their greed for perfection and immunity?

Bouncing off the issue of a slippery slope, wouldn't the suggestion for moral enhancement just pave way for genetic modifications? Since the concept of moral enhancement already involves invading in an agent's brain to physically alter some parts of the brain to make them moral, why then wouldn't we just suggest genetic modifications such as those the Clusters for Regularly Interspaced Short Palindromic Repeats (CRISPR) technology make for? While we are at it, why don't we just skip the mindfulness and attention training and move straight up to moral enhancement since it has been shown to be more accurate with a low risk of relapse? If the thought of surgery seems too extreme, then why don't we just advocate for the use of medication to increase the level of oxytocin in the hypothalamus of the human brain? Since oxytocin makes people cooperate more, wouldn't it be easier to just administer medication to improve this cooperation which will in turn make us more moral? The last and most disturbing counterargument to the proposition to fall to mindfulness and attention training as ways to improve morality is the belief that these practices are analogous to brainwashing. Would we really find it morally acceptable and ethical to brainwash agents into becoming moral?

Response to Counterarguments

Looking at the counterarguments outlined above against the moral enhancement, it is evident that proponents of these arguments did not take the fact that each of the proposed neural restructuring into account differs in its extent into consideration—they all do not have the same level of invasion, nor the same degree of effect, nor the same risks. Unlike the surgical restructuring of the brain as DARPA does, mindfulness training and attention training are not as radical. After all, the mindfulness training and attention training are similar to parents training their children according to some morals. Why don't we find it terrifying that society tries to change our way of thinking and relating to some things through laws and commonly held societal beliefs? If the proponents of the counterarguments observe the various systems proposed carefully, they will notice that the only system taking a step ahead of humanity's comfort zone is the one that has to do with surgically altering the brain. Hence, from the face value, proponents of the counterargument cannot raise any comprehensive objections

to the use of mindfulness meditation and attention training as moral enhancement forums.

Even with the DARPA style of moral enhancement, we can still respond that operations will be handled by an independent group of experts made up of neuroscientists, psychologists, philosophers and other relevant experts. It can be agreed on that such a delicate operation cannot be left in the hands of the government, the military or any other organized body that could have ulterior motives to just morally enhancing agents. If we have a trustworthy group managing such operations, then the fear of the slippery-slope as enacted through Gattaca should not exist.

The issues with CRISPR and opting for moral enhancement to be the only form of criminal reformation are very difficult to argue against since these processes they have been shown to be more accurate and effective than mindfulness and attention training. Tempting as it may be to just give in and accept these processes as our go-to solution to eradicate crime, we cannot do so because of the heavy ethical implications they come with. Hence, I can only suggest that CRISPR technology and moral enhancement should only be turned to when all else fails. With the CRISPR technology and its germline editing which has to do with altering the genetic modification of sperms and eggs, we can only permit such altering in extreme cases where we are absolutely sure that the child born from the fusion of such a sperm and egg will be born being disposed to indulge in criminal activity regardless of the environmental factors (Vidyasagar 2018). As to the issue with having moral enhancement be the prime way of eradicating crime, we cannot accept this because it will lead to the use of unnecessary invasion. If we were to only fall to moral enhancement to eradicate crime, then it would mean that an agent accused of lying or petty theft will have to have their brain surgically altered since this method is guaranteed to not lead to any relapse whatsoever. With this, I will stand my ground that moral enhancement only be used in cases where it will be pointless to try to use mindfulness and attention training to alter the brain. I will however permit a stipulation that moral enhancement be mandated for those agents who relapse to their old ways more than two times after being rehabilitated through mindfulness and attention training.

Should we just forgo all the technicalities with moral enhancement, mindfulness and adopt a system where agents can take medication to be moral? I will answer no. my main argument against the use of medication to morally enhance people

is the high risk of abuse. Drug abuse is an epidemic that has swept through many countries and claimed millions of lives while doing so. From experiences with drugs such as antidepressants, we can tell right from the get-go that placing a "morally enhancing" medication in the hands of people will immediately spiral out of control and cause more harm than the benefits it was intended to bring. Another objection to this easy way out is that taking medication to be moral does not reform the agent in any way. Here, the agent's morality is contingent on him taking the required medication, and if this is not done, all things will go bonkers. In talking about suggestions to make people moral so as to completely eradicate society of crime, the main purpose is to actually rehabilitate agents and reform their characters so that they do not relapse into their old ways. With this then, we can see that having one's moral state be dependent on a pill of some sort will not cater to this goal—it could rather lead to more grave consequences.

With the last counterargument as to mindfulness and attention training merely being brainwashing by another name, it will beg to differ. The processes and goals for mindfulness and attention training starkly differ from those of brainwashing. According to HowStuffWorks, brainwashing typically occurs in three stages—breaking down the sense of self, giving a possibility of salvation and rebuilding the self in a new, radical image (Layton 2009). According to the website, "mind-clouding techniques" such as starvation and sleep-deprivation are used to force an agent to deconstruct whatever image or beliefs he holds about himself or something. The agent is then guilt-tripped and led to question all he believes in—he is basically left with deep angst as to what to believe and what not to believe—and it is at this stage that the brainwasher offers a way out. With this way out being the only thing the agent can grasp on to, he then rebuilds his conceptions and beliefs through the new lenses that he has acquired (Layton 2009). Even from examining the process of brainwashing, we can see how starkly different it is from mindfulness and attention training—these processes do not seek to lead an agent to build an entire new conception of his self. Mindfulness and attention training rather focus on empowering the agent in question to take charge of his thoughts and actions where it might be difficult to do so. During the brainwashing process also, the brainwasher has absolute control over the functioning of the agent. With mindfulness meditation and attention training however, the agent has control over the entire process and only receives guidance on how to take charge of his thoughts and direct them. With this then, we can see that both the

purpose and the process involved in mindfulness and attention training bear no similarity to those of brainwashing, and hence, the two groups of practices cannot be measured up against each other.

CONCLUSION

In this paper, we have considered what morality and free will are and their sources. We have also looked at the relationship between neuroscience and morality, where it was shown that morality does exist and neuroscience simply defines and categorizes it. The relationship between free will and neuroscience was probably the most substantial part of the research, and it turned out that our free will is actually trimmed. With this information in hand then, we drew the relationship between the three and looked at what this relationship might mean. Here, we saw various arguments as to the role free will should play in the determination of responsibility of actions. We concentrated on free will skepticism which argued that although we might not have free will, that does not mean there can be no sense of responsibility, and hence, no justice. This theory showed that we can actually preserve justice without attributing guilt, which naturally comes along with blame and the notion that an agent deserves to reap the consequences of his actions. With free will skepticism, the proposed ways of dealing with agents who engage in immoral acts turned out to be more constructive as opposed to how our justice system works now—through retribution. Although neuroscience and the law seem to be miles apart, findings from neuroscience can actually help humans craft laws that will serve the good of all human beings as a whole, hence the uncanny relationship. With this then I will suggest that the various judicial systems around the world look into the free will skeptic view of responsibility as well as their suggestions for reformation.

REFERENCES

Alper, Mariel, et al. 2018. "2018 Update on Prisoner Recidivism: A 9-Year Follow-Up Period (2005–2014)." *Bureau of Justice Statistics.*

"Altruism." *Oxford English Dictionary.*

"Attention Training Technique." *MCT Institute.* 2018.

Caruso, Gregg D. and Owen Flanagan. 2018. "Neuroexistentialism: Third-Wave Existentialism." In *Neuroexistentialism: Meaning, Morals, & Purpose in the*

Age of Neuroscience, pp. 1–22. New York: Oxford University Press.

Churchland, Patricia Smith. 2018. "The Impact of Social Neuroscience on Moral Philosophy." In *Neuroexistentialism: Meaning, Morals, & Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, pp. 25–37. New York: Oxford University Press.

Focquaert, Farah. 2018. "Free Will Skepticism, Freedom and Criminal Behavior." In *Neuroexistentialism: Meaning, Morals, & Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, pp. 235–250. New York: Oxford University Press.

"Gattaca". 1997. Directed by Andrew Niccol. *Netflix.*

Greene, D. Joshua, et al. 2001. "An FMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (5537): 2105–2108.

Glannon, Walter. 2018. "Behavior Control, Meaning and Neuroscience." In

Neuroexistentialism: Meaning, Morals, & Purpose in the Age of Neuroscience, edited by Gregg D. Caruso and Owen Flanagan, pp. 146–161. New York: Oxford University Press.

Gross, Michael Joseph. 2018. "The Pentagon's Push to Program Soldiers' Brains: The Military Wants Super-Soldiers to Control Roots with Their Thoughts." *The Atlantic.*

Henne, Paul, and Walter Sinnott-Armstrong. 2018. "Does Neuroscience Undermine Morality?" In *Neuroexistentialism: Meaning, Morals, & Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, pp. 54–67. New York: Oxford University Press.

Layton, Julia. 2009. "How Brainwashing Works." *HowStuffWorks.*

Prinz, Jesse. 2018. "Moral Sedimentation." In *Neuroexistentialism: Meaning, Morals, & Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, pp. 87–107. New York: Oxford University Press.

United Nations Office of Drugs and Crime. 2019. "Why Promote Prison Reform?" UNDOC.

Vidyasagar, Apama. 2018. "What is CRISPR?" *Live Science.*

compos mentis

Euphemisms and Praxis: Degradation of Truth and Meaning

Logan B. Cross
Michigan State University

ABSTRACT
Euphemisms—soft, mild and indirect words or phrases that are used in place of harsher, more direct words or phrases—appear to be a ubiquitous phenomenon in the linguistic evolution exhibited in many modern cultures. By replacing harsh words such as "death" with softer terms like "passing away," euphemistic language can lessen the trauma felt by truths which are hard to bare without lying to oneself outright or averting one's attention away from one's problem's completely. In this essay, however, I will argue that the benefits of euphemisms come with a hidden price for cultures and individuals which use them. In particular, I will argue that euphemisms degrade the truth and meaning of statements by describing them through terms that are by and large devoid of emotional truth. Once the emotional truth has been removed from the statements, the praxis of a society—that is, how that society actively solves their problems and actualizes their ideals—is negatively affected by virtue of the fact that effective communication, which, I will argue, is compromised by the absence of emotional truth, is a vital component of the form of praxis.

KEYWORDS
Euphemisms, Philosophy of Language, Evolution of Language, Praxis, Meaning, Truth

PART I: THE EVOLUTION OF LANGUAGE

For the vast majority of linguists and philosophers of language, much like the biological realm, language itself is an ever-evolving entity. To recognize the results of the evolution of language, one need only compare an antiquated example of one's own language to a modern example and note how alien and different the former seems to the latter—the works of Shakespeare, for instance, seem almost indecipherable to modern readers of English literature, despite the fact that Shakespeare himself wrote in English. This notion is analogous to how, say, homo habilis, if observed today, would appear quite indistinguishable from the modern homo sapiens despite the fact that homo sapiens directly descend from homo habilis. Notwithstanding the similarities between the evolution of biological beings and the evolution, the comparison is not altogether analogous—at least one important difference exists. Dissimilar to the evolution of species and biological diversity, the evolution of language does not have a necessary and essential guiding mechanism. In biological evolution, despite the fact that the underlying processes which cause said evolution are essentially random, the end result is always guided in a specific direction. Only the lifeforms which had evolved traits that made them more "fit" to survive their environments survived to pass on their genes, thus increasing the likelihood that whatever beneficial evolved trait(s) that helped them to survive their environment in the first place perpetuated through future generations. This process is necessary for its own existence, as without this guiding mechanism the results of evolution that would either be totally random or directed towards some aim other than survival. In either case, it seems as though life, and therefore also the processes of biological evolution, would cease. In this way, the guiding mechanism of biological evolution is imbued within the essence of the process.

Because language is abiotic and without a physical form and thus does not have to contend with matters of survival, the evolution of language is not dependent on a necessary and essential mechanism which guides it. Language can evolve in so to speak any direction, and, more importantly, for any reason. As a result of this un-predetermined nature of the evolution of language, at least one important consequence arises for humanity, namely, that the evolution of language can be directly manipulated by people to go in a certain direction—whereas humans cannot directly decide what lifeforms are best suited to survive this or that environment (other than by changing the environment itself) and thus

cannot derail the processes of evolution (but rather only arrange ways to use it for their own benefit), the decision as to what terms of language are used seems able to be consciously manipulated. Indeed, the direction manipulation of the evolution language by a person or group of people is prevalent in advanced industrial societies. While the reasons behind such manipulations are diverse, one of the most common purposes seems to be for political correctness or to advance some political or ethical agenda—and one of the most common ways of manipulating the evolution of language so as to reach these desired ends are by employing euphemistic language. In the writing to come, an analysis of euphemistic language will be given, followed by an analysis of how euphemistic language affects philosophical praxis by altering our epistemological standings.

PART II: THE EUPHEMISM AND ITS EVOLUTION

The definition of a euphemism is a word or a phrase that is used to replace another word or phrase that one finds offensive or undesirable. One example of a commonly used euphemism found in the English language is to say that one who has died has "passed away." Here, the harsh reality of death, which brings with it the grim possibility that the one whom has died is gone into eternal unconsciousness and no longer exists, has been described not as "death" with the term "passed away," which suggests that the one whom has "passed away" has went away somewhere else, but seems to exclude the possibility of ceasing to exist. Passed away—away to where? More examples of euphemisms can be found in the military lexicon. Listen to them discuss their doings, and one will find that the military rarely "kills" or "murders" anyone, but instead "neutralizes" them. Here again, the grim reality has been stripped away from the term—whereas killing and murdering have thousands of years' worth of negative connotations and horrors to make the terms "kill" and "murder" near synonymous with evil, the term "neutralize" sounds modern, sterile, and morally 'neutral.'

As alluded to in the introduction of this work, euphemistic language is an evolution of language with a specific, human-controlled intention. When this intention has been established, the evolution often continues in the direction the intent had pushed it towards—this is to say, the euphemization of the term continues. In his book When Will Jesus Bring the Porkchops?, author George Carlin discusses euphemisms in detail, and in one short section tracks the progression of

one such evolution of a euphemized term. In a section titled "Euphemisms: Shell Shock to PTSD" he writes:

> "[T]he one thing euphemisms all have in common is that they soften the language. They portray reality as less vivid; they prefer to avoid the truth and not look it in the eye. I think it's one of the consequences of being fat and prosperous and too comfortable. So, naturally, as time has passed, and we've grown fatter and more prosperous, the problem has gotten worse. Here's a good example:
>
> There's a condition in combat—most people know it by now. It occurs when a soldier's nervous system has reached the breaking point. In World War I, it was called *shell shock*. Simple, honest, direct language. Two syllables. Shell shock. Almost sounds like the guns themselves. Shell shock!!
>
> That was 1917. A generation passed. Then, during the Second World War, the very same combat condition was called *battle fatigue.* Four syllables now. It takes a little longer to say, stretches it out. The words don't seem to hurt as much. And fatigue is a softer that shock. Shell shock. Battle fatigue. The condition was being euphemized. More time passed and we got to Korea, 1950. By that time, Madison Avenue had learned well how to manipulate the language, and the same condition became *operational exhaustion.* It had been stretched out to eight syllables. It took longer to say, so the impact was reduced, and the humanity was completely squeezed out of the term. It was now absolutely sterile: operational exhaustion. It sounded like something that might happen to your car.
>
> And then, finally, we got to Vietnam. Given the dishonesty surrounding that war, I guess it's not surprising that, at that time, the very same condition was renamed *post-traumatic stress disorder.* It was still eight syllables, but a hyphen had been added,

> and, at last, the pain had been completely buried under psycho-jargon. Post-traumatic stress disorder.
>
> I'd be willing to bet anything that if we'd still been calling it *shell shock,* some of those Vietnam veterans might have received the attention they needed, at the time they needed it. But it didn't happen, and I'm convinced one of the reasons was that softer language we now prefer: The New Language. The language that takes the life out of life" (Carlin 2004, 39–40).

PART III: EUPHEMISMS AND PRAXIS

If one analyzes Carlin's analysis of the euphemization of the term originally called shell shock to its currently used term post-traumatic stress disorder, one will notice a number of philosophical claims regarding euphemistic language. One of the more interesting and powerful of the claims to be found among Carlin's writing stems from his assertion that if society had still been calling post-traumatic stress disorder by its original name shell shock, then more Vietnam veterans afflicted with the condition would have been helped. If this assessment is true, it amounts to the proposition that euphemisms seem to have a tremendous effect on social praxis, or, in other words, on the way in which a certain social theory or philosophical system is practiced and realized. Within the context of our example, this idea suggests that, if we are assuming that we possess a philosophical or moral system that mandates that we ought to help victims suffering from shell shock, this system was, in practice, somehow nullified or disrupted by referring to the condition as post-traumatic stress disorder. Let us now turn to the question of how the process of such a disruption apparently works.

In order to understand how the disruption of praxis at the hands of euphemistic language described above really works, we will need a further understanding of the process by which a certain philosophical theory leads to a particular social praxis. To understand this process, I believe it will be helpful to break down the process (or form) of praxis into stages, beginning from its natural starting point while working our way toward the conclusion of praxis.

What, then, is the natural starting point of praxis? If praxis is said it be the process in which theory becomes practiced and actualized, then praxis has the

relational form of theory—actualization. Theory, of course, arises from a set of observations of the world, in conjunction with philosophical and logical reasoning about said observations—as such, the relational form of praxis can be expanded from theory—actualization to observation—reasoning—theory—actualization.

If our investigation were concerned solely with the idea of philosophical praxis in a general sense, then observation—reasoning—theory—actualization might suffice for an adequate description of the form of praxis. However, our investigation concerns the effect of euphemistic language on what I call the social praxis, or the way in which society applies their ideologies and values into a set of societal practices and policies. The major difference between praxis and social praxis is that while praxis exists in the individual, a social praxis exists on the societal level and thus must transmit from person to person. If we try to apply the observation—reasoning—theory—actualization to a concept like the social praxis, we will find it to be inadequate, because theory must somehow transmit from person to person to exist on the societal level. There is a gap between the theory—actualization portion of the form. The theory must be passed from person to person in order to actualize, and this passing must have a medium—communication. And so, the form of the social praxis can more adequately be described as observation—reasoning—theory—communication—actualization. Philosopher Calvin Schrag recognized this aspect of the social praxis, writing that "[p]raxis as the manner in which we are engaged in the world…is always entwined with communication (Miller, Ramsey & Schrag 2003, 21).

In what ways does the communication aspect of social praxis occur within society? Although there are multiple answers and possibilities to this question, the most pervasive and influential answer is language, both written and spoken. With this fact considered, we can analyze the effect of euphemistic language on the social praxis. If communication is a major aspect of the form of the social praxis, and this communication is normally mediated by language, then the clarity of this language would seem to play a highly important role in the social praxis. Indeed, this is my argument: euphemistic language, I contend, degrades the clarity of language, thus creating distortions in the social praxis.

In order to understand the claim that euphemistic language degrades the clarity of language thus leading to distortions in the social praxis, the way in which euphemistic language degrades the clarity of language must first be analyzed. Upon analysis, numerous methods are identified as to how this degradation in

clarity occurs. Firstly, as Carlin touched on, there is an aesthetic sense which euphemisms degrade clarity by 'watering down' the emotional salience of language and terms. 'Shell shock' "sounds like the guns themselves," and thus it certainly sounds emotionally salient enough to catch one's attention—comparatively, post-traumatic stress disorder sounds much less severe (everybody becomes stressed sometimes), and as such does away with the original emotional power of the term. In her work The Practical Study of Argument, philosopher Trudy Govier describes this diminishing of the emotional salience of terms by the use of euphemisms, writing that "[t]here is a sense in which euphemism is the opposite of emotionally charged language. With emotionally charged language, terms are more emotional than appropriate. Euphemism, on the other hand, involves a kind of whitewashing effect in which descriptions are less emotional than appropriate" (Govier, 2014).

A second, perhaps less apparent way euphemistic language degrades the clarity of language is by creating an additional detachment from the original conveying of the idea. Language attempts to convey an idea by taking an observation and describing it through a designated word or phrase. Any detachment from the original word or phrase carries with it the possibility of a distortion of meaning and clarity. This basic idea is often demonstrated to children in the child's game "telephone" in which one child tells something to another child, and that other child changes, slightly, what has been told to them and then tells another child the slightly changed message who repeats this changing process and so on and so forth for as many children are playing. As an example, suppose one child starts the game by saying "cheetah," and the next child changes this by saying "big cat," and the next child changes this to simply "cat." As can be seen, during every subsequent change in the phrase, the clarity degrades— "cheetah" which creates a very specific image in mind, whereas "big cat" is more ambiguous, and "cat" more ambiguous still. By changing the original phrase, euphemisms open themselves up to the possibility of these types of degradation of clarity and meaning.

With the ways in which euphemisms degrade the clarity of language, it may now be understood how euphemisms create distortions in the social praxis. How, then, do these distortions occur? I contend that euphemistic language distorts the social praxis through interfering with the link between communication and actualization in the form of social praxis described earlier.

The way in which I argue euphemisms interfere with the link between communication and actualization in the form of social praxis relates to philosopher Ludwig Wittgenstein's picture theory of language, as described in his work Tractatus Logico-Philosophicus. Wittgenstein's picture theory of language suggests that the world is comprised of a collection of facts and/or ideas that can be mentally pictured through language (Wittgenstein, 1922). By covering up the emotional salience of a term through employing a softer term, euphemisms degrade the luridness of the mental representations produced by the language we used, the effectiveness of communication is degraded, thus leading to problems with the actualization of the social praxis.

An example can help to illustrate the mode of interference described above. Let us analyze closer Carlin's example of shell shock being now described as post-traumatic stress disorder. Consider a nation at war that is experiencing a problem of soldiers experiencing this unfortunate condition and determines that helping these soldiers corresponds with their beliefs and ideology, and desires to adopt a praxis of treating them. If this adoption is to occur, during the communication phase of the social praxis, the military or relevant governing body must find a way to clearly and effectively communicate the nature of problem to the general public so that the public can understand the severity of the condition—only then will the issue be taken seriously enough for a solution to be adopted. Because it is emotionally reminiscent of war and guns themselves, the term "shell shock" clearly describes the severity of the condition it aims to describe, and therefore would be highly conductive of encouraging the actualization of helping those afflicted. The term "post-traumatic stress disorder," on the other hand, carries with it a mental representation that is far less lurid than "shell shock"—Wittgenstein says in his work Philosophical Investigations "uttering a word is like striking a note on the keyboard of the imagination (Wittgenstein, 1953). Euphemisms degrade the clarity of this "note," opening the door for misunderstandings to occur in the "language games" Wittgenstein suggests we play with each other (that is, the use of language to try to elicit a certain response or idea out of another). Because the term post-traumatic stress disorder creates such an abstract and indirect representation of the given mental affliction than the term shell shock, a person will be less likely to believe that mental affliction to be a problem worth the effort of solving, thus interfering with the actualization of a social praxis (the helping of those afflicted with shell shock, in this example).

Indeed, several philosophers have been concerned about the potential implications of euphemistic forms of language on the social praxis and individual behavior. Herbert Marcuse, social and political philosopher famous for his role in the Frankfurt School of Critical Theory, discussed some of these implications in his book One-Dimensional Man. In one instance, Marcuse writes on the abridged language that is commonly observed in the technical, scientific and military spheres of life. He writes:

> "Note on abridgement. NATO, SEATO, UN, AFL-CIO, AEC, but also USSR, DDR, etc. Most of these abbreviations are perfectly reasonable and justified by the length of the unabbreviated designata. However, one might venture to see in some of them a 'cunning of Reason'—the abbreviation may help to repress undesired questions. NATO does not suggest what North Atlantic Treaty Organization says, namely, a treaty among the nations on the North Atlantic—in which case one might ask questions about the membership of Greece and Turkey. USSR abbreviates Socialism and Soviet; DDR: democratic. UN dispenses with undue emphasis on 'united'; SEATO with those Southeast-Asian countries which do not belong to it. AFL-CIO entombs the radical and political differences which once separated the two organizations, and AEC is just one administrative agency among many others. The abbreviations denote that and only that which is institutionalized in such a way that the transcending connotation is cut off. The meaning is fixed, doctored, loaded. Once it has become an official vocable, constantly repeated I general usage, 'sanctioned' by the intellectuals, it has lost all cognitive value and serves merely for recognition of an unquestionable fact" (Marcuse, 1964, 94).

While not all abbreviations are euphemistic in nature, abridged language often seem to serve as epitomical examples of euphemisms. Consider the common abridgement of Post-Traumatic Stress Disorder, 'PTSD.' In this case, the pain of the term (which had already been greatly reduced from its original term shell shock, as Carlin argued for) has been factored out completely. Harsh words like trauma, stress and disorder have been reduced to single letters, empty of any

obvious meaning that could potentially be painful. This abbreviation, Marcuse would claim, alters the praxis by stifling the potential to raise questions towards the reality by presenting 'PTSD' as an 'unquestionable fact.' If this is so, it would certainly seem to create distortions in the communication stage of the social praxis, as that which is 'unquestionable' will ultimately be removed from the social discourse altogether.[1] Indeed, upon analysis the employment of the term 'PTSD' does seem to prevent several important, social praxis-relating questions from being raised. For example, the abbreviated term hides the fact the disorder occurs post-trauma and could therefore be practically eradicated in a world that is trauma-free, thus reducing the potentiality of one raising questions of how to free the world from trauma.[2]

PART IV: CONCLUDING THOUGHTS

In this investigation, I have analyzed the effects of euphemisms on the social praxis, arguing that euphemistic language creates an evolution of language that degrades the clarity of language and thus compromises the effectiveness of communication (an essential feature of the form of social praxis), ultimately having negative effect on society. What can be learned from this investigation? It is my belief that the most important lesson to be gleaned is that we ought to be very careful about what language we choose to utilize when discussing important messages—after all, a number of philosophers (most notably Wittgenstein) have suggested that language comprises, to a large extent, our entire reality, and directly determines what we are able to know epistemically speaking. In one part of Philosophical Investigations, Wittgenstein goes as far as to say that "[p]hilosophy is a battle against the bewitchment of our intelligence by means of language" (Wittgenstein, 1953). These effects of language happen even in contradiction to our will. Euphemisms, after all, are used for mostly good intentions but end up betraying the clarity of the message in the end, having devastating consequences for society. Would, for instance, killing in war persist if we were not so keen on referring to it as "neutralization?" It may perhaps continue, but I nevertheless believe that much fewer people would be willing to "murder than to "neutralize."

1. And what is worse is that this exclusion will appear to be completely rational—who in their right mind would question the unquestionable?

2. Questions some certain despotic states may hope to avoid, insofar as they are dependent upon war and other forces of mass trauma.

If this is so, people ought to employ euphemisms with extreme caution, as it would seem euphemisms could justify even the most heinous of atrocities.

REFERENCES

Carlin, George. 2004. *When Will Jesus Bring the Porkchops?* New York, New York, United States of America. Hyperion.

Govier, Trudy. 2014. *A Practical Study of Argument* (Sixth ed.). Wadsworth/ Cengage Learning.

Marcuse, Herbert. (1964) 1991. *One-Dimesional Man.* Beacon Press Books.

Miller, David James, Ramsey, Eric and Schrag, Calvin O. 2003. *Experiences Between Philosophy and Communication: Engaging the Philosphical Contributions of Calvin O. Schrag.* SUNY Press.

Wittgenstein, Ludwig. (1953) 1973. *Philosophical Investigations.* Translated by G. Ascombe. Pearson.

Wittgenstein, Ludwig. (1922) 2010. *Tractatus-Logico Philosophicus.* Translated by C.K. Ogden. Project Gutenberg.

compos mentis

A Defense of Common-Sense Deontology

Jacob Gordon
Northwestern University

BIOGRAPHY

I am a third-year undergraduate at Northwestern University, where I study Philosophy, Critical Theory and Political Science. I am currently pursuing an honors thesis on Nietzsche's Critique of Morality. After graduation, I plan on continuing my interest in moral and political philosophy through either a PhD program or a JD program.

