compos mentis

compos mentis

Table of Contents

Table of Contents (cont.)

compos mentis

In Defense of "I": A Defense and Revision of Psychological Continuity

Chris Abell

Michigan State University

ABSTRACT

Derek Parfit presents and argues for a psychological continuity account of personal identity. This account attempts to define identity in non-circular terms, only relative to psychology and not with the body of an individual. Marya Schechtman argues against this account, claiming that it cannot escape from circularity despite attempting to resolve this problem. I propose a revision of Parfit's original claims, such that quasi-psychological connections do not need to perform the work they were meant to do, in order to address the issues presented by Schechtman's objection, and to hopefully offer a better insight into what is important to defining identity.

Derek Parfit presents us with what can be called a psychological-continuity account of personal identity in his aptly named essay, *Personal Identity*. This account of identity claims that when attempting to define or talk about personal identity, or rather continuous survival of an individual across time, the language of psychological continuity is useful to us as a vehicle to do so. The language of personal identity should be understood here as the words and phrases commonly used to imply identity. These include "I", "Her", and descriptions of identity as something owned individually, across time.

Parfit prefaces his account of psychological continuity by claiming that identity is a one-one relation (Parfit 1971). As I understand Parfit, a one-one relation of identity would imply that one individual can only have one identity, or rather, it is the possession of one identity by, and only by, one individual. So, for example, I am the only person who has my identity, I am *this* Chris. Further, if my identity changes, or if I have multiple identities, they cannot co-present themselves simultaneously. I will only be one person at any given time. From this claim, Parfit argues that when we speak about identity, the language we use actually is actually implicitly about psychological continuity. This is because psychological continuity as described by Parfit is the continuation of an identity, composed of identical psychological component parts, across time (Parfit 1971). So, when I describe a memory that I believe to be my own, by describing myself in reference to the past, as myself, I am describing a psychologically continuous individual. Further, so long as identity remains a one-one relation, describing it through the vehicle of psychological continuity does actually provide a criterion for identity. The criterion is as follows: 'X and Y are the same person if they are psychologically continuous and there is no person who is contemporary with either and psychologically continuous with the other"' (Parfit 1971, 13). In other words, if Person X possesses the same psychology and psyche as Person Y, at a later point in time, they are the same person, yet only insofar as no other person also shares that psyche at the same time.

Since psychological continuity now provides a criterion for identity, Parfit sets out to illustrate how psychological connections can be conceptualized impersonally, without appeal to any particularities about personal identity. A psychological connection or relation is one of the two core ideas that Parfit uses to weave his argument. It is a relation between a previous experience or state of mind someone previously possessed and the associated individual's current state

of mind. For example, the memory of having gone to Vienna in 2002, and the experience of having actually gone. The connectedness between the memory and experience forms the psychological relation in this case. Memories are important to Parfit, so much so that he thinks they are the most important connection relating to identity. However, memories are not the only psychological connection that Parfit mentions. These connections also encompass concepts such as intentions, beliefs, attitudes, etc.

However, there is a strong attack known as the circularity objection commonly employed against psychological continuity accounts of personal identity. The objection used against psychological continuity theories was originally raised by Bishop Butler. The objection is roughly this: while memory may seem like an obvious candidate which to define personal identity by, it cannot achieve this as in order to define memory, someone must already have an understanding of personal identity, as by definition memory presupposes individual personal identities. Therefore, a theorist could not explain how to differentiate between delusional and non-delusional memories without reference to identity, making any attempt to define personal identity using memory ultimately circular, and therefore inadmissible for any theory of identity. Memories are considered by Parfit to be the most important psychological connection to his account (Parfit, 1971), he must present a solution to the fatal problem this objection creates for his theory.

Parfit proposes the idea of a *q*-memory, or a quasi-memory as a solution to the problems created by the objection for his theory. The definition of a quasi-memory is as follows:

> I am q-remembering an experience if (i) I have a belief about a past experience which seems in itself like a memory belief, (2) someone did have such an experience, and (3) my belief is dependent upon this experience in the same way (whatever that is) in which a memory of an experience is dependent upon it. (Parfit, 1971, 15)

A *memory belief* is an individual believing that the memories they possess are in fact their own (Parfit, 1971). The central claim is that these quasi-memories offer a non-circular way to characterize memories, by replacing memory "proper" (i.e. memories as we understand them), while retaining their function as a psychological

component of identity. Therefore, quasi-memories with reference to any previous experience will actually be in reference to other quasi memories in Parfit's system, in that, there is no reference to an understanding of personal identity in the definition of the quasi-memories. If this is the case, we can now use the language of psychological continuity to describe and formulate an understanding of personal identity.

Further, Parfit argues that quasi-memories, and by extension all quasi-psychological connections, are able to distinguish between delusional and non-delusional intentions, beliefs, and memories, *without* appealing to any facts or particularities about identity.

Finally, the last important aspect of Parfit's identity theory is the idea of psychological connectedness. It is defined as "… the holding of these direct psychological relations" (Parfit 1971, 20). In order for Person X to be psychologically connected with Person Y, both individuals need to possess the same direct psychological relations as one another, across time. These concepts form the core of Parift's psychological continuity account of personal identity.

Marya Schechtman, in her essay *Personhood and Personal Identity* (year), claims that Parfit's explanation of psychological continuity in terms of psychological "qausi-states" fails to avoid the circularity objection it was meant to address. Rather, Schechtman claims, quasi-states *cannot* avoid the circularity objection, as "… there is no way to capture what is relevant to personal identity in memories without presupposing identities" (Schechtman 1990, 79), meaning there is no way to avoid reference to particularities about whomever has the q-memory when actually remembering anything.

Schechtman's argument is that quasi-memories do not circumvent the circularity objection because they cannot distinguish between delusional and non-delusional memories in the way they are supposed to. Because, according to Schechtman, successfully distinguishing delusional from non-delusional memories is required in order that memories, or rather the quasi-memories, can be used as a basis of a psychological continuous account of identity, Parfit has not resolved the problems that the circularity objection has presented to his account. It is important to note, that while Schechtman only focuses on quasi-memories in her argument, this objection can be applied to any of the quasi-psychological connections presented by Parfit. Schechtman formulates her argument around an example memory presented by Edward Casey in his book, *Remembering:*

A Phenomenological Study. Casey's memory is seemingly simple enough, recounting an outing to the movies with his family to see a foreign film. However, upon closer inspection, an observer will begin to realize the intricacies of the relations woven into every detail in the memory. Emotions, related memories, reactions, relationships, and many other factors relating only to Casey are integral to the memory, suddenly making it impressively unclear how this quasi-memory would present itself if it is implanted into or experienced by another (Schechtman 1990). She claims that upon examination, we are left with two equally unappealing alternatives: Either, the memory will present itself as phenomenologically identical to Casey's without reference to any of his interpretations of it, or, it will present itself *exactly* as it did to Casey, with every relation and reaction that he experienced being understood as being the reactions of the other person (Schechtman 1990). The first alternative fails because, as evidenced by Casey's example, it is apparent that what constitutes a memory, qua memory, is not simply just the mental images it produces. The related associations and interpretations that it produces in the person remembering are just as critical to the memory *being the exact same memory* as the images producing them. Schechtman argues that without these associations, it seems impossible to say that an someone who experiences Casey's memory in this way would actually be sharing in the exact same memory as Casey, and goes so far as to claim that it seems unlikely that this phenomenon could even be called a memory at all, failing to capture "… what is relevant in the connection between a genuine memory and the experience remembered" (Schechtman 1990, 83). The second alternative likewise fails, however for different reasons.

Schechtman argues that while it may seem that the quasi-memory could be exactly the same as Casey's, with the same reactions and emotions associated with it, this is not actually the case. Instead, the quasi-memory would be altered by the psyche of the person who now possesses it. They would find Casey's family unfamiliar, the movie being watched peculiar, and all other associated interpretations of facts to completely alien to who they believed themselves to be. The quasi-memory takes on a completely different character to anyone who has it other than Casey, and Schechtman argues that in order to make it truly exactly the same, we would need to replace another person's psyche with Casey's. Further, if this were to occur, the memory would become non-delusional, according to Parfit's understanding, despite our knowledge that it actually is delusional. If

another person were to understand the memory as their own, there could be no way for them to distinguish from any other memory they possess, making it impossible to distinguish between the two kinds of memory.

In response to Schechtman's objection, I propose a revision of Parfit's position on the grounds that a quasi-psychological connection's ability to distinguish itself as delusional or non-delusional is irrelevant to the formation of psychologically continuous identity. As such, quasi-states would not need to operate the way Parfit had originally claimed they ought to, nullifying Schechtman's objection.

The first concept that needs to be addressed is what I will call the *Truth Relation* aspect of a quasi-psychological connection. I will use quasi-memories in my discussion of this idea; however, this relation can be applied systematically to any of the psychological relations presented and examined by Parfit and Schechtman. The truth relation of a quasi-memory encompasses its relation to the actual experience of the individual who possesses it. If a quasi-memory has a "true" relation to experience, we can say that this memory has an accurate relation to an experience that its owner actually had. This would be a non-delusional memory according to Parfit's account. For example, the "true" relation the memory of watching a documentary last night, and the experience of having actually watched the documentary. Conversely, if that quasi-memory has a "false" relation to experience, the memory does not relate to an experience the individual who possess it had. This could be a "false" relation of the memory of watching a documentary, without the experience of actually doing so, a delusional memory to Parfit. It is apparent that while Parfit and Schechtman disagree on how to distinguish between the two kinds of *q*-memories, all memories possess this relation as a component part, either as "true" or "false"; a single memory cannot possess both relations simultaneously, this would be contradictory. However, what is less apparent yet exceedingly critical to an understanding of personal identity, is that regardless of the truth or falsity of the quasi-memory, *it will still form a component part of an individual identity.* In the case of Casey's outing to the movies, if someone else had that quasi-memory implanted into their mind, I can concede that Schechtman's second scenario in her objection to Parfit is most likely to occur. The quasi-memory of an unfamiliar family seeing a strange movie together would be disturbing and confusing for its new owner. Yet, despite being delusional and not grounded in experience, this apparent memory is now a psychological relation in possession of this new owner. Whether this memory is

written off as delusion, hallucination, or accepted as their own does not matter when we consider that no matter what, it is now in their psyche.

Because of this additional insight, I propose that Parfit revise his original position about how to define psychological connectedness in his account. Recalling the initial definition of psychological connectedness, it was understood as a persistence of direct psychological connections across time. "Direct" here is to be interpreted as meaning non-delusional relations, with grounding in the actual experience of the person who possesses them. However, it is apparent that this is not actually the case, evidenced by the relationship of truth relations to identity. What I propose is that psychological connectedness not be constrained by only direct relations, but instead be understood as persistence of quasi-psychological connections across time, without regard for their being delusional or non-delusional. In other words, the need to have a way to distinguish between "true" memories and "false" ones *is not relevant to* an understanding of identity. Schechtman's circularity objection was meant to show how q-memories cannot avoid the necessary distinction of describing which individual possessed the memories when trying to distinguish between delusional and non-delusional. By removing the need to distinguish, grounded in the empirical experience of identity, q-memories can serve as a non-circular vehicle for conceptualizing identity. Further, this revision of Parfit's position works because of the nature of personal identity. It is based on belief, more specifically the individual's belief about who they are. Take for example the case of the madman who believes himself to be Napoleon. He claims to have memories, and believes they are his own, that he was actually at Waterloo leading the French forces. Despite this not actually being the case, there remains a persistence across time of identical quasi-psychological connections, direct or delusional, in the psyche of the madman. They remain foundational to his personal identity and therefore need to be accounted for in a psychological continuity account of personal identity.

However, this revision does not save Parfit's view from Schechtman's objection absolutely without further clarification. To reiterate, the modified circularity objection argued that quasi-states cannot distinguish between delusional and non-delusional memories without reference to the individual who possessed them, thus making them circular. My response, in defense of the psychological continuity view, was to revise the claims about quasi-states in relation to psychological connectedness, by eliminating the need for them to distinguish between "true"

and "false" memories. Yet, in doing so, the psychological continuity theorist is now faced with a new problem: If we redefine the scope of what quasi-memories can do, we now need a way to reliably distinguish between individuals who may possess the same q-memories. For example, suppose that I believed that I was actually my next-door neighbor. Somehow, we shared the same memories with identical attitudes, beliefs, and emotions accompanying them. Following this new proposed understanding of q-memories, they and I should be considered the exact same person at the exact same time, despite the obvious differences between us in other non-psychological elements, i.e. we live in different homes, have different families and possessions, etc. This would be absurd, as identity must be a one-one relation in order to be understood through psychological continuity. In order to resolve this, it is important to consider the nature of the objection.

Examined through the lens of psychological continuity, it seems as though it would be impossible to distinguish between myself and the neighbor. Yet, upon closer inspection I do not think this is actually the case. Looking at the case externally, it is obvious that we are not the same, given that we are different people. When referencing the madman who thought himself Napoleon for example, we call him "the madman" because we know he is not who he believes himself to be. I assume this position to be uncontroversial and will not explore it further. What is important then, is how to distinguish from an *internal* perspective. What we have explored so far is *personal* identity, how the individual distinguishes themselves from others. In this, I would argue that it would be possible to distinguish between seemingly identical persons, without regard for delusion or non-delusion. Individuals do not confuse themselves with others. An aspect of self-consciousness is an innate understanding of the self (I) being distinct and separate from other selves (them). If I believed myself to be my hypothetical neighbor, Terry, for example, then I would in this case also be Terry. Yet also, if I happen to encounter Terry out on a walk one day, despite sharing identical psychological connections, I would still comprehend myself as *this* Terry, while they are *that* Terry. Likewise, they would share a similar viewpoint. We are almost identical, save for this one dissimilarity, and this distinctness is found within the individual. In this sense, the one-one relation requirement of identity can be saved. Terry is not one person inhabiting two bodies, rather, there are two Terry's which are internally distinct, (perhaps also externally distinct), and separate, while sharing almost all psychological qualities with one another.

As such, if this revision is made to Parfit's position, I claim that quasi-memories and all other quasi-psychological relations do not actually need to be able to distinguish between delusional and non-delusional relations. Yet, they can still operate effectively in the formation of a non-circular, impersonal account of personal identity, fulfilling the criterion Parfit had originally hoped to categorize. Therefore, this modification of the account nullifies Schechtman's objection that they cannot differentiate delusionality from true relations, the work they were originally intended to do, as there would no longer even be a need for them to do so. While this does not resolve every issue in the continuity theory, I believe it can comfortably circumvent the circularity objection originally raised against it.

REFERENCES

Parfit, Derek. 1971. "Personal Identity." *The Philosophical Review* 80 (1): 3-27.

Schechtman, Marya. 1990. "Personhood and Personal Identity." *The Journal of Philosophy* 87 (2): 71-92.

compos mentis

The Inauthenticity of Memory Manipulation and the Ethics of Vulnerable Communication

Matt Denaro
Brooklyn College

ABSTRACT
Concurrent with an expanding medical science around the possibility of modifying and/or erasing memories has been philosophical questions concerning the ethicality of doing so, especially for trauma survivors and victims of PTSD. In this paper, I look to argue against such pharmacological methods as being inauthentic in a Sartrean sense. Following from this formulation, I turn to the work of Jürgen Habermas and synthesize it with the work of Jean Paul Sartre to discuss a possibility of using communicative rationality and an attention to authenticity as a means of mediating trauma in a more constructive way than memory manipulation. As a way of discussing the application of such an ethics, I discuss two disparate fields in which it could be applied: social media and food, drink, and the table.

I. INTRODUCTION

Recent developments in medical science conjoined with an elevation of awareness around memory related illnesses such as Post-Traumatic Stress Disorder have brought the possibility of memory manipulation, including erasure, to the forefront. There are debates concerning the ethicality of erasing or manipulating one's memories, even if those memories cause harm. However, before getting into these debates, it seems prudent to define what is meant by memory manipulation and erasure in reference to medical science's current and prospective ability to do so, as well as the nature of this manipulation, and what kinds of memories are in question.

Alexandre Erler presented a suitable understanding of memory manipulation as "all methods of modifying memory in a desirable way that do not involve enhancing it—at least not directly" (Erler 2010, 240). This understanding is inclusive of erasure of memories. It also heads toward what kind of memories are in question, that is traumatic or otherwise troublesome memories. The manipulation of trivial or happy memories will not be discussed, as it seems unlikely that people would pursue that, nor does it seem that a critical understanding of the manipulation of those memories would produce a different analysis of the ethicality of memory manipulation. As far as the medical science for memory manipulation goes, there is propranolol treatment in which someone could take that beta blocker shortly after experiencing a traumatic or troublesome event, numbing the emotional impact of the memory (Erler 2010, 240). There has also been research into memory manipulation via HDAC inhibitors by MIT neuroscientist Li-Huei Tsai in which that drug was used to "help mice extinguish a fearful memory of a traumatic event that took place in the distant past" (Johnson 2014). Although this drug has not been tested on human subjects, the possibility of such a thing is possible in the not so distant future.

As the medical science continues to develop, there has been a rising philosophical literature on the topic. One important work in the field was Alexandre Erler's article "Does Memory Modification Threaten our Authenticity?" in which he argues that it necessarily does threaten authenticity when it involves memory editing, but enhancement—something that will not be discussed here—does not necessarily threaten it (Erler 2010, 235). Another seminal work in the field of the ethics of memory manipulation is "The Normativity of Memory Modification" by

S. Matthew Liao and Anders Sandberg, in which they come to the conclusion that "it is up to individuals to determine the permissibility of particular uses of MMTs" provided that they "do not harm themselves or others… and there is no *prima facie* duty to retain particular memories" (Liao and Sandberg 2008, 96).

A third influential article in the field is Elisa A. Hurley's "Combat Trauma and the Moral Risks of Memory Manipulating Drugs." Hurley argues that the use of memory manipulating drugs may prevent the subject of combat trauma from understanding their role as perpetrators of violence in war, a term she refers to as the "state of grace." This preemption of the state of grace, "underwrites successful gestures of reparation toward those harmed by their actions" (Hurley 2010, 35). Hurley's position is made clearer in her article, "The Moral Costs of Prophylactic Propanolol." She states,

> prophylactic intervention that works by interfering with the laying down of trauma memories, propranolol threatens to permanently cut off access to the emotions experienced at the time of trauma, access that might be important for holding oneself and others accountable for moral wrongdoing (Hurley 2007, 35).

It seems that this line of thinking on accountability for moral wrongdoing also applies in reference to a narrative sense of identity as well. Also, in this there is a clear sense of responsibility for one's actions that will be discussed at length in relation to Sartrean authenticity.

Another important article in the literature of memory modification is Adam Kolber's "Therapeutic Forgetting: The Legal and Ethical Implications of Memory Dampening." In this article, Kolber argues that "memory dampening has the potential someday to ease the suffering of millions of people and that heavy-handed government restriction of memory dampening is inappropriate, it follows that we should have some limited right to therapeutically forget." He points towards a "freedom of memory," that people should have autonomy over what is and is not in their memory (Kolber 2006, 1567). While the proliferation of this research shouldn't be hampered, it does seem imprudent to approach memory from the perspective that its manipulation or erasure would not have an impact on others, something that Kolber's "freedom of memory" seems to do.

Memory manipulation and erasure is problematic for a few reasons. One reason is that it promotes an inauthentic form of being. Memory manipulation

is an act in bad faith, denying the authenticity of one's own lived experience, in favor of a presumably easier path forward through an event. Although, following from the honesty that authenticity should foster, a more compassionate form of resilience can be developed on the basis of vulnerable communication between empathetic subjects.

II. THE INAUTHENTICITY OF MEMORY MANIPULATION

In order to understand why memory manipulation is inauthentic, a conception of what it means inauthentic needs to be developed. Generally speaking, this conception of authenticity runs along existentialist lines, relying on Sartrean notions of it. This will not be a complete explication of the concept of authenticity, but a development of the relevant ideas contained in the concept, namely honesty and responsibility. For Sartre, the negation of authenticity is bad faith, which is "a lie to oneself within the unity of a single consciousness" (Sartre 1972, 800). From this definition of bad faith, a sense of an obligation of honesty towards every being can be implicitly understood. There is an *a priori* and universal obligation of authenticity. We owe it to each other to be the best selves we can be. People are constitutive of the human species and therefore equal architects of the human condition. By living authentically, which means to bear the responsibility of all actions and to face them honestly, one can fulfill this obligation and be the best person one can be.

Sartre describes this radical responsibility as a result of one's radical freedom, stating "man being condemned to be free carries the weight of the whole world on his shoulders; he is responsible for the world and for himself as a way of being" (Sartre 1972, 707). Since, for Sartre, we are the makers of our own reality, we have only ourselves to blame for when things go wrong. This means not running away from the hard things of life, for example trauma. Taking responsibility for one's actions—even what may happen to someone as a matter of contingency, which to some extent trauma may be—is an important part of Sartrean authenticity.

Not only must this duty of responsibility and radical honesty be upheld for others, but for oneself as well. This radical honesty towards self involves a critique of one's own actions and understanding if they truly reflect the best person one could be. Although this may seem harsh, and *prima facie*, it certainly is, there is a way a more compassionate system of vulnerable communication following

from this radical responsibility and honesty could be developed. This idea will be explicated in the next section of this paper. However, to be authentic, trauma survivors must confront their traumas as a means of reasserting control over their narrative, instead of running away from them with the use of pharmacological methods.