ACKNOWLEDGMENTS

I would like to thank Stephen White, an Assistant Professor of Philosophy at Northwestern, for advising this project. I would also like to thank Northwestern's Weinberg College of Arts and Sciences, whose Summer Research Grant funded my research.

ABSTRACT

In this essay, I attempt to defend common-sense deontological ethics from within a value maximizing framework that is often associated with consequentialism. I begin by exploring and defending what I take to be the two components of what is often called consequentialism's "compelling idea": first, a maximizing conception of rationality, and second, a commitment to the priority of value. After defending these two aspects of the "compelling idea", I argue that the proper conception of "value" is plural and agent-centered. Then, I use the value of "respectedness" to show that deontology can be a rational response to value, if value is conceived in a sufficiently complex way. My ultimate goal in this essay is to show that we can retain both the intuitions of deontology and the sensibility of the compelling idea, by understanding deontology as a rational guide to practical reasoning that arises from our basic commitments.

KEYWORDS

Consequentialism, Deontology, Value

INTRODUCTION

Deontological considerations have a substantial role in common-sense moral reasoning. Imagine that a woman named Jill watches a broken train barrel towards five people who are stuck on the track. Jill can push Tom, a stranger, in front of the train, and doing so will stop the train before it can reach the five. Jill will probably at least hesitate to push Tom. Even if she decides to push him, she will almost certainly be reluctant to do so. Jill, like most of us, thinks her actions should be restrained in some way, dependent not only on the simple calculation of maximizing life, pleasure, or any other apparent metric.

While many of us share Jill's deontological intuitions, these intuitions are notoriously difficult to justify. Critics argue that, if Jill's refusal to push Tom comes from an aversion to death, such an aversion should actually compel Jill to push Tom, because doing so will minimize the number of deaths Jill is able to produce. These critics contrast deontological commitments to a basic conception of rationality at the heart of consequentialism - a part of consequentialism's "compelling idea" - which states that, in Samuel Scheffler's words, "if one accepts the desirability of a certain goal being achieved, and if one has a choice between two options, one of which is certain to accomplish the goal better than the other, then it is, ceteris paribus, rational to choose the former over the latter"(Scheffler 1982, 414). If Jill accepts the goal of preventing death, she should simply produce fewer deaths, all else equal. Contrasting with this compelling idea, Jill's intuitive deontology irrationally "present[s] as desirable [a] non-relative goal whose maximum accomplishment it then prohibits" (Scheffler 1982, 416).

The first section of this essay will explore and defend the two parts of consequentialism's compelling idea: first, its maximizing conception of rationality, and second, its commitment to the priority of value. After defending the compelling idea, I will argue that the "value" we adopt should be plural and agent-centered. Then, I will show why deontology is a rational response to such plural and agent-centered value. Ultimately, this argument will show that we can and should retain both the intuitions of deontology and the sensibility of the compelling idea, by understanding deontology as a rational guide to practical reasoning that arises from our basic commitments. Finally, I will reflect briefly on how my argument relates to larger questions in moral philosophy.

1 - THE COMPELLING IDEA

The first principle at the heart of consequentialism's "Compelling Idea" is a maximizing conception of rationality. As above, Scheffler's conception of the compelling idea holds that "if one accepts the desirability of a certain goal being achieved, and if one has a choice between two options, one of which is certain to accomplish the goal better than the other, then it is, ceteris paribus, rational to choose the former over the latter"(Scheffler 1982, 414). For a moral theory to accord with this basic rationality, it must recognize that the better a certain action fulfills its stated goal, the more preferable it is.

This maximizing conception of rationality is incredibly intuitive. Making this point, Douglas Portmore has presented the obvious implausibility of a "deontological egoism", which forbids an agent from performing a single self-sacrifice even if doing so will minimize her total self-sacrifice. Insofar as an agent's goal is to minimize her self-sacrifice, such a deontological egoism, by forbidding the fulfillment of its stated goal, "seems paradoxical" (Portmore 2006, 14). Indeed, the maximizing conception of rationality merely states that an action A is preferable to an action B, from the perspective of a stated goal X, insofar as it better fulfills the stated goal X. A theory denies this only if it presents a goal X, admits that an action A will better fulfill X than action B, and still insists that B is preferable to A, without referring to any motivation besides X. The inconsistency of such a theory is undeniable.

Still, Scheffler's "ceteris paribus" clause opens the door to criticism from deontologists. "What if things are not equal, and there is a rule that simply forbids murder?", one could ask. There is nothing irrational, in Scheffler's sense, about such a rule. If taken as primary to our other goals, this rule could simply sway the prescribed action of our moral theory. In Jill's case, a primary and unconditional prohibition on murder would adequately forbid pushing Tom. Yet such a rule conflicts with the second prong of consequentialism's compelling idea: that a rule may only be justified if it advances a primary 'value'. While Scheffler's formulation of the compelling idea does not make this commitment clear, Jennie Louise's does. In her words, "Consequentialism, broadly construed, says only that agents should produce as much value as possible" (Louise 2014, 520). Consequentialism insists not only that one must be consistent in prescribing actions as to fulfill basic goals (as in Scheffler), but also that every rule or goal should be measured by its maximization of value.

This second prong of the compelling idea is, like the first, hard to deny. How might a deontologist who denies the primacy of value defend the rules she endorses? She cannot, given her denial of the primacy of value, argue that her rules are superior to others because they better achieve any given thing, because that something which her theory achieves would then represent a primary value. Further, she cannot even argue that there is value in her rules themselves being upheld, because doing so would admit that there is some value that morality aims to bring about, just that the following of rules constitutes that value. Admitting even this would lapse the deontologist into accepting the compelling idea. A theory's 'value' provides the reason for its prescriptions, so denying the primacy of value seems to entail denying that prescriptions have a reason. A theory admits the primacy of value insofar as its rules are rooted in their ability to further whatever that theory regards as basically important. It is hard to imagine a tenable normative theory that fails to root its prescriptions in value, construed as such.

To this point, I have defended the compelling idea's maximizing conception of rationality, as well as its insistence that value underlies moral rules. It might seem that this compelling idea does not dispel some classically "non-consequentialist" theories. Indeed, the literature examining which theories can and cannot be 'consequentialized', or made consistent with the compelling idea, is growing.1 My interest here, rather than to directly contribute to these higher-level distinctions, is to illustrate that certain rules - our common-sense deontological ones specifically - are not only consistent with the compelling idea, but are its reasonable offspring, if value is properly conceived.

2 - VALUE

2.1 - Potential Plurality and Agent-Centeredness

The two principles of the compelling idea, while illustrating that the moral status of an action should be a function of that action's promotion of value, say little about the nature of value. First, the compelling idea does not exclude the possibility of value plurality. As written, the compelling idea's rationality condition

1. See Doug Portmore's "Consequentializing Moral Theories", Jennie Louise's "Relativity of Value and the Consequentialist Umbrella", James Dreier's "Structures of Normative Theories", Mark Schroeder's "Teleology, Agent-Relative Value, and 'Good'", and Campbell Brown's "Consequentialize This", among others.

insists that it is irrational to hold that "an action A is preferable to an action B, though B better fulfills our only value, X, than does A." However, it is perfectly rational to hold that, "though B better fulfills X than A, A is preferable to B, because there is a value besides X, namely Y, which A better fulfills than B, and which takes priority over X." In other words, there being multiple intrinsic sources of value does not contradict the compelling idea, which only insists that we be consistent within our preference ordering, given our value. Consequentialism, as Louise writes, "does not say anything about what is to be regarded as valuable" (Louise 2014, 520).

Second, the compelling idea does not exclude the possibility of agent-centered value. A moral theory need not demand that all individuals maximize the same things. A simple example of a moral theory that utilizes agent-centered value, while abiding by the compelling idea's conditions, is ethical egoism. Egoism demands that every individual maximize only her own "good", and therefore formulates value in an entirely agent-centered way. Despite that, it still accords with both the rationality, and value-first reasoning, demanded by the compelling idea (Scheffler 1982, 416).

The fact that value could be plural and agent-centered does not mean that it actually is, but it does mean that the notorious spell-binding nature of the compelling idea does not eliminate theories of value that are plural and agent-centered. This point is significant because unity and agent-neutrality have often been sold as necessary accessories to the compelling idea. Scheffler begins his Consequentialism and Its Critics by selling this bundle. While presenting the compelling idea, Scheffler calls consequentialism "impersonal" (Scheffler 1988, 1), and insists that consequentialist theories "all share" the insistence that we "ought to…minimize evil and maximize good" (Scheffler 1988, 1). As shown, though, the conditions of agent-neutrality and of a unified "good" are not entailed by consequentialism's compelling idea.2 While I have not yet argued that agent-centered pluralist theories are preferable to their opponents, I have shown that they cannot be dismissed as non-consequentialist insofar as that label implies an inconsistency with the compelling idea. Answering yes to "Do I maximize value?" does not require one to answer a certain way when asked "Where does value lie?"

2. I have no problem with defining "consequentialism" to include only moral theories that offer an agent-neutral and unified theory of value, so long as we differentiate this condition from consequentialism's "compelling idea", which makes no such demand.

2.2 - Agent-Neutrality, Unity, and Alienation

The strengths of agent-centered and pluralist theories of value are best drawn by the weaknesses of agent-neutral or unified theories of value. Consider Peter Railton's "Alienation, Consequentialism, and the Demands of Morality", which begins by describing a husband, John, who treats his wife Anne as a perfect husband would, but who explains his behavior as follows: "I've always thought that people should help each other when they're in a specially good position to do so. I know Anne better than anyone else does, so I know better what she wants and needs…Just think how awful marriage would be, or life itself, if people didn't take special care of the ones they love" (Railton 1984, 135).

John's reasoning seems twisted and inhuman. As Railton puts it, "that he devotes himself to her because of the characteristically good consequences of doing so seems to leave her, and their relationship as such, too far out of the picture" (Railton 1984, 135). While John commits himself to a kind of rule-utilitarian practical reasoning, which aims at acting as will tend to make life better, his problem would remain if he were committed to a more directly utilitarian reasoning, through which he argued that "I should care for Anne, because doing so improves the world."3 John strays more basically, in actually valuing the wrong things. Instead of valuing Anne, on the most primary level, John values the entire world. John arrives at the conclusion to care for Anne second, only after calculating that he may care for the world best by caring for his wife. John is in this way "alienated" from those he should love. This kind of alienation is inevitable if we locate value in any agent-neutral place. If the thing I aim to bring about in moral action would always bind another agent as well as it binds me, then my actual relationships cannot serve as my primary motivations. This is a fundamentally inhuman way to value others.

Theories of unified value alienate similarly to those of agent-neutral value. Consider how Railton's John might respond when asked why he cares for his daughter. To satisfy us, his response would need to be about his daughter specifically. Alternatively, if John sees his wife and his daughter as different sources of the same value (as is necessary on a unified theory of value), he will be

3. Railton's essay ultimately argues that the alienation problem does not doom utilitarianism. I think his argument fails, but exploring that is beyond the scope of this essay, which is not capable of responding to every response to the alienation problem. If readers wish to explore defenses of utilitarianism against the charges I present here, they can read the second half of Railton's essay. I should note that I do not know why some feel so compelled to save utilitarianism in the first place.

alienated from each, only caring for each as a factor of her being instrumental for that value. To genuinely care for his wife and his daughter, John should love each irreducibly. Insofar as we are unsatisfied with any theory that fails to see John's wife and daughter as independent and irreducible sources of value for John, we will be unsatisfied with any theory that places value in a single, or agent-neutral, place.

2.3 - A More Faithful Alternative - Agent-Centered Plurality

Fortunately, we may view value in a different and more elegant way, as arising from our individual relationships. If we believe that John should care for his wife simply because he loves his wife, and care for a stranger simply because he values that stranger, then we believe value ought to be agent-centered and plural in nature. This admission allows us to easily avoid some of consequentialism's famous problems. If I need to choose between saving my child and saving two other children, I can simply save my child. Doing so is justified because the value of my child, to me, is irreducible to my caring about the world, and is simply greater (to me) than the values of the two other children. We retain this ability under an agent-centered and pluralist theory of value without having trouble explaining general benevolence. We ought to care for those who suffer simply because we value those persons.

This picture is more faithful to our genuine moral reasoning than any agent-neutral or unified theory of value can be. We are more inclined to answer, "why do you care about the world?" with "because I care about the individuals in it", than to answer, "why do you care about individuals?" with "because I care about the world." If this is correct, the foundation of value is in our individual relationships, and is therefore agent-centered, arising from our feelings of commitment, and plural, located in each individual, rather than in some mass of aggregate "moral stuff".

The plurality of value goes further even than that it is located in separate individuals; it seems located in different feelings and attributes of different individuals. A good friend wants to sponsor both her friend's autonomy and her joy. These two values are distinct, but each independently draws the good friend. It would be wrong to reduce autonomy to joy, or joy to autonomy.4 What we

4. The plurality of value, as I have described it, delivers a response to Mark Schroeder's allegation that "Agent-Relative Teleology" relies on the incoherent concept of "good-relative-to", which

are left with on this picture is appealing and simple - we ought to value what we actually value.

2.4 - Defending a Faithful Conception of Value

This conclusion might seem unsettling for a few reasons. One could ask, "Isn't the purpose of moral theory to tell us what we "should" value, rather than what we "do" value?" But how would we be convinced that we "should" value something new? Such an argument needs to be based on a set of shared premises, which we already do accept. An argument calling on us to change our understood sources of value is really an argument that we are already committed to accepting different sources of value. The goal of moral theory is to show us, at root, what we already are committed to, and to direct our moral beliefs and actions towards those commitments.

I have intended to show, largely through Railton's example of John the husband, that our primary sources of value are agent-centered and plural, meaning that no agent-neutral or unified theory of value can connect with our most basic commitments. We are individual relationships "all the way down." This is the foundational point of agent-centered and pluralist theories of value. These theories, aside from their more intuitive prescriptions which I will outline below, undoubtably accord better with our humanity.

It could still be argued that the picture of value I am drawing is too permissive. Is it true that all morality requires is for an agent act in accordance with her own value? If this is the case, does nobody ever act wrongly? I have two responses to this critique. First, my argument is not committed to accepting that morality is subjective in the alleged way. Even if we do accept a kind of value-objectivity, we should accept that value lies where it seems to given our most basic commitments, which is, as shown, in agent-centered and plural places.

Second, even if we do accept value-subjectivity, the concern about over-permission applies to far fewer cases than it may seem to, because an agent cannot simply choose whether an action accords with her deepest values. A son

is allegedly less clear than the concepts of "the good" or "good for" (Schroeder 2007). Even if Schroeder's critique is sound (I am inclined to think that all three of those concepts are incoherent), it presents no problem for the agent-relative pluralist theory of value that I have described. Perhaps we cannot understand what it means for an outcome to be "good-relative-to" a given agent, but we can understand what it means for an outcome to include more of an agent's friend's pleasure, or an agent's son's autonomy.

can tell his father that he was right not to bring a jacket to the park because he does not feel cold, but his insistence will have no effect on whether he really feels cold. Subjective value would act in much the same way. When we act against our values, nothing we say will make it otherwise. What this means is that the subjective agent-centering of value is not as terrifying a prospect as it might seem. "Coldness" is subjective, but we understand, for the most part, what causes others to feel cold. As we can reliably conclude when another person "should have brought a jacket", we can conclude then they should have acted differently. For example, if Jesse is rude to a stranger, even the subjectivist can criticize her rudeness, not because Jesse's rudeness fails to align with the critic's values, but because Jesse's rudeness probably fails to align with Jesse's own values. If Jesse looked into the eyes of the stranger, and heard about his life, she would surely regret being rude. To be clear, my theory is still entirely consistent with an objectivist view, but in case the agent-centering of value seems a slippery slope towards the 'subjectivizing' of value, I am not sure the terrain of that slope is so rough.

Lastly, it is worth noting that an agent-centered and pluralist theory of value is much better at responding to general moral skepticism than is an agent-neutral or unified one. A person with strong moral intuitions can always feel uncompelled to a utilitarianism, which calls her to accept a foreign theory of value. If she asks that pesky question of "why?", an agent-neutral unified theory often has no clear response. She cannot similarly ask "why?" to a theory of value whose structure is determined by the entailments of her most primary commitments.

In this section, I have attempted to illustrate that a plural and agent-centered theory of value is consistent with the compelling idea (in 2.1), that such a theory of value is faithful to our actual values (in 2.2 and 2.3), and that we are right to adopt, in moral reasoning, the theory of value that is faithful to our actual values (in 2.4). From this point, I will explain how an agent-centered and plural theory of value can reasonably lead to the deontological conclusions we intuitively hold.

3 - DEONTOLOGY

3.1 - Setting the Stage: Why Not Deontology?

I will begin this section by illustrating the alleged irrationality of deontology. Then, I will describe how the value of subjects' being respected, while being a

reasonable value for a pluralist agent-centered theory to adopt, also leads to deontological conclusions.5

Deontology appears irrational chiefly because of the difficulty it faces in responding to the "lesser evil" problem. If we accept that a certain value ought to be promoted, it seems irrational to establish rules relating to that value that neglect the net effect agents can have on that value's promotion. If we, for example, agree that people have a right not to be slaves, it seems that an agent's primary drive should be to minimize the total number of slaves, rather than to avoid using slaves herself. And if one is forced to use a slave to minimize the total number of slaves, it seems that she should use that slave. Deontology, by insisting that an agent never use a slave, even to minimize the number of slaves, allegedly endorses a worse outcome and forbids a better one. It therefore "present[s] as desirable [a] non-relative goal whose maximum accomplishment it then prohibits" (Scheffler 1982, 416), and fails to abide by the compelling idea, as above.

The proper analysis of value, however, shows that this critique over-simplifies the situations that draw on our deontological intuitions. Consider Jill, from the beginning of this essay. She must choose whether or not to push Tom in front of the runaway train. Contrary to the picture illustrated by the critic above, pushing Tom does not simply promote more of the value of "life" or "pleasure" than not pushing Tom, because, as established, each individual Jill is in a position to affect is a separate source of value, irreducible to the others she is able to affect. It would be different if, for example, Jill saw an unsuspecting Tom stuck in the middle of a road, with a bus coming toward him. There, since Tom's pleasure, or life, is a discrete and single source of value that can be maximized, Jill would undoubtably be right in acting as the "utilitarian" would towards him. She should push him out of the way of the bus, perhaps hurting him in the process, because doing so would maximize his net pleasure, or life. But, again, the case Jill faces to begin this essay is quite different, because Jill must choose what to do, given the plurality of values to which she is drawn. This point alone does not justify deontology, but it is significant to note, in order to make our depiction of "the lesser evil" problem more genuine. Even in cases that do not call on our deontology, such as one in which we must choose whether to save one drowning person or two other

5. This method is preferable to one that lays out a more complete list of the values that a proper theory would recognize for two reasons. First, I cannot claim to know what this complete list would look like. Second, drawing up such a list would likely spark a substantial amount of disagreement that is irrelevant for the purpose of justifying deontology.

drowning people, we always deal with discrete and independent persons, who should be recognized as such.6

3.2 - Deontology Through "Respectedness"

The more significant way in which critiques of deontology fail, though, is by under-recognizing subtle sources of value, particularly the value of subjects' being respected. Our intuitive deontology, only appearing to endorse worse outcomes, actually endorses better ones, given its acknowledgment of the value that lies in subjects' being respected, or in their "respectedness" (for lack of a better term). If respectedness is a credible source of value, and if its maximization leads to intuitive deontological conclusions, then those intuitive deontological conclusions, far from being irrational, are particularly compelling.

We ought to value respectedness because humans have real personal bonds to each other, which are inspired by mutual compassion for and understanding of others' experiences. Disrespect represents the harm done by the very breaching of one of these bonds, which is a harm that exists in addition to whatever harm is otherwise done to lead to that breach. Insofar as we value our very compassion for one another, we ought to accept that when human bonds are broken, real harm follows. If a man's arm is broken in a natural disaster, he may experience intense physical pain, but if that man's arm is broken willingly by another person, he will also face an accompanying emotional pain. Such a personal attack will, in a real way, "add insult to injury." Because its knowing intent causes an additional harm to the pain otherwise caused, human-driven suffering is generally worse than naturally-driven suffering.

Importantly, this value of respectedness magnifies harm more than it can magnifies 'help.' While a person helping another person might strengthen the human bond between them, and thereby represent an added benefit over and above the benefit otherwise given, that added benefit does not impact the subject

6. While some argue that we are not so clearly compelled to save, for example, two strangers rather than one (See John Taurek's "Should the Numbers Count?"), I will assume here that the solution to this moral problem is as straightforward as most suppose it is. It might seem that this de-commits me from the true plurality of value - that I have chosen in this assumption to treat three strangers' lives as simply countable. I reject this characterization, because I do not deny the real moral remainder in failing to save the one, even in saving the two. I acknowledge that hard questions sometimes have answers, but that does not put me in the same boat as those who deny that there are hard questions.

nearly as much as would that bond's breaking. Since the human experience deeply requires independence, we care more about not being unnecessarily hurt by others than we do about being unnecessarily helped by them. That our mutual bonds are composed chiefly of our not harming each other is exemplified by the fact that we are angrier when pushed to the pavement than we are relieved when helped off of it. Since each of us largely wants to be left alone by strangers, our being respected depends on our not being harmed by others much more than it depends on our being helped by them. Agents therefore respect others chiefly by not harming them.

What does maximizing respectedness look like in Jill's case? If we assume that Jill cares equally for all six people she may affect (between Tom and the five on the track), and if we accept the compelling idea, then whether pushing Tom is right will depend on whether Jill will bring about more value by pushing him or by not pushing him. The value in these six individuals, though, lies partially in their being respected. Jill can only maximize Tom's respectedness, as above defined, by refusing to push him in front of the train, since his death on her hands would break her human bond with him. Interestingly, though, Jill need not push Tom in front of the train to respect the five on the track. Since these five have found themselves in the path of a runaway train by accident, their deaths will not involve a lack of respect, because no person will have put them in a position to die, which would involve a unique harm over and above that of dying. Further, it is not as though Jill can greatly increase the five's respectedness by saving them, since their being respected is dependent, as above, chiefly on their not being harmed by another person, rather than on their being helped by another person. If Jill refuses to push Tom, she will maximize the net respectedness of the subjects she can impact. Thus, respectedness provides Jill with a perfectly justifiable value whose maximization is only achievable if she refuses to push Tom. Through respectedness, Jill can retain her intuitive deontological commitments, without betraying the compelling idea's maximizing conception of rationality.

Some might be unwilling to accept that Jill should worry about maximizing respectedness in a decision whose consequences involve multiple deaths. First, it should be noted that Tom's respectedness may not outweigh the lives of the five; perhaps Jill would be right to push Tom. My analysis of respectedness simply shows that our deontological intuition is right to recognize that something muddies the waters of Jill's choice. We are right that Jill's case is different than one in which she

must either save one drowning person or five drowning people. Second, I think respectedness becomes a more powerful value the more deeply it is considered. We do frequently judge situations that involve disrespect as worse than similar ones that lack disrespect. News of a person's being murdered is worse than news of their being killed by a fallen tree (even if the fallen tree may cause more physical pain than the murder), precisely because there is disvalue in someone's being disrespected. As so much of our lives are built on mutual trust and respect, the value of maintaining that respect should not surprise us. It may seem an odd value to systematize, as I have, simply because it is subtler than something like pleasure or pain. But the subtle importance of being respected may still rival the more obvious importance of avoiding physical pain. I write this to suggest that Jill's refusal to push Tom, on the grounds of maximizing his respectedness, really can be a perfectly reasonable choice. She could say to the five on the track "I am deeply sorry that nature has given you this fate, but the damage done when a person ends another person's life is so great that it should not be done."

Furthermore, the non-parallel relationship Jill maintains with Tom's respectedness relative to the respectedness of the five on the track would be partially retained if the runaway train were directed by a murderous conductor, rather than by a natural accident. Such a murderous conductor's actions, while more intentional than those of nature, are still likely less personal than Jill's. As this conductor is driven to murder in the first place, Jill may be right to assume that he lacks some basic moral apparatus that she has, and that he therefore never maintained the moral bonds whose breaching constitutes disrespect. Being killed by this murderous conductor, almost like being harmed by a child, lacks the degree of true and sophisticated disrespect that being killed by Jill would involve. We would rather be harmed by the school bully (the conductor), who lacks the basic capacity for compassion and understanding, than by another victim of the bully (Jill), who understands the harm caused by the bully, and who we thought we could trust. We can imagine that, in the path of the mad conductor, the individuals on the track might ask "why has fate befallen us this way?" as they would if struck by a natural disaster, rather than "why have you, another person, done this? Don't you care about me?", as Tom would ask Jill if she were to kill him. This means that refusing to murder to prevent more murders, in addition to preventing more deaths, can be justifiable on the grounds of respectedness.

It could be argued that this concept of "respectedness" is circular. Specifically, I rely on the fact that Tom would be disrespected if killed by Jill to argue that Jill would be wrong to kill Tom, even though Tom would perhaps only be disrespected in being killed by Jill if Jill would be wrong to kill him in the first place. I do not think this critique is correct, because Jill's personal bond with Tom exists independently of her bonds with the other five. As such, if Jill kills Tom, a real bond will be broken, and he will be disrespected, regardless of whether Jill saves the five. The fact that Tom will be disrespected if killed comes first, as an input to be weighed against the lives of the five, rather than second, as dependent on how compelled Jill is to save the five.

For the reasons above, it seems that maximizing respectedness can ground many of our deontological intuitions. I will not argue that our deontological intuitions can always be justified in this way, but it seems that they often can be. While Jill's case deals with the question of murder, it could just as well deal with questions of torture, stealing, or lying. Our common-sense deontology, I think, never actually forbids the lesser evil. It only prescribes rational action based on its recognition of value plurality, and of respectedness' importance.7

3.3 - Evaluating Agent-Centered Justifications of Deontology

I have attempted to justify deontological rules by citing plural values, rather than the agent-centeredness of decision making. Agent-centered justifications of deontology are possible, and frequently given, but they are often either circular or selfish. The circular argument for deontology supposes that we have a good reason to remain "morally pure" (meaning that we should, for example, refrain from killing one to save five), without first showing that refusing to kill is the "purer" choice. Certainly, a utilitarian would argue that "moral purity" is maintained by killing the one, because doing so is the right thing to do, and because moral purity should entail doing the right thing. The deontologist who relies on "moral purity" often presupposes that the rule she is defending is right in her argument that it is right.

7. It is possible that there is some value that overlaps with or exists separately from respectedness, which can do the same work for deontology that respectedness does. While I think respectedness alone can justify deontology, I do not think it must be alone in doing this. As noted in footnote 5, I welcome the possibility of new values, and do not claim to have a grasp on all the subtle values that would belong in a comprehensive moral theory.

Deontologists may make a subtler argument for moral purity, which defends refraining from causing harm, rather than refraining from doing 'wrong'. While it is possible that killing one to save five is not wrong, it certainly causes harm. And so, if refraining from causing harm is justified, a blanket prohibition on killing might be too. The problem with this position is that any argument for refraining from causing harm, that does not devolve into an argument for refraining from doing wrong, is selfish. Rather than being primarily about others, any such argument is primarily about ourselves, and the actions we would like to take. If we cared primarily about others, we would commit ourselves simply to do the right thing, even if the right thing caused some harm. For example, if I were forced to amputate someone's finger to save her arm, and I refused to amputate the finger because "I wished not to cause harm", my refusal would clearly be sourced in my own squeamishness, rather than in my care for that person. Perhaps these personal aversions can maintain marginal weight in our decision-making (that one might receive serious psychological trauma by killing someone is significant), but if these selfish considerations were the foundation of our deontology, that deontology would be indefensible.

This point detracts neither from my argument that value ought to be agent-centered, nor my argument that human-driven harm is generally worse than natural similar natural harm. Those two considerations affect our evaluation of consequences and impact our consideration of which actions are right. Meanwhile, justifications of deontology that cite the agent-centeredness of decision-making go one step further. They argue that, even once we have figured out what is right, we can refuse to act rightly, simply because of the actions we wish to take.

Alternatively, the justification of deontology that I have provided remains consistently about others. Besides being generally more defensible, this position tracks better with our intuitive deontological conclusions than does a system of deontological rules, like "do not kill", that address the actions themselves that an actor is allowed take. To show this, imagine that Tom's entire family is stuck to the train track, and that Tom asks Jill to push him in front of the train, which will still kill him, so that his family can live. Tom, here, can only get in the train's way with Jill's push, because he needs help getting past a guard-rail. Since Tom asks to be killed here, and since he has a good reason for his request, this case intuitively differs significantly from the case that began this essay. However, any deontology concerned only with the actions Jill can take will be unable to explain the clear

difference between pushing Tom here and pushing Tom in the original case, since the difference between the two scenarios lies in the subjects of Jill's acts, rather than in the acts themselves. Meanwhile, since my theory sources Jill's deontology in Tom's being respected, and since respecting Tom requires considering his requests, my theory is able to explain the difference between this case and the initial one. By focusing on subjects rather than actions themselves, my theory tracks with our intuitive deontology in a way agent-centered explanations of deontology do not.

3.4 - Addressing "Gimmicky-ness"

My goal in this essay has been to justify our intuitive deontology, by showing that it arises from a reasonable understanding of value. Some critics argue that justifications of deontology like mine, which attempt to demonstrate that deontology is consistent with the compelling idea, fail to do justice to the deontological commitments they claim to defend. Robert Nozick, for example, once accused these justifications of being "gimmicky" (Nozick 1968, 4). The core of Nozick's charge, I think, is that these justifications of deontology use reasoning that differs substantially from the reasoning that genuinely drives our intuitive deontology. This criticism could be substantial. Insofar as my goal is to justify common-sense, it would be deceptive for me to reach the same conclusions as common-sense, but from completely different grounds. If supposed justifications of deontology do this, they do not really defend our intuitions. Luckily, I think the reasoning I have used tracks well with our intuitive reasoning. By locating the importance of deontology in the importance of independent individuals being respected, I seem to have utilized the same sort of reasoning that Robert Nozick himself used in his defense of "side constraints": "Side constraints upon action reflect the underlying Kantian principle that individuals…may not be sacrificed or used for the achieving of other ends without their consent"(Nozick 1988, 138). Using an individual 'for the achieving of other ends without their consent' seems to be a particular way of disrespecting them, such that the image of value I have drawn captures, in a systematized way, the Kantian intuition expressed by Nozick. Furthermore, my theory's sincerity to our intuitive reasoning is reflected by the fact that its analyses of situations track better with our intuitions than do the analyses of common agent-centered justifications of deontology, as shown above. In sum,

I understand and respect the risk of "gimmicky-ness", but do not think that the particular picture of deontology that I have drawn is "gimmicky".

CONCLUSION

In this essay, I have sought to show that our intuitive deontological commitments rationally arise from a commitment to agent-centered and plural value, because the subtle value of "respectedness" compels us to act as our deontological intuitions also drive us to act. While my arguments address a number of topics, such as the viability of deontology and the nature of value, my chief motivation in writing this essay has been to explore whether our moral intuitions are defensible. I have argued that that they are.

If I am right, we can understand our common-sense morality as sincerely founded on a genuine interest people have in others' well-being. Among other things, this might indicate that moral theorists typically underestimate the complexity of the average person's moral reasoning (this essay might be a 7,000+ word explanation of a reasoning process we all make without even thinking). If I am wrong, we will be forced to accept that our moral common-sense is fundamentally inconsistent with the things we claim to most basically value. This would be a heavy charge to accept. It would force us either to abandon our deep moral intuitions, or, if we wish to keep these intuitions, to admit that our morality is fundamentally irrational, not actually arising from our care for others, but instead from something like selfish calculation or predetermined evolutionary psychology.

Our answer to the question of whether moral common-sense is defensible, therefore, informs and reflects our answers to moral and social philosophy's deepest questions. Regardless of whether my arguments have been compelling, this is a question that deserves to be asked.

REFERENCES

Brown, Campbell. 2011. "Consequentialize This". *Ethics* 121 (July 2011). 749–771.

Dreier, James. 1993. "Structures of Normative Theories". The Monist 76 (1993). 22–40.

Louise, Jennie. 2014. "Relativity of Value and the Consequentialist Umbrella". *The Philosophical Quarterly* 54 (October 2014). 518–536.

Nozick. Robert. 1968. "Moral Complications and Moral Structures". *Natural Law Forum*. 1–50.

Nozick, Robert. 1988. "Side Constraints". Printed in *Consequentialism and its Critics,* edited by Scheffler, Samuel. Oxford: Oxford University Press. 134–142.

Portmore, Douglas. Draft from April 2006. "Consequentializing Moral Theories".

Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality". *Philosophy and Public Affairs* 13 (Spring 1984). 134–171.

Scheffler, Samuel. 1982. The Rejection of Consequentialism. Oxford: Clarendon Press. revised edition.

Scheffler, Samuel. 1988. "Introduction". *Consequentialism and its Critics,* edited by Scheffler, Samuel. Oxford: Oxford University Press. 1–14.

Schroeder, Mark. 2007. "Teleology, Agent-Relative Value, and 'Good'". *Ethics* 117 (January 2007). 265–295.

Taurek, John. 1977. "Should the Numbers Count". *Philosophy and Public Affairs 6* (Summer, 1977). 293–316.

compos mentis

Sex Differences and Gender Bias in SSD

Eleanor Goulden
Occidental College

ACKNOWLEDGMENTS
First and foremost, I'd like to thank Aleksandra Sherman for all her excellent advising work that went into this project. Furthermore, I'd like to acknowledge all the chronically ill women who have faced misdiagnosis and whose activism helped inspire this paper.

ABSTRACT
Hysteria originally was characterized by a cluster of physical symptoms thought to be caused by the uterus moving throughout the body. One of the most modern iterations of hysteria in the DSM-V is Somatic Symptom Disorder (SSD), a psychiatric disorder characterized by excessive psychological response to physical symptoms. Though many changes have been made over the years, SSD, like its predecessor, is still diagnosed mainly in women. This diagnostic discrepancy can be explained by a combination of medical gender bias in the treatment of women's symptoms as primarily psychological in origin and differences in the underlying neurobiological mechanisms that govern pain response. Due largely to the societal stereotype that women are more emotional and psychologically unstable than men, many medical practitioners seem to assume a certain level of exaggeration and unreliability when listening to women describe their symptoms. This is evidenced by various studies showing that doctors and nurses tend to judge women's pain as less severe than men's and prescribe women antidepressants and referrals for psychotherapy for pain and men surgery, painkillers, and physical therapy. While evidence suggests women experience pain more frequently and intensely than men, this difference seems to be due mainly to underlying neurobiological differences rather than psychological ones. Sex differences have been noted in neural immune response to pain, as well as in the roles of sex hormones in pain analgesia. These differences, however, have only been known about for the last few decades due to a deficit in women's health research. As a result, practitioners have been overly focused on proposed psychological gender differences in pain experience that are not upheld by empirical research. Future medical practice must account for the unique neurobiology of pain experience in the sexes while also evading the long held belief that women's symptoms are more psychologically based.

KEYWORDS
Pain, Somatic Symptom Disorder, Gender Bias, Sex Differences in Pain, Medical Sexism, Hysteria, Sex as a Biological Variable, Illness

INTRODUCTION

Somatic Symptom Disorder (SSD), a disorder classified by disproportionately large distressing psychological reactions to physical symptoms, is diagnosed in women much more frequently than men, with an estimated ratio of 10:1 (Kurlansik and Maffei 2016). In this paper, I present multifaceted evidence to argue that this discrepancy can be partly explained by a historical bias in the treatment of women's symptoms as psychologically-based, as well as by neurobiological sex differences in physical symptom experiences.

I begin by providing a historical analysis of SSD. Specifically, I examine how the medical meaning of "hysteria" has changed over the past 3000 years and how we can trace these changes in meaning to today's definition of SSD. This historical analysis provides evidence for a gender bias in SSD diagnosis that seems to stem from a centuries-old tendency to treat women's symptoms as psychological in origin. I give a thorough explanation of the current diagnostic criteria for the disorder, offering a brief description of the five main theories for the cause of SSD. I then explore current gender bias in medical treatment, honing in on the continuing inclination to treat women's symptoms as psychologically-based and men's symptoms as physically-based that can partially explain the diagnostic disparity in SSD.

In order to also consider the possibility that women are genuinely at higher risk for SSD, I then look at differences in the way men and women experience symptoms, especially pain. Because pain is a multifaceted phenomenon involving both psychological and biological aspects, I consider whether there are gender differences in symptom experience relating to both facets. The research suggests that though there are differences in the way men and women experience symptoms, they are mainly neurobiological rather than psychological in nature. Specifically, if there exist differences at all, I argue that they are such that women experience more physical pain than men. This contrasts strikingly with how men and women are treated for their symptoms; the research suggests that men are often treated for physical pain, whereas women are often treated as psychiatric patients.

Finally, I review how the diagnostic discrepancy in SSD, a primarily psychological disorder, acts as an example for the failures and successes in how we factor gender and sex into medical diagnosis and treatment. I consider how research into sex differences in pain experience may inform future healthcare

practice while also stressing the danger of making assumptions about pain and psychological status based on gender and gender stereotypes**.**

SSD AS A WOMAN'S ILLNESS: A HISTORICAL ANALYSIS

Many scholars attribute the creation of hysteria to Hippocrates in Ancient Greece (Gilman 1993). While Hippocrates seems to be the first individual to coin the term "hysteria" in the fifth century BC from the Greek word for uterus "hysteros", hysteria actually has origins in Ancient Egypt with the Kahun and Eber Papyrus Scrolls, dating back to 1900 BC and 1600 BC respectively. These documents did not yet assign the malady a name, but described a series of disorders with various symptoms attributed to movement of an unhappy, sick uterus, thus diagnosed exclusively in women (Tasca et al. 2012). This idea was partially born out of the sense that women were lesser beings than men and thus uniquely afflicted by certain medical problems. This definition of hysteria, that physical symptoms in women arose from unhappy, moving uteruses, remained unchanged for centuries.

The idea that women were inferior beings prone to sin and sickness flourished in the middle ages as Christianity rose to power in Europe. Hysterical women became pariahs, witches who were plagued with an illness from the devil. If a physician could not understand or find the cause to physical illness, it was assumed that the devil was responsible. As a result, the focus of treating hysteria veered from helping treat the patient's symptoms to eradicating the devilish presence. This attribution of hysteria to witchcraft remained into the sixteenth century, until the Dutch physician, Johann Weyer, theorized that hysterical "witches" were mentally ill. Although hysteria continued to be associated with the uterus (and thereby with women), Weyer's theory was highly influential in shifting views of hysteria as a mental malady (Tasca et al. 2012). Hysteria began to take shape as a mental disorder resulting in physical symptoms that was unique to women.

In the 1800s, the idea that medically unexplained symptoms were attributed to the devil faded away in favor of the idea that medically unexplained symptoms were due to some underlying psychological abnormality. Similarly, the disorder was no longer associated with movement of the uterus but instead with some problem involving the nervous system or the brain. Jean-Martin Charcot declared it to be an inheritable nervous system disorder while Robert B. Carter assigned it two primary markers or criteria, the first being that hysteria arose from too much emotion in the nervous system and the second being that once afflicted, those

with the disorder had something to gain from being sick (Crimlisk and Ron 1999). As medical understanding increased, the idea that the disorder was caused by a migratory uterus became largely obsolete, but the disorder was still diagnosed almost solely in women. This new definition of hysteria strengthened the idea that not only were women's physical symptoms often psychological in origin, but that these reported symptoms might have been exaggerated to reap some sort of benefit, whether it be in the form of financial gain or simply more attention.

A significant change in the understanding of hysteria occurred in the latter half of the nineteenth century. Sigmund Freud revolutionarily crafted the idea of "male hysteria", even diagnosing himself with the disorder. Freud suggested sexual frustration as a possible cause of the disorder and described hysteria as the transformation of these unconscious desires into physical symptoms. He defined this transformation as the patient's "primary gain" because the patient was able to ignore his or her psychological traumas in favor of treating physical symptoms (Tasca et al. 2012). Despite these changes, hysteria still remained mainly a female disorder. With the World Wars of the twentieth century came mass epidemics of hysteria, still afflicting mainly women. The psychological stress of the war was thought to result in a peak in cases of the disorder. After the wars, however, the medical use of the term "hysteria" fell out of use. Some concluded that Freud had essentially cured the world of hysteria, but in reality, it still existed, just under other names (Gilman 1993).

The term "hysteria" was ruled largely obsolete in medical practice in 1980 when it was absent from the DSM-III. It instead reappeared under a various number of names and classifications, most clearly in the form of Somatization Disorder (APA 1980). Somatization Disorder in the third and fourth editions of the DSM was characterized by a list of physical symptoms rather than psychological ones. Diagnosis was generally only made when symptoms had no known medical explanation. The fourth edition specified that patients with the disorder tended to descibe their symptoms in overexaggerated, colorful terms with an unclear symptom timeline. It warned that patients tend to seek out the advice of many doctors and may undergo unnecessary medical testing and treatment (APA 1994). Somatization Disorder was diagnosed at a much higher rate in women than men, with some physicians even arguing it to be nonexistant in men (Golding, Smith, and Kashner 1991). In other words, some practicioners still openly viewed this

new iteration of hysteria, Somatization Disorder, as solely a women's disorder, even into the 1990s.

When crafting the DSM-V section of Somatoform Disorders, the task force seemed to keep in mind the critiques of the DSM-III and DSM-IV's diagnostic criteria for Somatization disorder; that they were too specific and strict, that they seemed to emphasize an archaic dualistic approach to mind-body separation, and that these problems made physicians uncomfortable with making a diagnosis (Dimsdale and Levenson 2013; Kurlansik and Maffei 2016). The focus shifted from the symptoms of the body to the symptoms of the mind. In doing so, the actual name of the disorder changed in 2013. Somatization Disorder, Pain Disorder, and Somatoform Disorder Not Otherwise Specified were dissolved into one disorder: Somatic Symptom Disorder. (Rief & Martin 2014).

Rather than focusing on the physical symptoms themselves, the diagnostic criteria for SSD focused on the psychological impact of physical symptoms. Criterion "A", for instance, requires that only one "distressing" physical symptom exist, without specification about bodily area affected, age of onset, or whether or not it has a known medical explanation. Criterion "B" consists of three manifestations of "excessive thoughts, feelings or behaviors" relating to the symptoms. These include having thoughts about the seriousness of one's illness or conditions that are disproportionate to the actual nature of the illness, having persistently high levels of anxiety about the nature of these symptoms, and devoting excessive time and energy to the symptoms. Only one of these manifestations need be present for a diagnosis of SSD. Severity of the disorder is dependent on how many of the manifestations are present (Rief and Martin 2014). Criterion "C" specifies that although the one particular physical symptom in Criterion "A" need not be persistent, the patient must be persistently symptomatic for more than six months (APA 2013; see Appendix 1).

Interestingly, there is currently no single etiological theory of SSD that is universally agreed upon. Based on the literature, there are at least five main theories for describing the causes of SSD: somatosensory amplification, the "vicious cycle effect", alexithymia, catastrophizing, and hypervigilance. Although there is some overlap in the description of these theories (e.g. somatosensory amplification is often co-morbid with alexithymia, and its symptoms and diagnostic are very similar to those of hypervigilance) there are also clear distinctions (Wise and Mann 1994).

Somatosensory amplification and the vicious cycle effect are reliant largely on the existence of proposed biological abnormalities in SSD afflicted patients. SSD may be caused by somatosensory amplification, during which an SSD patient perceives normal sensations as more noxious than healthy patients due to overactive sensory pathways (Harvey, Stanton, and David 2006). A patient with SSD may interpret a typically innocuous experience like digestion, for example, as more intense and painful than the average person. fMRI, PET, SPECT, and structural MRI imaging techniques have pointed to possible abnormalities in a variety of nervous system and cortical structures related to somatosensory processing in patients with SSD and related disorders. These include striatal and amygdalar abnormalities, bilateral caudate-putamen hypometabolism, decreased amygdalar volume, and possible differences in the lamina 1 spinothalamic cortical pathway (Perez et al. 2015). The next theory, the "vicious cycle effect", occurs when a patient's negative mood and unpleasant physical symptoms feed off each other, causing a worsening of both. A patient may feel sick and decide to stay in bed, leading them to isolate themselves, exacerbate their depression, and exacerbate their perception of their symptoms. Often, thinking about increased symptom experiences and worsened health continues the cycle (Cooper, Booker, and Spanswick 2003). Notably, this cycle is not unique to patients with SSD but has been identified in many patients with a variety of chronic health conditions and comorbid mental illness leading to negative affect, like depression or anxiety (Gatchel 2004; Katon, Lin, and Kroenke 2007; Teasdale 1983)resulting from a paradigm shift from an outdated biomedical reductionism approach to a more comprehensive biopsychosocial model, which emphasizes the unique interactions among biological, psychological, and social factors required to better understand health and illness. This biopsychosocial perspective is important in evaluating the comorbidity of mental and physical health problems. Psychiatric and medical pathologies interface prominently in pain disorders. Important topics in the biopsychosocial approach to comorbid chronic mental and physical health disorders, focusing primarily on pain, are presented. Though this biopsychosocial model has produced dramatic advances in health psychology over the past 2 decades, important challenges to moving the field forward still remain.","ISSN":"0003-066X","shortTitle":"Comorbidity of Chronic Pain and Mental Health Disorders","language":"en","author":[{"family":"Gatchel","given":"Robert J."}],"issued":{"date-parts":[["2004",11]]}}},{"id":320,"uris":["http://zotero.org/

users/3294891/items/8KCJUDPZ"],"uri":["http://zotero.org/users/3294891/items/8KCJUDPZ"],"itemData":{"id":320,"type":"article-journal","title":"The association of depression and anxiety with medical symptom burden in patients with chronic medical illness","container-title":"General Hospital Psychiatry","page":"147-155","volume":"29","issue":"2","source":"ScienceDirect","abstract":"Background\nPrimary care patients with anxiety and depression often describe multiple physical symptoms, but no systematic review has studied the effect of anxiety and depressive comorbidity in patients with chronic medical illnesses.\nMethods\nMEDLINE databases were searched from 1966 through 2006 using the combined search terms diabetes, coronary artery disease (CAD. Though there is a lack of research into the exact nature of the vicious cycle effect in SSD, researchers have suggested that negative affect is the most accurate predictor of a patient's reported symptom severity, more telling than the actual nature of the disease (Dimsdale and Levenson 2013; Van den Bergh et al. 2017).

The last three of these theories are rooted in proposed psychological and cognitive abnormalities in the way SSD patients process their pain and emotions. Alexithymia, defined as an inability to recognize and name one's own emotions (Taylor, Bagby, & Parker 1991), may lead SSD patients to have difficulty assigning a psychological meaning to their distressing emotions. Rather than being able to say he is feeling sad and alone, for example, a patient may instead complain of back pain. When asked if he can think of emotional or mental reasons for this pain, the patient is unable to name any. Past research has bolstered this, showing that patients with SSD identify themselves as having more difficulty giving their emotions name and cause than the average person (Erkic et al. 2018). Catastrophizing is based primarily on unhealthy thought patterns and abnormal cognition. Consistent with the "excessive worry" criterion for SSD, a patient who catastrophizes may take a small problem and "blow it out of proportion", giving it much more attention and worry than necessary. For example, a patient with intermittent chest pain caused by a known benign health condition may cause themselves undue stress by attributing their chest pain to heart attacks. Interestingly, multiple studies have connected catastrophizing to a low pain tolerance, low pain thresholds, and increased somatosensory cortex activation during painful stimuli (Rief and Martin 2014). Similarly, SSD patients are thought to have an increased awareness of their bodily functions caused by constantly "checking in" on how parts of the body are feeling and functioning, leading them to become obsessive and paranoid,

often attributing normal bodily functions to sign of disease. This is known as "hypervigilance", the third etiological theory of SSD. In fact, cognitive behavioral therapy (CBT) is effective at lowering hypervigilance in SSD patients by training the patient to stop this "checking-in" behavior and become more comfortable with "abnormal" sensations of the body. (Rief and Martin 2014). The success of CBT in lessening physical symptoms, disability, and psychological distress in SSD patients suggests hypervigilance could play a major role in the disorder (Kurlansik and Maffei 2016).

GENDER BIAS IN SOMATIC SYMPTOM DISORDER

For thousands of years, hysteria, now SSD, was a woman's illness. Because its very name comes from the Greek word for "uterus," it is difficult to detach from "female." Moreover, although the name has since changed, the ideas underlying it have remained essentially the same. Patients with SSD are frequently diagnosed with comorbid personality disorders, especially histrionic personality disorder. Other proposed risk factors include childhood neglect, sexual abuse, history of substance abuse, and "chaotic lifestyle" (Kurlansik and Maffei 2016). The latter factor is reminiscent of the middle ages idea that women were plagued with hysteria as a result of sinful behavior (Tasca et al. 2012). The 1800s idea of "secondary gain," that women with hysteria had something to gain from acting sick, seems to still be alive today in the description of attention-seeking, "doctor-shopping", histrionic behavior often attributed to SSD (Crimlisk and Ron 1999; Gerger et al. 2015).

The central critique of most papers on the topic of the DSM-V creation of SSD is that the diagnostic criteria are too vague. Although the change was intended to make the criteria more inclusive to a wider variety of physical symptoms and shift the focus to psychological symptoms, many argued that the change made the criteria too "loose", allowing for false diagnoses to be made (Frances and Chapman 2013). It is estimated that one in six heart disease patients, one in six cancer patients, one in four IBS patients, and one in four chronic widespread pain patients qualify for the diagnosis based on the current criteria, with a seven-percent false positive rate in the general population (Frances 2013). Because of the subjective nature of the criteria, the personal philosophy of the physician plays one of the biggest roles in whether or not a patient with a difficult case ends up receiving a medical diagnosis or a psychiatric one (Rief and Martin 2014).

This burden falls especially strongly on women, who are frequently judged as exaggerating their symptoms and pain (Frances 2013).

Moreover, due in part to societal stereotypes of gender, there is an inclination in medicine to believe that women are more emotional and thus have a tendency to exaggerate their symptoms while men are more stoic and thus have a tendency to describe their symptoms as less severe than they really are (Hoffmann and Tarzian 2001; Keogh 2018). A 2001 review paper stated that women were less likely to be admitted to the hospital, less likely to receive anesthesia, and less likely to have follow up tests than men. While men tended to receive referrals for specialty pain clinics from their general practitioner, women tended to have to go through the extra hoop of seeing a specialist. Doctors were found to prescribe less pain medication to women than men recovering from the same surgery and of a study of 300 nurses, the vast majority thought that women were less sensitive to pain than men (Hoffmann and Tarzian 2001). A 2019 Danish study of nearly seven million patients found that women consistently had to wait longer to receive a diagnosis than men, with an average of 2.5 years later for cancer and 4.5 years later for metabolic diseases (Westergaard et al. 2019).

Virtual patient studies have demonstrated that both physicians and nurses tend to judge the patient's pain differently and prescribe different medication depending on the patient's gender (Hirsh et al. 2014; Wandner et al. 2014). Even though all the virtual patients were coded to give the exact same description of pain quality, location, effect on daily life, and duration with the same facial expressions, participants judged the female patients as having lower pain intensities than the male patients. In addition, female patients were more likely to be prescribed antidepressants for their pain while men were more likely to receive pain medication like opioids (Hirsh et al. 2014). A 2015 study involving the medical records of 589 female and 262 male chronic pain patients found that women were significantly less likely than men to receive recommendations for further rehabilitation and medical testing. Men were more likely to receive surgery and physical therapy while women were more likely to receive medications and psychotherapy (Stålnacke et al. 2015).

Together, this evidence suggests the presence of a bias in how medical practitioners treat women's symptoms. While women's pain is often judged to be psychological in origin, men's pain is more frequently thought to be physical in origin. This tendency, in combination with the deeply gendered history of

SSD, indicates that the diagnostic discrepancy in SSD can be at least partially explained by a gender bias. Should there be differences in the way men and women experience symptoms, however, perhaps the bias is not a bias at all but a warranted difference in the way practitioners diagnose different genders. Because pain has psychological and physical aspects, it is important to look at the possibility of gender and sex differences in both facets, especially those differences relevant to the five proposed causes of SSD as mentioned earlier.

SEX-BASED NEUROBIOLOGICAL DIFFERENCES IN SYMPTOM EXPERIENCE

Many studies point to the fact that women seem to experience and report more physical symptoms with more severity than men overall (Fillingim et al. 2009; Hoffmann and Tarzian 2001; Keogh 2018; Ramírez-Maestre and Esteve 2014; Unruh 1996). Women also experience anxiety and depression more frequently than men, which could possibly act as a confounding variable to explain this difference in physical symptoms (Klonoff, Landrine, and Campbell 2000). However, even when comorbid mental illness is controlled for, women have been found to report an average of 1.1 more physical symptoms than men, regardless of if these symptoms have a known medical cause of not (Kroenke and Spitzer 1998). Despite the fact that women have a longer life span than men in the United States, they are significantly more prone to autoimmune diseases and other chronic conditions, especially chronic pain, than men and utilize healthcare services more frequently than men. 78% of autoimmune disease sufferers, including those afflicted by Multiple Sclerosis, Lupus, Rheumatoid Arthritis, and more, are women (Fairweather, Frisancho-Kiss, and Rose 2008). Women are effectively the "sicker sex"(Turabian 2017).

In addition to experiencing more pain in everyday life, laboratory studies have consistently shown women to have significantly lower pain thresholds and tolerances than men (Fillingim et al. 2009; Hoffmann and Tarzian 2001; Ramírez-Maestre and Esteve 2014; Unruh 1996). In addition, researchers have observed that women experience more activation in the contralateral prefrontal cortex, contralateral insula, and thalamus in response to painful stimuli than men (Hoffmann and Tarzian 2001). These studies have also pointed to a wide variety of possible neurobiological differences in how women and men experience symptoms. Firstly, estrogen seems to play a major role in both mechanisms of analgesia involving

opioid receptors and inflammatory response to pain. A woman's response and tolerance to ischemic pain changes throughout her menstrual cycle as her estrogen spikes and drops (Iacovides, Avidon, and Baker 2015). Overall low levels of estrogen have been linked to osteoporosis, tempromanidibular joint disorder, and other joint pain disorders (Hoffmann and Tarzian 2001). Similarly, rapid drops in estrogen levels have been associated with increased symptom severity in patients with Rheumatoid Arthritis (Sorge and Totsch 2017).

Estrogen-dependent analgesic mechanisms have been found in female mice, which may explain why pain tolerance shifts alongside hormones (Hoffmann and Tarzian 2001). Estrogen has even been found to have an effect on externally administered opioids. The hormone reduces available opioid binding sites on the cell membrane and can "uncouple" morphine binding, making it less effective (Averitt et al. 2018). Interestingly, progesterone seems to act as a counterpart to estrogen, possibly working as a therapeutic agent to neuropathic pain. Rat studies have shown progesterone administration to lessen the harmful electrophysiological changes in peripheral nerves relating to peripheral neuropathy and decrease the quantity and severity of neuropathic pain related behaviors (Coronel et al. 2016; Jarahi et al. 2014).

In addition to estrogen and progesterone mediated differences in pain experience, research has pointed to possible sex-based differences in parts of the PNS and CNS devoted to pain experience. Researchers have observed differences in tissue thickness and sensory receptor count in the peripheral nervous system of men and women, suggesting women may have lower pain thresholds and tolerances because they are actually sensing larger quantities of painful stimuli than men (Fillingim and Maixner 1995). A 2016 study found that immune response in the spinal cord, the over-activity of which is often associated with chronic pain, is primarily mediated by T-cells in female rats and microglia in male rats (Mapplebeck, Beggs, and Salter 2016). This difference may be associated with the fact that male rats in the study were able to recover from spinal cord injury much faster than female rats. Sex-based differences in immune response have also been observed in the rodent peripheral nervous system (Sorge and Totsch 2017). A 2019 study involving human dorsal root ganglion neurons found evidence for the existence of sex-differential gene expression that could possibly be related to sex-specific neuropathic pain (North et al. 2019). The female subjects in this study had a different set of spinal-cord injury related gene expressions when compared to the

male subjects that may have resulted in differing pain experiences. These studies provide strong evidence not only for disparities in the amount and strength of pain experienced by men and women but also for numerous notable differences in the underlying biological mechanisms of pain sensation and perception between the sexes.