The inauthenticity of memory manipulation is clear from this understanding of responsibility. If one were to manipulate their traumatic and/or troublesome memories, they would be shirking the responsibility of confronting them, and processing them. The act of taking propranolol is one of bad faith because it rejects true experience for a dampened one. Those who choose propranolol over being-in-itself choose to delude themselves and others concerning one of the most constitutive parts of who one is, memory. The prescription of memory manipulation seems to sell people's control over their own lives short. Using memory manipulation as a means of dealing with traumatic memories undercuts people's ability to overcome the worst situations, and still come out the other side. On the other hand, facing the terrifying responsibility of having one's own memories, and working through them by various forms of therapy, seems to be a more constructive way of going about processing trauma than forcing oneself to forget about it through medical means. Communicative methods of overcoming trauma such as therapy confront the trauma directly by making it have less control over one's being, whereas memory manipulation accomplishes that through a delusion of the self.

Even if the authenticity problem were not an issue, there would still be the issue of other people holding the narrative of the traumatic event in memory. It seems to be the ultimate shirking of responsibility in the sense that it offloads the emotional weight of trauma off of the victim and onto the others around them, while having nothing productive for spirit that a vulnerable communication could provide. This seems to be a great disservice and dishonesty towards one's fellow person. This is well illustrated in a scene of the film *Eternal Sunshine of the Spotless Mind* in which the two main characters Joel (Jim Carey) and Clementine (Kate Winslet) have an interaction after Clementine has the memory of their relationship erased. Joel approaches Clementine at her job and is racked with anxiety and emotional pain when it becomes clear she has no recollection of who he is, despite being in a loving relationship just recently (Gondry 2004). By erasing her memory of the

relationship, Clementine leaves Joel to hold the memory of their love in isolation, shirking her responsibility of radical honesty and responsibility towards others.

The notion of responsibility is expressed well in Hurley's conception of the state of grace. In retaining the traumatic memory of combat in war, perpetrators of violence are forced to confront the role they had in the violence they did, even if they are not directly responsible for the situation they are in. Although a Sartrean would not say that they are not directly responsible for being in a war situation, as they did choose to join the military or to not dodge the draft, the point concerning responsibility still shines through. Hurley is keen to point out that by embracing their responsibility for the situation they in part caused, perpetrators of violence in war are doing something constructive by enabling the possibility of making moral reparations to the ones that they wronged.

It is this obligation of responsibility and honesty that seems to be rejected by Liao and Sandberg's approval of MMTs on a case by case basis. We owe it to one another, assuming everyone is striving towards authenticity, to experience reality honestly, and express that experience truthfully, or at least to the best of one's ability. Liao and Sandberg's approval of MMTs on a case by case basis is only acceptable if one permits inauthenticity as a possibility of ethical living. Although, their approval of MMTs as a means of release from traumatic memory should be seen as compassionate, it is at the cost of honesty, responsibility, and authenticity, all of which potential MMT patients must be made aware of prior to the memory modification. However, from the honesty that authenticity necessitates, one can develop an equally compassionate way of coping with trauma, a form of resilience based in vulnerable communication facilitated by radical honesty and responsibility.

III. THE ETHICS OF VULNERABLE COMMUNICATION

At first glance, preventing people from modifying traumatic memories seems to lack compassion for trauma victims. It forces them to relive the trauma and hold it memory when this is painful. This is at the cost of honesty, responsibility, and authenticity, all of which potential MMT patients must be made aware of prior to a memory modification. This is the tension between the duty of authentic being and the duty of compassion. It seems that communication could mediate them. With

the honesty that authenticity necessitates, one can develop a compassionate form of resilience based in the vulnerability contained in radical honesty.

This kind of resilience is a communicative one. Through the communication of trauma to empathetic others, one can re-assert control over one's narrative, owning the trauma and forcing it to relinquish its hold over the victim. This is the argument Susan Brison makes in the preface of her book *Aftermath: Violence and the Remaking of a Self.* She states,

> The communicative act of bearing witness to traumatic events not only transforms traumatic memories into narratives that can then be integrated into the survivors' sense of self and view of the world, but it also reintegrates the survivor into a community, reestablishing bonds of trust and faith in others (Brison 2002, xi).

The kind of empathetic listening that is needed to bear witness to trauma involves vulnerability that should follow from the radical honesty that authenticity demands. If one lives authentically, then they are radically honest about their way of being, understanding their responsibility as human beings as determinates of the human condition as well as the individual's. This radical honesty means having the emotional fortitude to be vulnerable in the face of the worst possible events, i.e. trauma. It seems that this idea of empathetic listening is a more specific form of the loving perception that María Lugones discusses in "Playfulness, 'World'-Travelling, and Loving Perception." She argues that "travelling to each other's 'worlds' would enable us to *be* through *loving* each other" (Lugones 1987, 8). It seems that this world traveling is facilitated by that radical honesty and vulnerability located in authenticity.

In being radically honest with one another, one teaches the other about their world, so that the other can travel to it, and lovingly perceive. However, the loving perceiver doesn't proclaim to know the world or to feel it in its full effects. Instead, the loving perceiver, who is an empathetic listener, simply states "You are heard." From this loving perception, the other(s) can help the traumatized in loosening the grip the trauma has on the traumatized person, freeing them from the trauma in a more authentic way than memory modification. These situations of vulnerable communication facilitated by radical honesty seem much more productive than simply eliminating the problem like memory manipulation would. In fact, it seems that using memory manipulation would be a way of arrogantly perceiving, seeing

the traumatized person's world as something that is not worth travelling to and understanding. Instead of understanding the world of their trauma, they brush it aside and eliminate it. This empathetic listening and loving perception of the trauma narrative is only possible in a situation in which the listeners and speakers are radically honest with one another, something that MMT users would never have the opportunity to do. The question now becomes how can radical honesty and authenticity among all people be facilitated.

To be able to answer this pressing question, the ideas of communicative reason and action must be developed. This idea is central to the work of 20th century German philosopher Jürgen Habermas. In disambiguating instrumental reason from a communicative one, his ideas are able to be used effectively in trauma theory. Habermas also offers useful insights in terms of trauma theory as a result of his historical situation. The entry on Habermas in the Stanford Encyclopedia of Philosophy states "The Nuremberg Trials were a key formative moment that brought home to him the depth of Germany's moral and political failure under National Socialism" (Bohman and Rehg 2014). Witnessing this public exposition of the trauma of the Holocaust assuredly influenced his philosophy of communicative action.

After the Second World War, Habermas' mentor Theodor Adorno said "Hitler imposes a new categorical imperative on human beings in their condition of unfreedom; to arrange their thought and action that Auschwitz would not repeat itself" (Jeffries 2017, 747-748). In reference to this quote Stuart Jeffries, author of *Grand Hotel Abyss,* states "It is this thought, and this moral duty, that has impelled Habermas to work to ensure that human beings never stoop to such barbarism again" (Jeffries 2017, 747-748). This idea of a communicative rationality that can save human beings from the abyss of reason that Theodor Adorno and Max Horkheimer pointed out in *Dialectic of Enlightenment* is productive of an ethics that prevents this barbarism, an ethics of listening and being heard, of loving perception of the other, what Habermas calls the Theory of Communicative Action (TCA). Not only does it seem that his ethics can rescue reason from its barbarism, but it can also aid trauma survivors in overcoming trauma, or as Brison put it "reestablishing bonds of trust and faith in others" (Brison 2002, xi). The connection between Brison and Habermas' ethics is made clearer with Jeffries' description of the TCA as a situation "whereby participants in argument learn from others and from themselves and question suppositions taken for granted" and "like an

ongoing South African Truth and Reconciliation Commission" (Jeffries 2017, 774-775). In both their ethics, the importance of the speech act and communicating, which involves listening and being heard, has a great importance.

Many trauma theorists have pointed out that an approach towards trauma on the basis of rationality falls flat. However, this conception of rationality is only that in its instrumental sense. Instrumental reason is like that of Kant's. It is monological and isolated, based in a singular autonomous subject working through things according to universal laws. Habermas' reason, a communicative one, is dialogical. It is based in consensus that is brought about through communication and/or discourse (Jeffries 2017, 745-746). Where Kant's reason is subjective, but also universalizabile, Habermas' is intersubjective. Of course an instrumental reason is incapable of working through trauma for victims, for there are no rules or laws that can make sense of trauma. Trauma is a suspension of those ethical rules and laws as an objectification of the other, rather than equally recognizing the other as equally human, at least in person to person trauma. But a communicative reason is capable of mediating trauma because it is based in consensus that many people coming together form. This consensus is not an *a priori* and universal one, as a Kantian instrumental reason would be productive of, but it is arrived at universally by many subjectivities working in cooperation and solidarity. Habermas states this clearly states his Discourse Principle, "(D) Only those norms can claim validity that could meet with the acceptance of all concerned in practical discourse" (Habermas 1998, 41). It is not our duty to be communicatively rational, but it is to be authentic. Communicative rationality seems to facilitate this authenticity. Habermas articulates this facilitation well when he states "Discourse ethics defends a morality of equal respect and solidaristic responsibility for everybody" (Habermas 1998, 39). This is seemingly very analogous with the Sartrean conception of what is necessitated for action by authenticity, especially if one is willing to equate "equal respect" with a kind of radical honesty and responsibility.

Also, trauma theorists have pointed towards an unspeakable nature of trauma. Many cite the work of Jacques Derrida on the concept of hauntology in reference to this. For example, Justina Dillon and Michael O'Loughlin's "Questions Unasked: The Legacy of Childhood Trauma in the Life Narrative of a Lithuanian Woman Survivor of the 1941 Soviet Deportations" frame their understanding of trauma in reference to Derrida's hauntology. They state "we frame this work as an inquiry into hauntology, or the presence of ghosts or spectral presences that

while hidden from view, make their presence felt both in individual lives and in the collective psyche of a group or nation" (Dillon and O'Loughlin 2015, 175). This conception of trauma seems to point towards it being not really there, but a haunting presence, one that is unspeakable. However, it seems that through the power of communicative rationality, these specters of trauma could be exorcised, made real, and speakable.[1]

Habermas argued that communicative rationality was at its peak in the bourgeois public sphere of the Enlightenment, of which he identified a few constitutive institutions: the salon, the cafés, and table societies. (Habermas 1989, 30). Of course this was not limited to those institutions, but those institutions were certainly constitutive of some of the bourgeois public sphere in the Enlightenment. But, as Habermas argues, through the commodification of discourse through mass media, these sites of what he calls "ideal speech situations" throughout his works withered away, leaving us with little sense of communicative rationality. The unfinished project of modernity, which is another one of Habermas' major themes, it seems is to recover it as a way to reach consensus for society, i.e. build democracy into social relations (Bohman and Rehg 2014). To return to the discussion of mediating trauma, an ideal speech situation for that seems to be one in which the actors are able to be authentic, that is radically honest and vulnerable.

In *The Structural Transformation of the Public Sphere*, Habermas outlines three parameters for the existence of a public sphere in which communicative reason can thrive. The first criterion is a sense of equality among its members. Habermas writes, "They preserved a kind of social intercourse that, far from presupposing the equality of status, disregarded status altogether. The tendency replaced the celebration of rank with a tact befitting equals" (Habermas 1989, 36). This is the idea of equal recognition of the other's subjectivity, whose dialectic has its historical-philosophical basis in Hegel's Master-Slave dialectic. The second is that "the discourses of these institutions were not the interpretations of the courts or church, but their own" (Habermas 1989, 36-37). This points towards an idea of autonomy and full ownership of the ideas that one is espousing, which seems to be analogous to the understanding of radical honesty presented. Lastly, the

---

1. It is interesting that Derridean and Habermasian conceptions of dealing with trauma come into conflict considering their feud concerning Habermas' reading of Derrida in *The Philosophical Discourse of Modernity*.

third criterion is that institutions of a public sphere must never become entirely exclusive. He states "However exclusive the public might be in any give instance, it could never close itself off entirely and become consolidated as a clique" (Habermas 1989, 36-37).

These criteria are then universalized in *The Inclusion of the Other* with his four part statement concerning the features of argumentation. This reproduction of the criteria of the public sphere has added to it a greater sense of the lack of coercion and the "equal opportunity to make contributions" (Habermas 1998, 44). From this it is clear that subjects in a communicatively rational ideal speech situation must be active and empathetic listeners, as referenced in the first criteria, and radically honest, as referenced in the second. The third criterion seems to point towards a radical responsibility in the sense that each person owes it to the other and the self to maximize the perspective from which one hears from. This universalization of perspectives looks to understand the full effects of one's actions, and to communicate the effects of other's actions to them, allowing one to be better responsible for action, and thus more authentic.

Now, I would like to offer two novel sites of potential trauma mediation. The first is social media. Certainly, social media as it exists now does not exist as a public sphere of communicative rationality. Although it is nearly universally inclusive, and people do generally espouse their own ideas on the Internet, it is very clear that users of social media lack the equal recognition of others in their communications online. This is evidenced by the proliferation of cyber-bullying and misleading others on the basis of the internet's anonymity. Although such corners of the Internet could and do exist, it is very clear that this is not the dominant way of using the most radical means of communication ever developed in human history. It is important to remember that in the public sphere of the Enlightenment communicative rationality did not operate all the time and at all levels. The café was not always a site for rational debate, but oftentimes "these discussions would devolve into 'idle gossip'" (Robiquet 1965, 41). Just because particular portions of an institution negate its character as a public sphere, this does not universalize this quality. Nonetheless, communicative reason has a meager existence online currently.

An idealized social media would be good for facilitating communicative reason, and working through trauma. By not being bound by geographic space, the Internet allows for people to make connections, and communicate, across

great distance. Also, subjects in the trauma sharing situation are able to put more thought and time into what is being said. This is as a result of the lack of geographic and spatial bonds. Since one would not be in the same room with the other one is communicating with, there would not be an awkwardness in the silence while someone thinks of a cogent response. This would aid trauma sharing in the sense that responses could become more empathetic and clearer with this increased time. Not only this, but on the Internet, people can search out empathetic listeners that are ready and willing to hear one's trauma. Furthermore, the depersonalized aspect of the internet, that in some sense is the cause of the lack of equal recognition between subjects may also be its saving power for communicating trauma, as the victim doesn't have to feel the anxiety of being physically present with others while bearing one's trauma.

Interestingly, the advent of call-out culture in some sense offers a strong sense of the responsibility and honesty that would be contained in a vulnerable speech situation. People online who "call-out" others for their problematic behavior are holding the other responsible for their problematicism. This is despite the fact that this act of calling someone out, especially when that person is someone who holds some power, is difficult. This is exemplary of the emotional fortitude needed in radical honesty. This is not to hold call out culture as a prime example of mediating trauma but to point towards the possibility of using social media as means of exposing/naming trauma, holding people accountable for their actions (radical responsibility), and speaking using radically honest speech. What is lacking in it being helpful more mediating the named trauma, is that oftentimes is actual ramifications for the person called out, and/or a lack of recognition by the person called out of their wrong doing.

The second site of potential trauma mediation concerns food, drink, and the table. Food and drink have a remarkable way of bringing people together. So much of sociability is based around food and drink. It seems the necessary fact of maintaining existence by eating and drinking has resulted in food becoming a social fact. Food and drink were instrumental in the establishment of the bourgeois public sphere that Habermas discusses, an argument I've made more robustly in other work. More contemporarily, food and drink is constantly used to frame social activities such as lunch dates, business dinners, catching up with an old friend over a beer, etc. In its power of bringing people together, and being

conducive to conviviality, food and drink can be used as a way to organize the mediation of trauma with receptive and empathetic others.

One of the oldest symbols for making peace with another is breaking bread. This act of taking in a meal carries great symbolic weight, predicated on notions of shared trust and social bonds, oftentimes resembling that of family. In a rather animalistic, yet also beautifully human way, by choosing to break bread with someone, one places their trust in the other that each won't hurt the other in some way when they are vulnerable. It is this kind of thinking I wish to apply to trauma mediation at the table. By opening oneself up to taking a meal with someone, one chooses to share something with the other and this establishes a bond with the other. Even Habermas and Derrida could heal the wounds of their feud over a meal. Derrida's biographer Benoît Peeters states, "During a friendly lunch, Habermas did all in his power to 'wipe out the traces of the previous polemic, with an exemplary probity' for which Derrida would always be grateful" (Peeters 2013, 501). This bond can be used to facilitate vulnerable speech not only among spatting philosophers but even among the worst of enemies, for example, a rape survivor and her attacker. This certainly takes emotional fortitude on the part of the victim, but this kind of radical honesty towards what happened to oneself will exorcise the trauma, and direct this pain towards the attacker, forcing him to recognize the evil of what was done, holding him responsible.

In some sense, the use of food and drink to mediate trauma is already underway in present society. People, oftentimes women, gather together and drink wine and discuss their lives, oftentimes the troublesome aspects of it. In this case, food and drink loosens the tension and anxiety of trauma sharing, making it easier to speak of it. This also applies to more masculine dominated settings, albeit with considerably less vulnerability and actual communication involved.

IV. CONCLUSION

Memory manipulation is inauthentic. But survivors are not stuck in dealing with their trauma. In being authentic and taking responsibility for what happens in one's life, one has to be radically honest with themselves and others about what happened to them. This is true for all people, not just trauma survivors. This radical honesty is productive of vulnerable speech situations in which trauma can be mediated with empathetic and receptive others. This situation seems to point

towards a public sphere of trauma mediation that uses communicative reason to reclaim control over one's narrative with the presence and help of empathetic and receptive others.

Not only does this ethics seem to point towards an empathetic way of confronting trauma, and mediating it for the betterment of a fractured self, but also a way to prevent trauma from occurring. If discourse ethics conjoined by a Sartrean conception of authenticity is universally strived for, trauma situations seem as if they would be less likely to occur on the basis of equal recognition of individual's subjectivity. Trauma is the objectification of the other's subjectivity and the suspension of ethics, so if that tendency for man to objectify the other is transcended, it seems that a more ethical world can be developed for humanity by humanity on the basis of authenticity and communicative rationality.

REFERENCES

Bohman, James, and William Rehg. 2014 "Jürgen Habermas." Stanford Encyclopedia of Philosophy. Last modified August 4, 2014. https://plato.stanford.edu/entries/habermas/.

Brison, Susan J. 2002. *Aftermath: Violence and the Remaking of a Self*. Princeton: Princeton University Press.

Dillon, Justina Kaminskate, and Michael O'Loughlin. 2015. "Questions Unasked: The Legacy of Childhood Trauma in the Life Narrative of a Lithuanian Woman Survivor of the 1941 Soviet Deportations." In *The Ethics of Remembering and the Consequences of Forgetting*, edited by Michael O'Loughlin, 161–77. Lanham: Rowmann & Littlefield.

Erler, Alexandre. 2010. "Does Memory Modification Threaten Our Authenticity?" *Neuroethics* 4 (3): 235–49. https://doi.org/10.1007/s12152-010-9090-4.

Gondry, Michel, dir. 2004. *Eternal Sunshine of the Spotless Mind*. Universal City, CA: Focus Features.

Habermas Jürgen. 1989. *The Structural Transformation of the Public Sphere: an Inquiry into a Category of Bourgeois Society*. Translated by Thomas Burger and Frederick Lawrence. Cambridge: MIT Press.

Habermas Ju̇rgen. 1998 *The Inclusion of the Other: Studies in Political Theory*. Edited by Ciaran Cronin. Cambridge: The MIT Press.

Hurley, Elisa A. 2007. "The Moral Costs of Prophylactic Propanolol." *The American Journal of Bioethics* 7 (9): 35–36. https://doi.org/10.1080/15265160701518607

Hurley, Elisa A. 2010. "Combat Trauma and the Moral Risks of Memory Manipulating Drugs." *Journal of Applied Philosophy* 27 (3): 221–45. https://doi.org/10.1111/j.1468-5930.2010.00492.x.

Jeffries, Stuart. 2017. *Grand Hotel Abyss: the Lives of the Frankfurt School*. London: Verso, an imprint of New Left Books.

Johnson, Carolyn Y. 2014. "MIT Researchers Find a Drug That Helps Erase Traumatic Memories in Mice." Boston.com. Last modified January 16, 2014. https://www.boston.com/news/science/2014/01/16/mit-researchers-find-a-drug-that-helps-erase-traumatic-memories-in-mice.

Liao, S. Matthew, and Anders Sandberg. 2008. "The Normativity of Memory Modification." *Neuroethics* 1 (2): 85–99. https://doi.org/10.1007/s12152-008-9009-5.

Lugones, María. 1987. "Playfulness, 'World'-Travelling, and Loving Perception." *Hypatia* 2 (2): 3–19. https://doi.org/10.1111/j.1527-2001.1987.tb01062.x.

Peeters, Benoît. 2013. *Derrida: A Biography*. Cambridge: Polity.

Sartre, Jean-Paul. 1972. *Being and Nothingness: a Phenomenological Ontology*. Translated by Hazel Barnes. New York: Washington Square Press.

compos mentis

Aristotle's *Politics* and The Charge of Egoism

Cynthia Gonzalez

University of California, Santa Cruz

ABSTRACT

The recent revived interest in virtue ethics and, in particular, Aristotelian virtue ethics has also instigated a conversation between its proponents and opponents and while the criticisms against Aristotelian ethics are numerous and vary greatly, perhaps the most common is 'the charge of egoism'. This paper analyzes the 'charge of egoism' through Tom Angier's particular critique as well as Rosalind Hursthouse and Glen Pettigrove's general analysis of this criticism to show that the central objection in 'the charge of egoism' is that it is self-regarding, egotistic, and not other-regarding, altruistic. It then moves on to establish that Aristotle's Politics is a) a part of his ethical framework and b) the more foundational of his ethical treatises to demonstrate that, given the parameters of 'the charge of egoism', Aristotelian ethics is necessarily other-regarding and, therefore, not egotistic but altruistic. This paper concludes by considering whether this charge is less about what is self vs other regrading and more about the differences between the liberal and communitarian conceptions of human nature.