Somatosensory amplification, the idea that SSD is caused by the patient perceiving normal sensations as more intense than the average person, is based primarily in proposed neurobiological irregularities. There is little research into the question of whether the specific neural differences thought to be related to somatosensory amplification, such as striatal and amygdalar abnormalities and differences in the lamina 1 spinothalamic cortical pathway, are more common in women or men. There is evidence, however, that women generally experience pain more strongly than men and that this may be a result of sex hormone modulation of pain sensation and increased activation of pain-responsive areas in the brain (Hoffmann and Tarzian 2001). In this sense, this is evidence for natural somatosensory amplification of pain in most women as compared to men, independent of SSD.

Given that women exhibit both higher levels of chronic and everyday pain and higher prevalence of mood and anxiety disorders, it is not illogical to infer that women may be more prone to falling into the vicious cycle effect, another etiological theory of SSD in which a negative psychological mindset feeds into the worsening of physical symptoms and vice versa, than men (Riecher-Rössler 2017). These factors being considered, it is quite possible that women genuinely exhibit the symptoms of SSD more often than men. It is important to ask, however, how a person with comorbid depression and chronic illness would differ in presentation to a person with SSD. In addition, it is important to ask how a psychiatrist would differentiate between these two situations and how they would differently inform treatment. Might the gender of the patient determine their diagnosis?

GENDER-BASED PSYCHOSOCIAL DIFFERENCES IN SYMPTOM EXPERIENCE

Many of the proposed psychological differences between how women and men experience pain are reliant on gender roles and stereotypes. The social constructs of "man" and "woman" have vast and complex meanings that may vary across cultures, but generally, men are made to feel ashamed of their feelings, including

pain, while women are encouraged to be more vocal due to a community-based perception of the world (Keogh 2018; Kroenke and Spitzer 1998). As a result, men tend to suppress their feelings of pain and wait until they interfere significantly with work and daily life to seek medical attention and support. Women, meanwhile, are generally more socially and emotionally oriented, leading them to more readily seek support (Hoffmann and Tarzian 2001). In addition, there seems to be a societal assumption that pain is inevitably a bigger part of a woman's life than a man's due to childbirth and menstruation (Keogh 2018). As a result of these stereotypical differences, men may feel a pressure to hide their pain, causing them to falsely report lower pain tolerances and thresholds and ignore possible chronic pain conditions. Despite these assumed stereotypical differences in the way men and women experience symptoms psychologically, studies have shown mixed evidence as to whether or not these differences are upheld by empirical results.

For example, alexithymia, the theory that individuals with SSD are unable to name their emotions, is found to be more common in men than women (Unruh, Ritchie, and Merskey 1999). In a study of 2018 depressed patients, 891 male and 1127 female, male patients were significantly more likely to experience alexithymia than female patients. Among these patients with depression, 12.8% of men and 8.2% of women demonstrated alexithymia (Berger et al. 2005). Other studies have also shown alexithymia to be more common in men than women (Guvensel et al. 2018; O'Loughlin et al. 2018). Perhaps due to male gender role expectations that discourage men from showing and seeking help for their emotions, men are more prone to alexithymia than women, especially men who feel insecure in their masculinity (Berger et al. 2005). This evidence suggests alexithymia likely cannot explain the gender diagnostic discrepancy in SSD.

Catastrophizing, another proposed cause of SSD in which patients are thought to assign excessive worry to minor physical symptoms, has long been used as a theory to explain why women have lower pain tolerances than men but there is mixed evidence as to whether or not this is rooted in empirical evidence. Proponents of this view have hypothesized that women are especially prone to this cognitive dysfunction due to increased emotional vulnerability (Roth et al. 2005; Thorn et al. 2004). A 2004 laboratory study on this topic provided evidence that women may indeed experience catastrophizing more frequently than men. Even when comorbid mental illness was controlled for, women reported both more

pain and more catastrophizing than men during a thermal pain task. (Edwards et al. 2004). Other laboratory tests and field studies involving chronic pain patients have also shown women to have higher rates of catastophizing and poorer coping skills than men (Leung 2012).

A 2014 field study, meanwhile, found that men and women with the same chronic pain conditions reported no significant difference in catastrophizing mental behaviors when questioned on their coping strategies (Ramírez-Maestre and Esteve 2014). Multiple studies involving recovering whiplash patients actually found male patients to have higher levels of pain catastrophizing than female patients (Elklit and Jones 2006; Rivest et al. 2010). A 1999 field study found there to be no gender difference in emotional upset due to pain but simply an increase in the use of coping strategies in women (Unruh, Ritchie, and Merskey 1999). Based on this mixed evidence, it is unclear whether catastrophizing can account for the gender disparity in the diagnosis of SSD.

Like catastrophizing, hypervigilance, which occurs when SSD patients are overly aware of their own bodily sensations, has also been proposed as a possible model to explain why women seem to experience more pain with higher intensity. A 2003 theoretical review paper argued that hypervigilance, can be expressed through increased treatment seeking for pain, low expectations about one's own pain tolerance, and actual low pain tolerance (Rollman et al. 2004). According to this definition, because women have higher levels of pain treatment seeking and lower pain tolerances, they would overall have higher levels of hypervigilance than men. The researchers attribute this difference to a tendency for women to assign exaggerated meaning to their previous pain experiences (Rollman et al. 2004). Despite this theorization, however, there is little evidence that women actually have higher levels of hypervigilance or that low pain tolerance is a sign of hypervigilant mental behaviors. A 2014 study involving chronic pain patients found that men and women reported no significant difference in hypervigilant mental behaviors when questioned on their coping strategies. Women did, however, report higher levels of pain intensity, impaired daily functioning, and pain anxiety. This difference may also be related to overall higher levels of comorbid depression and anxiety in women and thus higher levels of negative affect (Ramírez-Maestre and Esteve 2014). Overall, there is not enough research around hypervigilance and gender to determine whether or not there is a correlation.

Although it is difficult to completely divorce the psychological aspects of pain from the physical ones, attempts to provide evidence for definite psychological differences in the way men and women experience pain have mostly been unsuccessful. More specifically, there is a lack of evidence that women's pain is in any way more psychologically influenced or oriented than men's. This includes studies aimed at examining differences in the way men and women process pain specifically related to the etiological theories of SSD, including alexithymia, catastrophizing, and hypervigilance. This is not to say that there are no psychological differences in the way men and women experience symptoms whatsoever, but that there is a lack of evidence for gender differences in the way men and women psychologically experience symptoms in relation to SSD.

CONCLUSIONS AND FUTURE DIRECTIONS

Though researchers have attempted to explain the differences in pain intensity and tolerance in men and women through cognitive differences (e.g. catastrophizing or hypervigilance), there is mixed evidence for these theories. A perhaps stronger explanation is the neurobiological differences in pain experience between men and women. The focus in research on attempting to find cognitive and psychological differences between men and women has detracted attention from the existence of neurobiological differences. As a result, differences in medical diagnosis and treatment have been based too strongly on theorized disparities in the way men and women think and process pain, many of which are at least partially based in unfounded gender stereotypes, and not strongly enough on differing underlying biological mechanisms of pain sensation and perception. The gender disparity in the diagnosis of SSD is a clear example of this.

Although women are the primary consumers of healthcare, they are not the primary target of healthcare research. Even in the second half of the twentieth century, the focus of medical research was men, specifically white, upper-class, older men. In 1993, legislation was created that mandated the inclusion of women and minorities in NIH research (Hoffmann and Tarzian 2001). This legislation was last amended in 2017 to better clarify what kind of studies were subject to the requirement. While this legislation is an important step forward, it has only existed for the last 20 years, suggesting there is a large gap in research between that focusing on white men's health and that focusing on everyone else's. Researchers have really only begun to identify that there could be differences

between women's and men's healthcare needs other than gynecological aspects in the last few decades. For example, in 2016, the NIH created a policy that all applicants applying for funding for studies using invertebrate animals must explain how their experiment will account for sex as biological variable. The reasoning for this policy stems from the general exclusion of female subjects in past animal research and the resulting lack of generalizability. This bias has been found to be especially prominent in neuroscience and pharmacology research. (Shansky and Woolley 2016)we present reasons to be optimistic that this new policy will be valuable for neuroscience, and we suggest some ways for neuroscientists to think about incorporating sex as a variable in their research.","DOI":"10.1523/JNEUROSCI.1390-16.2016","ISSN":"0270-6474, 1529-2401","note":"PMID: 27881768","journalAbbreviation":"J. Neurosci.","language":"en","author":[{"family":"Shansky","given":"Rebecca M."},{"family":"Woolley","given":"Catherine S."}],"issued":{"date-parts":[["2016",11,23]]},"PMID":"27881768"}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} Scientific research still has a long way to go to make up for these deficiencies. With these policies in place, hopefully future innovations in science will produce sex-specific pain therapies that account for neurobiological differences like sex hormones and neural mediation of immune response.

Perhaps more important than accounting for the newfound neurobiological differences in the way men and women experience symptoms is eradicating the centuries-old bias in treating women's symptoms as primarily psychological in origin. There is little to no evidence showing that women's pain is actually more often psychologically-based and influenced than men's pain is. More likely is that women generally experience more pain than men, and because the medical model has been historically based on men with the flawed assumption that women are essentially biologically identical, practitioners assumed the reported differences in pain severity and frequency were due to psychological differences rather than physical ones. As a result of this assumption, there is an ongoing tendency to treat women's pain reporting as exaggerated, influenced by supposed cognitive dysfunctions like catastrophizing and hypervigilance. This has led to an inevitable mistreatment of women's symptoms and pains as psychological in origin, a mistreatment that has lasted centuries. Future medical practitioners must work to undo this bias and treat women's pain as no less psychologically influenced than

men's while also working towards developing sex-specific pain treatment that will account for underlying neurobiological differences.

Pain is a multifaceted phenomenon involving both psychological and physical aspects, yet pain treatment and diagnosis tend to focus mainly on the psychological aspect in women and the physical aspect in men. As a result, individuals are facing misdiagnoses that can have much larger ramifications, including being denied access for necessary care. As described earlier, women must wait longer for a diagnosis than men and tend to be referred to specialists less often, instead receiving psychiatric treatment. Because it takes longer for women to get viable answers and treatment for their symptoms, they likely face a worsening of symptoms and possible escalation of underlying disease in the waiting period. This leads to larger healthcare costs in the long run and an overall worse quality of life. On the other hand, men who suffer from psychiatric disorders likely are not easily getting the treatment or diagnosis they require either. Pain and symptom treatment should be multifaceted, covering both the underlying disease and biology and the psychological distress that may arise. The increased specialization of healthcare means that patients are too often receiving care only for either the psychological aspect, affecting especially women, or the biological one, affecting especially men. In this way, the current Western medical system is failing to adequately treat both women and men.

APPENDIX:

1. The official 2013 DSM-V Criteria for Somatic Symptom Disorder are as follows:
    - A. One or more somatic symptoms that are distressing or result in significant disruption of daily life.
    - B. Excessive thoughts, feelings, or behaviors related to the somatic symptoms or associated health concerns as manifested by at least one of the following:
        1. Disproportionate and persistent thoughts about the seriousness of one's symptoms.
        2. Persistently high level of anxiety about health of symptoms.
        3. Excessive time and energy devoted to these symptoms or health concerns

C. Although any one somatic symptom may not be persistent, the state of being symptomatic is persistent (more than 6 months)

Specify if:

With predominant pain (previously pain disorder): This specifier is for individuals whose somatic symptoms predominantly involve pain.

Specify if:

Persistent: A persistent course is characterized by severe symptoms, marked impairment, and long duration (more than 6 months).

Specify current severity:

Mild: Only one of the symptoms specified in Criterion B is fulfilled.

Moderate: Two or more of the symptoms specified in Criterion B are fulfilled

Severe: Two or more of the symptoms specified in Criterion B are fulfilled, plus there are multiple somatic complaints (or one very severe somatic symptom). (APA 2013)

REFERENCES

American Psychiatric Association. 1980. Diagnostic and statistical manual of mental disorders (3rd ed.).

American Psychiatric Association. 1994. Diagnostic and statistical manual of mental disorders (4th ed.).

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders (5th ed.).

Averitt, Dayna L., Lori N. Eidson, Hillary H. Doyle, and Anne Z. Murphy. 2018. "Neuronal and Glial Factors Contributing to Sex Differences in Opioid Modulation of Pain." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, June.

Berger, Jill M., Ronald Levant, Katharine Kaye Mcmillan, William Kelleher, and

Al Sellers. 2005. "Impact of Gender Role Conflict, Traditional Masculinity Ideology, Alexithymia, and Age on Men's Attitudes Toward Psychological Help Seeking." *Psychology of Men & Masculinity* 6 (1): 73–78.

Cooper, R. G., C. K. Booker, and C. C. Spanswick. 2003. "What Is Pain Management, and What Is Its Relevance to the Rheumatologist?" *Rheumatology (Oxford, England)* 42 (10): 1133–37.

Coronel, María F., María C. Raggio, Natalia S. Adler, Alejandro F. De Nicola, Florencia Labombarda, and Susana L. González. 2016. "Progesterone Modulates pro-Inflammatory Cytokine Expression Profile after Spinal Cord Injury: Implications for Neuropathic Pain." *Journal of Neuroimmunology* 292 (March): 85–92.

Crimlisk, H. L., and M. A. Ron. 1999. "Conversion Hysteria: History, Diagnostic Issues, and Clinical Practice." *Cognitive Neuropsychiatry* 4: 165–80.

Dimsdale, Joel E., and James Levenson. 2013. "What's Next for Somatic Symptom Disorder?" *American Journal of Psychiatry* 170 (12): 1393–95.

Edwards, Robert R., Jennifer A. Haythornthwaite, Michael J. Sullivan, and Roger B. Fillingim. 2004. "Catastrophizing as a Mediator of Sex Differences in Pain: Differential Effects for Daily Pain versus Laboratory-Induced Pain." *Pain* 111 (3): 335–41.

Elklit, Ask, and Allan Jones. 2006. "The Association between Anxiety and Chronic Pain after Whiplash Injury: Gender-Specific Effects." *The Clinical Journal of Pain* 22 (5): 487–90.

Erkic, Maja, Josef Bailer, Sabrina C. Fenske, Stephanie N. L. Schmidt, Jörg Trojan, Annette Schröder, Peter Kirsch, and Daniela Mier. 2018. "Impaired Emotion Processing and a Reduction in Trust in Patients with Somatic Symptom Disorder." *Clinical Psychology & Psychotherapy* 25 (1): 163–72.

Fairweather, DeLisa, Sylvia Frisancho-Kiss, and Noel R. Rose. 2008. "Sex Differences in Autoimmune Disease from a Pathological Perspective." *The American Journal of Pathology* 173 (3): 600–609.

Fillingim, Roger B., Christopher D. King, Margarete C. Ribeiro-Dasilva, Bridgett Rahim-Williams, and Joseph L. Riley. 2009. "Sex, Gender, and Pain: A Review of Recent Clinical and Experimental Findings." *Journal of Pain* 10 (5): 447–

85.

Fillingim, Roger B., and William Maixner. 1995. "Gender Differences in the Responses to Noxious Stimuli." *Pain Forum* 4 (4): 209–21.

Frances, Allen. 2013. "The New Somatic Symptom Disorder in DSM-5 Risks Mislabeling Many People as Mentally Ill." *BMJ : British Medical Journal (Online); London* 346 (March).

Frances, Allen, and Suzy Chapman. 2013. "DSM-5 Somatic Symptom Disorder Mislabels Medical Illness as Mental Disorder." *Australian & New Zealand Journal of Psychiatry* 47 (5): 483–84.

Gatchel, Robert J. 2004. "Comorbidity of Chronic Pain and Mental Health Disorders: The Biopsychosocial Perspective." *American Psychologist* 59 (8): 795–805.

Gerger, Heike, Michaela Hlavica, Jens Gaab, Thomas Munder, and Juergen Barth. 2015. "Does It Matter Who Provides Psychological Interventions for Medically Unexplained Symptoms? A Meta-Analysis." 2015.

Gilman, Sander L., ed. 1993. *Hysteria before Freud*. Berkeley: University of California Press.

Golding, Jacqueline M., G. Richard Smith, and T. Michael Kashner. 1991. "Does Somatization Disorder Occur in Men?: Clinical Characteristics of Women and Men With Multiple Unexplained Somatic Symptoms." *Archives of General Psychiatry* 48 (3): 231–35.

Guvensel, Kan, Andrea Dixon, Catherine Chang, and Brian Dew. 2018. "The Relationship Among Gender Role Conflict, Normative Male Alexithymia, Men's Friendship Discords With Other Men, and Psychological Well-Being." *The Journal of Men's Studies* 26 (1): 56–76.

Harvey, Samuel B, Biba R Stanton, and Anthony S David. 2006. "Conversion Disorder: Towards a Neurobiological Understanding." *Neuropsychiatric Disease and Treatment* 2 (1): 13–20.

Hirsh, Adam T., Nicole A. Hollingshead, Marianne S. Matthias, Matthew J. Bair, and Kurt Kroenke. 2014. "The Influence of Patient Sex, Provider Sex, and Sexist Attitudes on Pain Treatment Decisions." *The Journal of Pain* 15 (5):

551–59.

Hoffmann, Diane E., and Anita J. Tarzian. 2001. "The Girl Who Cried Pain: A Bias against Women in the Treatment of Pain." *The Journal of Law, Medicine & Ethics* 28 (4_suppl): 13–27.

Iacovides, S., I. Avidon, and F. C. Baker. 2015. "Does Pain Vary across the Menstrual Cycle? A Review." *European Journal of Pain* 19 (10): 1389–1405.

Jarahi, M., V. Sheibani, H. A. Safakhah, H. Torkmandi, and A. Rashidy-Pour. 2014. "Effects of Progesterone on Neuropathic Pain Responses in an Experimental Animal Model for Peripheral Neuropathy in the Rat: A Behavioral and Electrophysiological Study." *Neuroscience* 256 (January): 403–11.

Katon, Wayne, Elizabeth H. B. Lin, and Kurt Kroenke. 2007. "The Association of Depression and Anxiety with Medical Symptom Burden in Patients with Chronic Medical Illness." *General Hospital Psychiatry* 29 (2): 147–55.

Keogh, Edmund. 2018. "Sex and Gender as Social-Contextual Factors in Pain." In *Social and Interpersonal Dynamics in Pain: We Don't Suffer Alone*, edited by Tine Vervoort, Kai Karos, Zina Trost, and Kenneth M. Prkachin, 433–53. Cham: Springer International Publishing.

Klonoff, Elizabeth A., Hope Landrine, and Robin Campbell. 2000. "Sexist Discrimination May Account for Well-Known Gender Differences in Psychiatric Symptoms." *Psychology of Women Quarterly* 24 (1): 93–99.

Kroenke, Kurt, and Robert L. Spitzer. 1998. "Gender Differences in the Reporting of Physical and Somatoform Symptoms." *Psychosomatic Medicine* 60 (2): 150.

Kurlansik, Stuart L., and Mario S. Maffei. 2016. "Somatic Symptom Disorder." *American Family Physician* 93 (1): 49–54.

Leung, Lawrence. 2012. "Pain Catastrophizing: An Updated Review." *Indian Journal of Psychological Medicine* 34 (3): 204–17.

Mapplebeck, Josiane C. S., Simon Beggs, and Michael W. Salter. 2016. "Sex Differences in Pain: A Tale of Two Immune Cells." *PAIN* 157 (February): S2.

North, Robert Y., Yan Li, Pradipta Ray, Laurence D. Rhines, Claudio Esteves Tatsui, Ganesh Rao, Caj A. Johansson, et al. 2019. "Electrophysiological and

Transcriptomic Correlates of Neuropathic Pain in Human Dorsal Root Ganglion Neurons." *Brain*. Accessed March 21, 2019.

O'Loughlin, Julia I., Daniel W. Cox, Jeffrey H. Kahn, and Amery D. Wu. 2018. "Attachment Avoidance, Alexithymia, and Gender: Examining Their Associations with Distress Disclosure Tendencies and Event-Specific Disclosure." *Journal of Counseling Psychology* 65 (1): 65–73.

Perez, David L., Arthur J. Barsky, David R. Vago, Gaston Baslet, and David A. Silbersweig. 2015. "A Neural Circuit Framework for Somatosensory Amplification in Somatoform Disorders." *The Journal of Neuropsychiatry and Clinical Neurosciences* 27 (1): e40-50.

Ramírez-Maestre, Carmen, and Rosa Esteve. 2014. "The Role of Sex/Gender in the Experience of Pain: Resilience, Fear, and Acceptance as Central Variables in the Adjustment of Men and Women With Chronic Pain." *The Journal of Pain* 15 (6): 608–618.e1.

Riecher-Rössler, Anita. 2017. "Sex and Gender Differences in Mental Disorders." *The Lancet Psychiatry* 4 (1): 8–9.

Rief, Winfried, and Alexandra Martin. 2014. "How to Use the New DSM-5 Somatic Symptom Disorder Diagnosis in Research and Practice: A Critical Evaluation and a Proposal for Modifications." *Annual Review of Clinical Psychology* 10 (1): 339–67.

Rivest, Karine, Julie N. Côté, Jean-Pierre Dumas, Michele Sterling, and Sophie J. De Serres. 2010. "Relationships between Pain Thresholds, Catastrophizing and Gender in Acute Whiplash Injury." *Manual Therapy* 15 (2): 154–59.

Rollman, Gary B., Jennifer Abdel-Shaheed, Joanne M. Gillespie, and Kevin S. Jones. 2004. "Does Past Pain Influence Current Pain: Biological and Psychosocial Models of Sex Differences." *European Journal of Pain* 8 (5): 427–33.

Roth, Randy S., Michael E. Geisser, Mary Theisen-Goodvich, and Pamela J. Dixon. 2005. "Cognitive Complaints Are Associated With Depression, Fatigue, Female Sex, and Pain Catastrophizing in Patients With Chronic Pain." *Archives of Physical Medicine and Rehabilitation* 86 (6): 1147–54.

Shansky, Rebecca M., and Catherine S. Woolley. 2016. "Considering Sex as a Biological Variable Will Be Valuable for Neuroscience Research." *Journal of*

Neuroscience 36 (47): 11817–22.

Sorge, Robert E., and Stacie K. Totsch. 2017. "Sex Differences in Pain." *Journal of Neuroscience Research* 95 (6): 1271–81.

Stålnacke, Britt-Marie, Inger Haukenes, Arja Lehti, Anncristine Fjellman Wiklund, Maria Wiklund, and Anne Hammarström. 2015. "Is There a Gender Bias in Recommendations for Further Rehabilitation in Primary Care of Patients with Chronic Pain after an Interdisciplinary Team Assessment?" *Journal of Rehabilitation Medicine* 47 (4): 365–71.

Tasca, Cecilia, Mariangela Rapetti, Mauro Giovanni Carta, and Bianca Fadda. 2012. "Women And Hysteria In The History Of Mental Health." *Clinical Practice and Epidemiology in Mental Health : CP & EMH* 8 (October): 110–19.

Taylor, Gj, Rm Bagby, and Jda Parker. 1991. "The Alexithymia Construct - a Potential Paradigm for Psychosomatic-Medicine." *Psychosomatics* 32 (2): 153–64.

Teasdale, John D. 1983. "Negative Thinking in Depression: Cause, Effect, or Reciprocal Relationship?" *Advances in Behaviour Research and Therapy*, Cognitions and Mood: Clinical Aspects and Applications, 5 (1): 3–25.

Thorn, Beverly E., Kristi L. Clements, L. Charles Ward, Kim E. Dixon, Brian C. Kersh, Jennifer L. Boothby, and William F. Chaplin. 2004. "Personality Factors in the Explanation of Sex Differences in Pain Catastrophizing and Response to Experimental Pain." *The Clinical Journal of Pain* 20 (5): 275.

Turabian, Jose Luis. 2017. "Is The Meaning of Symptoms the Same in Women And Men?" *Journal of Womens Health Care* 6 (3).

Unruh, Anita M. 1996. "Gender Variations in Clinical Pain Experience." *Pain* 65 (2): 123–67.

Unruh, Anita M., Judith Ritchie, and Harold Merskey. 1999. "Does Gender Affect Appraisal of Pain and Pain Coping Strategies?" *The Clinical Journal of Pain* 15 (1): 31.

Van den Bergh, Omer, Michael Witthöft, Sibylle Petersen, and Richard J. Brown. 2017. "Symptoms and the Body: Taking the Inferential Leap." *Neuroscience & Biobehavioral Reviews* 74 (March): 185–203.

Wandner, Laura D., Marc W. Heft, Benjamin C. Lok, Adam T. Hirsh, Steven Z. George, Anne L. Horgas, James W. Atchison, Calia A. Torres, and Michael E. Robinson. 2014. "The Impact of Patients' Gender, Race, and Age on Health Care Professionals' Pain Management Decisions: An Online Survey Using Virtual Human Technology." *International Journal of Nursing Studies* 51 (5): 726–33.

Westergaard, David, Pope Moseley, Freja Karuna Hemmingsen Sørup, Pierre Baldi, and Søren Brunak. 2019. "Population-Wide Analysis of Differences in Disease Progression Patterns in Men and Women." *Nature Communications* 10 (1): 666.

Wise, Thomas N., and Lee S. Mann. 1994. "The Relationship between Somatosensory Amplification, Alexithymia, and Neuroticism." *Journal of Psychosomatic Research* 38 (6): 515–21.

compos mentis

Accounting for Willful Hermeneutical Ignorance

Andrew Kesler
Eastern Michigan University

ABSTRACT
In this paper I argue that the most useful and important epistemic resources come out of the marginally situated position and these resources are the ones most subject to neglect. I also assert that willful hermeneutical ignorance develops out of cognitive missteps within both individual's and group's cognition. Epistemic resources are tools for understanding one's own experience, others' experiences, and furthering the development of useful and meaningful knowledge. These resources give way to entire bodies of knowledge, new ways to perceive the world, and they are necessary in order to highlight and subdue epistemic injustices. A marginally situated knower has more trouble affecting epistemic resources than a dominantly situated knower. Those dominantly situated in social positions are more likely exhibit what is called willful hermeneutical ignorance, or being blind and/or dismissive to certain concepts and ways of thought. Willful hermeneutical ignorance develops from cognitive missteps that lead to the dismissal or overlooking of epistemic resources. Cognitive missteps can include having a fundamental misunderstanding of one's own situatedness and a misunderstanding of others' as well.

KEYWORDS
Epistemology, Hermeneutical Ignorance, Epistemic Resources, Marginally Situated Knower, Dominantly Situated Knower, Epistemic Injustice

INTRODUCTION

In order to fully explain what willful hermeneutical ignorance means, I begin by defining epistemic resources and their role in the transfer of knowledge. I then use Miranda Fricker to introduce epistemic injustice and illustrate hermeneutical marginalization. Gaile Pohlhaus Jr. expands on Fricker's work to define willful hermeneutical ignorance, a type of epistemic injustice that is absent from Fricker's account. The main concepts necessary for understanding willful hermeneutical ignorance are situatedness and interdependence along with the relationship between the dominantly situated knower and the marginally situated knower. I discuss Jose Medina's explanation of epistemic vices in order to further define why and how knowers remain willfully hermeneutically ignorant. I assert that the most useful and important epistemic resources come out of the marginally situated position and these resources are the ones most subject to neglect. I also assert that this type of ignorance develops out of cognitive missteps within both individual's and group's cognition.

Epistemic resources are tools for understanding one's own experience, others' experiences, and furthering the development of useful and meaningful knowledge. These resources give way to entire bodies of knowledge, new ways to perceive the world, and they are necessary in order to highlight and subdue epistemic injustices. Because the spread of knowledge is a social endeavor, epistemic resources are social tools. A single individual cannot formulate useful epistemic resources successfully without conference with other individuals that share similar or somehow relatable knowledge. Even though these are social tools, not everyone has the same access and influence on epistemic resources. As I later explain in further detail, a marginally situated knower has more trouble affecting epistemic resources than a dominantly situated knower. Those dominantly situated in social positions are more likely to exhibit what is called willful hermeneutical ignorance, or being blind and/or dismissive to certain concepts and ways of thought.

WILLFUL HERMENEUTICAL IGNORANCE AS EPISTEMIC INJUSTICE

Miranda Fricker offers an account of two kinds of epistemic injustice that she calls testimonial and hermeneutical injustice. I mainly focus on hermeneutical injustice. In Chapter Seven of her work, Epistemic Injustice: Power and the Ethics of Knowing, she discusses hermeneutical marginalization. She affirms that, "When there is unequal hermeneutical participation with respect to some significant area(s)

of social experience, members of the disadvantaged group are hermeneutically marginalized" (Fricker 2007, 153). The marginalized group has a lesser ability to interpret experience which puts them at a disadvantage. The disadvantage on one hand is the unequal participation in forming meaning through experience, and on the other hand having to expend more effort for understanding while being subject to the advantage group for confirmation. Fricker defines this sort of injustice as, "Having some significant area of one's social experience obscured from collective understanding owing to hermeneutical marginalization" (Fricker 2007, 158). She makes a point that the various hermeneutical disadvantages one may face are not always clear and not always continuous, yet have lasting consequences. As a knower, one may slowly lose confidence over time, for instance, if people are consistently skeptical of her account of her own experiences. This can lead to self-doubt and inhibit the acquisition and spread of new knowledge and epistemic resources. A notable aspect of Fricker's account is the relation of power between the advantaged and disadvantaged. Because the disadvantaged are in a position of lesser power, they must be more aware of their position by experiencing injustices. The advantaged group need not accept the burden of responsibility for the position of the disadvantaged or even acknowledge their dominant position, thereby disregarding the power relations altogether. The advantaged group, I will argue, has a greater responsibility in educating themselves and noticing inadequacies in epistemic resources because they hold more power over the resources.

Fricker labels two kinds of epistemic injustice which are testimonial injustice and hermeneutical injustice. Her work has been expanded upon in order to highlight another form of injustice that she does not discuss. Gaile Pohlhaus Jr's Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance, provides another form of epistemic injustice that is absent from Fricker's account. Pohlhaus utilizes Fricker's example of Tom Robinson from Harper Lee's To Kill a Mockingbird to emphasize what Fricker's work is missing. According to Pohlhaus, Fricker does not go far enough in defining the real injustice that occurs during Robinson's trial.

In the following paragraphs I will detail Pohlhaus's work on willful hermeneutical ignorance in order to lay a groundwork to further her argument. Pohlhaus defines a unique kind of ignorance where the subject as a knower denies or discredits resources for knowledge acquisition, thereby discounting entire bodies

of knowledge. She calls this willful hermeneutical ignorance and defines it as, "The knower's continued engagement in the world while refusing to learn to use epistemic resources developed from marginalized situatedness" (Pohlhaus 2012, 722). The problem stems from the knower's continued engagement in the world, because by refusing to learn to use certain resources the knower is navigating with outdated, inefficient epistemic resources. This has the possibility to lead to distortions of one's worldview.

Pohlhaus points out an interesting dialectic relationship between an individual's situatedness and interdependence. Situatedness depends on how one is positioned with relation to others as knowers. Interdependence refers to the social aspect of epistemic resources. As knowers, we operate and manufacture epistemic resources collectively. These two ideas create a tension which can be observed in the asymmetrical relationship between marginally situated knowers and dominantly situated knowers.

The situatedness of the dominantly situated knowers is determined by their ability to affect epistemic resources with more influence than the marginally situated, and also to have these resources recognized and accepted. For the marginally situated knower, her situatedness is determined by her lesser ability to affect epistemic resources. Yet, the marginally situated knower is at an epistemic advantage, according to Pohlhaus, in that she more aptly notices injustices and hermeneutical gaps that are deeply entrenched in social norms (Pohlhaus 2012, 720). This is because her marginalized position provides her with a certain worldview and what Pohlhaus describes as a vulnerability. When an individual is vulnerable to others, "She must know what will be expected, noticed by, and of concern to those in relation to whom she is vulnerable" (Pohlhaus 2012, 717). This is an example of how power relations within different social positions affect individuals' worldviews. Due to the position of the marginalized knower in such relations of power, she experiences much more subtle and nuanced injustices that are not common experiences for the majority of knowers. Furthermore, because the experiences are not so commonplace, the epistemic resources commonly used by the majority of knowers are not going to be as suitable for the marginally situated. The advantage of being positioned to notice deeper epistemic inadequacies comes with the disadvantage of not relating to other epistemic resources that come from the dominantly situated. This leads to tension between situatedness and interdependence. How knowers are situated in relations of power affect the

epistemic resources that are available to them and affect the transfer of knowledge between differently situated knowers.

The example that Miranda Fricker uses to further her ideas on epistemic injustices is Tom Robinson's trial from To Kill a Mockingbird. On Fricker's account, Robinson experiences testimonial injustice because the jury does not take his claims seriously due to the color of his skin and position in society. These aspects can be analyzed as his situatedness. Fricker explains that this is the jury giving Robinson deflated credibility on account of identity prejudice. Pohlhaus furthers this example by explaining that Robinson also experiences hermeneutical injustice, and the jury exhibits willful hermeneutical ignorance. Robinson experiences hermeneutical injustice in that the jury "Consistently misinterprets his words" (Pohlhaus 2012, 725). All the while Robinson is aware of what is happening, yet he cannot make this clear to others. Pohlhaus states, "The economy of hermeneutical resources preempts Robinson from transferring that knowledge to the jury" (Pohlhaus 2012, 725). The jury's misunderstanding goes deeper than Fricker's assertions of identity prejudice. In this case, Robinson is hermeneutically marginalized. Robinson is unable to communicate the jury's misunderstanding and social position. Furthermore, according to Pohlhaus, the jury can be held culpable for distorting Robinson's credibility and not utilizing epistemic resources that would help the jury understand Robinson's position. The case is not that the epistemic resources for understanding are unavailable to the jury, but rather the resources used by the jury are faulty and wrongly distort their views of the world.

Pohlhaus evaluates this example in terms of situatedness and interdependence. The jury fails to enter into a cooperative interdependent relationship with knowers that are outside the juror's experienced world i.e. Tom Robinson. She explains, "The individual jurors' past and continuing failure to enter into cooperative epistemic interdependence with marginally situated knowers results in a current structural problem with regard to the transfer of knowledge" (Pohlhaus 2012, 725). The structural problem with the transfer of knowledge is that social inequalities inhibit the formation of cooperative relationships. If the jury were to analyze the situation with the epistemic resources developed out of Robinson's situatedness, thereby allowing interdependent relations between knowers, then there would be a much greater possibility that the jury would not successfully misinterpret Robinson's testimony. Because of his marginalized position as a knower, Robinson has a heightened sense of awareness about concerns of those to whom he is

vulnerable. This is the reason that he is able to recognize the jurors' simultaneous misinterpretation of his testimony, whereas the jurors see no fault in their own reasoning. Being the dominantly situated knowers, the jurors need not worry about Robinson's concerns because they are not vulnerable to Robinson in that same way that he is to the jurors. Robinson's credibility is deflated because of identity prejudice, he is unable to communicate his experience and others' misunderstand his words, and the jurors neglect to even consider alternate epistemic resources when evaluating his case.

Pohlhaus is making normative claims about the spread of knowledge in society. First, that situatedness and interdependence determine the formation of epistemic resources, and second, that epistemic resources that come out of marginalized positions are neglected or not even acknowledged. Situatedness is defined by Pohlhaus as, "How relations with others position the knower in relation to the world" (Pohlhaus 2012, 717). One's social position warrants her a certain amount of power over accepted epistemic resources, but this power is also determined by social positions of other knowers in relation to her. This is where tension between situatedness and interdependence can be observed. Differently situated individuals may have trouble properly communicating experiences because they may not share the same epistemic resources. This would make forming a cooperative interdependence more difficult, but because the tension is greater, that is all the more reason to enter into honest interdependence.

She explores some reasons why dominantly situated knowers are more prone to exhibit willful hermeneutical ignorance. Some dominantly situated knowers, or people in positions of power, choose to actively resist epistemic resources. This lack of utilization can be attributed to an absence of need to enter into a cooperative interdependent relationship with knowers in a different position. This reluctance is the case for both the marginalized and dominant positions. Some dominantly situated knowers may not necessarily have motivation to seek out such a relationship. I will detail reasons behind this motivation later. The marginalized might be hesitant to enter into a cooperative relationship due to a lack of trust, or possibly a fear of not being taken seriously. Individuals and groups of people may fail to recognize the social and cultural relevance that such resources hold. Another possibility is that some dominantly situated knowers prematurely disregard marginalized knowers as authorities of knowledge, or assume that they are unable to properly understand their own situation. The knower may be

confident in the epistemic resources that forms her worldview, so she sees no need for correction. All of these responses can be attributed to identity prejudice, unfounded skepticism, or a lack of information and/or misinformation that leads to a fundamental misunderstanding of epistemic resources.

Interesting complications arises when considering how knowers come to utilize faulty epistemic resources and do not realize the faultiness, deny it, or are indifferent about it. Faulty epistemic resources can be partially attributed to structural norms that, for example, carry historically racist, classist, or sexist attitudes. As norms, these attitudes may go unnoticed and thus remain uncorrected. Considering faulty epistemic resources, determining the degree of culpability onto a knower who exhibits willful hermeneutical ignorance comes into question. Pohlhaus explains that the jury sees the world through certain frameworks, and these effectively block out other perspectives. If the jury were to view the world from some framework other than classist, racist, sexist, white supremacist, etc. then they could partially open themselves up to Robinson's experienced world. The problem is that the jury fails to open themselves up to alternative frameworks, thereby making them culpable for the injustice done onto Robinson and culpable for not correcting the relevant frameworks. For cases that are not so obvious in determining the reasons for individuals and groups exhibiting willful hermeneutical ignorance, the party's culpability may not be so clear.

I have thus far explained Pohlhaus's argument for willful hermeneutical injustice. To further consider what motivates knowers to remain willfully hermeneutically ignorant, I will discuss Jose Medina's explanation of epistemic vice and epistemic virtue, while focusing mostly on epistemic vices as reasons for willful hermeneutical ignorance.

Jose Medina's The Epistemology of Resistance can be analyzed in order to expand on Pohlhaus's account of willful hermeneutical ignorance, and further explain the motivation behind willful hermeneutical ignorance. In Chapter Two of The Epistemology of Resistance titled "Resistance as Epistemic Vice and as Epistemic Virtue," Medina explores the epistemic vices of privileged groups, epistemic virtues of oppressed groups, and how resistance and responsibility play a role in our cognitive processes as knowers. Medina makes clear that the vices and virtues are not generalized uniformly throughout an entire group of people, nor does being part of a certain group automatically grant an individual such conditions. He argues that being in a certain social position leaves one

more prone to specific epistemic virtues and vices. The privileged knower and the oppressed knower on Medina's account can be respectively compared to the dominantly situated knower and the marginally situated knower on Pohlhaus's account. The epistemic vices Medina connects with the privileged knower are epistemic arrogance, epistemic laziness, and close-mindedness. The epistemic virtues connected to oppressed knowers are epistemic humility, intellectual curiosity, and open-mindedness. Again, not all privileged knowers exhibit these vices and not all oppressed knowers exhibit these virtues, but they are very helpful in understanding the motivation behind ignorance.

The epistemic vices that Medina labels can further our understanding about why individuals and groups of people are willfully hermeneutically ignorant. Epistemic arrogance explained by Medina involves, "Indulging in a delusional cognitive omnipotence that prevents him from learning from others and improving" (Medina 2013, 31). Here, the knower has a superfluous and unwarranted amount of credibility, and this also prevents any possible resistance against his provided knowledge from others. Epistemic arrogance works to block the formation of interdependent relationships between knowers of different social positions. This also sustains the dominantly situated position because no motivation is pushing one to seek out new perspectives. For epistemic laziness, Medina argues that, "A habitual lack of epistemic curiosity atrophies one's cognitive attitudes and dispositions. Continual neglect creates blinders that one allows to grow around one's epistemic perspective, constraining and slanting one's vantage point" (Medina 2013, 33). This can be exhibited by knowers who do not see relevance of other perspectives, or do not even acknowledge the existence of other perspectives outside one's worldview. Epistemic laziness as a vice is an attitude of not seeing certain knowledge as necessary, when in fact certain knowledge would be beneficial in some way for the knower and for others. The third epistemic vice Medina covers is close-mindedness. Close-mindedness, as Medina states, "Involves the lack of openness to a whole range of experiences and viewpoints that can destabilize (or create trouble for) one's own perspective" (Medina 2013, 35). This closed-off perspective discriminates against other perspectives in order to maintain a privileged social position. Closed-mindedness also selfishly asserts the attitude 'I already know everything that I need to know.'

These three epistemic vices work to expand Pohlhaus's view of situatedness and interdependence. They can also all be tied to reasons for not accepting and

acknowledging epistemic resources. Arrogance, laziness, and close-mindedness all contribute to the formation of worldviews that inhibit the spread of knowledge. As previously explained, the dominantly and marginally situated can be looked at as the privileged and oppressed respectively. The dominantly situated knowers are more prone to exhibiting these epistemic vices, which will help make clear the reasons that knowers maintain willful hermeneutical ignorance. Some dominantly situated knowers disqualify other marginalized knowers as authorities of knowledge. Those dominantly situated may not take others' experiences as true due to skepticism, or because they simply believe that others do not know what they are talking about. These various responses can be categorized as epistemic arrogance. Some dominantly situated knowers fail to recognize the relevance of other perspectives in their own lives. They may believe they have all the experience necessary for them and not search for anything outside of that framework. Such responses can be categorized as epistemic laziness and close-mindedness. All these vices can be seen as motivation to not enter into cooperative interdependent relationships with knowers who are differently situated.

To have blinders over one's epistemic perspective is a useful analogy to understand how willful hermeneutical ignorance is maintained. One may have blinders that block out ways to view the world and not even realize that one has blinders, or they recognize the blinders and maintain the closed-off perspective knowingly. Willful hermeneutical ignorance does not have to be actively acknowledged by the knower who is exhibiting it, which makes accounting for ignorance and working against it much more difficult.

EXPANDING ON POHLHAUS

In the following paragraphs I will explain what I believe to be missing from Pohlhaus's account of willful hermeneutical ignorance. First, Pohlhaus does not explicitly state this but all knowers are subject to exhibiting willful hermeneutical ignorance, it is not exclusive to the dominantly situated knowers. This is because a knower can be dominantly situated with some aspects of knowledge, yet marginally situated in other categories. Also, marginally situated knowers can fail to recognize and acknowledge epistemic resources that were developed out of a similar situatedness. Different areas of situatedness work together in order to determine one's position in relation to epistemic resources. Pohlhaus somewhat simplifies the two classes of knowers, but situatedness can be viewed as a complex

composition of multiple power relations that exist through one's social identity and relationships.

Secondly, Pohlhaus does not talk much about how, as knowers, we can determine the necessary relevance of other's experiences. This is a complicated question when searching for the importance of our own experience as well. For example, consider the two following scenarios. A young woman in college is at a social gathering, and a young man under the influence of alcohol approaches her. She may be interested, but throughout the night he insists on making comments and inappropriately feeling her, making her uncomfortable and a bit threatened. This young woman has never encountered a situation such as this and she is unsure how to navigate it. Now compare this with a different scenario. A middle-aged male actor is at a gathering with other individuals who are associated with riches and fame. A well-known movie producer approaches him and compliments his looks. The producer proceeds to tell him that he would look great in one of his films and that he could provide him ample opportunity. The producer begins to make inappropriate sexual comments and advances, making the male actor very uncomfortable and confused. He would like to accept the job to further his career, but he would be subjecting himself to possible further harassment. Even though these two individuals are very differently situated, the young woman and the male actor, they still can be said to share similar experiences of some sort of unwarranted sexual advances and social pressures. Yet, the question of how much they can relate to one another's experiences comes into play. Can the young woman see her situation through the male actor's predicament, and vice-versa? Both are at risk of being hermeneutically marginalized when sharing their experiences, and I can imagine a case where the hearer(s) exhibit willful hermeneutical ignorance. Such a response would entail the denial of any sexual harassment and the passive assertion that says that is just how things operate in the real world. Both responses neglect the real concern and show how easy misinterpreting experience can become. Though these situations share some similarities of individuals being taken advantage of, complicated questions arise of how much can be taken from both scenarios in order to further our understanding of experience. Relating certain experiences is a context-dependent task, which is why remaining open to new perspectives and alternative viewpoints is so important.

Pohlhaus explains that dominantly situated knowers may experience difficulties in recognizing and accepting alternate epistemic resources, and it may be rather

easy to refuse certain epistemic resources. Yet she makes note that, "Such a refusal is not an inherent inability, but rather a willful act," (Pohlhaus 2012, 729). One reason she offers for the difficulty involved in recognition and acceptance is that for some dominantly situated knowers, learning of alternate, marginalized perspectives is disorienting. She states, "It opens one's eyes to aspects of one's situatedness with which it is not easy to contend" (Pohlhaus 2012, 721). In some sense, becoming aware of any new perspectives can be disorienting because the new experiences are shaping the way in which the world is viewed. This sheds light on the importance of recognizing and becoming aware of one's own perspective. Understandably this is a complex idea, and may be impossible to fully recognize every single aspect of one's social situatedness. Yet, working to actively shape and expand one's worldview can make entering into cooperative interdependent relationships with those differently situated a much easier process. This can work to alleviate willful hermeneutical ignorance. As previously explained, marginally situated knowers are more aware of their positions because they are vulnerable to those dominantly situated and often experience injustices. For the dominantly situated knower, there is nothing motivating her to investigate other parts of the world in light of others' concerns (Pohlhaus 2012, 721). Because of this, the dominantly situated knowers may have more work to do in order to further their understanding of their own situatedness. This does not excuse them of the responsibly to educate themselves, rather more responsibility rests on them to learn of injustices to which some marginally situated are subject.

So far I have provided accounts of Fricker's epistemic injustices, Medina's epistemic vices, and Pohlhaus's overview of willful hermeneutical ignorance. With the example of Tom Robinson from To Kill a Mockingbird, Pohlhaus utilizes Fricker in order to show what she was missing from her explanation on epistemic injustices. Pohlhaus introduces willful hermeneutical ignorance as another type of injustice that Fricker fails to take into consideration. Marginally situated knowers are, "In a position to notice inadequacies in our epistemic resources that are more entrenched" (Pohlhaus 2012, 720). The problem is that some dominantly situated knowers do not acknowledge or consider the resources needed to grasp marginalized individuals' experiences. They may consider the experience of others, but by not utilizing proper epistemic resources, the consideration may be through a faulty framework, causing a misunderstanding of perspective. Jose

Medina's work on epistemic vices helps further understanding of the motivation behind willful hermeneutical ignorance.

The most important and essential epistemic resources are conceptualized by marginally situated knowers. This would include people facing structural injustices, such as women, African Americans, immigrants, disabled, etc. The epistemic resources that come out of marginalized situatedness are the most important because they highlight structural injustices and social injustices that are more subtle and under the surface. The marginally situated knowers do not have as much influence over implementation and recognition of epistemic resources, so the dominantly situated knowers, in a sense, have a certain degree of power over the spread of knowledge, more-so than the marginally situated knowers. Because of their epistemic situatedness, dominantly situated knowers should have a greater responsibility over epistemic resources and should be continuously correcting for injustices when possible. Unfortunately, their epistemic situatedness also provides them with the option of ignoring and/or not acknowledging epistemic resources for their own benefit whether the ignorance is exhibited knowingly or unknowingly.

Willful hermeneutical ignorance develops from cognitive missteps that lead to the dismissal or overlooking of epistemic resources. Cognitive missteps can include having a fundamental misunderstanding of one's own situatedness and a misunderstanding of others' as well. To completely understand one's situatedness is basically asking an individual to be wholly aware of her social position with relation to everyone in her life under consideration of all experience. This does not require the knower to completely understand her situatedness, but she should be open to consideration and being aware of epistemic vices. Also, one should acknowledge the importance of useful epistemic resources. Another cognitive misstep is not recognizing the use-value of certain resources. In other words, an individual or group of people may not see any real importance in the acceptance of other perspectives or frameworks. They may fail to recognize the social and cultural significance of useful epistemic resources. Because of this, they overlook whole bodies of knowledge and ways of understanding. Some people may also be content or indifferent about certain knowledge and perspectives. This tracks back to the vices of epistemic laziness and close-mindedness. This is a worldview which asserts that no new perspectives are necessary and other ways of understanding are unimportant. Such a worldview is very epistemically closed-off.

Pohlhaus explains that a reason that some dominantly situated knowers do not acknowledge or accept certain epistemic resources is because learning of one's social position is disorienting and not especially easy. I argue that becoming aware of one's own privilege is necessary in order to form cooperative interdependent relationships with others whom are differently situated. If a dominantly situated knower is self-delusional and considers himself oppressed as a knower, then he will be constantly misinterpreting his own experiences and others'. Yet, determining the necessary epistemic resources needed to grasp certain experiences which may be foreign to an individual seems difficult. Constantly considering the epistemic resources that are available to oneself is an arduous task and may not be so straight forward. The first step is understanding the concepts of situatedness and interdependence, and how they interact within one's own life.

The structural aspects of willful hermeneutical ignorance can be correlated with structural inequalities in society. Even though the example of identity prejudice from To Kill a Mockingbird is a bit outdated, still racism, classism, and sexism are all relevant issues when considering identity prejudice. Yet, these certain prejudices may be even more deeply imbedded in a social worldview in a much more nuanced way. Prejudices can affect cognition underneath the surface without us noticing the effects. That is a cognitive misstep that needs critical attention to manage because some norms that have been widely accepted in ways of thought have proven to cause epistemic harm.

In her conclusion, Pohlhaus offers her solution, she states, "The solution is for dominantly situated knowers to catch up and learn to use epistemic resources they lack by forging truly cooperative interdependent relations with marginally situated knowers" (Pohlhaus 2012, 733). When individuals fail to form honest interdependent relations with other knowers, they effectively maintain their own ignorance and deny the usefulness of epistemic resources. Yet, dominantly situated knowers may not see importance in forging an epistemic relationship of this kind. So then, the question arises of how to motivate and convince the dominantly situated knower to enter said epistemic relations, and thus take responsibility for correcting his faulty epistemic resources into his own hands. The source of motivation can be argued to come from nowhere else other than the knower himself. The duty of eliminating willful hermeneutical ignorance lies on the shoulders of those exhibiting the ignorance. Marginally situated knowers should

not have the responsibility to push dominantly situated knowers into cooperative interdependent relations, yet they should be helpful in forging interdependence.

Because learning of new epistemic resources is a sort of challenge for a knower's worldview, the new perspective may be difficult to accept. Yet, the restructuring of one's epistemic framework is necessary in order to step out of ignorance. When one has a worldview built upon faulty, out-of-date epistemic resources the reconstruction of one's framework is necessary, regardless of how disorienting that it may be. Furthermore, because epistemic resources are social tools, a social responsibility exists surrounding willful hermeneutical ignorance. This means that dominantly situated knowers that continue to engage in the world while lacking important epistemic resources indirectly affects all of us as knowers. Culpability is a complicated issue when discussing willful hermeneutical ignorance, but more importantly the focus should be on the causes of the ignorance and the ways in which knowers, both dominantly and marginally situated, can work against ignorance.

CONCLUSION

Willful hermeneutical ignorance develops out of cognitive missteps within one's own worldview, and out of a lack of necessary knowledge of one's own situatedness. The most useful and important epistemic resources come out of the marginally situated knowers' position, yet these are the resources that are most neglected. Fricker provides groundwork for epistemic injustice and labels hermeneutical injustice as one of the two injustices done onto knowers. Her account of injustice lead to Pohlhaus forming the concept of willful hermeneutical ignorance as another separate form of epistemic injustice. Medina offers three epistemic vices, arrogance, laziness, and close-mindedness, which furthers understanding of the motivation that sustains willful hermeneutical ignorance. Pohlhaus asserts that situatedness and interdependence are essential in the understanding of willful hermeneutical ignorance, and are essential to grasp the relationship between marginally situated knowers and dominantly situated knowers. Her explanation shows how power relations affect the transfer of knowledge. Furthermore, because epistemic resources are necessary social tools that help give way to new ways of understanding, they help us access entire bodies of knowledge and further understand knowledge as a social entity. The most attainable solution I believe is the solution in which Pohlhaus offers, which is

the dominantly situated knowers must become more open-minded and work to forge honest, cooperative, interdependent relations with the marginally situated knowers so to develop and become aware of necessary epistemic resources.

REFERENCES

Fricker, Miranda. 2007. *Epistemic Injustice: Power & the Ethics of Knowing*. New York. Oxford University Press.

Lee, Harper. 2006. *To Kill a Mockingbird*. New York. Harper Perennial Modern Classics.

Medina, Jose. 2013. *The Epistemology of Resistance*. New York. Oxford University Press.

Pohlhaus, Gaile Jr. 2012. *Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance*. Hypatia vol. 27 no. 4. (715–733).

compos mentis

Being Human

Caitlyn Lecour
Augustana College

ABSTRACT
The purpose of this project was to examine and explain the conception of being human from a philosophical perspective. I argue that our conception of being human is based on a standard of mind. After establishing that our conception of being human is based on a standard of mind, I describe the three aspects, or pillars, on which this standard of mind is based: The Organic Pillar, The Cognitive Pillar, and the Social Pillar. My goal in pursuing this topic is to provide a comprehensive framework in the face of advancing artificial intelligence technologies and research into non homo sapien consciousness. I draw influence from Confucianism, Taoism, Hindu philosophy, and Western Philosophy. The framework I provide is meant to take on a cross cultural perspective so that it may be applied broadly.

KEYWORDS
Being, Human, Personhood

I. INTRODUCTION

The notion of being human is one that permeates all cultures. This notion has undergone many changes that have left open the question of what it is to be human. Often the notion of being human is equated to the notion of person. However, I am reluctant to use the term 'person' as it denotes a conception of 'is' and 'is not' which is passive. The conception of being human that I describe is active in that it is dependent upon one's striving for balance and focuses on the actions of an individual and the motivations behind said actions. For example, a non-artificially intelligent coffee maker is passive in that it can only do what it is made to do and does not have any goals or volitions which may result in its striving for improvement nor in its striving to maintain some standard. It is a coffee maker not because it tries to be but because it can be nothing else. In addition, prior conceptions of personhood have been based on arbitrary checklists of qualities which typically include some reference to intelligence or awareness of certain abstract ideas (Farah and Heberlein 2007). The conception of being human I describe is not necessarily based on any quotient of intelligence nor awareness of any abstract concepts. In this paper, I argue that to be human is not based on some innate feature of homo sapiens as a species but, rather, to be human is a standard which one must strive for in order to retain their humanity. This is similar to Charles Pasternak's Curiosity and Quest (Pasternak 2007, 114-132) in that it focuses on an active feature which does not necessarily exclude non homo sapien forms of life. However, my account differs in that rather than focusing on four innate features of homo sapiens which allow for more intense searching for knowledge, it focuses on the mind, the standard thereof, and one's ability to strive for balance among three range-based criteria1. After establishing being human as a standard of mind, I will attempt to describe the nature of this standard. In recent times, there has been a movement away from a heavily Eurocentric point of view. Nevertheless, although it is important to give voice to other points of view, it is also important

1. A range-based criterion, as I define it, is any criterion which allows for degrees of its defining concept rather than a passive criterion which relies on an arbitrarily imposed checklist of qualities. For example, identifying number X as the lowest possible IQ for a full person is arbitrary. Why number X and not number Y or number Z? There is no obvious reason to choose number X over numbers Y or Z. In addition, if individual A were to have an IQ of X while individual B were to have an IQ of X-1, then A would retain full personhood while B is denied it despite the vast similarities between the two with regard to IQ. The range-based criteria I describe allows for a more fluid approach which avoids this problem.

that the more commonly heard perspectives are not silenced, especially when the topic pertains to a conception which permeates homo sapien culture broadly. The standard I describe is influenced by key principles of Confucian, Daoist, Indian, as well as Western thought and I will describe each where appropriate.