Virtue ethics, after being on the decline for centuries, has seen an interesting revival since the mid-twentieth century (Baril et Hazelett 1894). And, in spite of the fact that virtue ethics can have various instantiations (Stoic, Epicurean, Platonic… etc.), it seems that this revived interest in virtue ethics is largely rooted in the Aristotelian tradition, particularly, Aristotle's *Nicomachean Ethics*. Unsurprisingly, this renewed interest in Aristotelian virtue ethics has also instigated a debate between supporters and detractors. Criticism related to the feasibility or application of Aristotelian virtue ethics, as well as concerns related to its purported relativism and subjectivism, are among some of the common objections. Arguably still more common, however, is, what is often referred to as, 'the charge of egoism'.

The so-called 'charge of egoism' argues, much like the name asserts, that virtue ethics is egotistical and should, on that ground, be renounced as an ethical framework. Of course, the claim that Aristotelian virtue ethics is egotistical naturally leads one to ask what 'egoism' is understood to be and whether or not Aristotelian virtue ethics is, in fact, 'egotistical'. Additionally, however, one may ask if perhaps, the 'charge of egoism' is within the same framework as that of Aristotelian virtue ethics and, if it is not, whether that criticism can hold any worth. This paper will deal with the former set of questions, that is, what egoism is, and whether Aristotle's ethical framework can be understood to be egotistical. The aim of this paper is to present 'the charge of egoism', as it is often related, in order to demonstrate that this charge fails even within its own parameters.

UNDERSTANDING THE CHARGE OF EGOISM

As it was stated earlier, 'the charge of egoism', is perhaps one of the more common criticisms against Aristotelian virtue ethics and, because of the prevalence of this criticism, it may be difficult to find what the charge is exactly. In light of this, this paper will utilize both T.P.S. Angier's paper "Aristotle and the Charge of Egoism" as well as Rosalind Hursthouse and Glen Pettigrove's analysis on the objection of egoism. The Angier paper provides both a concrete example of criticism against Aristotle, as well as a functioning definition of egoism directly from a philosopher who characterizes Aristotelian virtue ethics as egotistical. On the other hand, Hursthouse and Pettigrove are able to provide an analysis of the objections associated with egoism that are more general and can, therefore, serve

to demonstrate that the objections made by Angier are not outliers, uncommon, or uncharitable but are instead fairly common criticisms.

According to Angier "actions... desires and feelings" can all be categorized as "both 'egotistic' and 'altruistic'" but, due to issues of scope, he focuses on egoism as it relates to "actions" (Angier 2018, 459). He goes on say that "the 'holy grail of moral philosophy" is "finding an argument to defeat egoism, and [to] show the rational necessity of altruism" (Angier 2018, 459). Additionally, he states that "altruism [is] fundamentally other-regrading action… while egoism, [is] fundamentally self-regrading action" (Angier 2018, 459). If the objection against Aristotle is that his ethics is egotistical, and if egoism is "self-regarding action", then, it would seem that the objection against Aristotle, as far as Angier is concerned, is that Aristotle's ethics is focused on the self, instead of another or others (Angier 2018, 459). This understanding of Angier is further supported by his definition of egoism which he states is "the view that one is never justified in acting for others' sake alone, and that the well-being of the self must constitute one's ultimate or basic end" (Angier 2018, 460). He does qualify that egoism "is consistent with acting for the sake of others…[but] that the egoist will shun action that does not contribute to his own well-being" (Angier 2018, 460). From there the paper analyzes Aristotle's *Nicomachean Ethics* in various ways to demonstrate how it is that this ethical theory is 'egotistical'. [1] However, for our purposes it is enough to know that the issue at hand for Angier is that Aristotelian virtue ethics is *primarily* concerned with the self instead of others.

Hursthouse and Pettigrove's analysis of common objections against virtue ethics appropriately includes a section on 'egoism' that coincides greatly with the objections raised by Angier. They note that the "egoism objection has a number of sources" and that among them is that "the virtuous agent… acts as she does because she believes that acting thus on this occasion will help her to achieve eudaimonia" (Hursthouse and Pettigrove 2018). This view seems to parallel Angier's conception of egoism as someone who may concern herself with others, so long as it contributes to her own well-being. Hursthouse and Pettigrove go on to say that "a lingering suggestion of egoism may [also] be found in... [a] distinction between" what is "'self-regarding' and 'other-regarding'" (Hursthouse

---

1. This is a very condensed summary of Angier's paper, however, as can be easily discerned from Angier's title the objection is that of egoism and so his conception of the problem itself and how it compares to Hursthouse and Pettigrove's analysis is, for the purpose of this paper, more important than how he arrives at this view.

and Pettigrove 2018). Once again, Hursthouse and Pettigrove's analysis of common objections against virtue ethics has a readily apparent parallel in Angier's own objections. Given Angier's own characterization of the 'charge of egoism' as well as Hursthouse and Pettigrove's wider analysis of general objections we are able ascertain that Angier's own objections have wider, and well enough established, implications to warrant consideration.

Of course, none of this is to say that this is the *only* way of understanding 'the charge of egoism'. Much like Hursthouse and Pettigrove state, 'the charge of egoism' "has a number of sources" (Hursthouse and Pettigrove 2018). Angier as well as Hursthouse and Pettigrove, for example, both mention the issue of 'self-effacement' as it relates to virtue ethics.[2] However, it seems that even that objection itself is rooted in a 'self vs other' concern. But, due practical considerations, such as length and scope, the secondary objections will be placed aside in order to better address the central objection: self-regarding vs. other-regarding ethics.

ARISTOTLE'S POLITICS AS AN ETHICAL TREATISE

In the case of Angier's "Aristotle and the Charge of Egoism" the objection that Aristotle's ethics is egotistical is based on an analysis of the *Nicomachean Ethics*. This is not that surprising, since much of the conversation concerning "whether Aristotle is an egoist or an altruist…has focused on the *Nicomachean Ethics*" (Ray, n.d.). Still, understanding that the *Politics* is not only a part of Aristotle's ethical theory but is also the more foundational text will allow us to better address the claim that Aristotelian ethics is 'egotistical'. This section will provide a brief argument in favor of understanding Aristotle's *Politics* as a part of Aristotle's ethical framework before moving on to argue that the *Politics* is 'other-regarding'. It is worth noting, that Aristotle's 'ethics' can also be found throughout other works, such as the *Eudemian Ethics*, but seeing as much of the scholarship on Aristotle's ethics is focused on the *Nicomachean Ethics* and since 'the charge of egoism' is often brought against the *Nicomachean Ethics* as well, this paper will only endeavor to show the connection between the *Politics* and the *Nicomachean Ethics*.

Understanding that the *Politics* is an ethical work may be difficult for contemporary thinkers for various reasons. The modern tendency, for example,

---

2.   See Angier p.472 and Hursthouse and Pettigrove section (e) .

to sperate ethical courses from political science courses, may perhaps be both a symptom and a cause for the conceptual separation between ethics and politics. Or perhaps, as Adkins notes, the issue may be rooted in languages themselves; "Greek ideas are transmitted from Greek words" and since "not all of [Greek terms] are readily translatable into English" (Adkins 1984,76-77) we may have some trouble understanding the philosophical concepts being relayed. Whatever the reason for this separation, an analysis of the structure and the word choice in the *Politics* and the *Nicomachean Ethics* can serve to demonstrate that the *Politics* is a part of Aristotle's ethics.

The *Politics* opens with an inquiry into the *polis* itself.[3][4] Aristotle notes that "every *polis*" is a "*koinonia*"[5] and that "every *koinonia* aims at some good" (Aristotle et al. 1894, I.1252a1-2).[6] This analysis continues on as he states that the highest good is that which pertains to "what is called the *polis* or the political *koinonia*" (Aristotle et al. 1894, I.1252a1-2).[7] The ethical nature of the *Politics* can already be seen in the original text's repeated use of *agathos* a particularly value laden term, as Adkins notes in "Aristotle's *Ethics* and *Politics*". The *Nicomachean Ethics* opens in a similar fashion, with Aristotle arguing that every *praxis* is "for some good" (Aristotle, n.d.), for some *agathos*.[8][9] Here, the similarities between the words being used becomes easily observed. In addition to the similarities between the words themselves, the structure between the two openings can

---

3. *polis*, or πόλις, is often translated as city or city-state; however, due to some common connotations found in the English that are, arguably, not found in the original Greek I have opted to only to transliterate this word.

4. All translations from Ancient Greek are my own.

5. *koinonia*, or κοινωνία, is often translated as partnership or community; however, much like *polis*, these English translations often come with certain connotations that are not found in the Greek.

6. "πᾶσαν πόλιν ὁρῶμεν κοινωνίαν τινὰ οὖσαν καὶ πᾶσαν κοινωνίαν ἀγαθοῦ τινος ἕνεκεν συνεστηκυῖαν" (Pol.I.1252a1-2).

7. "δῆλον ὡς πᾶσαι μὲν ἀγαθοῦ τινος στοχάζονται, μάλιστα δὲ καὶ τοῦ κυριωτάτου πάντων ἡ πασῶν κυριωτάτη καὶ πάσας περιέχουσα τὰς ἄλλας. αὕτη δ᾽ ἐστὶν ἡ καλουμένη πόλις καὶ ἡ κοινωνία ἡ πολιτική" (Pol.I.1252a3-7).

8. Praxis, or πρᾶξίς, is often translated as practical, but perhaps better understood as 'action'; for reasons discussed in the previous terms it has only been transliterated.

9. "πᾶσα τέχνη καὶ πᾶσα μέθοδος, ὁμοίως δὲ πρᾶξίς τε καὶ προαίρεσις, ἀγαθοῦ τινὸς ἐφίεσθαι δοκεῖ" (NE.I.i.1094a1).

also be seen in that both texts open with an inquiry into "some good" before moving on to discussing various understandings of the 'good' being described. This analytic "method" is likewise explicitly referenced in the original Greek texts but may, otherwise, be lost in translation.[10]

The connection between the *Politics* and the *Nicomachean Ethics* is perhaps most obvious in that Aristotle explicitly notes that the science of finding the supreme good is that which belongs to "politics" (Aristotle, n.d.). In fact, the *Nicomachean Ethics* makes explicit reference to the *Politics* numerous times throughout the text. He states "at the beginning of the Ethics that politikê" is "the science of the practical good" (Adkins 1984, 75) and mentions from I.1094a28-b3 that one should study political science since, as Mulgan notes, "political science is the 'architectonic' or master discipline" (Mulgan 1997, 3). Adkins goes further, noting not only the references to the *Politics* in the *Nicomachean Ethics* but the references to the *Nicomachean Ethics* in the *Politics* as well.

On the subject of the relationship between these two texts, Adkins notes that Aristotle, in the *Politics*, characterizes "the polis [as] an association [*koinonia*] of like people for the sake of the best life, or eudaimonia" which is "the same characterization of *eudaimonia* as in the ethics" (Adkins, 75-76). It is reasonable, then, given the similarities between the two texts with respect to style and content, to see that the *Politics* is an ethical treatise.

Now that we have established that the Politics is an ethical work, it is worth asking where the *Politics* stands in relation to Aristotle's ethical theory. Namely, which is the more foundational text: the *Politics* or the *Nicomachean Ethics*? In the following section I will argue that, under an Aristotelian framework, the *Politics* is essentially prior to the *Nicomachean Ethics*.

ON THE PRIMACY OF THE POLITICS

Aristotle famously says in the *Politics* that "man is by nature political animal" (Aristotle et al., I.1253a2-3) .[11] Some have used this quote to argue that man "is or ought to be a being who is politically active" (Kullman 1991, 1) . However, this

---

10. See "δῆλον δ᾽ ἔσται τὸ λεγόμενον ἐπισκοποῦσι κατὰ τὴν ὑφηγημένην μέθοδον" (Pol.I.iii.1252a17-18) and "πᾶσα τέχνη καὶ πᾶσα μέθοδος, ὁμοίως δὲ πρᾶξίς τε καὶ προαίρεσις, ἀγαθοῦ τινὸς ἐφίεσθαι δοκεῖ" (NE.I.1094a1).

11. "ὁ ἄνθρωπος φύσει πολιτικὸν ζῷον" (Pol.I.ix.1253a2-3).

statement, read within the context of the subsequent statements leads one to understand that Aristotle means to say that the *polis*, the *koinonia*, is essentially prior, and therefore more foundational, than the individual man.

Following the declaration that "man is by nature political animal", Aristotle notes that a man that is without a *polis* is like a draught without its game (Aristotle et al 1894, I.1253a2-3) .[12] Understanding what this analogy could mean would prove somewhat elusive if it were not for the later section where he states that this is in the same way that the *polis/koinonia*, is prior to the house (Arisotle et al 1894, I.1253a13-19). He continues to say that this relationship is the same as that between whole and part: "the whole is necessarily prior to the part" (Aristotle et al 1894, 1253a20).[13] The discussion of *polis* and mereology is brought to a close when he concludes that "the *polis* is prior to each person" (Aristotle et al 1894,. I.1253a25).[14] The analogy of the draught can then be completed in light of the last remark.

In the same manner that the draught is a part of the game, the individual is a part of the *polis*. If a pawn, for example, were to be without the game of chess, the pawn would be unable to exercise its characteristic function, the game on the other hand can continue without a pawn, or any given particular piece. Perhaps, however, it would be best to explain the analogy through fractions. One fourth, for example, can only exist if there is already a *whole* to be divided into that fraction. Even if one imagines that one may take four separate fourths and place them together to make a whole, it would still be necessary to take those separate fractions from *other* wholes.

Now that the priority of the whole over the part has been better explained, we can revisit the text and observe that, for Aristotle, the *polis* functions as the whole, while each person is the part. The statement "man is by nature political animal" (Aristotle et al 1894, I.1253a2-3), then, is not a statement of 'being politically active', but a statement of the foundational nature of human beings, where the *polis* takes priority, so that "the 'political' is the fundamental human characteristic from which the *Politics* proceeds" (Kullmann 1991, 112). Given that information, it would seem that the *Politics*, which "exists for the sake of 'the good life' [or *eudaimonia*] of the *polis/koinonia*, would have a necessary priority

---

12. "ἅμα γὰρ φύσει τοιοῦτος καὶ πολέμου ἐπιθυμητής, ἅτε περ ἄζυξ ὢν ὥσπερ ἐν πεττοῖς" (Pol I.1253a10-11).

13. "τὸ γὰρ ὅλον πρότερον ἀναγκαῖον εἶναι τοῦ μέρους" (Pol.1253a20).

14. "ὅτι μὲν οὖν ἡ πόλις καὶ φύσει πρότερον ἢ ἕκαστος, δῆλον" (Pol.I.1253a25).

over the *Nicomachean Ethics*, which is concerned only with the *eudaimonia* of the individual.

ARISTOTELIAN VIRTUE ETHICS AS 'OTHER-REGARDING'

Now that we have established that the *Politics* is not only an ethical work but is, necessarily, the foundational ethical text for Aristotle, we can revisit the charge that Aristotle's ethics is egotistic on the ground that it is self-regarding. If we recall, the charge of egoism is often understood as being foundationally 'self-regarding' as opposed to foundationally 'other-regrading'. As Angier states, for an ethics to be altruistic instead of egotistic there needs to be a starting point of others. An altruistic ethical framework could, however, "be consistent with acting for one's own sake" so long as "the altruist shun[s] action that is for no one's sake except his own" (Angier 2018, 460). As we have previously established, the *Politics* is the more foundational ethical text for Aristotle. And, given that the *Politics,* like the rest of Aristotle's ethical works, is a 'practical text' not a 'theoretical' one, the *actions* concerned with the *agathos* of the *polis/koinonia* would hold priority over the *agathos* of the individual. To restate this in a different manner, for Aristotle, the whole precedes the part so that the good of the whole would likewise, and necessarily, precede the good of the part.

The good of the *polis/koinonia*, however, is unambiguously other-regarding in that it is concerned with the community, with others, and not with the individual, the self. Since 'man is by nature a political animal' it would be impossible, under an Aristotelian view, for anyone to act merely for his "own sake" (although perhaps, one may *mistakenly* believe that one can act and affect solely one's self) in the same way that it would be impossible for a pawn to act as a pawn without a game, without others. Even if one assumes that the *eudaimonia* of the *polis/koinonia* would necessarily imply the *eudaimonia* of the individual, this would not make it egotistic since the altruistic person is able to act "for his own sake" so long as he is primarily other-regrading which, given the priority of the *Politics* over the *Nicomachean Ethics,* would necessarily be the case. And so, even within the parameters expressed by 'the charge of egoism' as related by Angier and generalized by Hursthouse and Pettigrove the Aristotelian ethical framework must be considered other-regarding, must be considered altruistic.

CONCLUSION

In the beginning of this paper, I stated that it would perhaps be worth investigating whether the concepts of egoism/altruism are at all applicable to Aristotle's ethical framework. I stated in the previous section that, if the priority of the *Politics* is to be taken seriously as an aspect of Aristotle's ethics, it would be impossible for any human to ever do anything without others. This view is perhaps what leads Aristotle to state that "a man without a polis" is either "a beast or a god" (Aristotle et al 1894, I.1253a28-29), the implication being that someone who is ever without others, without partnerships, without associations, without community, is something other than human; perhaps beast, perhaps god, but most definitely not human.[15]

This conception of what it is to be human may, potentially, strike those inclined toward a more liberal understanding of human beings as strange. Liberal thought does traditionally, after all, present an understanding of human beings that is more individualistic than perhaps the ancient mind could have conceived. Where Aristotle offers the naturalness of the *polis/koinonia* the liberal tradition offers the artifice of society, a view easily found in proponents of social contract theory. Perhaps it would very much be worth exploring whether 'the charge of egoism' is less about egoism vs altruism, self vs others, and more about liberalism vs communitarianism. At any rate, what does remain clear is, that if Aristotle's ethical theory is analyzed within the limits set by those who object to Aristotle's ethical theory on 'the charge of egoism' it is best categorized as altruistic.

REFERENCES

Adkins, A. W. H. 1984. "The Connection between Aristotle's Ethics and Politics." *Political Theory* 12 (1): 29–49. JSTOR, www.jstor.org/stable/191380.

Angier, Tom Peter Stephen. 2018. "Aristotle and the Charge of Egoism." *The Journal of Value Inquiry* 52 (4): 457–75. https://doi.org/10.1007/s10790-018-9632-2.

---

15. "ὁ δὲ μὴ δυνάμενος κοινωνεῖν ἢ μηδὲν δεόμενος δι᾿ αὐτάρκειαν οὐθὲν μέρος πόλεως, ὥστε ἢ θηρίον ἢ θεός" (Pol. I.1253a28-29).

Aristotle, Franz Susemihl, and R. D. Hicks. 1894. *The Politics ...: a Revised Text ...* Place of publication not identified: Macmillan.

Aristotle, J. Bywater, Ed. *Nichomachean Ethics.* http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0053.

Baril, Anne, and Allan Hazlett. "The Revival of Virtue Ethics." *The Cambridge History of Philosophy, 1945–2015*, 2019, 223–36. https://doi.org/10.1017/9781316779651.019.

Hursthouse, Rosalind and Pettigrove, Glen. 2018. "Virtue Ethics."In *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/.

Kullmann, Wolfgang. 1991. "Man as Political Animal in Aristotle." In *A Companion to Aristotle's Politics*, edited by David Keyt and Fred Drycus Miller.

Mulgan, R. G. 1991. *Aristotles Political Theory: an Introduction for Students of Political Theory.* Oxford: Clarendon Press.

Ray, Carolyn. "Egoism in Aristotle's Nicomachean Ethics." Egoism in Aristotle's Nichomachean Ethics. enlightenment.supersaturated.com/essays/text/carolynray/aristotleegoism.html.

compos mentis

The Beauty Within the Individuality of the Self

Dakotah Kinsella
Augustana College

ABSTRACT

Individuality (the self) encompasses a single point of view and is subjective. Each self has a specialized mental state which emerges out of conscious experience. While individuals could share experiences, no two selves could share all the same experiences nor could they ascribe the same representations to their experiences. Every self is authentic in that through one's identity, they can develop their own stance which serves as their inner voice. The self is complex due to the nature of identity. Through private conscious experiences one's identity develops which enables one to take an individual stance. This inner voice makes the self authentic and allows each person to ascribe different meanings to their experiences. Therefore, I propose that the self is beautiful due to its complexity which allows the self to be authentic. Beauty is a representation each self can ascribe to things based on prior experiences. The first-person perspective which is a defining characteristic of an individual allows each person to have a distinct way of perceiving beauty. I believe that the self is beautiful due to its individualistic nature. The body projects our appearance into the world, and the self emerges out of the body. The mystery of the self and the meanings one ascribes to their experiences is extraordinarily beautiful.

What does it mean to be a self? No two humans are alike; each human is individually unique. Although identical twins may have the same physical characteristics, this does not mean that they share the same experiences. These different experiences make each unique in that each will attribute different meanings to their experiences. If we take the self to be a collection of unique experiences and perspectives, then the representations that this specific self ascribes to their experiences cannot be fully understood by other individuals. The mystery of the self, and the meaning (representations) that the self ascribes to their experiences is inherently beautiful.

As a non-reductive materialist, I believe that while the self and the body coexist, they are distinct; the self cannot be reduced to physical properties. Since the self is subjective and the body is objective, these entities cannot be explained using the same mechanisms despite their coexistence. I will argue that the self cannot be reduced to physical properties. Let it be clear that I do not identify as a dualist. I do not believe that the self is anything more than the body it arises out of. Without the body, you cannot have the self.