II. DEFINITIONS

In this paper, there are three terms of relevance which each have a distinctive definition and basis with regard to morality: human, person, and being human. Human, in the noun sense, is too often conflated with the biological human, or homo sapien. There is nothing applicable with respect to morality in biological determinations. An entity is determined to be biologically human or not in much the same way as a book is determined to be written in English or not. Although homo sapien DNA is a necessary trait for being classified as a biological human, it is not sufficient for the distinctions made with regard to morality or the rights ascribed to moral agents. This sense of the term 'human' is not to be confused with 'human' in the adjectival sense. Human, as used in the adjectival sense, refers to what is morally applicable, or what is able to be involved in moral determinations. Person, or personhood, refers to a set of traits or qualities which supposedly entitle an entity to certain rights or privileges. This notion does have a moral basis, although a one-sided one. The moral basis for personhood is only with regard to how one should be treated, not how one should treat others. There is no accountability necessarily contained in this notion. Not all persons behave morally with regard to the treatment of others. For example, abusive partners in domestic violence situations may beat, rape, threaten, and commit a plethora of other heinous acts against their victims and not face any consequences with regard to their being recognized as full persons. Being human, or the conception thereof, is based on a spectrum in which the behaviors of an individual oscillate about a central point which determines the degree to which one embodies a sense of humanness, or proximity to the center-most point of the Three Pillars. This conception has a two-way moral basis in that it focuses not only on how one should be treated but also how one should treat others. This moral basis is elaborated on later in this paper with respect to the social pillar.

III. PART 1

Before describing the standard by which our conception of being human is based, I must first provide some proof of its existence and that it is something that is clear and consistent so as to be recognizable. If there is not a clear and consistent conception of being human, then there is nothing to separate being human from not being human. This type of distinction is logically sound in that it is not possible to sort entities into two separate groups without knowing what qualities are attributed to each group or at least what qualities are attributed to one group. For example, if Alice wishes to sort a collection of books based on some distinguishing feature of the books, then she could only sort the collection into two groups if all the books are not identical in their ISBN identifications. With regard to distinctions made of conceptual natures, this notion remains true. In the Advaita Vedanta there is discussion of the two notions of Brahman: Saguna Brahman and Nirguna Brahman. Saguna Brahman is defined as Brahman with qualities whereas Nirguna Brahman is Brahman without qualities. Having qualities merely means that one is able to make assertions about what descriptions can be made with regard to the entity whereas not having qualities is the lack of this ability. It is pointed out that Saguna Brahman is a contradiction by nature due to the fact that whenever one ascribes one quality to Brahman, one must also ascribe its opposite given that Brahman is an entity which is all encompassing by nature (Deutsch 1969, 14). If an entity is both X and not X, then the entity is also neither X nor not X. This results in Nirguna Brahman, or Brahman without qualities, being the conception of Brahman that is most consistent with the core definition of Brahman as an entity which is all-encompassing. This explanation of Brahman serves to highlight the fact that in order for there to be one thing and not another, there must be some distinction to separate the two notions, just as there is between the two notions of Brahman.2

The first objection that one may be tempted to assert is that the distinction between being human and not being human is merely a semantic distinction rather than a pragmatic one. I concede that the objection is correct to say that the distinction is heavily rooted in semantics. However, it is best not applied here since the implications of this distinction are pragmatic. For instance, if entity A were merely considered less than human, then entity A would be treated as such

2. If one needs another example of this notion, then one might refer to Benedict Spinoza's argument for substance monism.

with no thought about ethics. This was the case for the indigenous populations of the Americas as well as for African peoples brought into the country as slaves. One may object that such practices are a thing of the past and assert that the distinction between being human and not being human currently plays no part in the world today. Nevertheless, such an assertion would be mistaken for two reasons. The first reason is such an assumption seems to assert that an expanded view on what is human is the same as an abolishment of the distinction altogether. This is simply not the case at all. For example, expanding one's circle of friends does not entail that one is friends with everyone or considers objects to be friends. It merely means that more entities are held in the category of 'friend'. Second, to this day, non-human specimens are fundamentally treated differently as can be seen in cases involving laboratory animals and other cases which will be discussed later. It remains to be seen to what degree morality is rooted in this idea of being human. But, it is undeniable that it plays some key role in the distinctions that we make.

Given the evidence that the distinction between being human and not being human is a distinction that has some effect in the world we live in and the ways in which we conduct ourselves, it follows that there must be a clear and consistent conception of being human. This conception, however, must have some basis, some essential feature, that would constitute the essence of what it is to be human that would distinguish it from what it is to not be human. The conception of being human can only be based on one of three things: homo sapien DNA, homo sapien form, or what I will call human mind3. I will address each of these in the following paragraphs.

3. The reason I say homo sapien DNA and homo sapien form rather than simply DNA and form is due to the fact that simply having DNA or a definite form or both, of any type or combination, has never been a criterion for personhood. The fact that an animal bled or screamed did not stop ancient peoples from hunting or sacrificing it despite having a definite form and clear biological similarities. Any solid object has a definite form and yet no one of a sane mind greets and converses with a basic (i.e. non artificially intelligent) refrigerator despite having routine interactions with it. Publications on the discovery of DNA did not stop scientists from continuing animal experimentation, which continues to this day, nor did it result in all organic life being attributed with personhood.

IV. PART 2

If our conception of what it is to be human were based on whether or not one has homo sapien DNA, then any substance which contained even the slightest trace of this type of DNA would be considered fully human. However, this is not the case, and it would be absurd if it were since traces of homo sapien DNA can be found in fluids and secretions of the body many of which are thought of as nothing more than waste to be flushed down a drain. One may object that cellular waste products are not living entities with homo sapien DNA and cannot be considered as such since they lack living processes. However, one need not look further than the history of cancer research to find a case of living cells with homo sapien DNA that were treated as nothing more than disposable lab specimens. These cells are what are known as HeLa cells. The cells have survived for decades and have contributed greatly to cancer research. Nevertheless, they receive no accolades, no honors, no merits for their sacrifices, and they shouldn't simply because the cells themselves are not fully human despite having full strands of homo sapien DNA. So, if having homo sapien DNA is not enough to be considered fully human, then it follows that homo sapien DNA is not the essential feature on which we base our conception of being human.

Although having homo sapien DNA is not enough to be considered human, one may argue that having the form of a homo sapien is the essential feature of being human. However, this is also not the case since statues and mannequins are not considered to be fully human, if at all. This can be seen in cases in which such figures are damaged. Let's say that Jerome works in a beauty parlor with several mannequins strewn across the room. One day, Jerome is on a tall ladder near one of the mannequins and loses her balance, falling on and breaking the mannequin so severely that it must be replaced. Jerome is not going to be charged with manslaughter for breaking the mannequin since the mannequin is not considered to be human. One may argue that the mannequin, though it has the form of a homo sapien, does not have homo sapien DNA and, thus, is not considered to be human on the grounds that it needs both the DNA and form of a homo sapien in order to be considered human. However, this is also not enough to be considered human since corpses are not considered to be human. If an individual were in a situation in which one had to choose between a living being and a dead one to save from falling into a deep chasm never to be recovered, the individual would always choose the living one given that the individual making this decision is a

sensible one. In addition, if having both homo sapien form and DNA were enough to consider one as fully human, then there would be no difference with regard to the treatment of a living being as opposed to a dead one. This is plainly false considering the fact that these entities are fundamentally treated differently. A corpse is burned, buried, left to rot, or made to undergo some other ritual with the end goal of its disposal whereas a living being is its own agent to do onto others and react to things done onto it. Rights of living are ascribed to living beings whereas funeral rights are ascribed to dead ones. If homo sapien DNA and form were enough to be human, then it would not make sense to treat homo sapien corpses as things to be disposed of and living homo sapiens as agents of themselves. A family would be completely sensible, if not required out of respect, to leave a deceased relative in the exact same location as the one in which the relative died, especially if that location is in the family's home. However, that is not considered sensible, in any case. So, neither homo sapien form nor homo sapien form with homo sapien DNA is the essential feature that distinguishes what it is to be human.

If what it is to be human is not distinguishable via homo sapien DNA, homo sapien form, or some combination of both, then our conception of being human must be based on human mind. However, one runs into the problem of discerning what is the essential characteristic of a human mind as opposed to a non-human mind. Afterall, if there is something to separate being human from not being human and our conception of being human is based on mind, then it stands to reason that there must be a distinction between merely having a mind and having a human mind.

V. PART 3

Although in the past many have assumed that only homo sapiens have the capacity for thought characteristic of having a mind, this notion is mistaken. It is clear that some thought must occur in order to solve problems and create structures such as those which are commonly observed in nature. To assert that other animals are mindless machines while at the same time describing them as running on some program is contradictory since the mind is a kind of program which dictates action and thought-based planning. Although the nature of this program is something that has been debated extensively for years, that much is clear. If it were true that only homo sapiens had the capacity for mind in the

sense that one is capable of higher thinking, then there would not be cases of non-human persons, natural or artificial. However, there are and have been cases of non-human persons in the world, such as Sofia the robot and Koko the gorilla, and many more cases of non-biologically human characters in media being perceived as nothing less than human, as in the adjectival sense described earlier. Nevertheless, there are also cases in media of non-human minds, particularly in the horror genre. For example, it is not uncommon to observe monsters and demonic possessions in film. The viewer typically never feels remorse at the demise of these antagonistic entities. For example, in the film, The Possession of Hannah Grace, a young woman is possessed by a demon which kills in order to gain power and heal itself. During the exorcism, the woman dies, ending the continuity of her mind. Nevertheless, the demon remains in her body and continues to control it. If there were no distinction between the non-human mind of the demon and the human mind of Hannah Grace in the sense of its humanity, then the film would lose much of its overall plot as a horror film and become a scientific fantasy about either a supernatural being occupying a natural being's body or a regular being undergoing a profound shift in appetite and gaining some powers to be used in a morally questionable manner. Given the fact that it is considered a horror film, the distinction with regard to humanity is clear. But, how is that distinction made?

All distinctions are made via some standard. As was discussed previously, it is not possible to separate one group into multiple smaller groups without having something to determine which smaller group each of the members of the one group belong. Given the distinctions between being human and not being human and the argument that these distinctions are based on mind, it follows that there must be some standard on which we base our conception of mind. If it is true that our conception of being human is based on a standard of mind, then the question remains regarding the nature of this standard.

VI. THE THREE PILLARS

There are three pillars on which the conception of being human rests: The Organic Pillar, the Cognitive Pillar, and the Social Pillar. Each pillar relies upon the others for the maintenance of one's proximity to the center. In other words, it is not possible to have one or several but not all the pillars and maintain the full scope of one's humanness. I will take time to describe each and refer to previous and new examples to point out the shortcomings of one of these essential pillars

that reduce the humanity for the entity in question. I say reduce simply since, cinema and media aside, there are few if any cases in real life in which one is able to separate oneself entirely from any one of or all of the pillars of humanity.

The simplest of the three pillars is the Organic Pillar which relates to one's ability to respond to the conditions and events of the surrounding world. This is not to say that an individual is less human when asleep than awake since the potential to respond remains and the individual is still able to respond reflexively. Nevertheless, it does explain the fundamental difference between a living human and a dead one. A corpse can never respond to the world but simply be an object within it. One must note that this trait is not unique to homo sapiens and that it is a fundamental trait of living things. In other words, as far as the Organic Pillar represents humanness, other animals and anything which responds to the conditions and events of the world may be considered just as human as homo sapiens are. Nevertheless, that does not necessarily entail thoughts or emotions regarding those conditions or events.

The Cognitive Pillar is concerned with one's awareness of the world, emotions, and abilities to reason and analyze information thereof. This pillar has often been the focus in the past4, being set out as the defining feature of human beings, or homo sapiens, that sets them apart from other species. However, the notion that homo sapiens are the only species that exhibit some connection to this pillar is vastly mistaken due to the fact that other species have exhibited the problem-solving capabilities and emotional awareness that is characteristic of this pillar. I would like to clarify the fundamental difference between what I am describing and past notions of reasoning so as to not be misconstrued. When I refer to one's capacity for reason, I am not referring to the notion of wisdom described by Plato in The Republic. Reason, as I intend to describe it, is one's capacity to think about the information one is presented with and make decisions based on that information. This is not an ability unique to homo sapiens since other species exhibit this capacity in their interactions with the world. For example, a predator, such as a leopard, must make decisions while hunting that may increase or decrease its likelihood of catching its prey in addition to making the decision to pursue its current prey or redirect its focus to an easier target. Of course, it

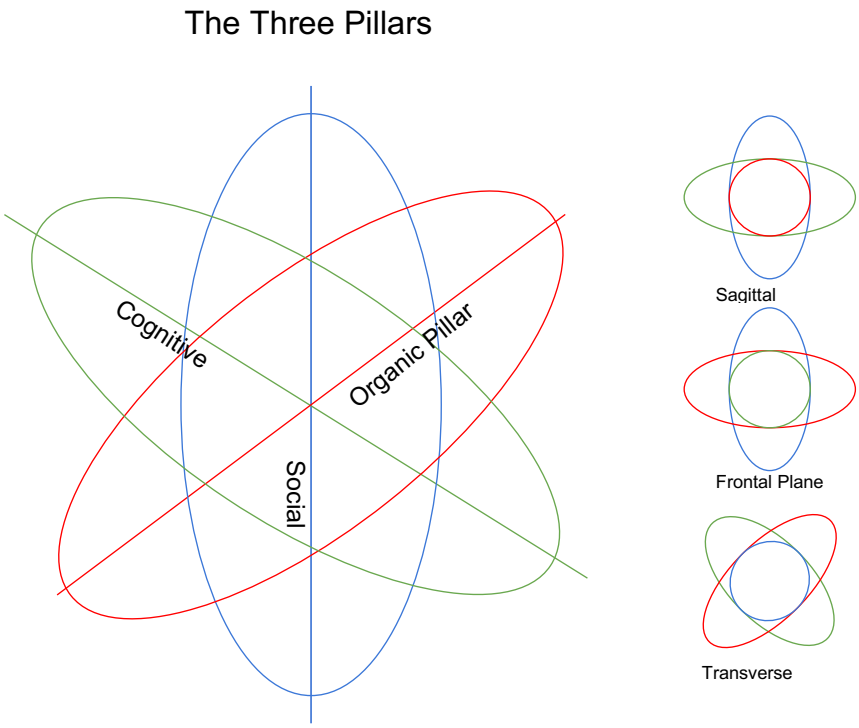
4. The Cognitive Pillar contains within it part of what may be referred to as consciousness. However, it must be noted that it does not necessarily contain the full scope of consciousness. Due to the uncertainty and widespread disagreement with regard to what consciousness is and where it occurs, it is best not dwelled upon here.

cannot be overlooked that there are discrepancies between species due to the scope of information that may contribute to one's decision. Abstract concepts such as value, virtue, emotion, and meaning may also play a role in one's decision. There is not much that can be said with regard to the abstract concepts that other species may have given that, as Thomas Nagel pointed out in his bat argument (Alter and Howell 2012, 15-23), we as humans do not have an applicable concept of the lives and experiences of other species, only our imaginative guesses thereof. Nevertheless, such concepts differ from culture to culture and even from one individual to the next. So, it is important to clarify that it is not the particulars about these concepts and notions of value that determine the connection one has to the Cognitive Pillar, but that one is able to apply information while making decisions. The difference between decision making and reflexive responses lies in the differences in complexity and the variation in the results between individuals. For example, regardless of the background, beliefs, or other mental phenomenon of an individual, when the individual touches something hot and experiences pain, that individual will pull away as a reflexive action. However, when making a decision, an individual may consult with one's background knowledge, beliefs, or other mental phenomenon in order to construct some course of action. In other words, there is usually some thought accompanying decisions which reflexes seem to lack.

The Social Pillar concerns itself with the connections and potential to connect that one has with others and the world. At the individual level, this relates to one's ability to interact with others by forming social connections and living by some moral code. It is important to note that these connections are bidirectional. In other words, the connections that individual A has with individual B can affect the humanness of both. For example, if Nora holds Ashram, whose humanness is in question due to events that Ashram did not contributed to directly, to be nothing less than human, then Ashram is more human than otherwise. This is typically only applicable to the unborn5 but can also be applied to those in comatose or degenerative states. On the opposing side of the spectrum, if Carja treats Arcee as if Arcee were less human than in actuality, then Carja has the potential

5. I would like to clarify that a woman's choice to have an abortion is permissible under this view. She is not necessarily more or less human for having undergone such a procedure due to the fact that there are too many reasons for getting an abortion to pin the act itself with some inherent moral value. In addition, as will be discussed later, having compassion for the unborn entity is not a justification for lacking compassion for the would-be mother.

of reducing his own humanity given that he takes cruelty too far and drifts too far in one direction of the pillar. Of all the pillars, the Social Pillar is the one which focuses the most on our conception of being human as socially embedded. It does not necessitate a particular view on the morality of one's actions, but it does necessitate a degree of compassion and understanding thereof. As Confucius supposedly said, "To impose the death penalty without first understanding words is to be cruel… [a man] has no way of judging men unless he understands words" (Confucius [500 B.C.] 1979). Although one may argue that anything done in excess, even something seemingly virtuous, quickly becomes a vice when there is a lack of moderation, it must be noted that true compassion and understanding are found at the point of balance. The fact that an individual has compassion for and understands why another did some act which brought harm to another does not excuse the act. To dismiss the harmful act is to lack compassion for those who are harmed by the act. On the other hand, having so much compassion for those who are harmed that one inflicts too great of a punishment on the perpetrator is also to have a lack of compassion, especially given recent neuroscience findings which imply less free will than previously thought (Caruso & Flanagan 2018).



VII. BALANCE

The goal of each of the pillars is balance. If one were to imagine a three-dimensional figure to represent the Three Pillars on which our conception of being human is based, then one may use the three-dimensional plane commonly found in mathematics as a framework for this model. Each axis is a pillar which has a common intersection with all the other pillars. One may reside on either side of this central point with regard to one's humanness. However, although each of the pillars has a positive and negative side, there is no value ascribed to either side with regard to which is the superior side due to the fact that the center is the point of balance which we strive for and any push toward one side indefinitely results in an extreme which reduces this balance. One may relate this to certain Daoist principles. Imagine a wheel with thirty spokes all joined to one hub. The further the hub is from the center of the wheel, the more unstable the wheel is and the rougher the ride. The Tao Te Ching emphasizes this notion of balance throughout but especially in the lines "There is no crime greater than having too many desires; there is no disaster greater than not being content; there is no misfortune greater than being covetous. Hence in being content, one will always have enough" (Lao Tzu [600 B.C] 1963). One may object that one source emphasizing balance is not a strong enough defense. However, this notion, that balance results in a superior outcome than otherwise, is not unique to Daoism. Aristotle also toted this notion in his description of the excellences. For Aristotle, each excellence was found at the point of balance between two vices (Aristotle [340 B.C.] 1987).

One may attempt to argue that, with regard to the model I presented earlier, it is always better to be on the positive side of each pillar. However, this is plainly false since when Kant's universalized maxim is applied, the result is either stagnation or chaos. Imagine a society in which the positive side of the Social Pillar were enacted universally; let's call this society the Land of Gentlemen. In the Land of Gentleman, everyone strives to act altruistically to always benefit the other person more than or rather than oneself. However, such a society fundamentally could not function since everything would stagnate. For instance, if Cayo, an average but stubborn citizen from the Land of Gentlemen, were to seek to trade some amount of cabbages for an amount of bananas at the local shop run by Caiyu, another average but stubborn citizen, then an altercation over the value of each supply would result in nothing being bought since each person would assert that the other deserved more for what was being offered. In addition, if

one were to expand upon this mindset, it is probable that in this type of society some individuals would eventually starve themselves to death as a result of this altruistic extreme due to the fact that in our world, one in which complete altruism is not the norm nor the standard way of life, there are individuals who become vegetarian or vegan for moral reasons that pertain to the suffering and death of the creatures they would otherwise consume. One may argue that a society such as the Land of Gentleman is not so bad and that the scenario described is merely an example of an excess in stubbornness. However, I respond that although such a society functions in theory, it is not probable that it could function in practice since in our world people differ in degrees with regard to altruism and adherence to social norms. Further, such an extreme of altruism disregards the values and attachments of individuals. One may be accused of selfishness for not wanting to give up something arbitrary which that individual holds meaning in. If two individuals are biologically compatible and likely to produce offspring which would benefit society greatly, then according to this altruistic society's norms, these two individuals are morally required to produce a child regardless of the feelings of each individual involved. To hold an individual to the standard of sacrificing one's own morals, thoughts, and emotions for the sake of the wants and desires of others is unreasonable. For example, if there is a group of individuals who want to engage in sexual activity with one individual simultaneously, then the individual does not necessarily have a reason to say no in an altruistic society. Sexual activity itself is not a harm and provides many benefits. However, if an individual does not desire to engage in sexual activity but is pressured into it, then it is rape, which is always a harm. The opposite side of the Social Pillar is no better.

Imagine a society in which selfishness was considered a virtue that everyone strived for. Let's call this society the Land of Narcissists. In this society, if you could call it that, everyone strives to get the most for oneself through any means necessary. There is no social contract among people since it is in no one's best interest to sustain such a contract. One might be reminded of the Prisoner's Dilemma in which two individuals each must choose whether or not to betray the other person. If the one talks and the other doesn't, then the one receives the greatest benefit to the greatest detriment of the other. However, since both have the incentive to betray the other, both talk and receive less than if they collaborated silence. The fundamental condition of any contract lies in trust. However, in a society in which everyone seeks to benefit oneself without any of

the constraints that trusting or relying on others would bring, such a society would be nothing more than a free-for-all of back-stabbing and throat cutting in both the literal and metaphorical sense. It would be chaos.

VIII. THE CENTER

The extremes of each side have shown to be unstable and unsustainable. Complete and universal altruism leads to starvation, stagnation, or a standard of disregarding the individual in favor of the group whereas complete and universal selfishness leads to universal mistrust, murder, and a standard of disregarding the group in favor of the individual. However, the question remains with regard to how one might achieve the central point of the Three Pillars of Being Human. Frankly, achieving the centermost point of the Pillars is not achievable, but the fact does not entail that it is not worth striving for. Before defending the value of striving for the center, I must first explain why the centermost point is not achievable. There are several reasons.

First, for one to achieve the centermost point, one must know what the centermost point entails. One would need to know an absolute truth, an unchanging, stagnate truth in all its dimensions and details. Since living things are limited by their own perceptions and do not have direct access to the perceptions of other living things as to form a network, this is not possible. One could imagine looking at a three-dimensional figure. Although one may clearly and distinctly perceive one side of the figure, one cannot see the other hidden sides without moving to reveal them. Even when all sides are revealed, the details may remain unseen, such is the case with living tissues. To the naked eye, the tissue of the brain appears to be uniform and solid. However, if one were to use a light microscope, one may observe the neurons that make up the tissue. Nevertheless, there remains more to be seen. If one were to use an electron microscope, then one may reveal the Golgi cells that assist in the maintenance of the neurons, the organelles within the neuron, and the compositions of the synaptic cleft between neurons. Yet, the individual elemental compounds that make up the neurotransmitters, compounds, and atomic structure of each remains hidden and can only be represented by probable theories. Probable truth is not actual truth.

Second, to achieve the center, it is necessary for one to remain fixed in that position. To remain fixed in a given position, one must have equal forces acting on it. The only way in which one moves along any of the pillars is by acting in the

world. This includes all actions and non-actions. Given that a single entity cannot enact opposing actions, such as inhaling and exhaling simultaneously, it is not possible for one to stagnate at a fixed point on one of the pillars. So, it is not possible to achieve the center.

Now that it has been explained why the center is not achievable, it is important that I explain why striving for it is nevertheless valuable. If one were to release oneself from striving for the center, then one would inevitably fall to one of the extremes. As shown previously with regard to the Social Pillar, this is unsustainable and results in universal death. One might object that this notion assumes that life is inherently valuable. However, I disagree since I have not made any claim with regard to the value of life nor the value of death. The only argument I have made about value is the argument for balance. Since balance requires both life and death, one cannot hold one as more valuable than the other. In addition, for the present argument to be applicable, there must be entities which make distinctions. It is not possible to make any distinction, or act for that matter, without having an existence which makes it possible for one to do so. Therefore, the objection regarding value assumptions on life or death does not apply to this argument.

IX. BORDERLINE CASES

An objection that one might attempt is that there are some homo sapiens who lack the faculties to comprehend the world around them, such as those with anencephaly, and ask whether or not that results in their being less human and, thus, less worthy of rights. What rights are being referred to exactly? The only rights necessitated by this paper are dependent upon the social pillar which only entails the right to be treated with compassion and understanding. However, treating one with anencephaly with compassion and understanding is neither more or less important than treating with compassion and understanding those without the condition who may accept or reject the task of caring for such a being. As for being less human, although individuals with anencephaly, or any neurodegenerative condition, lose or lack much of their ability to make decisions and respond to the world, they still have connections to the world and to others; they still respond, albeit reflexively. The only pillar that may be affected is the Cognitive Pillar. Given that one does not have that direct access that would allow one to know exactly what another is aware of, it is not possible to ascribe an absolute negation of the Cognitive Pillar to those with anencephaly, especially

since such individuals may survive for several years in the care of willing family members who develop attachments to them. In other words, the neurological condition that one is born with or later develops cannot affect one's humanness, especially if there are others willing to vouch for one's humanness.

The only feature that can affect one's humanness is one's connection to the center. It is this connection which upholds one's humanness. Although some may reject striving for this center and in turn reject their humanity, they still retain some connection to this center and some aspect of their humanity since they retain the capacity for change. As the saying goes, every saint has a past and every sinner has a future. Even an individual who tries in earnest to be a bookshelf is limited in this endeavor by biological restraints, such as matter consumption or excretion and muscle fatigue or atrophy. In addition, such an individual retains the capacity to change, to become something other than or more than a bookshelf. Although this individual strives to be an object, the ability to strive is characteristic of all subjects and our conception of being human. Thus, even the most adamant rejection of one's humanity is not enough to sever ties with it.

X. CONCLUSION

One may wonder at the implications of this argument. In short, our conception of being human is not a matter of the arrangement of one's physical material but a matter of being connected to a state of balance among three range-based criteria. This connection to the center is maintained in the actions of the individual. Much like a piece of silver, one's humanness is something which must be repeatedly attended to in order to prevent it from dulling and tarnishing. However, whereas silver is merely polished periodically for its maintenance, humanness must be continually attended to for its maintenance because every action, or non-action, affects one's relation to the center. The conception of being human is as much about being as it is about humanness. One must be human in order to be human.

REFERENCES

Aristotle (340 B.C.)1987. *A New Aristotle Reader*. Edited by J. L. Ackrill. New Jersey: Princeton University Press.

Confucius (500 B.C.) 1979. *The Analects*. Translated by D.C. Lau. England: Penguin Books.

Deutsch, Eliot. 1969. *Advaita Vedanta*. Honolulu: University of Hawaii Press.

Farah, Martha J. and Heberlein, Andrea S. 2007. "Personhood and Neuroscience: Naturalizing or Nihilating?". *The American Journal of Bioethics*, 7(1): 37–48

Gallagher, Shaun; Morgan, Ben; and Rokotnitz, Naomi. 2018. "Relational Authenticity". In *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, 126–145. New York: Oxford University Press.

Lau Tzu (600 B.C.) 1963. *Tao Te Ching*. Translated by D.C. Lau. London: Penguin Books.

Nadelhoffer, Thomas and Wright, Jennifer Cole. 2018. "Humility, Free Will Beliefs, and Existential Angst: How We Got from a Preliminary Investigation to a Cautionary Tale". In *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, 269–297. New York: Oxford University Press.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?". In *Consciousness and the Mind-Body Problem: A Reader*, edited by Torin Alter and Robert J. Howell, 15–23. New York: Oxford University Press.

Pasternak, Charles. 2007. "Curiosity and Quest". In *What Makes Us Human?*, edited by Charles Pasternak, 114-132. Oxford: Oneworld Publications.

The Possession of Hannah Grace (2018). Written by Sieve, B. Directed by Van Rooijen, D.

compos mentis

Birth, Natal Anxiety, and Possibility

Andrew Lee
Haverford College

ABSTRACT
In Heidegger's philosophy and especially outlined in "Being and Time," death delinates the possible for Dasein. Once Dasein understands the unavoidable nature of death, it becomes a freedom toward death whose possibilities are given meaning. However, it is a mistake to look only at one end of Dasein, its death, and not also give attention to the birth of Dasein. Just as there is an anxiety about death, there is an anxiety over one's birth, that one was born at all, and that there was once a time before one was born. Heidegger mentions birth in "Being and Time" but it does not prove to be much at all. Philosopher Anne O'Byrne's work gives us some resources to discuss Dasein's birth. Ultimately, I want to argue that Dasein's birth is individualizing as Dasein's death.

KEYWORDS
Heidegger, Natality, Dasein, Death, Anxiety, Anne O'Byrne, Being and Time, Freedom Towards Death

Martin Heidegger's chief philosophical project is an investigation of the meaning of Being. He believes that it is of utmost importance as it is the grounding for all studies, from the ontical sciences to other areas of philosophical inquiry. This question is often framed as "Why is there something rather than nothing?" (O'Byrne 2010, 26). In Being and Time, he frames Dasein as the focus of ontological or metaphysical study. This is because Dasein is an entity who is concerned by Being. Any insight into Dasein would give an insight into the meaning of Being itself. Heidegger moves onto an exploration of Dasein's death so that he can understand the whole of Dasein. Death emerges as a possibility that is unavoidable and unique to every Dasein. And once Dasein realizes that death is its ultimate fate, it becomes a freedom towards death whose possibilities are given meaning by death. However, it would be a mistake to focus only on one end of Dasein without giving an adequate analysis of the other end of Dasein. While Heidegger gives a cursory glance over Dasein's birth, Anne O'Byrne gives a much more thorough account of birth in the chapter entitled "Historicity and the Metaphysics of Existence: Heidegger" in her book Natality and Finitude. In this chapter, we are introduced to the concept of natal anxiety, an analogue of Heidegger's anxiety, in that it is a realization of Dasein's orientation around its birth. Dasein can get lost in the question of why it was born in the first place, just as easily as it can when it faces the reality that it will one day die, but once it realizes this lack of reason behind why it was born, it can then create possibilities. O'Byrne ultimately gives a more compelling and richer account of birth than Heidegger does and her concept of natal anxiety seems more pertinent to answering the question of Being than Heidegger's anxiety about death. However, while O'Byrne seems to get much correct about how we ought to think about the birth of Dasein, she is mistaken in saying that birth does not individualize Dasein.

Heidegger introduces Dasein as the object of study to understand the meaning of Being. Dasein is differentiated from all other living beings in that it is inherently concerned not only about the question of Being, but also the question of its own Being. He says, "Rather it is ontically distinguished by the fact that, in its very Being, that Being is an issue for it" (Heidegger [1927] 1962, 32). When he says Dasein is ontically distinguished, this is to say there is something factically, within its existence as a physical being, that makes Dasein ontological or concerned with its Being. He then asserts that Dasein is aware of its ontologial character because of its ontical character. He says, "Dasein always understands itself in terms of its

existence--in terms of possibility of itself: to be itself or not itself" (Heidegger [1927] 1962, 33). Here, Heidegger hints at how Dasein's possibilities are indeed made possible by its existence.

Heidegger believes that the whole of Dasein ought to be studied to understand the meaning of Being. The whole of Dasein implies an investigation of the ends of Dasein and one end of Dasein is death. For Heidegger, death is the ultimate possibility for Dasein that enables all other possibilities. This is because, even though Heidegger asserts Dasein as being-with-others in that in any possible action, another Dasein can perform the same action, no Dasein can die for another. He says, "Thus death reveals itself as that possibility which is one's ownmost, which is non-relational, and which is not to be outstripped" (Heidegger [1927] 1962, 294). Here, Heidegger establishes the characteristics of Dasein's death. Death is non-relational in that while we can imagine another Dasein in our place as we perform any other action, another Dasein cannot die for us. This non-relational characteristic thus individualizes Dasein (Heidegger, 308). Death can also not be "outstripped" in that it is inescapable. Dasein cannot choose to not die and no one can take death away from it. Finally, death is Dasein's "ownmost" in that it is only possible for that particular Dasein (Heidegger [1927] 1962, 307).

Heidegger argues that as death emerges as Dasein's inescapable and unique possibility, it ought to embrace it. He calls this proper orientation towards death "anticipation" (Heidegger [1927] 1962, 306). Dasein should not flee from death as it will never be able to do so. Dasein should instead embrace death as the only sure possibility in its life. In a word, it should anticipate its inevitable death. Only then can Dasein achieve a "freedom towards death" (Heidegger [1927] 1962, 311). What does anticipation and a freedom towards death do for Dasein? Heidegger argues that not only is death the ultimate possibility of Dasein, it makes all other possibilities meaningful for Dasein. He says, "Only by the anticipation of death is every accidental and 'provisional' possibility driven out" (Heidegger [1927] 1962, 435). This is not to say that death makes all other potentiality possible for Dasein. Rather, out of all the possibilities in front of Dasein, death shines a light on the possibilities worth pursuing. The possibilities that remain after this filtering effect death has proven to be meaningful because it is what Dasein chooses to do in its finite existence.

This conception of death as the utmost possibility of Dasein does work for Heidegger in terms of understanding Dasein's temporality and its Being. However,

there ought to be more investigation in whether he privileges one end of Dasein more than the other. Death is one end of the whole of Dasein, but it would be a mistake to think of this end as the death of Dasein, the end of its factical existence. When we draw a line segment from point A to point B, we can say that point B is one end of the line segment. But it would also be correct to say that point A is the other end of line. To bring this analogy to Dasein's factical existence, point A would be Dasein's birth. Heidegger himself addresses Dasein's birth. He says, "Understood existentially, birth is not and never is something past in the sense of something no longer present-at-hand… Factical Dasein exists as born; and as born, it is already dying, in the sense of Being-towards-death" (Heidegger [1927] 1962, 426). Here, Heidegger introduces the concept of birth to illustrate historicity and thrownness. We should understand birth as thrownness because it parallels birth in the following quote: "Thrownness and that Being towards death in which one either flees it or anticipates it, form a unity; and in this unity birth and death are 'connected' in a manner characteristic of Dasein (Heidegger [1927] 1962, 426–427). Here, "thrownness" parallels "birth" and "Being towards death" parallels "death". Every present moment of Dasein is shot through with its past and stands before its future. Dasein's past is thrown into every moment and certainly, its birth is also in its path. Thrownness is an essential characteristic for Dasein to have possibilities that will then be made meaningful by its death.

But is Heidegger correct in this characterization of Dasein's birth as making way for Being towards death? In this conception and just in the amount of pages he dedicates to his discussions of birth and to his discussion of death, birth seems to less important than death. O'Byrne gives us resources to develop the birth of Dasein. She first establishes birth as the source of Dasein's thrownness and then "natality--the condition of our having been born--appears as that thrownness" (O'Byrne 2010, 16). The concept of natality then seems to be opposed to Heidegger's concepts of anxiety and anticipation. Anxiety and anticipation are Dasein's state of minds as a result of death standing before it as a possibility. They are a result of Dasein's facticity which effect its ontology as the fact Dasein will die enables it to question its Being and then makes its possibilities meaningful. As she terms the state of mind of Dasein that is a result of its birth, O'Byrne prepares us for importance of Dasein's birth for its ontology. So when O'Byrne then takes issue with Heidegger's framework, as outlined above, how birth is ultimately incorporated into the picture of Dasein as a Being towards death

(O'Byrne 2010, 17), she wants to highlight how birth has a power, independent of death, on Dasein. She says, "our being thrown into a world is overshadowed by our thrownness toward death" (O'Byrne 2010, 17). She will argue that the impact of our thrownness is discounted by Heidegger. It should be given a thorough exploration.

O'Byrne gives a characterization of a "natal anxiety" (O'Byrne 2010, 26) analogous to Heidegger's concept of anxiety. It is precisely this state of mind that is constituted by Dasein's having been thrown. She says, "Natal anxiety is the experience of the groundlessness of our finite existence. It is one thing for Dasein to grasp that it will one day die but another for it to understand that it once came into existence… It is the difference between realizing that Dasein's existence is limited and realizing that it might never have existed at all" (O'Byrne 2010, 26). Heidegger's anxious Dasein is concerned by its inevitable and inescapable death. It is a realization that its "existence is limited" and this limitation makes its possibilities meaningful. However, natal anxiety points us in the other direction. Not that Dasein will one day die, but that it might not have the chance to die in the first place. The fact that Dasein had to first come into existence opens up the idea of a world without this particular Dasein and then the scenario where it "might never have existed at all". There is the questioning of why this Dasein came into being. O'Byrne says this question can be asked as "'Why was I born?'" which she notes is just the "existential version" of one of Heidegger's formulations of the question of Being, "'Why is there something rather than nothing?'" (O'Byrne 2010, 26). With the question "Why was I born" so closely resembling the question of Being, and this question stemming from natal anxiety, then it seems as if an exploration of birth rather than death would have served Heidegger better.

Anxiety of death led Heidegger to the concept of freedom towards death and the limiting of possibilities. Where does natal anxiety lead us? O'Byrne says, "being-toward-death might drive us to a project, to have or concoct possibilities for ourselves, but it is our being in the world--a world that was there in all the variety and complexity of its being and having been before we came--that is the wellspring of those possibilities" (O'Byrne 2010, 32). Now, this idea that birth opens up an infinite range of possibilities does not seem at all different from what Heidegger had to say about the birth in Being and Time. But O'Byrne wants to argue that not only does birth give us the range of possibilities, it also lets us choose. She reappropriates the "moment of vision" that "Heidegger describes

the moment when Dasein pulls itself back from falling" (O'Byrne 2010, 33). The moment of vision for Dasein is when it realizes death as its utmost possibility and becomes a freedom towards death. This is the product of Heidegger's anxiety. The result of natal anxiety would be a moment of vision also, where instead, Dasein pulls away from the fact that it once did not exist and that it might not have existed at all. She says, "the moment of vision is the moment of openness in which newness becomes possible. It makes it possible for us each to be born, for there to be new beginnings, for each of us to act" (O'Byrne 2010, 33). Dasein in the moment of vision does not become consumed by the wonder of the question of why it was born, but instead, realizes that things can come into being. There is the possibility for things to be "new". When we are born, we come into being and are new in the world. Dasein can then act as it realizes it has the ability to make "new beginnings" for itself.

O'Byrne then illustrates how in birth Dasein is with others, while in death, Dasein is separated with others. When Dasein chooses and acts on its possibilities, with its creative power of making possibilities into new actualities, it "disrupts" the world (O'Byrne 2010, 34). Dasein's birth and its actions disrupt the world because it is born into a world that has been given meaning by others (O'Byrne 2010, 34). While disrupt may seem like a negative connotation, it seems to just mean the emergence of new things which were previously not there. Birth and natal anxiety then contribute to the idea of Dasein as Being-with-Other in that it emphasizes when Dasein historicizes, it does so in a world, old with already present meaning. She also says, "Death may Dasein's ownmost non-relational possibility, separating Dasein from all others, but birth is precisely what puts us in relation with others since, while we each may die alone, we could not have been alone at birth" (O'Byrne 2010, 35). Her argument for birth over death is that birth empathizes Dasein as Being-with-Others. Death individualizes and separates Dasein from all other Dasein because no other Dasein can die for another. In that sense, Dasein dies alone. However, O'Byrne argues that birth has the opposite effect. When Dasein is born, it is immediately thrown into a world with meaning constituted by others. Furthermore, we get the sense that Dasein cannot have been factically born alone while we can indeed imagine Dasein dying alone, away from other Dasein. A mother, another Dasein, gives birth to Dasein. It is thus impossible for a Dasein to be alone at birth.

We have examined opposing illustrations of anxiety and what makes possibilities meaningful to Dasein. Heidegger's concept of anxiety makes death a sort of delineating power; that Dasein has a wide range of possibilities in front of it and that death helps Dasein see which among them is important. Natal anxiety gives Dasein a creative power. Dasein can choose from among the possibilities that come from its birth and turn the possibility into an actuality, something new in the world that other Dasein can interact with. Firstly, it seems very much that these two conceptions of anxiety are compatible. We can imagine death delineating what possibilities are before Dasein and then the creative power of natality allowing us to then actually act on these possibilities. But Heidegger set off on his exploration of Dasein's death to get an understanding of the whole of Dasein, so that he can then learn something about the meaning of Being. We can then think about whether birth or death is more helpful in answering the question of Being. When the question of Being is phrased as 'Why is there something rather than nothing?', then it seems to me that birth is the more helpful conception. We can also think about this question as "Why did something come to be rather not?" because something cannot be there unless there was an initial becoming. This strengthens the appeal of natal anxiety as answering the question of Being. Dasein's birth has the same mysterious question of why it was born rather than not being born. This is because birth is the threshold for Dasein of it having not existed and then it existing. Furthermore, the creative power of natal anxiety that gives Dasein the ability to act on our possibilities is based on a sort of understanding of the question of Being. There is something rather than nothing, things can come into being and be new things in the world, and we as Dasein can do this exact thing, create new meaning by acting our possibilities. This new meaning also builds on an old world of meaning that Dasein is born into. While this may be possible in Heidegger, it is made explicit in O'Byrne. Furthermore, birth has a priority over death. This may be obvious because one's birth comes before one's death. But before an anticipation towards death can delineate what is meaningful for Dasein, there must first be a wealth of possibility that stands before Dasein. Heidegger himself mentions that birth gives Dasein this infinite possibility, but as O'Byrne argues, it seems as if he prioritizes death to the extent that birth is swept away and made less significant.

However, while O'Byrne's conception of birth, natal anxiety, and how it relates to the question of Being is compelling, there is some worry about Dasein

as Being-with-Others at birth. The argument is that while death individualates Dasein, Dasein cannot be alone at birth due to its coming into a world made meaningful by others and also that a mother always birthed Dasein (O'Byrne 2010, 35). However, the argument for why death individualates Dasein can also show that birth individualates Dasein. Indeed, this is how it ought to be thought. As said many times above, Dasein's death is non-relational in that it "must be taken over by Dasein alone" (Heidegger [1927] 1962, 308). This is to say that a particular Dasein must face its death and cannot escape it. One Dasein cannot die for another. If a Dasein sacrifices its life for another, that Dasein will still die at some point in the future. This same argument can be made for Dasein's birth. Just as no one can die for me, no one else can be born for me. Dasein's birth must be unique because if another Dasein would be born in my place, there would be no 'me' to begin with, only this other Dasein. Every Dasein was born and that birth cannot be taken away from it. So, while O'Byrne's argument seems to be correct in that we are indeed born into a world of meaning constituted by others and thus we are with others in that sense, Dasein's birth remains its own in that no one else could have been born for it.

Heidegger ultimately turns to Dasein to understand the meaning of Being. To achieve a full picture of the meaning of Being, he would need a full picture of Dasein. He investigates one end of Dasein, its death, and establishes death as personal and inescapable for Dasein. Death then becomes critical for Dasein as it gives its possibilities meaning. While Heidegger briefly explores the other end of Dasein, its birth, it is only to establish the limitless possibilities that stand before Dasein that will then be filtered when Dasein becomes a freedom towards death. O'Byrne gives a more thorough picture of birth. She establishes birth as the source of Dasein's thrownness and rather than have birth be consumed in an anxiety about death, she appropriately characterizes an anxiety about one's birth. The anxiety about birth is the question of the fact that Dasein once did not exist and that it might not have come into existence at all. Rather than fall before this anxiety, Dasein learns that things can come into being and that it itself has a creative power. While O'Byrne gives a compelling account about Dasein's birth and how an anxiety about birth shines more insight onto the question of Being, she is wrong about how birth does not individualate Dasein like how death individualates Dasein. We may be born into a world of meaning constituted by

others, but no one can be born for us. My birth is my own and that constitutes my Being.

REFERENCES

Heidegger, Martin. (1927) 1962. *Being and Time*, trans. by John Macquarrie and Edward Robinson. San Francisco: Harper & Row.

O'Byrne, Anne. 2010. "Historicity and the Metaphysics of Existence: Heidegger", *Natality and Finitude*. Bloomington: Indiana University Press. 15–45.

compos mentis

Semantics, 'Strong' AI, and the Chinese Room Argument

Ameer Sarwar
University of Toronto, St. George

ABSTRACT
The main purpose of this paper is to defend Searle's (1980) classic Chinese room argument against a number of objections. Searle takes his argument to show that semantics do not inhere in formal symbols. Consequently, since 'strong' AI concerns itself solely with the implementation of formal symbols over recursive syntactical rules, its inability to account for inherent meaning precludes it from being established as a viable research program in cognitive sciences. Two major strands of objections and sub-objections are reviewed against Searle's argument, but it is ultimately concluded that they both fail. The 'disjoint personalities objection' fails primarily because there can be no change in the personality of the room without a change in the personality of the inhabiting symbol manipulator. The 'other minds objection' fails because it engages in reverse causality: it concludes from manifestations of intelligent behaviour that the thing behaving intelligently is thereby intelligent. My attempts at demonstrating the failure of the two objections rescue Searle's argument, and therefore, the problem of original meaning remains a thorn in the philosophical foundations of 'strong' AI.

KEYWORDS
Chinese Room Argument, 'Strong' AI, Turing Test, Semantics, Formal Systems, Joint Personalities, Other Minds, Original Meaning

INTRODUCTION

This essay attempts to establish that Searle is correct in arguing that semantics do not inhere in formal symbols, and so, there is no inherent understanding in formal computational systems, thereby brining into serious doubts the prospects of 'strong' AI as a research paradigm. I begin the paper by explaining what I mean by 'strong' AI, what the Chinese room argument is, and how the latter causes problems for the former. I then consider a series of objections—which are divided into the 'Disjoint Personalities Objection' and 'Other Minds Objection'—and try to diffuse them all in order to ultimately conclude that Searle's argument survives. The implication is that the prospects of 'strong' AI become doubtful; specifically, it cannot explain the problem of original meaning in terms of implementing formal programs.

"STRONG AI" AND THE CHINESE ROOM ARGUMENT

'Strong' artificial intelligence (AI) is the idea that an instantiation of a formal program is an instance of genuine intelligence. A formal program refers to a string of purely abstract symbols or tokens that are syntactically manipulated in a recursive or iterative fashion. The individuation of symbols occurs not via their immanent semantics but based on their orthographic characteristics or the functional roles they play within a formal system (Rescorla 2017). Hence, if a physical computer or a Turing machine can implement a formal system, then the implementation is taken to be a case of bona fide intelligence. Importantly, the physical substrate implementing the formal system is seen as secondary, if not altogether irrelevant. As long as the physical system is sufficiently complex to implement a given formal system, the latter can be realized by the former. Accordingly, formal systems are multiple realizable in a multitude of appropriate1 physical systems, and therefore, understanding the brain is secondary to understanding the computational mechanisms it implements. Proponents of this view think that implementations of formal systems constitute duplications, not mere simulations ('weak' AI), of real intelligence.

1. It is often argued that for some physical system to be an 'appropriate' realizer of a formal system, the physical system must play the same causal roles (i.e., have the same initial states, undergo the same state transitions, and produce the same output states) of the formal system in an isomorphic manner that is counterfactually-supported.

Alan Turing proposed in 1950 an 'imitation game' (or Turing test), which entails a human judge conversing with another human and a universal Turing machine.2 The judge's task is to correctly determine which of his interlocuters is a machine and which is a human. The probing conversation takes place over written text, so that voice, physical characteristics, gestures, and other non-linguistic elements do not tip the judge into finding the right answer. If the judge is fooled by the machine into thinking that it is a human, then the machine is said to have passed the test; it has successfully managed to imitate a human (Turing 2009). More often than not, the proponents of 'strong' AI take the passing of the Turing test as good reasons for believing in the existence of genuine machine intelligence (see, Oppy and Dowe 2019, for details).

Let me now explain the (in)famous Chinese room argument (cf. Cole 2019), followed by an explanation of its implications for 'strong' AI. John Searle (1980) invites us to imagine a monolingual Englishman situated in a room. He receives through an input slot a piece of text that is foreign to him; he then receives another piece of foreign text. Later, he finds a set of instructions, written in English language, that he can comprehend. These instructions tell him to place, say, symbols x after a, y after b, and so on. He diligently follows these instructions to string together sets of intricate symbols whose meanings he does not understand. Finally, the instructions tell him to insert in the output slot set p after set q and so forth. This is, in cursory terms, the experience of the man in the room: he is reading a book in English that tells him how to identify (based on physical characteristics) certain symbols and where, in the strings of symbols, each one belongs and when it is appropriate to put each symbol-string in the output slot. Despite becoming adept at following instructions, the Englishman has no idea about the meanings of the symbols.

Unbeknownst to him, the symbols actually belong to the Chinese language. Outside the room, there are native speakers of Chinese that are inserting the first batch of symbols, which may be understood as a story written in Chinese. Then,

2. Two results from mathematical logic lead to universal Turing machines. First is the Church-Turing thesis, which states that for any possible algorithm, there exists a Turing machine that can, at least in principle, implement it. Second, the Turing thesis states that a universal Turing machine can imitate any given Turing machine. Hence, a universal Turing machine is able to implement, at least in principle, any and all possible algorithms (Searle 1990). A universal Turing machine, then, is a good candidate for possessing domain-general cognitive capacities, because it can simultaneously imitate a number of different Turing machines with domain-specific capabilities.

they insert a second batch of symbols that is analogous to asking questions about the story they initially presented. The Englishman then manipulates the symbols in line with the English instructions, which are analogous to the abstract program the man is implementing. The strings of symbols, whose meaning I should emphasize he does not understand, that he places in the output slots of the room are interpreted by the outside native Chinese speakers as answers to the questions. Given how proficient the Englishman had become at manipulating symbols, from the perspective of the native speakers whatever is answering the questions inside the Chinese room 'black-box' understands Chinese very well.

However, as the experimental set-up makes clear, the man has no understanding of Chinese whatsoever. For him, the symbols may have belonged to Japanese, Dutch, C++, or no language at all. The important point that Searle takes the experiment to show is that since semantics do not inhere in the symbols, simply implementing a computational program, which is nothing more than a set of formally defined symbols that are syntactically manipulated, is not sufficient for understanding. The man clearly does not understand Chinese even though he can effectively perform syntactical manipulations with such adroitness that even the natives think that the 'processing' in the black-box (the Chinese room) is of the nature that there is understanding of the Chinese language.

The conclusion of the Chinese room argument—namely, that semantics do not inhere in symbols, which are defined formally and manipulated syntactically—is used as premise three in the following overarching argument that Searle (1984) makes against 'strong' AI: (P1) programs are defined purely formally or syntactically; (P2) human minds have mental content or semantics; (P3) syntax by itself is neither constitutive of nor sufficient for semantic content (Chinese room argument); (P4) so, programs are neither constitutive of nor sufficient for semantics; (P5) universal Turing machines implement abstract programs that are purely syntactical; (P6) thus, there are no inherent semantics in computers; (C) thus, 'strong' AI is not an instance of genuine cognition.3 As this deduction shows, the crucial premise in the argument is (P3), which is established by the Chinese room argument. I think it is imperative for the proponents of 'strong' AI to refute this premise in order to have a philosophically sound basis for procuring a computational research paradigm in cognitive science. Debates around the soundness of Searle's (1980)

3. The underlying, though plausible, assumption is that genuine cognitive agents have original intentionality. See the section on 'Other Minds Objection' for details.

argument will be the focus of the rest of this paper. I will review some major attempts at refuting (P3), and I shall respond on Searle's behalf to show that they all fail, thus preserving the argument laid out in this paragraph to conclude that 'strong' AI chronically suffers from the problem of original meaning.

DISJOINT PERSONALITIES OBJECTION

Searle's (1980) original paper anticipates a 'Systems Reply' according to which the room as a system understands Chinese even though the Englishman as its constituent does not. Searle simply replies that if the person memorizes the rule-book and all the information necessary and sufficient to effectively manipulate tokens, he would still have no understanding of Chinese even though he would have in his mind everything that the 'system' also has. Some interesting modifications were later made to the 'Systems Reply,' and I review and respond to them below.

Cole (1991) considers a thought-experiment in which the Englishman inhabits a joint Chinese-Korean room. In the morning, he may be implementing the program such that the 'answers' he gives are provided to Chinese speakers, and in the afternoon, he may be running the program in a way that the 'answers' are given to Korean speakers. Again, the Englishman is ignorant of the meanings of the symbols he is manipulating in accordance with a rule-book; what is more is that he does not know that he is manipulating two different types of linguistic tokens at different times of the day. Now, suppose that the 'answers' given to the two types of speakers display completely different psychological profiles: in the one case, the profile may be very amicable and polite, while in the other case, it may be aggressive and hostile. (Also suppose that the answers are given in such a way that an onlooker is convinced that the black-box does not understand a language other than that of the onlooker, e.g., by denying knowledge of the other language.) Suppose also that the Chinese and Koreans who attend this, say, 'festival' of sorts converse with each other later at night; they talk about their experiences of visiting the room and the attitude displayed by the 'answers' from the room. The behavioural evidence available to the speakers of both languages is markedly different, and so they conclude that there are two non-identical minds in the room. Since these minds have mutually exclusive psychological properties, they "cannot be identical [with each other], and ipso facto, [they] cannot be identical with the mind of the implementer in the room" (Cole 2019,

§ 4.1.1). Maudlin similarly observes that "Searle has done nothing to discount the possibility of simultaneously existing disjoint mentalities [that are different from each other and from that of the syntactical manipulator]" (1989, 414-15). This argument shows that since there can be psychological personalities different from that of the token manipulator, there is something in the room or the system as a whole that is not entailed by the psychological make-up of the Englishman inhabiting it. Accordingly, Searle is premature in asserting that the man's inability to understand Chinese constitutes that there is no understanding of Chinese.

I have three responses to this argument (presented in increasing degree of strength). The first is that the man in the room is manipulating symbols that he still has no understanding of. Instead of using tokens that were only in Chinese, he is now manipulating Korean tokens as well. This, no doubt, will produce different understandings in the native speakers who independently observe the room from the outside, but Searle's original claim that the man understands nothing still stands. It is similarly pertinent to observe that, rhetorically speaking, there is a certain element of magic associated with understanding being created in the "system as a whole." I am not sure where such understanding inheres if not in the mind of the only conscious and intentional entity present in the room (the Englishman).

The second reply I have asks the reader to imagine a trilingual man capable of speaking English, Chinese, and Korean. He meets the Chinese speakers in the morning and the Korean speakers in the afternoon, and just like the man in the joint-room, he exhibits (for whatever reason) different personality traits to the two types of speakers. When the Chinese and Koreans talk at night about meeting an Englishman that day, they do not think that they are talking about the same person (for how one person can be amicable in the morning and bellicose in the afternoon is difficult to comprehend) but about two people with different personalities. Given that they can erroneously and unknowingly think of the same person as having two different personalities, it does not follow that the Englishman indeed has two distinct personalities. If we are to consistently apply Cole's and Mauldin's arguments, it follows that the two personalities exhibited by the person are different from who he is. This strikes me as prima facie absurd to say that his two different personalities are not his own simply because the onlookers thought so (admittedly, based on evidence). Hence, I am inclined to reject their arguments.