I. THE DISTINCTION BETWEEN THE SELF AND BODY

I propose that the self is subjective. The self consists of a person's thoughts, feelings, and beliefs. The self will be explained through the mind, a component of the self. Mental states, which are subcomponents of the self can be defined as having a sense of what it is like "to be that organism" (Nagel 1974, 436). This sense can only be described by the individual that is this organism. A mental state is an "inward cognitive [perspective] that is specialized for each individual" (Metzinger 2003). This means that an individual's mental state is private and cannot be directly accessed by others. Therefore, mental states consist of a "subjective phenomenon" that is associated with a single point of view (Nagel 1974, 437). I agree that if the self is subjective, then it consists of a single point of view.

For a conscious experience to exist, there must be "something it feels like to be the subject of our thoughts, actions, and perceptions" (Prinz 2011, 147). In other words, a self must have a sense of what it is like to be who they are to have conscious experiences. While individuals could share common experiences with each other, I believe that no individual can possess all the same conscious experiences as another, because each individual has a private single point of

view. Therefore, if the self is subjective, then each person has their own conscious experiences that no other individual can completely possess.

If the self consists of a single point of view, this means that we are restricted to the information within our own minds which does not allow us to understand "what it is like to be" anyone besides our self (Nagel 1974, 438). Consider a dog for example. While we do have information that could help us predict how a dog might behave, this does not mean that we know what it is like to be a dog. Due to the informational restrictions within our minds, we are unable to understand wholly what it is like to be this dog. Additionally, because we do not have the same conscious experiences as this dog, we do not pay attention to the same information, and we therefore cannot know what it is like to be this dog. The single point of view that the mental self operates on restricts the amount of information that we absorb. It is physically impossible for an individual to take in all information that exists in the world. There are conscious experiences that we have through which our brains "interact with the environment" to produce experiences that we are unaware of (Metzinger 2003). Experience is not objective like the objective information our world portrays. Experience is subjective, because we attribute representations to these experiences which therefore gives these experiences subjective meaning. This results in a single subjective self which no other self can fully understand.

In contrast, I believe that the body is objective. Let us consider the brain (body). The organization of the human brain "is more complex than any other system in the universe: yet its basic ingredients are…simple" (Zeman 2008). The body is objective in that the science of the body "remains just that, a science... it works with concepts that have been carefully defined in terms of observations that anyone, with the right expertise and equipment, can make-concepts like… synapses" for the brain (Zeman 2008). If the body is objective, then it is not limited to the first-person point of view. This means that "anyone, with the right expertise" can fully understand the objective physiological processes that occur within the body (Zeman 2008). If the body is not limited by its point of view, we are not restricted by information. In this case, information we receive from the world is accessible for us to use.

II. IS ROBUST SUBJECTIVITY SUFFICIENT FOR SELF-AWARENESS?

In order for an individual to have a subjective and private point of view, the individual must be able to identify themself as being themself (Zahavi 2002). Once one can determine that they are themselves and are therefore distinct from other entities, they have self-awareness which allows them to experience the world (Zahavi 2002). This determination is made possible through "the linguistic ability to attribute…first-person reference to oneself" (Baker 2000 cited in Zahavi 2002, 10). The capacity of self-awareness develops over the course of one's life through "concepts and language" that they are exposed to (Zahavi 2002, 10).

I propose that while the self is subjective in that individuals have a unique sense of what it is like "to be that organism", one cannot be a self without being inducted into a language (Nagel 1974, 436). Individuals rely on language to connect them to the world and others. It is only through acquiring the language where children can gain knowledge from the world through other individuals. If one can only be a self by acquiring a language, then an integral part of being a self is being connected to others who teach us this language. The only way for us to learn a word is "through my and others' experience of these being objects for us, in some common space" (Taylor 1989). One day children will be equipped with vocabulary that they have experienced which can be utilized to explain to others what they are thinking about. This means that information we receive from the world is accessible for us to use. An individual's "perception involves co-perception of self and of environment" (Gibson 1979 cited in Zahavi 2002, 11). The experiences we have in the world allow us to gain knowledge through which we acquire new words to add to our mental dictionary (the mind). This knowledge we gain from the world does not solely manifest in the mind (a component of the self). We can outwardly project the knowledge we have gained through linguistic descriptions. For example, by describing what an apple looks like, one is projecting to the world that they have knowledge about an apple. Through language, we can put into words the experiences we have, and inform others about the information we have gained. Without language, we would have no way to identify ourselves as our interaction with the world and others is "essential to… achieving self definition" (Taylor 1989).

III. DISTINGUISHING PERSPECTIVE FROM POINT OF VIEW

Within the discussion of inner cognitive functions, perspective and point of view could be taken synonymously, however, the distinction between the two is conceptually important. Perspective and representations in this context mean the same thing. Through information we absorb from the world (through our experiences), we are exposed to vocabulary which we can use to describe our experiences. We use this vocabulary gained from previous experiences to ascribe representations to an experience that we are explaining. These representations are generated through vocabulary which must be within our cognitive capacity to be utilized. "All knowledge [that we absorb from the world] is perspectival in character" (Searle 1998). This means that knowledge is assessed through a framework in which individuals use conceptual resources (vocabulary) to describe their inner mental states to the world (Searle 1998). This vocabulary can be shared with others, and thus many individuals could use the same representations to describe their experiences. Where point of view comes into play, is that an individual can actively choose which representations (perspectives) they ascribe to their experiences. Perspectivism is "always mediated by point of view" (Searle 1998). While an individual can develop many perspectives through vocabulary, these perspectives are mediated when an individual chooses which perspectives to ascribe to their experiences by executing their singular point of view. Should this approach be executed, an individual could develop their own voice through which they could explain to others what their experiences mean to them. Individuals can have many perspectives, but they can only have one point of view. No matter how hard one tries, it is impossible for them to portray all the information depicted in the world. This means that we are restricted to the information within our own minds which does not allow us to understand "what it is like to be" anyone besides our self (Nagel 1974, 438). No individual absorbs the same information as someone else. We could utilize the perspectives one uses to explain their experiences to better understand the person, but these perspectives only serve as a guide to help us predict things about this individual. We do not have access to all the information this individual has access to, and therefore, we cannot fully understand what it is like to be this individual.

IV. THE IMPORTANCE OF EXPERIENCE AND THE NATURE OF CONSCIOUS EXPERIENCE

For thoughts to be represented through the use of language, we must first experience the thing that we are trying to describe. For example, for me to describe what a dog is to someone, I first need to experience a dog for myself. Before representations can be explained to another person using language, we have non-linguistic representations that privately exist in our minds. These non-linguistic representations are registered neurophysiologically in that one is privately aware of them; however, these representations cannot be expressed linguistically to others. I argue that if the self consists of a single point of view, then each person is subject to their own conscious experiences.

A conscious experience is defined as "something it feels like to be the subject of our thoughts, actions, and perceptions" (Prinz 2011, 147). While individuals could share common experiences with each other, I believe that no individual can possess all the same conscious experiences as another due to each individual having a subjective point of view. Therefore, if the self is subjective, then each person has their own conscious experiences that no other individual can completely possess. If the self consists of a single point of view, this means that we are restricted to the information within our own mind which does not allow us to understand "what it is like to be" anyone besides our self (Nagel 1974, 438). Experience is not objective like the objective information our world portrays. Because we attribute representations to our experiences, these experiences have subjective meaning. To have an experience "means that there is something 'it is like' for the subject to have that experience" (Zahavi 2002, 14). This likeness of the experience is only possible if one is aware "of the experience itself" (Flanagan 1992 cited in Zahavi 2002, 14).

Both self-awareness and experience are necessary for "subjectivity [to] reveal itself" to an individual (Zahavi 2002, 13). For example, I am self-aware when I am conscious of my perception of a bird in a tree. When an experience is presented in this manner, this experience becomes my subjective experience of which I am self-aware of. An individual has a subjective first-person perspective of the things they are experiencing. First person perspective is distinct from third-person perspective in that through first-person perspective, "we are acquainted with

our own subjectivity in a way that differs radically from the way in which we are acquainted with objects" (Zahavi 2002, 14).

V. THE COEXISTENCE BETWEEN THE SELF AND THE BODY

While I believe that the self (subjective) and the body (objective) are distinct, I argue that the self and body coexist. The body can be explained in objective terms such as through a chemical state which in turn can "provoke changes in mental state" (Gazzaniga 1988). If the mind (self) and the brain (body) depend on each other in order to exist, then they must therefore influence each other. "Since the mind is derived from brain tissue, the state of brain tissue ought to affect the mind" and vice versa (Gazzaniga 1988). This finding supports my claim that if the self and body influence each other, they must coexist.

VI. THE DISTINCTION IN MECHANISMS THAT DEFINE THE SELF AND THE BODY

While the self and the body coexist, because the self and the body cannot be explained using the same mechanisms, the self is distinct from the body and therefore cannot be explained objectively, nor can it be reduced to objective terms that define the body. Recall that the self is subjective. If the self is subjective, then it embodies a single point of view. Additionally, if the self has a single point of view, then objective (materialism) cannot explain the self. I agree with Nagel when he states that "it is useless to base the defense of materialism on any analysis of mental phenomena", because materialism does not deal with the subjective nature of mental phenomena (Nagel 1974, 437). Therefore, if the self is subjective, then objective (materialism) terms cannot be used to explain the self, nor can the self be reduced to objective (physical) terms that describe the body. Because the objective terminology used to explain the body cannot be used to explain the self, the self and the body cannot be explained in the same mechanisms and are therefore distinct from each other.

VII. THE EMERGENCE OF THE SELF OUT OF THE BODY AND PROJECTION

If the self and the body are distinct, then the self can emerge out of the body. This does not mean that the mind (self) can exist without the brain, but rather that the mind "can have emergent properties that become active in guiding the workings of the [body] that gave rise to them" (Gazzaniga 1988). "The emergent properties of the brain [body], the operating rules of the system we call the mind [self], can push information around…[so] that the actual functioning of the nerves can be influenced by what the mind does" (Gazzaniga 1988). While the self and the body coexist, the body is what projects our appearance into the world. "We know the self only through phenomena" which is directed outward facing the world (Prinz 2011, 148). This unique identity is shaped through the experiences we gain which allows us to develop our sense of self. As Metzinger states, "consciousness is the appearance of a world…if you are conscious, a world appears to you" (Metzinger 2009). Through this world, we appear, because our self emerges from the body. "Consciousness is a very special phenomenon, because it is part of the world, and contains it at the same time" (Metzinger 2009). Our body is a part of the world, and through our sense of self we can contain this world. Through consciousness, "a reality appear[s] within itself. It creates inwardness; the life process has become aware of itself" (Metzinger 2009). It is this sense of inwardness, the single point of view of the self that makes the self and body distinct. Through the self, we can be consciously aware of objects in the world by perceiving them through our own lens. For example, as Merleau-Ponty states, "it is not the contingent aspects of my bodily make-up…which force me to see the surroundings vaguely if I want to see the object clearly. Even if I knew nothing of rods and cones...to look at an object is to plunge oneself into it" (Merleau-Ponty 1962, 24). "In other words: to look at an object is to inhabit it" (Merleau-Ponty 1962, 24). Through this process, the self gazes at an object, and makes perceptions based on the way they view the object. We can view objects, because they "form a system [of our] world" (Merleau-Ponty 1962, 25). If we choose to pay attention to objects, then we can make our own individual perceptions of them. This is what it means for the self to have a conscious experience: when "a single and unified reality becomes present" to an individual (Metzinger 2009). Conscious experience is an internal affair that is subjective, meaning that each person's conscious experiences are

private. Because the body is an object, it projects shadows onto the wall of the world. The body gives us our appearance in the world, and the self emerges from the body therefore activating conscious experience. "Our conscious experience of the world is systematically externalized because the brain constantly creates the experience that I am present in a world outside my brain" (Metzinger 2009). Our thoughts, emotions, perceptions, and "bodily sensations" are all "integrated into the self model" (Metzinger 2009). Through this inward perspective, the self can form perceptions about anything, and no other individual can fully understand this perception, because they do not have the same self that emerges out of the body and into the world. Therefore, because the self and body are distinct, the self emerges out of the body.

Let us explore what is meant by no individual having the same conscious experience. Consider an individual who is unable to identify facial expressions. While this individual can study all the intricate details of a face, they cannot identify the face as a whole and do not recognize who the face they are studying belongs to. Someone who has no neurological trouble with identifying faces and expressions would have a completely different experience interacting with people than this individual would. Because a person would not have the same conscious experiences as this individual described above, they could use information to predict how the individual described would behave, however, this does not mean that they know what it is like to be this person. If the person described above and a random person were asked to both explain what the same face looked like, their responses would be completely different, due to their diversity in experiences. This diversity in experiences allows each self the ability to maintain their own sense of identity through the conscious experiences they are exposed to.

VIII. THE BODY, PERCEPTIONS, AND SUBJECTIVITY

For us to experience objects in relation to ourselves, our bodies perform necessary movements (Zahavi 2002). There is a relationship between perception and bodily movements in that I "can only perceive and use objects if [I am] a body" (Zahavi 2002, 19). This means that if I know the position of an object relative to my subjective self, then I can pick up this object. My perception of the object must contain some information about myself, so that I can act on it (Zahavi 2002). For example, when I am aware of my interaction with a pencil, I can

then use this pencil to write. When I experience a door and its position from me, I am self-aware of the relationship between my personal perception of the door and its existence in the world. The body is "present in every project and in every perception" (Zahavi 2002, 21). Our body is a part of the world, and through our sense of self we can contain this world. The body allows us to exist in the world as an entity that others are physically aware of. Through the body, I can perceive objects using my first-person perspective. By existing in the world, the body "is present…as myself" (Zahavi 2002, 21). Therefore, the self and the body present my subjective self to the world which is different from any other self.

IX. INDIVIDUALITY AND THE AUTHENTIC SELF

In order to understand what it means to have a sense of self the concept of individuality must also be explained. Individuality is "thoroughly determined from or by the totality of its logical, historical, social, and psychic conditions" (Heller et al. 1987). Individuality is subjective. What I mean by this, is that "individuality… withdraws from…objectification" (Heller et al. 1987). Individuality encompasses a single point of view and is subjective in that "each [person] carries a productive uniqueness within… the core of his being" (Heller et al. 1987). This means that no other person can attribute the same representations that this individual has to their own personal experiences. The individual's sense of self is inaccessible to all others due to the restriction of access to information within our minds.

There is something beautiful about being an individual. This beauty is known as authenticity which gives individuals their own sense of being in this world. Authenticity can be defined as "each of us [having] an original way of being" (Taylor 1992). Authenticity and individuality are directly correlated in that "there is a certain way of being human that is *my* way. I am called upon to live my life in this way, and not in imitation of anyone else's" (Taylor 1992). Through our original point of view, "each of our voices has something of its own to say" (Taylor 1992). By listening to our inner voice, we maintain our unique identity through which we can discover ourselves. The self-discovery of an individual "passes through a creation, the making of something original…[and the individual] become[s]" what they have in them to be themselves (Taylor 1992). This understanding of what it means to be this self is ineffable to all others. This means that the self has a single point of view, and conscious experiences that no other person can possess.

X. MENTAL REPRESENTATIONS AND THE SELF

If the self consists of a single point of view, then each person is subject to their own conscious experiences. If each person is subject to their own conscious experiences, I argue that everyone is authentic in that no person can ascribe the same mental representations to their experiences. Mental representations can be defined as "a process by which some biosystems generate an internal depiction of parts of reality" (Metzinger 2003). Mental representations are internal states through which an individual can ascribe meaning to their experiences. "This content can only be accessed in a special process" "because the causal properties making it available for conscious experience are only realized by a single person" (Metzinger 2003). This means that the self has a private point of view which no one else has access to. Therefore, if the self has a single point of view, then each individual is authentic in that no person can ascribe the same mental representations to their experiences.

To better conceptualize this, let us refer to mental representations as aesthetics. Aesthetics in a broad sense is the appreciation of beauty by an individual. One example of aesthetics is taste which can be defined as "the capacity of appraising the beautiful" (Cohen & Guyer 1982). Taste is an aesthetic judgement that individuals use to choose what they consider beautiful. Through taste, an individual can use their imagination and ascribe mental representations based on their experiences to describe what is beautiful to them. Beauty can be appraised by reflecting on "a plurality of representations with one another in relation to a concept" (Cohen & Guyer 1982). These representations must be within an individual's cognitive capacity, meaning that they must have experienced this concept before to ascribe representations to it. I believe that while beauty emerges out of a biochemical material (the body), beauty itself is not materialistic. Beauty is a representation that individuals can ascribe to an object based on prior experiences. I believe that each individual ascribes a different meaning to what they find beautiful. Beauty in this case is meaning. Individuals can describe meaning (beauty) to people, places and things based on their personal experiences. The information that individuals choose to take in depends on the experiences they have encountered which shapes the way in which they ascribe representations (meaning) to objects, people, or situations. Therefore, because no two individuals

can share the same experiences and point of view, no one can ascribe the same mental representations to things that they encounter.

Consider an individual whose visual experiences may differ from your own. This individual has good "visual acuity" "in that they have no difficulty seeing a pin on the floor, though sometimes [they] missed [an object] if it was placed to [their] left" (Sacks 1985). In addition, this individual has no problem explaining minute details in a painting such as color, brightness, and shapes, however, they fail to see the painted scene as a whole. Would this individual say that the painting is beautiful in the way that you would? In other words, would they ascribe the same mental representations to this painting as you would? Because this person has unique experiences that no one else possesses (including yourself), they would not describe the painting the same way as you, nor would they attribute the same mental representations (beauty) to this painting as you would. This means that both of you have a different idea based on past experiences of what beauty is, and the mental representations you attribute to this painting cannot be understood by anyone else because they do not have all of the same experiences of beauty as you. This is why beauty is ineffable. No other individual could ever fully understand the individual's reasoning behind why they find something beautiful. Beauty is subjective in that the representations of beauty that individuals ascribe to an object are dependent on the experiences of an individual.

XI. AESTHETIC THEORY AND OBJECTIVITY

Recall that aesthetics is the appreciation of beauty by an individual. When one engages in this practice, this "involves deployment of a certain amount of theory" through which an individual becomes aware of what they are observing (Kraut 2007). To reflect on an experience, one must engage in a higher-order thought process that involves both semantic and syntactic theory (Kraut 2007). This theory is subject to inquiry which through reflection an individual can improve and better explain. Through syntactic theory, an individual could use vocabulary to ascribe representations to a painting they are observing, and through semantic theory, an individual could determine what this painting means to them. Think of aesthetic theories in terms of the purpose of a dictionary. Just as dictionaries "encode information" to describe the majority view of an object or entity within the world, theories seek to describe, codify, and to articulate an experience one

is engaging in (Kraut 2007). Through aesthetic theories, individuals can ascribe representations to a painting that they are observing, however, aesthetic theories do not seek to legitimize. Individuals using aesthetic theory have the right to their own opinion of the work they are observing despite whether others view the work to be beautiful. This opinion (which is neither true nor false because aesthetic theory does not legitimize) is simply a stance one takes given something they are experiencing. One does not need to have experienced this painting before to call it beautiful, they must simply have vocabulary from prior experiences at their disposal to choose which representations they want to ascribe to this painting. Through language, individuals can justify the reason behind why they described a painting they experienced in a certain way, but language is the only way this person's views of the painting could be projected to the world. Language is an imperfect form of communication. While language may connect us to others, it also distances us due to the restrictions we all have when it comes to accessing knowledge of the world. While individuals have the right to their opinions regarding a piece of art, the artworld is also subject to objectivity. "Objectivity is the feature that marks the contrast between what is in the world and what is not" (Kraut 2007). There is a sense in the artworld that there is a "standard of correctness" for how to interpret a piece of art" (Kraut 2007). However, aesthetic theory is not concerned with the correctness of these interpretations. Aesthetic theory is a reflective way that individuals can determine the kind of meaning that experiencing a specific entity has for them. The beauty that individuals find in entities cannot be directly accessed by others due to the inability of language to capture all the knowledge that exists in the world.

"What it is to be a self ([an individual]) is difficult to conceive" (Taylor 1989). The mystery behind the understanding of the self makes each individual beautiful. The self is complex due to the nature of identity. Identity "fundamentally orientates" oneself in the world and provides a framework through which one can determine "what is good…or what [they]…oppose". Identity "is the horizon within which [one is] capable of taking a stand" (Taylor 1989). Knowing where we stand helps us determine "what meaning things have for us" (Taylor 1989). Taking a stand allows one to construct an inner voice, maintain their identity, and create their authentic way of being. "We are not selves in the way that we are organisms…We are living beings with these organs quite independently of our self…-interpretations, or the meanings things have for us" (Taylor 1989). The body projects our appearance

into the world (us as organisms), and the self (interpretations) emerges out of the body. "One is a self only among other selves. A self can never be described without reference to those who surround it" (Taylor 1989). We are present in the world while still being outside of it at the same time. "Our conscious experience of the world is systematically externalized because the brain constantly creates the experience that I am present in a world outside my brain" (Metzinger 2009). Through these conscious experiences, we develop an inward perspective that no other individual can fully understand simply because they do not have the same emergent properties. The mystery of the self and the meanings that the self ascribes to their experiences is extraordinarily beautiful.

XII. CONCLUDING THOUGHTS

This paper points us in the direction of additional areas to explore for future work. Within the scope of this work I emphasize that the first-person perspective which is a defining characteristic of an individual allows each person to have a distinct way of perceiving beauty. From this I conclude that the self is beautiful due to its individualistic nature. I recognize that further research could be helpful in explaining why having a distinct first-person perspective causes the self to become beautiful. I also state that the self is complex due to the nature of identity. Through private conscious experiences one's identity develops through which one can take an individual stance. This inner voice makes the self authentic and allows each person to ascribe different meanings to their experiences. Therefore, I propose that the self is beautiful due to its complexity which allows the self to be authentic. Further research is needed to demonstrate what makes complexity beautiful for the claim regarding the self being beautiful to be strengthened. Future research could highlight a potential opposing view to my claim that the self is beautiful due to its individuality. I acknowledge that others may find the self to be beautiful for reasons other than its individuality, and I hope to address this view in my future work.

REFERENCES

Baker, L. R. 2000. *Persons and Bodies.* Cambridge: Cambridge University Press.

Cohen, Ted, and Paul Guyer. 1982. *Essay's in Kant's Aesthetics.* The University of Chicago.

Flanagan, O. 1992. *Consciousness Reconsidered.* Cambridge, MA: MIT Press.

Gazzaniga, Michael. 1988. *Mind matters: how the mind and brain interact to create our conscious lives.* MIT Press, Bradford Books.

Gibson, J. J. 1979/1986. *The Ecological Approach to Visual Perception.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Heller, Sosna, and Wellbery. 1986. *Reconstructing individualism.* Stanford University Press.

Kraut, Robert. 2007. *Artworld Metaphysics.* Oxford University Press.

Merleau-Ponty, Maurice. 1962. *Phenomenology of perception.* London: Routledge. 15-34. http://jraissati.com/PHIL256/01_MerleauPonty_PhenomenologyOfPerception.pdf

Metzinger, Thomas. 2003. *Being no one.* Massachusetts Institute of Technology.

Metzinger, Thomas. 2009.. *The ego tunnel: The science of the mind and the myth of the self.* Basic Books.

Nagel, Thomas. 1974. "What is it like to be a bat?" *The Philosophical Review* 83(4), 435-450. https://warwick.ac.uk/fac/cross_fac/iatl/study/ugmodules/humananimalstudies/lectures/32/nagel_bat.pdf

Prinz, Jesse. 2011. "Wittgenstein and the neuroscience of the self." *American Philosophical Quarterly* 48(2), 147-160. https://www.jstor.org/stable/pdf/23025084.pdf

Sacks, Oliver. 1985. *The man who mistook his wife for a hat.* Gerald Duckworth.

Searle, John. 1998. *Mind, Language, and Society.* Basic Books.

Taylor, Charles. 1989. *Sources of the Self: The Making of Modern Identity.* Harvard University Press.

Taylor, Charles. 1992. *The Ethics of Authenticity*. Harvard University Press.

Zahavi, Dan. 2002. "First-person thoughts and embodied self-awareness: Some reflections on the relation between recent analytical philosophy and phenomenology." *Kluwer Academic Publishers* (1), 7-26. https://link.springer.com/content/pdf/10.1023/A:1015568518276.pdf

Zeman, Adam. 2008. *A Portrait of the Brain.* New Haven: London: Yale University Press.

compos mentis

Missing the Matter of Fact: Justification in Testimony

Jacob Kokoris
University of Illinois Urbana-Champaign

ABSTRACT

In this paper, I argue that reductionism fails as an adequate account of testimonial justification. In order to do so, I will offer Jennifer Lackey's objections against reductionism. I challenge Lackey's objection and argue that reductionism fails because of asymmetry availed by the positive reasons thesis. In all, I argue that reductionism fails because of the ambiguities and resulting epistemic asymmetry of the positive reasons thesis.

How do we acquire justified beliefs through testimony? Answers to this question generally fall into two camps: reductionism and non-reductionism. Reductionism and non-reductionism argue, differently, that certain conditions must be satisfied in order for a hearer to acquire a justified belief from a bit of testimony. In this paper, I will be largely focused on the reductionist account of testimonial justification. Reductionism, largely credited to Hume in his *Of Miracles* (1748), is the thesis that testimonial justification just reduces to other epistemic faculties, including induction and memory.

In *Learning from Words* (2008), Jennifer Lackey articulates the reductionist view of testimony. More specifically, she contests that reductionism fails to provide a satisfactory account of testimonial justification. In other words, reductionism fails to provide conditions that explain how a hearer may acquire a belief from a speaker through testimony. In this paper, I will argue that reductionism fails as an adequate epistemology of testimony. However, I will also argue that Lackey's objection to reductionism is unsuccessful. My objection to Lackey's analysis will serve to elucidate what I find faulty in the reductionist thesis- the ambiguity and lack of restrictions on what is called the positive reason requirement. My process in reaching this conclusion is as follows: 1. I will articulate the reductionist thesis, as offered by Lackey. Of import in this section will be the matter of fact or the questions of "what reduces to what"; 2. I will argue that Lackey's example, UNNESTED SPEAKER, fails to refute the Reductionist thesis; 3. I will expand on the failing of UNNESTED SPEAKER, and incorporate the analysis of C.A.J Coady's, *Testimony: A Philosophical Study* (1992), in order to explain what I think the real failure of Reductionism is. In all, I argue the reductionism fails because there is a lack of restriction/generality on the positive reason requirement, a problem with devastating epistemic consequences.

Before articulating Lackey's analysis and my objection, I'll characterize reductionism in testimony. The general thesis of reductionism states that a hearer is justified in believing the words of a speaker just in case they have independent, non-testimonial reasons for doing so. Perhaps my friend, who I know to be a very reliable testifier, tells me that the latest train from Chicago to Champaign departs at 9:50. Here, I am justified in accepting his report because I know him to be a reliable speaker. My reasons for believing his testimony *reduces* to my belief in his reliability, in other words, the positive reasons I have for believing him. The belief I acquire, that the train leaves at 9:50, is justified by observed instances of my

friend's reliability. Reductionism is the general view that testimony is a species of other sources of evidence since the justification for testimonial beliefs typically lies in inductive inference, memory, or perception (Coady 1992, 80) This is in contrast to the non-reductionist, who claims testimony is a source of knowledge on par with those other sources.

There are two specific sects of reductionism, *global* reductionism, and *local* reductionism. We can frame the difference between both global and local reductionism in terms of their epistemic relata—how each answers the question of "what reduces to what?" According to Lackey, global reductionism "is (the view) that the justification/warrant of testimony as a source of belief reduces to the justification of sense perception, memory, and inductive inference, (Lackey 2008, 145) and further, "...in order to justifiably accept a speaker's report, a hearer must have non-testimonial based positive reasons for believing that testimony is generally reliable," (Lackey 2008, 145) Global reductionism requires a hearer have a good reason to believe that testimony, as a source of belief, is generally reliable. Global reductionism is "global" because our justification for some testimonial belief *reduces* to the reliability of testimony as a general source of knowledge. This kind of requirement demands a hearer to evaluate a wide range of reports in order to establish the general reliability of testimony. I won't discuss objections to global reductionism in length, but there are a few important ones to consider. One challenge to the global reductionist view is the near impossibility of establishing the general reliability of testimony. Coady (2002, 82) remarks that it "seems absurd to suggest that, individually, we have done anything like the amount of fieldwork that reductionism requires." Another challenge to the global reductionist thesis is that, in order to evaluate the reliability of testimony, one must accept a variety of testimonial reports. However, we aren't justified in accepting testimony unless we have reason to believe its general reliability. So, a circularity in justification arises: we need to accept testimony to evaluate it, but we can't accept it unless we know testimony is reliable. These two objections are articulated in some fashion by both Lackey (2008) and Coady (1992), but Lackey presents a third unique objection to global reductionism.

The third objection goes something like this: let's say I establish the general reliability of testimony, in accordance with global reductionism. Presumably, this sort of general reliability is based on an expansive variety of reports. Some reports may be about science, some about sports, some about farming, perhaps. The

general reliability of testimony depends on a massive collection of specific types of reports. If we want to establish general reliability, we want a general set of reports. But what does this established reliability say about the accuracy of a given report? Lackey expresses this challenge in an example she offers (Lackey 2008, 145). According to Lackey, let's I establish that 70% of all reports are true. Next, I'm confronted with a report about my friends' child's accomplishment, and reports about friends' child's' accomplishments are only accurate 15% of the time In this case, the general reliability of testimony says very little about the general reliability of a given report. In other words, global reductionism is a poor justificatory system because it says very little about the reliability of specific types of reports.

Considering these shortcomings, a more favorable, version of reductionism, *local* reductionism, emerges. According to Lackey, "The second version of reductionism-often called local reductionism- is that the justification/warrant of each particular report or instance of testimony reduces to the justification/warrant or instances of sense perception, memory, and inductive inference." (Lackey 2008, 148) Local reductionism evades the problems that plague global reductionism. One must possess positive reasons for believing a speaker on a *particular* report. Issues including the reliability and competence of individuals are of import, now. The obvious draw for local reductionism is that justification for accepting testimony is easier to come by. It's easier to establish the reliability of individuals or specific reports than it is to establish the general reliability of testimony. Lackey articulates two distinct ways of understanding the local reductionist's positive reasons thesis. "Positive reasons thesis" simply refers to the requirement that a hearer have positive reasons for believing a speaker. The two versions read:

1. PR-N: Appropriate positive reasons are necessary for testimonial justification

2. PR-N&S: Appropriate positive reasons are necessary and sufficient for testimonial justification. (Lackey 2008, 148)

Lackey thinks that PR-N&S is a better way to capture the positive reasons thesis. Since the local reductionist wants to argue that the testimonial belief is *reducible* to the positive reasons for belief, the epistemic status of each, positive reason and testimonial belief, must be equivalent. You can't have a positive reason with high epistemic status, and a testimonial belief with low epistemic status. In that

instance, the testimonial belief isn't reducible since it enjoys a different epistemic status than the positive reason in question. PR-N&S is better suited to capture the need for this equivalence. To say that positive reasons are sufficient for testimonial justification is just to say that each enjoys the same epistemic status- they are symmetrical. Lackey goes on to argue that PR-N&S is false in her rejection of reductionism. In order to reject PR-N&S, Lackey must show that positive reasons for believing a speaker are not always sufficient in acquiring justified testimonial belief. Otherwise put, there must be a case where positive reasons for belief are epistemically excellent, yet the testimonial belief being reduced is unjustified, or of low epistemic status. Lackey offers an example, NESTED SPEAKER, to make her case (Lackey 2008, 149). In the interest of clarity, I've re-formatted Lackey's version of NESTED SPEAKER.

> Cast:

> **Fred**   Friend of Helen, who believes Pauline based on Helen's word.

> **Helen**   Fred's friend, a highly reliable testifier

> **Pauline** Helen's friend, a highly unreliable testifier

1. Fred has excellent epistemic reasons for believing that Helen is a reliable testifier. She has always been truthful in her reports to Fred, on a wide range of topics.
2. Helen tells Fred that Pauline, her friend, is a highly trustworthy person, especially when it comes to information about wild birds.
3. Fred, then, believes Pauline when she tells him that "Albatrosses have the largest wingspan among wild birds"
4. Even though Helen is generally reliable, she made a mistake on this occasion. Pauline is an incompetent, insincere speaker, *especially* regarding info about wild birds.
5. Pauline is correct in her report about albatrosses, but she came to hold this belief by wishful thinking, not any reliable method. (Lackey 2008, 149)

Fred's positive reason for believing Pauline	Testimonial-based belief that Fred acquires from Pauline
↓	↓
Helen's excellent reliability **(Justified/High Epistemic Status)**	Albatrosses have the largest wingspan of wild birds **(Unjustified-poor Epistemic Status)**

The diagram above is designed to show the different epistemic status of 1) the positive reason Fred has and 2) the belief that Fred acquires. The problem is that Helen's reliability is an excellent positive reason for believing Pauline, but the belief Paul acquires is an unjustified true belief. Pauline gained the true belief by sheer luck, rather than the use of any reliable capacity. It just so happens her belief turns out to be true, and she relays it to Fred. The best way to think of NESTED SPEAKER is as a counter-case to PR-N&S. Recall the PR-N&S thesis means that possession of a positive reason necessarily puts one in contact with a true, justified belief. Fred possess excellent reasons for belief, but those excellent reasons do not put him in contact with a justified belief. This example, Lackey thinks, proves the PR-N&S thesis false, which, by extension, demonstrates the inadequacy of local reductionism as a theory of justification. Lackey thinks that this example displays an asymmetry between the positive reason and the epistemic status of the belief. Since NESTED SPEAKER is a counter-case to PR-N&S, and PR-N&S is vital to the local reductionist, local reductionism fails to provide an adequate theory of justification. To summarize this point, Lackey says, "For, if the Reduction thesis were correct, there wouldn't be any difference between the epistemic status of the testimonial belief being reduced and the positive reasons doing the reducing," (2008, 151).

I do not think NESTED SPEAKER is a fatal objection to local reductionism. More specifically, I believe the PrN&S survives the example of NESTED SPEAKER. Here is the broad scope of my argument: Lackey's refutation of local reductionism depends on her refutation of PrN&S. Her refutation of PrN&S is based on a single case where excellent positive reasons is not sufficient for testimonial belief. However, NESTED SPEAKER wrongly assumes that "Helen's reliability" is the only logically permissible positive reason available to Fred. There are other positive reasons in NESTED SPEAKER, which do not lead to asymmetry. So, Lackey's refutation of local reductionism hinges on a single case that fails to explore

alternative positive reasons. The availability of other logically permissive positive reasons is important because it alters the question of "what reduces to what?" NESTED SPEAKER shows an asymmetry between a single positive reason but fails to narrow the scope on positive reasons. Going forward, I'll explore the epistemic consequences of ignoring the availability of other positive reasons.

First, recall the third objection Lackey provided to the *global* reductionist thesis: the general reliability of testimony says little about the accuracy of specific types of reports (2008,1 47). The attractiveness of local reductionism, in contrast, is the fact that the positive reason says a lot about the accuracy of testimonial report. Local reductionism provides an account where the positive reason is closely related to the acquired testimonial belief, simply because the positive reasons employed are specific, sensitive to the report in question. What this means is that the spirit of local reductionism requires a positive reason as closely aligned to the acquired belief as possible. In other words, the more attuned the positive reason for believing is to the report in question, the higher chance of symmetry between positive reason and acquired belief.

We can extend Lackey's objection to NESTED SPEAKER. In Lackey's example, Helen's general reliability is Fred's positive reason for believing Pauline's testimony. But, like the way global reliability of testimony says little about a given report, general reliability may say very little about reliability with respect to a specific topic. Consider that while Helen is generally reliable, she consistently misrepresents the bird-related expertise of others. For whatever reason, Helen reports that certain people are bird-experts when they really aren't, and vice versa. Let's say this weakness isn't due to deception, or any kind of feature that would challenge general reliability. Lackey specifically writes that "Helen told him that Pauline... Is a trustworthy person, *especially when it comes to information regarding wild birds,*" (Lackey 2008, 149, emphasis added).

Let's reconsider the positive reason/ acquired belief chart, now with knowledge that Helen is unreliable in testifying about the bird-related knowledge of others.

Positive reason	Acquired Belief
↓	↓
Helen's testimony regarding Pauline's knowledge of birds, given Helen is unreliable in her reports about others' degree of expertise regarding birds	Pauline's testimony that Albatrosses have the largest wingspan of wild birds

How does this reformulation of the positive reason/acquired belief chart help the local reductionist avoid the asymmetry charge? Asymmetry refers to a difference in the epistemic status between positive reason and acquired belief (Lackey 2008, 148). Epistemic status refers to the justification and proximity to truth. A belief with high epistemic status is justified and proximal to truth. In the original formulation of NESTED SPEAKER, the positive reason for belief is of high epistemic status, and the acquired belief very low. Now, we have a positive reason with a poor epistemic status. Therefore, it would be unwise to believe Helen's testimony in this specific case, because she is unreliable in this specific domain. Further, the acquired belief is the same as in the example. It's a belief with poor epistemic status. There is no asymmetry now, because both the positive reason and acquired belief hold a poor epistemic status. The positive reasons thesis is not threatened by this formulation, because there are no excellent positive reasons to begin with. In a sense, the question of "what reduces to what" never gets off the ground floor; we aren't discerning the quality of the acquired belief because we lack the positive reasons to even consider the truth of the statement. If we frame this worry in the context of NESTED SPEAKER, if Fred believes Pauline on the basis of Helen's word, then he is a bad epistemic agent. He employs reasons with low epistemic status to justify his acceptance of an unjustified belief. In this sense, the reformulation of positive reasons erases the charge of asymmetry and brings a new charge: epistemic irrationality on the part of Fred.

The crux of this investigation deals with the ambiguity of "positive reasons" We saw that the difference between global and local reductionism dealt with the scope of the positive reason. I don't think local reductionism has a properly fleshed notion of positive reasons. There is no requirement as to which positive reason must be used for justification. So, Lackey's (2008, 151) use of the "general reliability of Helen" does not prove the PRN&S false. Instead, it shows that there is an asymmetry between Helen's reliability as a specific positive reason, and

belief acquired through testimony. I think that Lackey's objection misses the mark because it depends on a specific positive reason.

Lackey's response to my concern may go something like this: Fred has established Helen's credibility through a long series of matching her reports to facts about the world. Fred has done enough to establish Helen as a reliable testifier. Fred has not independently verified Helen's reliability in her reports about other's knowledge of birds. Regardless, Helen's general reliability suffices as an excellent reason to believe her reports. As such, the positive reason of Helen's general reliability is available to Fred, and thus appropriate in justifying his acceptance of her word. Thus, Helen's general reliability is the *most* appropriate positive reason, because it has been verified by Fred's experience. If we accept the general reliability as an excellent positive reason, we see an asymmetry between positive reason and acquired belief. There may be a rare case where some speaker is generally unreliable in a specific domain, *and the* testimonial belief acquired is in that domain. General reliability is a strong predictor of the truth of a given report, and its generality allows a hearer/agent to test it. It is unrealistic to superimpose all these positive reasons that lie in a practically inscrutable domain: who could ever possess the requisite experience to check for the reliability of some speaker's bird references? This response is after a bigger question. It seems like local reductionism needs a condition that restricts the generality of positive reasons. This kind of condition would cast a limit to the number of potential positive reasons used in justification, which may alleviate concerns about asymmetry in justification.

The question that emerges from NESTED SPEAKER and my counterexample, is to question whether there are constraints on the positive reasons employed in justification. We've seen the effects of this uncertainty: in NESTED SPEAKER, using "general reliability" as a positive reason yields an asymmetry and refutation of local reductionism. In my counterexample, NESTED SPEAKER provides no asymmetry, and local reductionism is unscathed. These two starkly different consequences are owed to a difference in the positive reason used in justification. One aspect of the problem is the PrN&S is potentially vague. A more specific articulation than "positive reason" is necessary to understand the matter of fact, and provide a clear answer to "what reduces to what?" To disentangle this problem, it will be helpful to visit C.A.J Coady's treatment of the same issue. In his seminal work *Testimony: A Philosophical Study* (1992) C.A.J Coady remarks on a similar

confusion. In chapter 4 of *Testimony*, Coady investigates global reductionism, and local reductionism. His account is largely a Humean one, which means that his articulation of reductionism is based on Hume's *Of Miracles (1748)*. Coady denies the global reductionism Hume endorses, and formulates a favorable local reductionist argument:

> We rely upon testimony as a species of evidence because each of us observes for himself a constant and regular conjunction between what people report and the way the world is. More particularly, we each observe for ourselves a constant conjunction between kinds of report and kinds of situation so that we have good grounds for expecting this conjunction to continue in the future. (Coady 1992, 82)

Before proceeding, it'll be helpful to flesh out these "constant and regular conjunctions" Thinking of justification in terms of constant conjunctions is just a different way of capturing the positive reason requirement. I am entitled to believe a speaker if the conjunction of (my experience of her reports) and (accuracy of those reports) is constant and regular. The strength of these connections amount to something like reliability: past experiences of truthfulness are a good indicator of future truthfulness. I'll be referring to these conjunctions as "justification conjunctions" for the rest of the paper.

Referring back to Coady, the justification conjunction he's referring to is: (Kind of Report) and (Situation indicated by report). Another way to capture this is the following: I am justified in believing a speaker provided that I have witnessed a past conjunction between the kind of report offered, and the accuracy of that report. Justification amounts to having a track record of accuracy. Roughly speaking, Lackey's local reductionist argument tracked the reliability of the speaker. Framed differently, the observed success of this conjunction, (Speaker's report) and (Situation indicated by report), amounts to positive reasons for belief (Lackey 2008, 148). The important takeaway is that I challenged Lackey's local reductionism on the grounds that her positive reason requirement was vague. Now, it appears Coady has presented a reductionist account that is even more vague. The fact that each account struggles to clarify their positive reason requirements, namely "Kind of report," could mean many things, and avails us of an even more confusing view than Lackey's local reductionism. I won't elaborate on the vagueness of "kinds of

report," because Coady investigates it the vagueness. Keep in mind he intends to refute reductionism, much like Lackey. Regarding the ambiguity of "Kind of Report" Coady writes that "It seems to me that "kind of report" may be meant to refer either to the kind of speaker who gives the report or to the kind of content the report contains," (1992, 83).

So, on the one hand, "kind of report" may refer to the reliability of the speaker, or reliability of the content of the report. To capture the difference, accept Coady's simple example: "My friend testifies to me that "There is a sick lion in Taronga Park Zoo,"" (Coady, 1992, 84). If we take "kind of report" to mean reliability of the speaker, then I'm accessing my friend's general reliability, and the accuracy of his reports. If I take "kind of report" to mean the content of the report, then I'm tasked to find out whether reports about lions are generally reliable. Moreover, I may be tasked with figuring out whether reports about sick lions are reliable, or whether reports about Taronga Park Zoo are reliable. Perhaps reports about Taronga Park Zoo are highly reliable, and I have confirmed this reliability through my own experience- when people report about Taronga Park Zoo, the reports are accurate. However, I have little experience in confirming reports about sick lions, and thereby lack the grounds to justifiably believe the report in question.