My third reply rests on a distinction between a stronger and a weaker version of the Chinese room argument. The stronger version, which Searle proposes, maintains that understanding or personality of the room is constitutive of or identical with that of the symbol manipulator; understanding or personality of the room is nothing 'over and above' what the man in the room understands or exhibits, respectively.4 The weaker version, which we can consider here for the sake of argument, does not maintain a relationship of constitution or identity but that of supervenience. On this account, understanding in the room supervenes on the understanding of the Englishman. So, while the room may have a change in its understanding only if there is a change in man's understanding, it thereby does not mean that the two understandings are identical. So, the joint-room may exhibit personalities that are numerically non-identical from yet causally dependent on that of the person. Despite maintaining the weaker relationship of supervenience, it cannot be shown that the joint room has had any change in understanding or personality without a corresponding change in the Englishman. For the supervenience relationship to work, the lower-order organization (the symbol manipulator) must change in order for it to cause a change in the higher-order organization (the room as a whole) (McLaughlin and Bennett 2018)5; yet, even in the weaker version, the man not understanding the languages shows the untenability of Cole's and Mauldin's critiques, namely the room cannot have a change in its understanding or personality without a change in these characteristics with respect to the Englishman, thus vindicating Searle's arguments against these attacks.

OTHER MINDS OBJECTION

This objection essentially states that because we rely on behavioral information to attribute mentality or cognition to other people and animals, the native speakers observing the Chinese room or the human judge conversing with the Turing machine should likewise ascribe mentality to them due to their access to only the behavioral information. If we are to apply our epistemology

4. Indeed, this is the crux of Searle's (1980) response to the 'Systems Reply.'

5. A classical example is that of a painting, which has aesthetic properties organized at a higher-order and physical properties organized at a lower-order. The former properties supervene on the latter properties, so any change in the aesthetic qualities of a painting is brought about only through a change in its physical constituents.

consistently, the argument goes, then all entities, whether humans or machines, should be treated in the same manner—knowledge based on which we consider other humans as mental should similarly be sufficient to deem the Chinese room as cognitive. Otherwise, we are engaged in anthropocentric chauvinism.

My response to this objection is that there is first and foremost a difference in the degree/ amount of behavioral information available to the human being in charge of ascribing mentality. In the cases of the Chinese room and Turing test, the information that is available to us is in the form of language. While language is no doubt an important part of cognition, it should not be identified with it. When we ascribe mentality to other humans, however, we implicitly rely on a whole range of behavioral information at our disposal, including language, gestures, eye gazes, facial expressions, intonations, and so forth. The manifold data make it far easier to think that other people are conscious, but the machines may be pre-programmed to spit out certain linguistic phrases in light of questions. There is simply not a sufficient amount of behavioral evidence available that can justify the ascription of mentality to the Chinese room or the Turing machine.

The interlocuter may, however, rightly protest that it is 'just' a matter of time or technological advancement that we will create robots that are capable of producing behavioral outputs that are just as complex as those of humans. Films on artificial intelligence like Ex Machina already exploit our intuitions in this respect. So, a philosophical argument must rely on a difference in principle, not on a difference in degree, to explain why we cannot, assuming we have equal amounts of behavioral evidence for robots and humans, ascribe mentality to both. Furthermore, this argument about applying epistemology consistently can be extended from the pragmatic to the scientific domain. Dennett (1997) has argued that in the Turing test we should utilize what he calls the 'quick-probe' assumption. The idea is that since a machine must choose from a number of different possible responses that may be given to questions of an interrogative judge, it cannot utilize brute-force computation, because at each linguistic answer-node, a number of other topics are opened up that may need to be addressed, leading to combinatorial explosion. Thus, he argues that the act of providing an intelligent response without brute-force computation is good reason for thinking that the Turing machine has some kind of cognitive capacities. And from here it is not unreasonable to generalize that the machine may also be capable of exhibiting other mental abilities; this assumption, then, is used to quickly probe the mental

capacities of the Turing machine. Likewise, the native Chinese speakers looking at the room from the outside may construe it as a Turing machine answering questions; there is no reason for thinking that the room has no understanding, because the linguistic (or behavioral) outputs of the room are no different from those of other people. If our epistemology is applied consistently, then the Chinese room has internal meaning just as other people do (or else, other people are not cognitive either!).

I respond to this objection by first pointing out that there is a conflation between nomological or metaphysical facts with epistemological facts. When we are concerned with comprehending whether a system truly has understanding, we are interested in uncovering the mechanisms of its internal processing (Block 1981). The notion of ascribing mentality to a system as opposed to discovering it as inhering within it are two different things—only a behaviorist would be content with thinking that behavioral dispositions are all there is to having cognition. It is quite possible for things to behave intelligently (a parrot mimicking human language) without actually being intelligent (a parrot not understanding the words it uses). I suspect there is a fallacy of reverse causality underlying the interlocuters' claims: from the fact that intelligent things behave intelligently it does not follow that all those things that behave intelligently are thereby intelligent. The causality here is unidirectional: only intelligent things behave intelligently, not vice versa.6

6. An analogy is that a disease should not be confused with its symptom. Surely, the removal of a disease leads to the removal of its symptoms, as the former is the cause of the latter. However, just because the symptoms are removed, one cannot think that the disease is also gone (even though the symptoms may be used as indicators for the existence of the disease). There is no biconditional in this case. One may represent this formally as a modus ponens argument. Let the disease (or cognition) be p and the symptom (or intelligent behaviour) be q.

$p \rightarrow q$
p
$\therefore q$
The interlocuters are committing the fallacy of 'affirming the consequent':
$p \rightarrow q$
q
$\therefore p$

(Curiously, one sees in this modus ponens argument that the same symbols, p and q, can at one time stand for disease and symptoms, and they can at a different time stand for intelligence and intelligent behaviour. The rules of inference are the same irrespective of which semantics are ascribed to the symbols. So, even this argument shows the correctness of Searle's observation, namely that the meaning does not inhere in the symbols; it is simply ascribed to them.)

One may at this point say, "well, how do we then know that something really is intelligent if not due to the manifestations of intelligent behavior?" This leads me to the key point of the argument. We know for certain that what is going on in the Chinese room is nothing more than manipulations of formal symbols; we also know that the outsiders have access only to the linguistic outputs which they interpret as being a result of genuine cognitive activity. Since we, as philosophers thinking about the Chinese room thought-experiment, are already aware of what is taking place in the room (i.e., we have access to the matters of facts of the room's internal processing), we have no reason to believe that there is any understanding taking place in the mind of the Englishman or in the room as a whole.

Unlike in the case of humans, where we try to discover the nomological facts about the workings of the brain through neuroscience etc., in the case of computational machines implementing programs we are already aware of the principles underlying their workings. So, there is no reason to think that in the machine there is anything 'over and above' what we already understand. Searle (1984) is correct in pointing out that in the sciences we presuppose the existence explicandum. Here, in trying to understand the basis of cognition, we presuppose that human minds exist just as physicists presuppose the existence of physical things. The 'other minds' objections seem to place the cart before the horse. We already know how the computational systems work; we do not know how the human minds work, and this is something that needs to be explained. There is no reason to place two things on the same epistemic grounding when the metaphysics of both are asymmetrically known.

CONCLUSION

I have tried to show that Searle's argument that semantics do not inhere in formal symbols successfully survives the objections I have considered herein. Consequently, the plausibility of his argument is threatening to the prospects of 'strong' AI as a viable research programme, because one of the features that is crucial to human minds—and, therefore, something that any scientific theory of the mind must explain—is that they have inherent meaning.

REFERENCES

Block, Ned. 1981. "Psychologism and Behaviorism." *The Philosophical Review* 90 (1): 5–43.

Cole, David. 1991. "Artificial Minds: Cam on Searle." *Australasian Journal of Philosophy* 69 (3): 329–33.

———. 2019. "The Chinese Room Argument." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University.

Dennett, D. C. 1997. "Can Machines Think? Deep Blue and Beyond." *Icca Journal* 20 (4): 215–23.

Maudlin, Tim. 1989. "Computation and Consciousness." *The Journal of Philosophy* 86 (8): 407–32.

McLaughlin, Brian, and Karen Bennett. 2018. "Supervenience." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2018. Metaphysics Research Lab, Stanford University.

Oppy, Graham, and David Dowe. 2019. "The Turing Test." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2019. Metaphysics Research Lab, Stanford University.

Rescorla, Michael. 2017. "The Computational Theory of Mind." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2017. Metaphysics Research Lab, Stanford University.

Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–24.

———. 1984. *Minds, Brains and Science*. Harvard University Press.

———. 1990. "Is the Brain a Digital Computer?" *Proceedings and Addresses of the American Philosophical Association* 64 (3): 21–37.

Turing, Alan M. 2009. "Computing Machinery and Intelligence." In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, edited by Robert Epstein, Gary Roberts, and Grace Beber, 23–65. Dordrecht: Springer Netherlands.

compos mentis

Self-Deception

August Smith
Wheaton College

ABSTRACT
In this essay I examine the possibility of bona-fide self-deception considered as an intellectual vice. I first briefly survey traditional construals of self-deception, concluding with Donald Davidson that self-deception necessarily consists of a set of contradictory beliefs that exist simultaneously in the mind through the presence of a sort of incorporeal "cognitive barrier." Like every vice, self-deception stems from epistemic pride due to a vicious distortion of one's pursuit of truth through deficient desires arising from a malformed will. This habitual failure of willpower often causes one to reject evidence that rebuts a view that they previously held or wished to be true—such repetitive, vicious ignorance produces this "cognitive barrier" that gives rise to self-deception. One's will might arrive at this deficient state through an inordinate desire for comfort or attempt to avoid the acknowledgement of an unpleasant truth, such as a cheating spouse or discreet alcoholism. Self-deception is opposed not only to the virtue of self-knowledge (as is prima facie true), but also strength of the will insofar as one's will is directly involved in a conscious avoidance of a self-deceptive cognitive state. Furthermore, while the vice of self-deception is inexorably related to vices of wishful thinking and willful naïveté, it is important to note that it is qualitatively distinct. Finally, self-deception does seem to come in degrees, and can become permanent through repeated failure of self-examination and exacerbation of a weak will. Ultimately, though all people are prone to self-deception, it can be avoided through a cultivation of a strong will and self-examination with the aid of others.

KEYWORDS
Epistemology, Self-deception, Virtue Theory, Vice, Cognitive Barrier

In this paper I will discuss the intellectual vice of self-deception, as well as its corresponding causal ancestry, characteristics, and methods of avoidance. When it comes to intellectual vices, it doesn't get much trickier than self-deception. For one, epistemologists do not agree on the necessary and sufficient conditions for legitimate cases of self-deception, nor do all even admit that there are instances in which bona fide self-deception is possible. For the purposes of this essay, I will assume that it not only possible (for reasons which I will explicate below) but rife in both past generations and contemporary society. John, in his first epistle, agrees. The Biblical author chides his audience: "If we claim to be without sin, we deceive ourselves and the truth is not in us" (1 Jn. 1:8). He argues that all people are influenced by the corruption of sin—thus, to believe that we are unaffected by such is to deceive ourselves, and to therefore be without the truth. Ultimately, like all intellectual vices, self-deception is a deficiency of intellectual virtue, consequently depriving one of some epistemic good; therefore, one should strive to avoid self-deception at all costs to promote a virtuous and flourishing intellectual life.

IS SELF-DECEPTION POSSIBLE?

The first task is to define self-deception. This duty, however, proves more difficult than initially expected. Self-deception is often popularly construed as simply "lying to oneself," yet this definition presents a problem: "to be self-deceived one must at some time have known the truth, or, to be more accurate, have believed something contrary to the belief engendered by the deception" (Davidson 2010, 4). To lie involves both believing a true proposition and expressing a proposition that is contrary to this true one with the intention to garner another's belief in such falsity. However, in the case of "lying to oneself," there is a clear inconsistency; how can one both know the truth of a proposition and simultaneously be self-engendered to believe a contrary proposition?

According to a more traditional unified and integrated sense of the self, this inconsistency proves fatal to the possibility of self-deception, as Amélie Rorty notes: "If the self is essentially unified or at least strongly integrated, capable of critical, truth-oriented reflection, with its various functions in principle accessible to, and corrigible by, one another, it *cannot* deceive itself" (McLaughlin 1998, 13, my emphasis). This internal, rationalistic coherentist picture of the self therefore disallows the prospect of legitimate self-deception. On the other hand, some conceive the noetic structure as a simple amalgamation of atomic

beliefs that "can be individually added, changed, and deleted without regard to their propositional environment" (Davidson 2010, 5). This model allows and even encourages intellectual inconsistency. However, this conception of the mind seems somewhat inept; the self is ostensibly more than a simple conglomeration of atomic beliefs.[1] Thus, a more appropriate way to conceive of the notion of self-deception is needed.

Others have postulated that self-deception be thought of in solely self-actualizing moral terms, as any sort of analysis of beliefs is entirely unhelpful. "[…] Self-deception cannot be eliminated by checking whether there are contradictions in my beliefs, for even if I were successful in that endeavor I may yet deceive myself, for the interest with which I approach these beliefs and this endeavor might be self-deceptive" (Strandberg 2015, 49). While I admit that we may deceive ourselves in intellectual analysis, if we carefully examine held beliefs, we may nonetheless come to realize present inconsistency. The simple fact that self-deception itself may often hinder our attempts to thwart the vice does not prevent the essence of self-deception from essentially consisting in an incompatible set of beliefs. Therefore, for the purposes of this paper, I will refer to self-deception as a case in which two or more inconsistent beliefs are held simultaneously as result of physiological, cognitive, or as in our case, habitual deficiency. In other words, self-deception consists of a "contradiction of beliefs."

DEFINING SELF-DECEPTION

It remains to be answered, however, how two or more conflicting beliefs may be held simultaneously. Donald Davidson admits that this problem is significant but not insurmountable. He begins by asserting that it is indeed impossible for an intellectual agent to believe (p and not-p). For example, it is entirely incoherent for someone to say that they both believe in God and do not believe in God. Yet, on Davidson's account, it is *not* impossible for an intellectual agent to believe (p) and (not-p) simultaneously (2010, 198). To explain this apparent contradiction, Davidson states "that people can and do sometimes keep closely related but opposed beliefs apart. To this extent we must accept the idea that there can be boundaries between parts of the mind" (211). He quickly reminds his readers that these are not literal psychological barriers, but rather conceptual objects to aid understanding.

1. See Dupuy 1998 for an extensive discussion of this issue.

How does this barrier affect the right functioning of the mind and promote self-deception? On this point it is worth quoting Davidson at length:

> We should not think of the boundaries as defining permanent and separate territories. Contradictory beliefs [about a correlated subject] must each belong to a vast and identical network of beliefs about [this subject] and related matters if they are to be contradictory. Although they must belong to strongly overlapping territories, the contradictory beliefs do not belong to the same territory; to erase the line between them would destroy one of the beliefs. I see no obvious reason to suppose one of the territories must be closed to consciousness, whatever exactly that means, but it is clear that the agent cannot survey the whole without erasing the boundaries. (Davidson 2010, 211)

This model provides a helpful solution to the problem; an intellectual agent can hold belief (p) and simultaneously hold (not-p) if the two beliefs are part of an intricately related web of beliefs, but do not occupy identical cognitive "areas." This allows for a sort of intellectual barrier to arise (for reasons discussed below) which inhibits conscious cognitive contact between the two contrary beliefs, an event that would mandate the destruction of one. An agent whose mind has developed or does develop these cognitive barriers is one who is viciously self-deceived.

CAUSAL ANCESTRY

We have now completed the arduous task of defining the vice of self-deception. How then does this vice come about? Like all vicious habits, self-deception arises out of pride. When one begins to value subjective desires over the objective intellectual goods of truth and self-knowledge, intellectual self-deception may result. "The worry, rather, is that what people will or desire may produce patterns of belief formation that undermine their ability to weigh evidence, assess claims, and evaluate behavior (especially their own)" (Floyd 2004, 60). It becomes clear that the will plays an active role in the development of self-deception. Upon evaluating information and attributing probative weights to certain propositions, a pure, unadulterated desire for truth is almost never attainable. Our wills, rather, are often influenced by external and internal factors (arising out of pride) that

increase our personal intellectual prejudices and inhibit our search for truth through the development of these cognitive barriers.

I acknowledge that there are instances in which withholding the truth from *others* provides a greater benefit. Take, for instance, the German woman who lies to Nazi forces to prevent the horrific murder of the Jews hiding in her basement—certainly this is morally commendable. However, it is not unreasonable to claim that we should nonetheless pursue truth for *ourselves* in all circumstances. However, "truth and rationality aren't the only things we value, of course, and often the other things we value influence our inquiry in such a way to make believing what's true less likely" (Elshof 2009, 27). These values often cause us to fall prey to faulty inference, incomplete survey of available evidence, lack of thorough diligence, or sympathy in our perceptual evaluations (Davidson 2013, 42). Ultimately, self-deception results when we allow our wills to become subject to our own desires to the extent to which our cognitive abilities to effectively determine truth and apprehend contradictory beliefs are severely hindered, especially in regard to self-knowledge and assessing our own intellectual states.

CHARACTERISTIC THOUGHTS AND ACTIONS

These values may come about through various different processes. One of the most prominent ways in which a non-truth-conducive cognitive habit is formed occurs when one rightfully adopts a true belief based on their circumstances but refuses to relinquish or adapt said belief once circumstances change insofar as to not allow for the original belief to remain tenable (LePore & McLaughlin 1992, 30-31). Say, for example, Sheila rightfully believes that her husband is not cheating on her, as he has remained faithful to her throughout their marriage thus far. However, after a couple of years, empirical evidence begins to suggest otherwise. Her husband is constantly returning home later than usual, he is overly protective of his phone and email, and proves defensive and shifty in conversation. The evidence mounts—it is apparent that her husband is cheating on her, and most unbiased witnesses would agree if presented this data (Mele & Rawling 2010, 247). And yet Sheila maintains her now *false* belief that her husband remains faithful because she has habituated herself to believe what was once true—this causes a cognitive barrier to materialize in her mind. The belief that her husband remains faithful and knowledge that he is cheating are intricately related but

clearly inconsistent. Yet these beliefs do not "cognitively encounter" one another because of this barrier, and Sheila therefore remains in a state of self-deception.

What causes her to maintain this false belief amidst rising evidence to the contrary? In this instance it proves to be fear of coming to terms with a belief that will most likely cause significant pain and discomfort (Davidson 2010, 209). This fear demonstrates the vicious habituation of external values affecting the will—her concern for her own comfort (which stems from pride) causes her to misevaluate the evidence with which she is presented and consequently deceive herself. The same fear provides a motivation for vicious habituation in other characteristic actions of the self-deceived, of which addiction is one of the most common. One can easily imagine the alcoholic who believes that he only drinks in moderation—he implicitly fears that, if he rejects this false belief, he will be forced to acknowledge his addiction and (depending on his moral values) attempt to assuage his vicious habit, which will undoubtedly result in much distress.[2] This means he does not recognize his addiction, only serving to propagate his self-deception further (Coleman 2007, 4). His own pride utilizes self-deception to prevent his will from motivating him into a proper evaluation and confrontation of the inconsistent beliefs.

OPPOSING VIRTUES AND SUPPORTING VICES

Therefore, it is clear that self-deception is not solely opposed to self-knowledge, as is prima facie the case, but to strength of the will as well. Davidson explains: "An agent's will is weak if he acts, and acts intentionally, counter to his own best judgement; in such cases we sometimes say he lacks the willpower to do what he knows, or at any rate believes, would, everything considered, be better" (Davidson 2013, 21). Furthermore, Roberts and Wood describe a properly functioning will (or willpower, rather) as that which moves one to act based on a proper construal of themselves and the situation at hand in accordance with their intellectual good (2012, 63). Self-deception inhibits this process by *improperly* construing the situation at hand, causing the will to fail to move one to appropriate action at a suitable time. Thus, self-deception is opposed to the virtue of a strong will and promotes the development of the vicious weak will.

One may claim that both cases above are simply illustrations of wishful thinking, rather than self-deception. This objection raises the question: which

2. For a thorough discussion of this claim, see Leeuwen 2009.

vices support and closely relate to self-deception? It is only fitting to start with wishful thinking. Wishful thinking and self-deception are not one in the same, as often construed. Rather, wishful thinking proves to be a species of self-deception, as it involves a mental state of self-deception that is achieved through specific circumstances: namely, when one is already in possession of knowledge that a said belief is false, but purposefully suppresses that knowledge in order to make room for a contrary belief that appears more attractive to him or her. For example, say Tom has failed his last seven math quizzes. He knows this fact and its propensity to discourage belief in a future passing grade, but that does not keep him from choosing to believe that he will, in fact, pass the next quiz, despite a complete failure to sufficiently prepare. Thus, it can be seen that wishful thinking is not a separate vice, but one that falls into the category of self-deception.

However, a vice that is closely related to but distinct from self-deception is willful naïveté. Thomas Aquinas, in his Summa Theologiae speaks on a related vice, which he calls "blindness of mind." The Angelic Doctor states:

> Sometimes it is due to the fact that a man's will is deliberately turned away from the consideration of [the] principle [of intelligibility] according to Psalm 35:4, "He would not understand, that he might do well": whereas sometimes it is due to the mind being more busy about things which it loves more, so as to be hindered thereby from considering this principle, according to Psalm 57:9, "Fire," i.e. of concupiscence, "hath fallen on them and they shall not see the sun." On either of these ways blindness of mind is a sin. (Aquinas 1969, II.2.15.1)

Aquinas, in this excerpt, explains that people often turn away from self-realization and intelligibility because of willful neglect, as well as absent-minded business. While this vice is starkly similar to self-deception, it is actually quite different; rather than consisting in a state of mind that concurrently holds to two contrary beliefs, willful naïveté involves a purposeful neglect of the dutiful pursuit of evidence which may in turn support a belief which one does not wish to hold. Finally, I must note that self-*induced* deception is not an instance of self-deception either. Self-induced deception, as distinct from self-deception, consists in consciously convincing oneself of a false belief, which does not necessarily admit of a corresponding inconsistent belief in the mind once adopted. The possibility of

self-induced deception is itself polemical, but that is unrelated to the topic at hand.

PERMANENCE AND DEGREES

Is it possible for a state of self-deception to become permanent? It seems so. One can imagine a belief that admits of such depth of ingression that to remove it would be to essentially raze the believer's entire cognitive framework. Similarly, it seems possible that a set of related albeit inconsistent beliefs could become so ingrained within one's noetic structure that it practically cannot be removed. Take, for example, Jeff. Jeff believes that the United States is in imminent danger of international nuclear war with China. Despite all the evidence to the contrary (namely, the fact that the United States is *not*, in fact, in imminent danger of transcontinental nuclear war), he nonetheless has deceitfully convinced himself of this danger insofar as to construct an underground nuclear bunker in his back yard, in which he now lives. He has, rather literally, built both his life and perceptual framework upon the foundation of this self-contradictory set of beliefs. It seems that no amount of reasonable discussion or solicitation of evidence will convince Jeff that he entertains a set of beliefs that are, at least partially, self-contradictory (the U.S. is in danger and that it certainly seems that the U.S. is not in danger). Therefore, it is evident that the permanence of self-contradictory beliefs hinges on their depth of ingression, or how many beliefs are supported by at least one member of the inconsistent pair.

Consequently, it follows that the vice of self-deception can also come in degrees. The man we discussed above is ostensibly more self-deceived than Tom, the student mentioned earlier who refuses to admit that he will most likely fail his next math quiz. This discrepancy can be attributed to the fact that the man who fears nuclear war is no longer open to a change of his beliefs, while Tom most likely still possesses the willpower to right his inconsistent set of beliefs. Where then do these two diverge? As noted above, self-deception begins to solidify when more beliefs are built on top of the aberrant pair. But in addition to this, self-deception arises when one permits or even mentally encourages further self-deception upon the act of self-examination (Strandberg 2015, 49). Jeff, upon self-examination, will most likely conclude that his mind admits of no contradictory beliefs. Rather, while he cannot dispose of the belief that almost all available evidence discourages his belief that the U.S. is in danger, he will attempt to wrongfully justify this belief by

asserting that all this evidence is, in fact, misguided. This process only serves to deepen the depth of ingression and promote further self-deception.

PREVENTION

Thus, it can be seen that honest self-examination is a crucial ally in the battle against self-deception. However, there is more to the evasion of self-deception than honest self-reflection. To this point Augustine weighs in: "[Charity] produces humility and forestalls the tendency to exaggerate our own goodness…According to this view, then, the remedy for self-deception lies not in a person's own self-reflective capacities; it comes by way of a moral transformation made possible by charity" (Floyd 2004, 77). According to Augustine, Floyd explains, *charity* is the key to avoiding self-deception. True charity is constituted by a reordering of one's desires (with the help of the Holy Spirit) to pursue the proper objects, which leads to a mitigation of self-interest and pride. And, as I discussed earlier, pride is the root of self-deception. Thus, reordering one's loves through a significant and continually ongoing "moral transformation" increases wisdom and diminishes pride, and enables one to fairly examine their own beliefs. This practice can and should be abetted through the assistance of others, as outside sources will almost always exhibit a more charitable approach to one's cognitive structure. Finally, we must seek to practice the virtue of a strong will—this enables us to courageously and authentically pursue true belief in any circumstance and avoid those vicious states of self-deception.

In the last analysis, self-deception seems to us a vice that always afflicts those around us (particularly those who exasperate us), but never one that affects our own mental life. And yet, ironically, to believe this notion is to deceive ourselves, as each one of us has been affected by a state of self-deception at some point. In this paper I discussed the possibility of self-deception and concluded that it is not only possible, but prevalent in noetic convention. I then defined self-deception as a habitual tendency to allow two or more self-contradictory beliefs to exist in one's mind due to a sort of cognitive barrier. Moreover, I discussed its causal ancestry, characteristic thoughts and actions, related vices and virtues, as well as its permanence and degrees. Finally, I discussed strategies of avoidance: namely, the reordering of loves to humble us and permit the fair investigation of our mental state. Ultimately, we must remember that we deceive ourselves all the time—yet this is no cause for despair. Rather, we should acknowledge our vicious states, and

do everything in our power to diminish them through honest introspection and persistent authenticity.

REFERENCES

Aquinas, Thomas. 1969. *Summa Theologiae*, edited by Thomas Gilby. Garden City: Image Books.

Coleman, Mitchell Carl. 2007. "Contribution of Thomas Aquinas's Treatise on Temperance to the Contemporary Effort to Understand and Treat Addiction." MA diss., University of Iowa.

Davidson, Donald. 2010. Problems of Rationality. Oxford: Clarendon Press.

Davidson, Donald. 2013. *Essays on Actions and Events*. Oxford: Clarendon Press.

Dupuy, Jean-Pierre. 1998. *Self-Deception and Paradoxes of Rationality*. Stanford: CSLI Publications.

Elshof, Gregg Ten. 2009. *I Told Me so: Self-Deception and the Christian Life*. Grand Rapids: Eerdmans.

Floyd, Shawn D. 2004. "How to Cure Self-Deception: An Augustinian Remedy." *Logos: A Journal of Catholic Thought and Culture* 7 (3): 60–86.

Leeuwen, D. S. Neil Van. 2009. "Self-Deception Won't Make You Happy." *Social Theory and Practice* 35 (1): 107–132.

LePore, Ernest, and Brian P. McLaughlin. 1992. *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. New York: Blackwell.

McLaughlin, Brian P., and Amélie Oksenberg Rorty. 1988. *Perspectives on Self-Deception*. Berkeley: University of California Press.

Mele, Alfred R., and Piers Rawling. 2010. *The Oxford Handbook of Rationality*. Oxford: Oxford University Press.

Roberts, Robert Campbell, and W. Jay Wood. 2012. *Intellectual Virtues: an Essay in Regulative Epistemology*. Oxford: Clarendon Press.

Strandberg, Hugo. 2015. *Self-Knowledge and Self-Deception*. London: Palgrave Macmillan.

cognethic.org