Another way of framing this issue is that there is no clear classification requirement for "kinds of report" and there are blatant epistemic consequences for this classification issue. Whether we classify the report as a zoo report we are justified in accepting, but if the report is classified as a "sick lion" report, it appears we are unjustified. So how would we go about classifying the kind of report, and by extension, our justification be believing some report? The concerns that arise from the Coady example start to resemble the concerns I showed with NESTED SPEAKER. There must be a way to classify "kind of report" because the myriad of ways we are able to determine the scope, reference of "kind of report" has clear consequences. Regarding this issue, Coady writes:

> Since either classification is logically permissible, then it seems to be purely a matter of whim whether Jones (the hearer in the zoo example) has or has not good reason for accepting the report. Clearly some sort of non-arbitrary restriction on the scope of "report of a kind of situation' is required to make this notion of any real value in the elaboration of RT'. (1992, 84-85)

RT' refers to Coady's reformulation of Hume's global reductionism and can be understood as a local reductionist account. In this passage, Coady acknowledges that there are several logically permissible ways to classify our justification conjunction. Put simply, there are several logically permissible ways to justify belief of some report. The problem amounts to something like this: Reductionism lacks a condition regarding classification of the report. In other words, there is no way of determining the appropriate positive reasons for a given report, and numerous positive reasons are logically permissible. The problem here is that while numerous positive reasons are logically permissible, there are different epistemic consequences depending on which reason you accept. Certain types of report are unreliable, others are reliable. Furthermore, the reliability of certain reports is unknown or inaccessible to the hearer. If we classify the zoo report as a report about zoo animals, then our hearer is tasked with knowing the general reliability of that specific report. If the hearer lacks requisite experience to check such reliability, they are unjustified in accepting the testimony. The reverse is true if we think of the "kind of report" as referring to the kind of speaker. In this case, the hearer may have requisite experience for determining the speaker's reliability, and by extension is justified in accepting the report. It should be noted that Coady does not pursue this objection very far. He offers the issue of generality but leaves it to the reader to pursue further.

There's another issue that arises with Coady's example. Let's grant that we find a way of determining the kind of report, and the justification needed to accept the report. In many cases, the necessary positive reasons will be beyond the speaker's experience. Perhaps we classify the Taronga Park zoo example as a report about sick lions. Our hearer lacks the requisite experience to determine the reliability of the report, given that there is a restriction on the justification. The hearer has no experience with reports about sick lions. However, our hearer has plenty of experience with the reliability of the speaker, his friend. It seems intuitive that the reliability of the speaker justifies the hearer's acceptance. But since we have classified the report as a "sick lion report" it's unclear just how far the speaker's general reliability will take us in terms of justification. The hearer must be justified with regard to the kind of report, and that kind of report is "sick lion reports" not, "the reliability of the speaker's reports." So, the problem is that finding a way of classifying restrictions on reports makes justification a whole lot harder to come by. Further, acceptance of testimony would require a lot of fieldwork that most

people simply couldn't do. This objection mirrors Coady's point that "many of us have never seen a baby born, examined circulation of the blood..." (1992, 82). Coady brings up this point before arguing for the "kind of report" ambiguity. I think the practical objection regarding field work/observation holds even if we find a way to classify kinds of reports, and thereby demand specific kinds of positive reasons. Given this restriction, and the requirement of more work, perhaps a hearer is entitled to employ general positive reasons. For instance, lacking insight on the accuracy of "sick lion reports," a hearer is entitled to opt for something like the general reliability of the speaker. Although general reliability may not speak to the accuracy of the given classification, it's surely a predictor of arriving at true beliefs. The alternative, given a rigid classification and lack of positive reasons, would be a suspension of judgment.

So far, I have argued that Lackey's refutation of local reductionism is unsuccessful and have objected to local reductionism through the lens of Coady. Exploring the "kinds of report" ambiguity will provide better grounds for challenging Lackey, specifically on her asymmetry charge. Lackey argues that local reductionism fails because of the PRN&S condition, that possession of excellent positive reasons for believing the speaker justifies the hearer in believing the speakers testimony. The PRN&S failed because NESTED SPEAKER showed a case in which a hearer possessed excellent positive reasons for believing yet acquired an unjustified belief. The high epistemic status of the positive reasons compared to the low epistemic status of the acquired belief is the gist of the asymmetry charge. Because of the asymmetry, PRN&S fails, and since PRN&S is vital to local reductionism, local reductionism fails. My objection to this argument is that if you change the positive reason for belief in NESTED SPEAKER, then you encounter a situation in which there is no asymmetry between the positive reason and acquired belief. Considering Coady's discussion of the "kind of report" ambiguity, it's clear that local reductionism lacks a clear restriction regarding classifying reports. These restrictions, or ways of classifying reports, yields different positive reasons (A report classified as "sick lion report" is justified by the accuracy of "sick lion reports"). Let's revisit NESTED SPEAKER with the classification ambiguity in mind.

Let's once again assume we have a way of classifying reports according to Coady. Let's say NESTED SPEAKER will be classified as a "report about birds." We can call this case one. As such, Fred (hearer) must possess some general assessment of the reliability of "bird reports." The justification conjunction here

would look like (Past reports about birds) ^ (Accuracy of those past reports). If Fred has witnessed high correlation between these conjuncts in the past, he is justified in believing Pauline's word. This example is no different than NESTED SPEAKER as given, since we have some excellent positive reason for belief, and the acquisition of an unjustified belief. The only difference between case one and NESTED SPEAKER is that the excellent positive reason focus has shifted from Helen's reliability to "bird reports" in general. Given this classification, the asymmetry charge holds. Now case two: NESTED SPEAKER is classified as a "report about Helen's testimony regarding other speakers' knowledge of birds" The appropriate justification conjunction is: (Times Helen has testified about others' knowledge of birds) ^ (Times those people demonstrated knowledge of birds).

This is an attractive conjunction. Because of the extremely narrow restriction on the kind of report, positive reasons for belief should be based on Fred's observation of Helen's testimony in that extremely narrow domain. If we accept that a kind of report requires excellent positive reasons within that report's domain, then we clearly have a situation where Fred lacks positive reasons. On the rare chance he has observed Helen's testimony in this regard, then he is presented with her unreliability in this regard. On the probable chance that he lacks an observance of her testimony in this domain, then he simply lacks positive reasons for belief, because he has no reasons. But there are two possible readings of the "kind of report" confusion. The first has been established: types of reports requires types of reasons within the domain of the report. The second was brought up above. If a hearer lacks positive reasons in the specific domain, they may favor a more general positive reason. In the context of NESTED SPEAKER: If Fred lacks observations of Helen's testimony in this specific domain (the conjunction above) he may rely on her general reliability as a speaker as reason for/ against believing. Since, as Lackey writes, Helen is reliable, then Fred can defer to that general reliability if he lacks the specific, domain aligned positive reason. The positive reason of Helen's general reliability leads to asymmetry, as established. But the mere availability of a more general positive reason, and a demonstration of asymmetry, is not enough to defeat local reductionism. If we accept a restriction on kinds of reports, then we accept that excellent reasons for believing that report lie in the domain of the report and are as closely aligned to the kind of report as possible. In the event that such reasons escape the hearer, opting for the general reliability of the

speaker as a positive reason is generally truth conducive, and in the absence of those positive reasons, even more general ones may be available. Reductionism seems to fail because there's an indeterminacy in the identity of "kind of report." In NESTED SPEAKER, Lackey inadvertently answers the question of "what kind of report" is in question by offering "general reliability" as the positive reason for belief. In this sense, submitting "general reliability" opens the window for the viable use of different positive reasons. In other words, the asymmetry charge is avoided because Lackey doesn't rule out the use of other positive reasons.

A better objection to local reductionism deals with the ambiguity regarding "what reduces to what?" There is no obvious restriction/scope to the justification we seek. As demonstrated, the permissibility of different sets of positive reasons avail different epistemic consequences. Two logically permissible sets of positive reasons, as shown, can lead to an asymmetry of justification, or fail to provide an asymmetry. In NESTED SPEAKER, we have two available positive reasons. Choosing "general reliability" led to asymmetry and choosing the one I proposed avoided that charge. We have widely different consequences depending on which route we choose, and both appear permissible. A lack of restriction means a lack of consistency in application: the local reductionist must accept justification, and a lack of justification as permissible for a single bit of testimony. While I agree with Lackey that reductionism fails to provide an adequate epistemology of testimony, I don't think her objection, NESTED SPEAKER, provided a successful objection to the reductionist. Lackey's claim that PrN&S fails depends on a faulty assumption that there is only one permissible positive reason. A better objection to the local reductionist should *start* with the problem of multiple permissible positive reasons. On this note, local reductionism fails to provide an adequate epistemology of testimony because it fails to provide consistent justificatory explanations. The ambiguity of "positive reasons" presents a major epistemic challenge to the reductionist. Different positive reasons lead to different verdicts on justification.

REFERENCES

Coady, C. A. J. 1992. *Testimony: A Philosophical Study*. New York: Oxford University Press Inc.

Hume, David, and Tom L. Beauchamp. 2000. *An Enquiry Concerning Human Understanding: A Critical Edition*. New York: Oxford University Press Inc.

Lackey, Jennifer. 2008. *Learning from Words: Testimony as a Source of Knowledge.* New York: Oxford University Press Inc.

compos mentis

On "Kant Meets Cyberpunk": An Analysis of Science Fiction and Transcendental Idealism

Vivian (Xiao Wen) Li
University of Toronto

ABSTRACT

In this paper, I argue that it is possible to conceive of non-material systems producing phenomenal spatial experiences, provided that we can factor in the notion of the transcendental ideal nature of time. I will first examine Eric Schwitzgebel's "Kant Meets Cyberpunk," outlining his argument that transcendental idealism may be true if an immaterial Cartesian soul underlying our current reality is creating virtual reality spaces that immerse us within the spatial environment around us. Schwitzgebel also argues that we are unable to access this immaterial soul because it exists on a base level of reality, while we live on a simulated reality plane. Moreover, I will examine David Chalmers and Nick Bostrom's notions of the matrix as well as the notion that we are currently living in a simulation. I will conclude with my argument for the transcendental ideal nature of time, as well as consider possible counterarguments and the implication if we accept my view.

ON SCHWITZGEBEL'S "KANT MEETS CYBERPUNK"

In order to explain the argument I will be proposing about the transcendental nature of time, it is important to obtain background from Eric Schwitzgebel's paper, "Kant Meets Cyberpunk." In the article, he argues that transcendental idealism, the proposition introduced by Kant that the world is made of phenomena (the appearances of objects) and noumena (things which are in themselves), could be accurate if conceived of along with ideas in Cyberpunk and science fiction. Moreover, he argues for conceiving of non-spatial systems organizing and bringing about spatial phenomenal experiences in the world. He claims that Kant's transcendental idealism consists of two theses, one, that spatial properties of objects depend on our minds, and two, that the fundamental nature of how the world works cannot be understood by us. The characteristic of spatiality, then, is something we bring to objects, and properties such as length, breadth, and depth are accessible to us because of the empirical evidence we obtain from our environments (Schwitzgebel 2019, 5). However, this is not to say that we completely create the illusion of spatiality in our minds, rather, Schwitzgebel is stating that there exists an underlying layer of reality which we interact with empirically and gives rise to the spatial properties we attribute to objects.

In both William Gibson's book, *Neuromancer,* and *The Matrix* (1999), a virtual reality (a computer generated space where a spatial environment can both be interacted with and is immersive, making the user feel like they are really in the scene) is created where the people submerged inside these realities can visually act and react to these spaces. In the former, hackers link up into Internet and navigate within the network of various computer programs, while in the latter, people are unaware that they are being held in warehouses and fed inputs from a computer device (Schwitzgebel 2019, 7). Hence, in *The Matrix,* people who have not awakened to the reality of the situation believe that they are drinking coffee or relaxing in Miami when their real bodies (at least in the warehouses) are motionless. Schwitzgebel also states that Cyberpunk is centered on the view that if we are submerged within our immersive spatial environments (something that makes us feel we are really within the scene), even if we can interact with empirical objects around us (objects we can touch/see/taste/move around), we cannot tell the difference between the current spatial environment and the fundamental environment which is controlling these phenomenal experiences. In other words,

it is entirely possible that we are living in a spatial manifold (spatial environment where each spatial part is related to another part) which is entirely based on another spatial manifold, the latter of which we have no control over. For instance, our brains may be strapped to machines in a warehouse, but our phenomenal experiences of the world are that of a dance party and drinking wine (with the empirical objects we interact with being the lights at the party and the wine glass, which exist within that specific spatial manifold). Schwitzgebel insists that it is conceivable we are currently in a world not unlike *The Matrix* and *Neuromancer* because everyday people can conceive of the aforementioned situation. I believe that he does not provide a good reason for why we might believe this is the case, a point I will bring up later in the paper.

Schwitzgebel, after pointing out the property of *Sims* (where an artificial entity is sustained by a computer and living in a shared virtual reality space with other artificial entities) in Nick Bostrom's paper, also argues that there is a difference between the base level of reality and the simulated level of reality. For instance, the base level could be that brains are stored in a warehouse and connected through wires, while the simulated level of reality could be the large and colorful world people live in. In order to create an argument for a non-materialist generator on the base level of reality, he moves to consider the Turing machine. If conceived of in a x-y plane, the Turing machine is an strip of tape that extends infinitely in either x or y directions (Schwitzgebel 2019, 13). It has a tip that can move in both x-y directions, which reads the alphanumeric characters on each square on the tape. According to if-then rules in its system, the machine can then erase the old characters on the tape and write new ones and move one square to its left or right, or stay in place. To complete his move of portraying a non-spatial system as an underlying spatial manifold operating at base level for humans, Schwitzgebel introduces a specific version of the Cartesian soul, which he names Angel. The soul can think, have conscious experiences, has causal abilities, and perfect memory, but does not have spatial positions nor traits such as length, breadth, or depth (Schwitzgebel 2019, 14). Schwitzgebel focuses on the notion that the Turing machine can be conceived of in a non-spatial manner, for instance, by replacing the symbols of the machine with musical pitches (low A or middle C) and the tape with integer numbers. Angel can then use his memory to move from one integer number to another, associating his current mental states (ranging from extreme sadness to happiness) with musical pitches, and rewriting the pitches

associated with the integer numbers in much the same way as a Turing machine would rewrite the alphanumeric character on the tape.

Schwitzgebel also considers whether the substrate on which the base level of reality is operating matters, for instance, whether or not transistors in parallel series could impact the speed a system processes things. However, he proposes a scenario with Kate and Peer, attempting to show that the idea of a non-spatial system creating spatial phenomenal experiences is entirely possible. Kate and Peter are two artificial intelligent beings who are enjoying the day, however, Schwitzgebel argues that a Turing machine could be responsible for their experiences. He also states that since it is possible that Kate and Peter could be run by Angel, thus, if Kate and Peter's existence is conceivable, then the notion that a non-material computer could be controlling their mental states and processes is also possible. If we accept his above points, then it is possible that beneath our experiences, there is a Cartesian soul which is immaterial and has instructions to execute and create a virtual reality space for us to live in, where it is impossible to access the soul because it is on a different level of reality than us. Also, Schwitzgebel states that we may experience spatiality only in terms of tracking our interactions and experiences with this immaterial soul (Schwitzgebel 2019, 20). Thus, if we can conceive of the aforementioned possibility, we can see how transcendental idealism may be true (Schwitzgebel 2019, 21).

ON NICK BOSTROM AND DAVID CHALMERS

Since Schwitzgebel employs a specific version of Kant's view in his paper, examining the historical Kant's notions of transcendental idealism may not help us much in analyzing the former's argument. Meanwhile, the notion of simulations and the existence of Sims is an important move towards a nonmaterial program generator at the base level of reality. Hence, I will now introduce and examine Nick Bostrom's article, "Are We Living in a Computer Simulation." In the paper, Bostrom argues that at least one of the following three statements is true: (1) the human race will extinguish before reaching the posthuman age (a posthuman is a being that exists beyond being a human, for instance, a cyborg with neurons artificially connected to processing units) age, (2) a posthuman is unlikely to run ancestor-simulations (which encompasses all mental belief-states of every human being that ever lived), or (3) we are living in a computer simulation at the moment

(Bostrom 2003, 1). I will focus on (3) as this is the section most pertinent to Schwitzgebel's argument.

In the case of (3), the complete mental history of humans will be captured on a computer and multiple ancestor-simulations would be executed without the posthuman needing to spend many resources (Bostrom 2003, 247-248). If (3) is true, and we are living in a computer simulation at the moment, Bostrom argues that there would be levels of reality, as the computer which is running the simulation may not have the same physical laws as the one we are viewing at the moment (Bostrom 2003, 253). The people living in simulated worlds could also become posthuman and create ancestor-simulations based on their reality. Since someone that is being simulated by a computer could also become the simulator of a new civilization, thus, if we (in our current reality on Earth) become capable of creating ancestor-simulations, we would arrive at the conclusion that we live in a simulated reality (rejecting (1) and (2)). Hence, it is likely that we are also living in simulator-realities created by simulated beings, while they are also living within the simulated environments of another group of posthumans.

While we could be ancestor-simulations living in another simulated world, it seems unlikely to me that there exists such a layering of realities. Bostrom seems to be assuming that every posthuman civilization will behave in the same manner, however, this may not be the case. For instance, civilization A could decide that they are interested in exploring ancestor-simulations, and create such a world. However, civilization B may decide, arbitrarily, that it is not necessary to explore these simulations because they wish to access alternate universes and do not want to use up even a fraction of their resources running simulations. Hence, I do not believe that even accepting (3) would lead to the conclusion that we are currently living in a simulation. Moreover, technology may be completely able to simulate belief-states of humanity, however, this might not lead to a completely individual experience of being in such simulations. In other words, while it is possible to simulate mental processes using expert nanotechnology and neuroscience, even this may not lead to complete individual thinking on the part of the ancestor-simulations. This is a point that Schwitzgebel does not address in his discussion on the incorporation of cyberpunk themes with Kant's notions of transcendental idealism.

Bostrom's simulations are also elaborated on by David Chalmers in "The Matrix as Metaphysics." Chalmers argues that the Matrix Hypothesis is not a skeptical

hypothesis (one where many of our beliefs would be incorrect if it was true, and is also impossible to strike out) (Chalmers 2005, 3). To give some background on the Matrix Hypothesis, it is the notion that we have always been a brain in a vat (or a brain being connected to a giant computer simulation of the world and receiving inputs and outputs from the simulation, thereby creating the illusion of beliefs and sensing) and continue to exist as such (Chalmers 2005, 2). Since we cannot be certain we are not within a matrix, many of our beliefs can also be questioned. We could, for instance, think that we are drinking coffee in Miami but this might not be the case, as we could be in a matrix and thus would not be in Miami (assuming the vat cells are not placed in Miami). In essence, we do not know for certain that we are not in a matrix, even though we may believe we are drinking coffee in Miami. If we are in a matrix, we are most likely not in Miami. But we do not know that we are not in a matrix, and thus we do not know if we are in Miami (Chalmers 2005, 2). This skepticism (that is, we believe certain things, but we are not certain of them) can then be applied to everything we believe about the world.

However, Chalmers rejects the notion that the Matrix Hypothesis is a skeptical hypothesis and argues that it is a metaphysical hypothesis, or one that is concerned with the philosophical elements of underlying reality (Chalmers 2005, 3). As such, the hypothesis that we are simply brains in a vat is a conjecture about the nature of our minds and of the fundamental nature of relaity. Furthermore, if the Matrix Hypothesis is accepted, it does not lead to the conclusion that we are not in Miami drinking coffee, since we are interacting with the bits that are representing Miami as well as the coffee and the coffee mug (Chalmers 2005, 4). Although not explicitly stated, Chalmers' points relate to Schwitzgebel's argument since the latter notes that there are distinct spatial manifolds, as well as different levels of realities. I am not particularly convinced by Chalmers' argument about interaction on different levels, as it does not explain which exact bits of reality and virtual reality correlate in the matrix. This is a problem since it seems that reality and virtual reality can be simply reducible to bits and would copy exactly into another, although the process of how this is done is unclear. For instance, the coffee mug I use in the external world would, in the matrix world, correlate to an aggregate of bits in a coffee mug formation. In response, Chalmers would argue that even in our current world, we are unsure of which quantum particles or waves exist in the coffee mug. The reasoning he would present, however, seems to point at a lack

of knowledge in our current reality rather than answering the question of how it is possible virtual reality and reality can easily correlate into each other.

REBUTTALS AGAINST SCHWITZGEBEL

Delving into the analysis of Schwitzgebel's points, he argues that it is conceivable for our phenomenal experiences to be completely controlled by a base level of reality constituting of machines and systematic mechanisms. In other words, he argues that it is possible we exist in spatial manifolds separate from our biological brains or bodies, and live in a shared virtual reality space. However, the only reason he gives for this point is that people are able to conceive of the possibility of this happening, as seen by the success of cyberpunk narratives (Schwitzgebel 2019, 10). This, to me, is not a good enough reason to believe in his argument, as he employs the use of "conceivable" as almost a means of establishing truth, or at least one possibility that potentially contains truth. However, many people can conceive of similar things without them being true at all (for instance, a unicorn). Hence, the reason he gives at this point in his argument appears to be an appeal to the masses fallacy.

Another issue with his argument is within the Kate and Peter scenario, which he uses as a stepping stone to argue for the conceivability of Angel. To reiterate, he stated that since it is possible that Kate and Peter could be run by Angel, if we can conceive of Kate and Peter existing, thus, it is possible that the system running them is non-material (Schwitzgebel 2019, 19). However, this may be a case of circular reasoning, as he first assumes it is possible that Angel runs the system without justifying Angel's existence. Then, he moves to state that Kate and Peter are conceivable, with the implicit assumption that their phenomenal experiences are being generated by a nonmaterial system. At the end of his argument, he states that since Kate and Peter are conceivable, it is possible their experiences are being generated by Angel. His argument moves full circle, as he is ultimately trying to prove his first assumption (that Angel runs the system).

Schwitzgebel also decides to assign the base level of reality as the one and only accurate level of reality. However, it appears to me that following his logic of a base level of reality generating phenomenal experiences for people in the same spatial manifold, there is also the implication another level could also be generating and controlling the base level on which our current reality operates.

As well, it is difficult to state the meaning of "levels," a problem that manifests more in a materialist setting than an immaterialist (like Angel, for instance). For Kant, it might be difficult to state the physical properties and interactions between phenomena and noumena in terms of atoms, quarks, and physical laws. Meanwhile, for Schwitzgebel, there may be a problem regarding how descriptive the base level of reality needs to be to produce a vibrant simulated reality, or even how the base level of one programming entity can generate so much variety and uniqueness, potentially creating multiverses, and can keep itself from overloading with information.

Although there are a few rebuttals that could undermine Schwitzgebel's argument, his claim is that transcendental idealism only *might* be true if incorporated with Cyberpunk themes. Hence, I would like to continue with this line of thought (thinking that transcendental idealism might be true) to see if it is possible to address my aforementioned rebuttals regarding the base level of reality and a vibrant simulated reality. I will explore the possibility of transcendental idealism with the notion that time could also be transcendental, and conceived of with immaterial programming systems existing on different levels of reality.

THE TRANSCENDENTAL IDEAL NATURE OF TIME

It appears that the transcendental ideal property of time may be necessary when approaching the view of an immaterial system creating phenomenal experiences for human beings on the simulated reality level. The reason for this is because following Schwitzgebel's reasoning and logic, there is no reason to stop at just the base level of reality being the fundamental system that functionally controls and operates the world. In fact, there is also a possibility that the spatial manifold the non-spatial system operates is based on another base level, which in turn is functioning to control this reality. (I am drawing inspiration from Christoper Nolan's "Inception" and the notion of a dream within a dream. Instead of a dream, I am proposing that there exists a system within a system, or a reality within another reality). This would solve the problem of how a single immaterial programming unit could create such variety and colour in multiple universes of simulated reality. The transcendental ideal nature of time (arguing that time is subjective and not a feature of things independent from us) would be necessary to start to understand the different functioning realities within each other, since

they cannot all be experiencing the passing of time at the same rate. For instance, since a system A would be the base level of reality for System B, and System B for C, with some time needing to pass before System B can be influenced by System A, it follows that time would not exist independently at all, but rather be subjected to the different Systems which are running the spatial manifold a subject X is living in.

I now consider a few counterarguments:

> (1): One may argue that there must exist a basic operating system (such as in *The Matrix*), as my scenario may be too idealist. For instance, it would be difficult to define why something exists in System Z if we have to map back to System A (following Chalmers, the bits that constitute the coffee mug in System A may become completely warped by System Z). Furthermore, if there is an infinite series of worlds, there has to be an end (or a base reality) to it, or else the idea seems too bizarre.

> (2): Another possible counterargument is the question of how these non-spatial systems could interact with our spatial minds, as it seems easier to explain this with *The Matrix*, given the spatiality of the machines. This is harder to understand in the context of Angel, even without considering Angel 2.0, 3.0, or any other iterations of Angel in my scenario (on Systems from A to Z).

I would respond with the following:

With regards to (1): Within a world as complex as *The Matrix*, it is equally as unlikely to map back specific parts of code within the system to the phenomenal experiences of the characters. Moreover, similar to what Schwitzgebel proposed, in that there is a base level of reality of the brain in the warehouse, where there are other empirical objects located within it, it is likely that the world we conceive of existing could also be conceived within another world. Then, the move that Schwitzgebel makes in his paper would be the same move I make, with the only difference being the added proposal that the base level of reality, including the non-spatial notion of it, could be operating within another simulation that both we and Angel (the Cartesian soul that seems to mimic the Turing machine) are unaware of. As for the bizarreness of the infinite series, I argue that this

scenario could be possible, not that it is entirely the case. It is, moreover, equally conceivable to me as Schwitzgebel's conception of the non-spatial systems creating spatial experiences for human beings on a day-to-day basis. We also live within dimensions that are more complex than just three-dimensional (according to Physicists and Mathematicians), dimensions that we cannot always view and visualize. However, just because we have limited mental capabilities, we cannot then determine that it is definitely not the case further dimensions do not exist outside of our comprehension. This is also the case with realities embedded within realities. By realities within realities, I mean that there are immaterial generators that create virtual reality spaces, which in turn create further generators, with each step requiring time to run differently in every system. Of course, if someone asks me to prove my point with scientific and empirical evidence, it would be difficult to do, as Science seems to view the world as reducible to formulas and fundamental waves or particles. However, I do believe that my view is possible within the frame of Schwitzgebel's argument, and is in fact a logical next step according to Schwitzgebel's points.

(2): Moreover, although we could be experiencing spatiality in our current reality, we may really be non-spatial entities living on a virtual reality space. If so, then there does not seem to be a problem with a non-spatial system interacting and simulating another non-spatial system, at least on a theoretical basis.

CONCLUSION AND IMPLICATIONS OF
ACCEPTING MY PROPOSED VIEW

One implication of accepting my aforementioned view is having non-spatial mechanics structure the phenomenal experience of empirical objects. For instance, it may be possible for machines to create, by themselves, non-spatial mechanics which would structure an entirely new being into existence (as seen by the entities chasing Neo in *The Matrix*). In that case however, we would most likely not realize this occurring, as it would happen on a level of reality that is simulated from our reality, and which (at least following the argument of Schwitzgebel) would rely on our reality also being a simulated one that is currently generated by an immaterial and non-spatial machine (perhaps someone named this machine Angel down the road).

Thus, while I agree with Schwitzgebel (that it is possible that a non-spatial system could create spatial phenomenal experiences), I also note that this could only be possible provided that time can also be seen as transcendentally ideal (not existing independently from us) and that there exists multiple base levels underlying each simulated level of reality (or for us, our current reality as we experience it).

REFERENCES

Bostrom, Nick. 2003. "Are We Living in a Computer Simulation?" *The Philosophical Quarterly* 53 (211): 243–255. JSTOR, www.jstor.org/stable/3542867.

Chalmers, David J. 2005. "The matrix as metaphysics." In *Philosophers Explore the Matrix*, edited by Christopher Grau, 132. New York: Oxford University Press.

Schwitzgebel, Eric. 2019. "Kant Meets Cyberpunk." *Disputatio* 11 (published online ahead of print) 11: 1–25.

Perceptual Assumptions: An Argument Against Epistemic Immediacy

Isabella Maguire

University of Akron

ABSTRACT

Does objectivity live in perceptual states? In this paper, I address the question of whether accurate knowledge can arise from subject/object relations alone, given the occurrence of a perceptual state. I explore the direct realist view of immediate perceptual knowledge with a discussion of mind-dependent and mind-independent sensory objects and the interplay between subjective experience and environmental factors. I consider the possibility of epistemic perceptual immediacy under regular circumstances and in cases of hallucinatory or illusory experiences. I argue against the possibility of immediate accurate knowledge isolated from psychology and dispel the direct realist view. Returning to the continued debate about the acquisition of knowledge, my discussion looks to the resurfacing popularity of Cognitive Penetration Theory as a neuroscience-backed avenue for answers, and attempts to discern how much psychological factors aid in the development of our beliefs about reality.

INTRODUCTION

As I sit here typing this essay, I notice that there are currently several things occupying space in my conscious awareness. First, there is my laptop computer, which desperately needs to be plugged into an outlet. Secondly, I notice my mug of tea with steam rising from the rim of the milky white glass. This I will forget about, and it will get cold. There is also a candle softly flickering in the corner of my left eye. The list goes on.

Of the things that are present, I notice that I can make two general sorts of distinctions. The first is that I can distinguish between each of these things and recognize them immediately. The second, is that I can distinguish these individuals from the experience of myself. For example, there are other things that I am aware of, like the fact that my head hurts or that my wrist itches. I *know* my headache is part of my experience in a way that my laptop is not. Furthermore, I come to consider items like my laptop and the candle as existents in and of themselves. Where does this knowledge arise?

In this paper, I conclude that there must be an objective reality to which we have indirect access. We can derive partial knowledge from this objective reality by way of our subjective experiences. We cannot step outside of our own subjectivity. Consequently, we make evaluations about the nature of reality based on the subjective nature of perception. Because of this, I attempt to refute the direct realist stance on perceptual knowledge. I argue that we do not have direct immediate access to knowledge through perceptual states alone, as a direct realist view would assert. I argue that knowledge does arise from our interactions with this objective reality. I offer Cognitive Penetration Theory, or the view that mental states and psychological processes influence perceptual states, as a viable middle ground between direct realism and idealism. In this case, an idealistic view manifests as complete perceptual skepticism. Perceptual skepticism holds that knowledge is unobtainable through perceptual states because knowledge is dependent on the subjective mind. If idealism holds true, there is no reality outside of our own subjective experience for perceptual states to correspond to, and therefore, no objectivity at all. By introducing cognitive penetration's role in perception, I aim to support the claim that there is an objective reality to which our subjective experience corresponds, but thoughts and beliefs will always influence the perceptual process.

The naive realist, or common sense realist, maintains that the experiences I describe above amount to a direct relation of my awareness to the objects within its figurative grasp (Crane and French 2017). There is something about the objects themselves that, without involving my intentions, directly relates the object to my experiences of them, and conforms them to my mind. This directly gives me knowledge of the particularity of each object and enables me to experience their objectivity in relation to myself. Sense-data[1] is not needed to distinguish objective reality from my own experience (Crane and French 2017). In other words, I seem to possess (1) immediate accurate knowledge of the objects in my awareness, (2) distinguish (a) the experiential particulars of myself from that of (b) things less immediate to myself (like the candle), and (3) I seem to *know* that both of the latter categories (a and b) have equal existence in relation to each other.

The above view is tantamount to how we think about reality under familiar conditions. We do this everyday. Nobody needs to coax me into my belief that I exist, and that I exist with some separateness to the other objects which I experience. I can easily recognize and parse the objects in my experience and maintain a high degree of accuracy about what they are. For example, I know that the object in my pocket is a phone with a high degree of certainty. Were I to hand it to someone desiring to make a phone call, I am fairly confident that they would immediately dial a number and not question me as to why I handed them this particular object rather than any other. This is due to my belief that they are having a shared experience of the object as a phone, and as possessing all of the qualities of a phone. In short, I tend to believe that I and others can see the world around us objectively and without bias under typical conditions. Although it is advantageous for most of us to live this way, philosophers want to know what this suggests about the nature of perceptual states.

Returning now to the prior example of my experience typing this essay, the process of perception as described above seems to admit some knowledge. This case raises the question of whether immediate perceptual states contain knowledge in themselves, or whether perceptual states can give rise to knowledge only with the addition of the judgments made about them. My goal in writing this essay aims to address this question. In addition, I aim to confront the naive

---

1. Sense-data theory is a 20th Century Analytic theory which maintains that there are mind-dependent objects called sense-datum, which are exactly as they appear to be and of which we are directly aware when in a perceptual state.

realist and dispel this view generally. I do not believe that perceptual immediacy contains knowledge and will attempt to argue this point.

In order to approach the above claim, Section I of this paper will first introduce immediacy in perceptual states generally, and then explain two distinct notions of perceptual immediacy in greater detail. In Section II, I will focus on perceptual immediacy as it relates to knowledge and present a more formal version of the general naive realist view. Following this, in Section III, I will present an original argument contra the naive realist showing that epistemic immediacy is false. In Section IV, I will elaborate on my claims and attempt to justify them. Finally, in Section V, I will return to the central question: where does knowledge arise, and does objectivity live in perceptual states?

I. TWO NOTIONS OF PERCEPTUAL IMMEDIACY

Generally, the concept of perceptual immediacy may be understood in a more or less straightforward manner. By perception, I have in mind a state of awareness about the world (Crane and French 2017) and the presumed presentation of mind-independent objects to that of a knowing subject. For example, my visual experience of the colour blue and my auditory experience of a loud and sudden noise are both perceptual states (Antony 2009, 557–558). Perceptual states involve the processing of the signal relay in the neural network of the brain.

With respect to perception, the relationship between the subject and object has immediacy when the content of a given perceptual state is present to the mind independently of other factors and is without the intervention of another object or agential power.

When referring to *perceptual immediacy*, this expression could be describing a number of different notions. Todd Buras (2008) describes an immediacy relation as one where "the existence of an object distinct from the relata is not a necessary condition of the relation obtaining" (Buras 2008, 604). Buras differentiates between a notion of *absolute* immediacy and a notion he calls *qualified* immediacy (Buras 2008, 604–605). The distinction between these two notions being that *absolute* immediacy is a claim that two given relata need no intermediaries through which a relation can obtain. On the other hand, *qualified* immediacy allows for the intermediaries necessary where relation may obtain through sensory organs and nerves with a given relata. The latter notion is understandably the type of

immediacy modern naive realists were interested in, and also the preliminary kind of immediacy I am going to assume firsthand in the following discussion of perceptual immediacy. For the purposes of this essay I will be discussing two distinct notions of immediacy grounded in the perceptual process as presented by philosopher Georges Dicker (2006, 517–35).

Dicker presents a *psychological* notion and an *epistemic* notion of immediacy. He argues in the paper "Berkeley on Immediate Perception: Once More Unto the Breach" that the two notions are often conflated, which Dicker sees as misguided, and thus leading to major misrepresentations, especially in relation to the views of philosophers such as Berkeley and Hume. As such, Dicker takes special care to clarify and distinguish between his two presented notions of perceptual-based immediacy.

The *psychological* notion of perceptual immediacy, adapted from Dicker, claims that (an object) *x* is immediately perceived if that *x* is perceived without the perceiver performing any conscious inference pertaining to *x* (Dicker 2006, 518). In this case, psychological immediacy is said to obtain in a perceptual state when the subject does not have to make any adjustments in the way an object is perceived in order to perceive its true objective nature. This would imply a shift in the perceptual state rather than an assimilation of new knowledge about the nature of the perceptual object, given the case where psychological immediacy does not obtain.

Comparatively, the *epistemic* notion of perceptual immediacy maintains that (an object) *x* is immediately perceived if *x* is perceived in such a manner that both its existence and true nature can be known completely. This true nature is known on the basis of the subject's present perceptual experiential state (Dicker 2006, 518). This notion of perceptual immediacy asserts that perceptual states themselves admit of knowledge about the perceptual object with no additional qualified mediate referential (Lyons 2017).

While the prior case of psychological immediacy is interesting, I am presently more concerned with Dicker's second distinction of epistemic immediacy as relating to the questions treated in this paper. The notion of whether a subject can have epistemic immediacy in perception returns to the question I originally posed: does the immediacy of a given perceptual state give us objective knowledge in itself, or is accurate knowledge always the result of some relation to the perceptual process? The naive realist believes that the immediacy of a given perceptual state

provides objective knowledge, so, in the following section I will dive deeper into the naive realist's views on perception and knowledge.

II. EPISTEMIC IMMEDIACY AND THE NAIVE REALIST

Now, I would like to return to the naive realist view, which I briefly sketched in the introduction. For the naive realist, or direct realist, there is psychological immediacy, but also epistemic immediacy. As I mentioned earlier, the naive realist is committed to immediate accurate knowledge through the direct subject-object relation alone. Naive realism maintains that this relation admits to absolute truth about the nature of reality through the direct contact of the perceptual object with the knowing subject. This leads to the conclusion that the world simply *is* the way I directly perceive it to be, and that I always have direct access to the objective nature of reality, which I can accurately distinguish from that of my own subjective mental states.

It should be noted here that the direct realist view differs on a few key points from a similar view, the *indirect* realist view, or *representationalism*. Indirect realism generally holds that, perceptually, I am directly aware of my own subjective experience. My experience is the lense through which I am indirectly aware of an existent objective reality (Lyons 2017). I can interact with the objective particulars that I have awareness of, and so, I am able to assert that there is an objective reality apart from my experience that I am able to access. The direct perceptual immediacy relation asserted by the direct realist holds that I am directly aware of the external world and immediacy in my perceptual states give me access to reality *per se*.

As stated in Section I, Todd Buras (2008) notes that this direct immediacy relation, with respect to perceptual states, is not simply a logical relation. The object-subject relation is not free from obvious intermediaries, namely the sense organs and neural pathways, and is therefore also a kind of *qualified* immediacy (Antony 2009, 557–558). This is a necessary distinction, as I wish to point out that the naive realist believes that the obtaining immediacy relation is simply free from intermediary objects of thought, rather than a relation completely free from the perceiver's subjective experience. Experiential immediacy is foremost. The subject is always bound by environmental and bodily factors.

Returning to the notion of epistemic immediacy, Buras points out that most direct realists admit to epistemic immediacy in perceptual states. He writes that modern philosopher Thomas Reid, a direct realist himself, is committed to a kind of epistemological immediacy. Although, he notes that Reid never explicitly distinguishes between his own epistemic commitments and those of other immediacy aspects (Buras 2008, 615).

With all of these distinctions in mind, Buras presents the following formalized version of the naive realist's view (Buras 2008, 612):

> *a* (the subject) immediately perceives *b* (an object) if and only if *a* perceives *b*, and it is not the case that if *a* perceives *b*, then there is an *x* distinct from *a* and *b* which is an object of thought for *a*.

The above is the version of the naive realist's claim which I will address with a counterclaim in the following section. For now, I would like to take up the position of epistemic perceptual immediacy and explore this assertion in detail.

When I am in a state of awareness, I must also simultaneously be in a perceptual state, as a state of awareness must be awareness *of something*. It seems as though the knowledge of my experience, therefore, has simultinuity with the experience itself. The content of my experience contains the features of my awareness about something. The thought of my observation of the mug of tea on the table is the *aboutness* of what it is like to currently have the perceptive experience of the mug on the table (Kind 2010, 902–903). In other words, this thought is the only necessary thought I must have in relation to my perceptual experience of the mug in order to experience its particular objective reality *as* a mug. But, do I even need this? In fact, the features of my awareness seem to be present and direct whether I have any thought about them at all.

Any given perceptual relation may be informed by a given thought or belief about it,[2] but the relation in itself obtains in both (a) the event that my belief changes, and (b) a non-belief state. For example, if I am presented with an ambiguous image, I may be unable to discern what it is, but it would be absurd to assert that I am not aware of the image itself as an image contained in my perceptual state. I am aware of the image whether I can discern it or not.

---

2. Cognitive Penetration Theory suggests that there is an influence of cognitive factors on sensory aspects of the perceptive process.

So, what epistemic commitments can I have, if any? If the direct realist also wishes to assert epistemic immediacy, then the non-belief based perceptual relation must admit objective knowledge. To refresh, a commitment to epistemic immediacy entails that perception is immediate whereby a given object is perceived in a such a way that, both its existence, and its true nature may be known on the basis of the perceptual state alone. Therefore, if the perceptual state gives me direct access to reality, then I will also have knowledge of the perceptual object by necessity. In this sense *to perceive* may be equated with what it is *to know*.

Under every perceptual condition I, therefore, must also have epistemic immediacy. If I have the conformity of my experience with the actual state of reality, only then am I in a perceptual state. Thus, it seems that in order to maintain this epistemic assertion, the direct realist must maintain that we are either always perceiving directly, or that we are not always in a state of perceiving.

In this case, a state of awareness cannot be one and the same with a perceptual state. If this is the case, then awareness is not always awareness *of something*, as I asserted earlier. For the naive realist asserting epistemic immediacy, it seems that a basic state of awareness may obtain without a perceptual state. I would like to briefly dwell on this point, as it seems to be one way to explain how I could have awareness of an ambiguous image, but be unable to discern what it is, if a direct perceptual state should give me knowledge. If I am to follow through with this epistemic commitment, then it seems that awareness and perceptual states cannot be referring to the same conscious state.

If I am wrong about what I am aware of, then, in this case my perceptual state will be directly influenced. This influence would lead to mistakenness and false-beliefs, which come about from inadequate knowledge, or a lack of immediate knowledge. So, this leaves the direct realist view in an interesting place with respect to commitment to epistemic immediacy.

Perhaps, to say that I can have a false perceptual belief is to, as Dicker points out in Section I, conflait epistemic immediacy with the distinct notion of psychological immediacy. The conclusion, then, for the direct realist is that a perceiver only has knowledge when they are experiencing a perceptual state. If I am mistaken about *what* is in my field of awareness, then I am not *in* a perceptual state.

Finally, in order for the direct/naive realist to assert epistemic immediacy, the following must hold up. Epistemic perceptual immediacy asserts that (a) states

of awareness are not always perceptual states, and that, (b) if the perceiver is mistaken about the existence and nature of a given perceptual object, then (c) that given state of awareness is not also a perceptual state.

III. CONTRA EPISTEMIC IMMEDIACY AND THE NAIVE REALIST

I will now offer some opposition to the direct realist. I believe the naive realist is incorrect in all of their assertions, but am here only aiming to refute the general naive realist view, addressing the specific falsity of epistemic immediacy within perceptual states.

For the naive realist to assert epistemic immediacy in a direct subject-object relationship, all of the conditions listed in Section II must be met. Therefore, to effectively dispel the notion of epistemic perceptual immediacy, a perceptual relationship between a knowing subject and a given perceptual object must be shown to obtain in a state of awareness where (1) the subject is mistaken about the true nature of the perceptual object, or (2) when the subject is mistaken about the objective mind-independent existence of the perceptual object. If, either (1) or (2) obtains in a state of awareness, then states of conscious awareness must be *of something*. If this is the case, then states of awareness are always perceptual in nature, and a subject's experiential state is also a perceptual state. If true, then the subject's perceptual state does not imply immediate objective knowledge of a given perceptual object, and epistemic immediacy is false.

In response to the naive realist position, I will now offer a formal counterargument to epistemic immediacy as follows:

> **P1**. If a perceptual state obtains with incomplete immediate perceptual knowledge, then epistemic immediacy is false.

> **P2**. If a perceptual state obtains in a state of awareness in which the subject is mistaken about *either*, the true nature, or mind-independent existence of an object, then perceptual states obtain despite incomplete immediate knowledge of an object.

P3. Perceptual states *do* obtain when the subject is mistaken about the true nature of an object.

P4. Perceptual states *do* obtain when the subject is mistaken about the mind-independent existence of an object.

P5. Perceptual states obtain, *both* when the subject is mistaken about the true nature, and mind-independent existence of a given object (conjunction, P3, P4).

C1. Perceptual states obtain with incomplete immediate perceptual knowledge (*modus ponens*, P2, P5).

C2. Therefore, epistemic immediacy is false (*modus ponens*, P1, C1).

As I have asserted throughout the development of this paper, naive realism entails that a direct subject-object relation allows for the subject to experience and to know mind-independent reality. This knowledge obtains in the moment of subject-object conformity alone. Thus, perceiving is equated with *knowing* in the sense that perceptual states grant access to the true nature of reality whereupon a given perceptual object is apprehended. For the naive realist, what it is to know reality is simply the conformity of the subject to the object within a perceptual state. When I am in a perceptual state, I am simultaneously in a state of *knowing*.

In one way, my defense may highlight an issue with the kind of perceptual immediacy the naive realist wishes to maintain, but in another way, this exercise might simply serve to exemplify the problems that arise where epistemic and psychological immediacy are confused. Despite this risk, I argue that the kind of subject-object relationship the naive realist asserts entails epistemic perceptual immediacy. Therefore, if perceptual epistemic immediacy is false, then naive realism/direct realism is also, arguably, false.

The following section is devoted to the defense of my counter argument, primarily focusing on the justification of premises three and four.

IV. EXPLANATION AND JUSTIFICATION

Accurate knowledge of and about objects in my field of awareness seems perceptually immediate under familiar, or 'normal' perceptual conditions. But, when perceptual conditions are novel or ambiguous, the absence of accurate immediate knowledge becomes clear. Jennifer Church (2010) expresses a similar observation in her paper "Seeing Reasons." Church states that "we experience objectivity only when we discover consistency across perspectival change" (Church 2010, 641). Conditions in which my state of awareness is also *of something* must admit of a perceptual process. This is the process by which I am able to make sense of what it is that I am aware of.

In premise (3) of my argument, I say that a given perceptual state obtains within a state of awareness regardless of whether the subject is mistaken about the true nature of a perceptual object. It would be absurd to claim that I am not aware of an object that I consciously recognize, but is it not also absurd to claim that my awareness *of* an object is absent of any perceptual state?

To illustrate my point, I will return to the example of an ambiguous figure briefly mentioned in Section II.
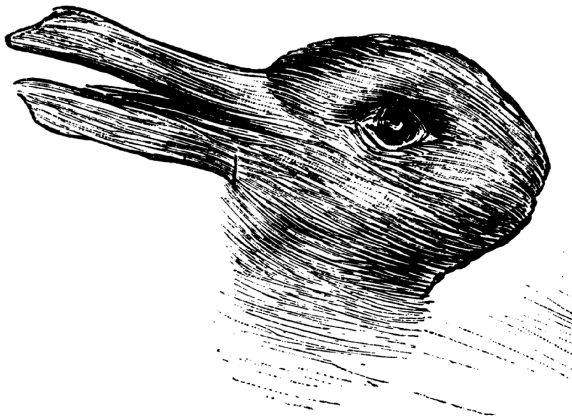

Figure 1.1 Duck-rabbit illusion, Anonymous Illustrator, 1892. ‚Welche Thiere gleichen einander am meisten?' Fliegende Blätter. Braun & Schneider.

The ambiguous figure pictured above is a common example of a visual illusion. By simply looking at the image it is difficult to tell whether it illustrates a duck or a rabbit. I can capture the image as an object in my field of visual awareness, but I do not have immediate knowledge of the true nature of the object. I do not immediately know *what* the object is, I just know *that* it is an object in my awareness. My experience of the object is unfamiliar and, once I recognize the object as an object, my perceptual state begins the process of discerning what the object is. If I were presented with this image and was told it is a sketch of a rabbit, I would then be more likely to justify it as being a rabbit, and so, it would no longer be ambiguous in that sense.

Apart from this, my perceptual state is absent of epistemic immediacy about the object's true nature. Therefore, I can consciously switch back and forth between the presentation of, either a rabbit, or a duck, without knowing whether it is either. It would be absurd to assert that I am not in a perceptual state upon becoming aware of the image. It is in my visual system. I can take it in, imprint it in my memory, and recall it later with the same confusion about what it is that I had been looking at. It is my perception of the image which allows me to do this. The realization that I can do all of these things shows me that (1) I am in a perceptual state, (2) I have imperfect knowledge about *what* the object truly is, and (3) my perception of the object can be informed by something outside of my perceptual state. Point (3) touches on the thoughts of Church, who believes that prior perceptions about a given object allows the subject to have an accurate and demonstrated thought about a given perceptual object.[3]

Premise (4) asserts that a perceptual state obtains in awareness despite mistakenness about the mind-independent existence, or objectivity, of a given perceptual object. In other words, I might know that my head hurts, and this is immediate to me as part of my self-experience. This is a subjective experience. Whether the cold mug of tea, or my laptop, really exist apart from me is another story entirely.

I can be mistaken about whether I know exactly what an object is upon perceiving it, but I also assert that perceptual states obtain in awareness whether there actually is a mind-independent object at all. To explain this, I will employ another example.

---

3.   See note 2.

If I am driving down a dark road on a clear night, I might believe that my perceptual state is epistemically immediate where I am aware of the road stretching out in front of me. If I were not clearly consciously aware of the road, I would fail to maintain a straight path for very much longer. If I were to suddenly hit a person who happened to be walking across the road, the presence of the person was not included in my conscious awareness. I am aware of the road, but clearly not the person I just massacred. In this case, there is a perceptual deficit.

Clearly, I am aware, as I see the road and other features of my awareness. I am also in a perceptual state, as I am aware *of* objects and actively discerning them in order to maintain my driving. My perceptual state is consistent. So, it is therefore, in a sense objective, as Church points out. I am aware and I am perceiving, and yet I made a mistake about the mind-independent features of reality and smashed a person, whom I mistook for a feature of the dark road. Therefore, perceptual states seem to obtain in a state of awareness despite mistakenness about the objective existence of its features.

In another, positive example, I could be in a perceptual state where I believe something to be present that is actually absent, rather than a deficit in my perceptual state, whereby I miss an object actually present. I could be hallucinating a perceptual object. I believe this perceptual object exists mind-independently, but it does not. If I were to hallucinate a doughnut on the table in front of me, I would be aware of that particular perceptual object. In this case, I can observe the doughnut, and describe *what* I am aware of. I can pick out its particular features. Maybe, it has pink icing and purple sprinkles. Based on my experience of the doughnut, I believe it has objective mind-independent reality. If, however, I attempt to bite into the doughnut, and bite through thin air, I would discover that I am mistaken about the existence of the doughnut completely. Nonetheless, I seem to be reacting to stimuli in my field of awareness, and am thus *in* a perceptual state. Despite my mistakenness, I am perceiving something, and am simply mistaken about the objective nature of *what* it is that I am aware of.

Based on the examples provided, it seems that I can have perceptual states without the immediate accurate knowledge assumed by the naive realist. I perceive a given perceptual object of which I am aware without knowing its true objective nature upon initially perceiving it. From these observations, it follows that epistemic immediacy in perceptual states must be false.

V. RETURNING TO THE QUESTION OF PERCEPTUAL KNOWLEDGE

Finally, I will return to the question posed at the beginning of this paper: where does knowledge arise, with respect to perceptual states, and can objectivity live in perceptual states at all? I will begin this discussion by first addressing the second portion of this question.

In the previous section, I established that perceptual states are present in awareness regardless of immediate knowledge of a given perceptual object. With this in mind, it seems that I know when I am experiencing a mind-independent object. While the question of whether we can directly perceive any objectivity remains open, I believe that, in some sense, I must say that the answer is yes. I do have some objective knowledge of a perceptual object provided by subjective perceptual states. In order to escape a sort of Humean idealism, I must be able to say that my experience of the laptop in front of me differs from the subjective experience of hallucinating a perceptual object, as in the example of the doughnut.

The interplay between objective reality and subjective experience is what provides me with knowledge about the nature of the world. The combination of sensory experience and a perceptual state tells me that there is an object in my awareness, and provides me with some knowledge about whether it exists and what it could be. I do not know exactly *what* its true objective nature is, but I always know that something is present. I can have imperfect knowledge of what is present, and while this knowledge may not be complete, it is accurate for the most part, allowing me to successfully navigate my surroundings and communicate with others. To return to the point discussed in Section IV, Jennifer Church directly states "we experience objectivity only when we discover consistency across perspectival change" (2010, 641). It is only when consistency is absent that we question objective reality. This could be due to purposive vagueness in the object itself, as in the case of the illusory image example.

Based on the short discussion above, I will now address the first half of the question. How does knowledge arise, if there is some knowledge present in perceptual states? One possibility, as Church states, is that "it is perception that secures the particularity of the objects of our thoughts, not the other way around" (2010, 639). From this assertion, I could conclude that perceptual knowledge gives us some idea of what there is, objectively speaking. Further, objective knowledge may follow from perceptual states rather than solely from cognition about the

content of perceptual states. Conversely, cognitive penetration theory, which assumes the influence of cognitive processes on perceptual states, suggests that perceptual states are influenced by prior and co-occurring cognition and psychological factors.

Either way, we seem to need something objective, whether we begin with reference to objective reality, or end with an approximation of it. But, is this a satisfactory answer to the knowledge question? In current discourse, the final answer of this heavily debated topic is still open, prompting a possible refashioning of views like Cognitive Penetration Theory, which has new empirical backing (Vetter and Newen 2014, 62–75). and suggests that higher level cognitions, like thoughts and beliefs, directly influence sensory perception. If this is the case, then the presence of an individual's experience and beliefs about a given perceptual object may prevent the veridicality of knowledge in perceptual states completely. On one hand, it seems that some knowledge must be present in perceptual states. This allows for the correspondence with an objective reality existing apart from my experience of it. On the other hand, the possibility of cognitive penetrability calls into question whether the subject is ever actually experiencing an object as it is. If perceptual states rely heavily on prior psychology, how could it occur to the subject that the regularity in their environment had shifted? Upon returning to my apartment after a month, it might suddenly seem to me that the furniture had been re-arranged by my roommate. This could be because prior experience of my apartment does not match its current state, or I am able to match my prior experience to the sensory experiences of my current perceptual state. Possibly, I simply expected it to be a certain way and forgot, which is why I tripped over the coffee table upon entering.

As I previously argued, an objective reality of some sort is necessary in order to avoid falling prey to idealism manifesting as perceptual skepticism. I do not hope to create a false dichotomy between idealism in the extreme sense and the sole possibility of one remaining option. I merely wish to direct attention to the renewed sense of hope in Cognitive Penetration Theory. The theory raises doubts about whether perceptual experiences grant us access to an outside world. It raises the possibility that mediate (Carrier 1969) perceptual states are always and completely permeated by both prior and cocurrent psychological processes.

REFERENCES

Antony, Louise. 2009. "The mental and the physical." In *The Routledge Companion to Metaphysics*, edited by Robin Le Poidevin, 555-567. New York: Routledge.

Buras, Todd. 2008. "Three Grades of Immediate Perception: Thomas Reid's Distinctions." *Philosophy and Phenomenological Research* 76 (3): 603–32.

Carrier, L. S. 1969. "Immediate and Mediate Perception." *Journal of Philosophy* 66 (July 3): 391–403.

Church, Jennifer. 2010. "Seeing Reasons." *Philosophy and Phenomenological Research* 80 (3): 638–70.

Crane, Tim and French, Craig. 2017. "The Problem of Perception."In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.

Dicker, Georges. 2006. "Berkeley on Immediate Perception: Once More Unto the Breach." *Philosophical Quarterly* 56 (225): 517–35.

Kind, Amy. 2010. "Transparency and Representationalist Theories of Consciousness." *Philosophy Compass* 5: 902–913.

Lyons, Jack. 2017. "Epistemological Problems of Perception." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.

Vetter, Petra, and Albert Newen. 2014."Varieties of cognitive penetration in visual perception." *Consciousness and Cognition* 27: 62-75.

compos mentis

Autonomy and Sense-Making: Problematizing the Individual in Social Cognition

Eli McGraw
Beloit College

ABSTRACT

This paper examines the overarching trajectory of the social cognition conversation with particular focus placed upon the use and role of the individual throughout. Traditional theories posit that individuals apply knowledge about their own internal states to something or someone else in order to infer what that person or thing is experiencing. Embodied and enactive approaches pay closer attention to the individual's bodily experience, as well as interaction processes themselves. As more research is done in regards to both traditional (Theory of Mind Theory, Simulation Theory) and contemporary theories (Interaction Theory, Participatory Sense-Making), the role and use of the individual becomes increasingly convoluted. By calling attention to relevant points of agreement, problematic concessions and obvious inconsistencies, this paper aims to shift the conversation towards a more homogeneous use of the individual throughout social cognition. Inconsistent use has led to inconsistent research. With a better defined individual, contemporary theories like Participatory Sense-Making can effectively move beyond the shortcomings of the ongoing discussion.

KEYWORDS

Participatory Sense-Making, Interaction Theory, Autonomy, Enactivism, Social Cognition

Social cognition attempts to determine how we go about understanding others as well as understanding *with* others. To understand another person is a complicated process. Theorists of many disciplines and dispositions have understood the issue in several ways. One piece integral to the current conversation in the discipline is the use and role of the individual.

Traditional approaches to social cognition tend to lean on individual cognitive functions in explaining how we go about coming to terms with other people. In such approaches, the individual applies knowledge about their own internal states to something or someone else in order to infer what that thing or person is experiencing. Those approaches fail, however, to consider the extent to which cognition arises from bodily interactions with the world and others. Embodied accounts more seriously consider the experience of an individual in terms of their particular perceptual and bodily capabilities. Those capabilities are combined with high level concepts like reasoning, language, memory, and all other aspects of mental life, to construct and perpetuate an identity in the world (Shapiro, 2019, p.56). Embodied approaches still take seriously individual cognitive functions, but begin to place emphasis, too, on social interaction. Finally, an enactive account of embodied social cognition takes the interaction process much more seriously. In such accounts, interaction becomes an explicitly measurable entity. Attention is given to the interaction process as a whole which in itself incorporates, affects, and is affected by, the internal cognitive mechanisms fundamental to other approaches.

Throughout these theories, the individual plays an important, but convoluted role. The current debate rests on the individual in a way that has yet to have been sufficiently recognized by theorists contributing to the conversation. While the individual is taken seriously, the inconsistencies attached to its use and role in relevant theories hinder the overarching progress of social cognition. This paper will examine the trajectory of the social cognition debate and delve more specifically into the ways in which differing conceptions of the individual, as put forward by newer theories, tend to complicate the conversation in an unproductive manner.

TRADITIONAL APPROACHES TO SOCIAL COGNITION

To best understand the distinctions drawn between differing theories surrounding social cognition, looking at how standard approaches go about contemplating the problem will be useful. There are two key approaches which dominated the field of social cognition for years: Theory of Mind Theory (TT) and Simulation Theory (ST). Each of these approaches stem from methodological individualism. Broadly understood, methodological individualism posits that "... social phenomena must be explained by showing how they result from individual actions, which in turn must be explained through reference to the intentional states that motivate the individual actors" (Heath, 2015, p.1). Evidently, the individual is immensely important in these traditional theories. To understand the world and the people in it, individuals must utilize their internal cognitive mechanisms. While the theories in themselves are genuinely complex, this paper will focus less on the intricacies within, and instead pull out relevant information about the use and role of the individual in these standard approaches. These approaches serve to contextualize the discussion surrounding social cognition, while exemplifying the shortcomings of methodological individualism.

According to Theory of Mind Theory, individual cognitive functions are integral to the study of social cognition. In describing TT in his paper "Two Problems of Intersubjectivity," Gallagher states that "we use a theory about how people behave (folk psychology) to infer or 'mindread' (or mentalize) the beliefs, desires, intentions of others" (Gallagher, 2009, p.290). Folk psychology in this context is understood as the ability to treat agents as the proprietors of unobservable mental states. Those mental states are subsequently used to explain and predict the behavior of agents (Apperly, 2008, p.3). In TT, individuals process something in front of them. When taking an interaction between two people into consideration, TT would suggest that there is in some sense a conceptual wall that must be breached in order for each individual to understand what the other is feeling, intending and desiring. The presupposition behind that conceptual wall is that our intentions as individuals are hidden from other people. We must therefore figure out another person from what we observe of their behavior. That behavior, however, isn't in itself linked directly to their internal states. Rather, inferences are necessary using our own internal mechanisms to determine how another's behavior might be linked to their internal states.

Simulation Theory puts forward a similar conception of the individual to TT. ST accounts of social cognition were developed in the first place as a skeptical response to TT's claims about the way individuals use theory to explain and predict the behavior of others. "Simulationists note that biology ensures that our own minds will have processes for the fixation of beliefs… desires… and other processes involving mental states that are essentially similar… to the same processes in the minds of others" (Apperly, 2008, p.5). Put simply, because the cognitive states and functions of human brains are similar to one another, much of the work involved in considering another mind can be done using one's own mind as a model. Humans have the capacity to put themselves into the shoes of others, using their own mind to create 'as if' beliefs, desires, and intentional states which are then projected into the mind of another person to explain and predict their behavior (Gallagher, 2009, p.290). Like TT, there is again a presupposition attached to the way in which individuals go about projecting simulated understandings *of* one another *to* one another. The presupposition is that other people's minds are hidden until a simulation stemming from one's own mind can be projected into the other.

The basic suppositions behind methodological individualist theories such as ST and TT can be synthesized to make clear the relevant aspects of these traditional approaches to social cognition. Gallagher neatly lays out these suppositions. First, each theory conceives of the individual as processing the world through the lens of their own subjective cognitive functions. Individualist theories posit that we cannot directly perceive another person's thoughts, feelings or intentions. The mind and the body are separated here in a Cartesian sense. Moreover, due to this disconnect, extra cognitive processes such as theorizing or simulating are necessary in making inferences about another person's mind (Gallagher, 2009, p.291). Second, TT and ST both use the process of observing another person's behavior as the starting point for making those theoretical inferences or simulations which in turn can be used to explain and predict future behavior. Third, both standard theories posit that these "…mentalizing processes constitute our primary and pervasive way of understanding others" (Gallagher, 2009, p.291). TT and ST make clear that individual mechanisms provide the most useful way of understanding the problem of social cognition. These traditional approaches fail to recognize the significance attached to the embodiment of the individual in social situations. To properly take into account the pieces of social cognition

which methodological individualist theories leave out, Gallagher puts forward an embodied approach to social cognition which he calls Interaction Theory (IT).

EMBODIMENT AND INTERACTION

Embodied approaches to understanding the mind hold that "...the body is crucial for cognition" (Di Paolo & Thompson, 2014, p.68). To embodied theorists, the mind arises from the nature of our brains, bodies and bodily experiences. Cognition is therefore not confined to the functions of the brain, rather it is influenced and in some ways determined by our experiences in the physical world. This extension of the mind into the surrounding environment, as well as the ways in which each factor can determine things about the other, is important in beginning to understand the individual's multifaceted role in social cognition.

Gallagher's IT, an embodied approach to social cognition, directly challenges the base suppositions attached to TT and ST. He argues that the dualistic framework within which TT and ST are understood is problematic. To Gallagher, other minds are not "...hidden away and inaccessible…" (2009, p.292). Rather, he points to evidence from his research in phenomenology and developmental psychology to put forward the claim that we can and do directly perceive another person's intentions, feelings, and desires through their embodied behavior. In his book "The Phenomenological Mind," he states, "Before we are in a position to theorize, simulate, explain, or predict mental states in others, we are already interacting with them and understanding them in terms of their expressions, gestures, intentions, and emotions, and how they act toward ourselves and others." (Gallagher, 2012, p.210). To illustrate this idea, consider an infant. An infant, presumably without the intervention of theory or simulation, can perceive bodily gestures as goal directed intentional movement. This idea at once calls into question methodological individualism's claim that other minds are inaccessible, while bolstering the thought that in everyday interaction, no theoretical or simulated inference is necessary. Gallagher further claims that "...we are not primarily spectators or observers of other people's actions; for the most part we are interacting with them on some project, or in some pre-defined relation" (Gallagher, 2009, p.292). Rather than the observational stance offered by methodological individualist theories, Gallagher points to a second-person stance which focuses on the embodiment of interactors. He claims that mindreading, which is that process by which inferences

can be drawn through theory and simulation, is not in fact, the primary and pervasive way of understanding others (Gallagher, 2009, p.292). According to Gallagher, "...in most intersubjective situations we have a direct understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions, and mirrored in our own capabilities for action" (2005, p.224). Put simply, we understand others due to our own embodied experience, with reference to other bodies.

Throughout his work, Gallagher does well at distancing his ideas from a purely methodological individualist standpoint. He puts clear emphasis on the embodiment of the individual within interaction processes. However, according to Di Paolo, Rohde, and De Jaegher, "...proposals of embodied cognition like these have a drawback: they often presupposed coupling between persons" (2010, p.61). Coupling here is understood to be "The influence between a system's variables and another system's parameters" (De Jaegher, Di Paolo & Gallagher, 2010, p.441). Think of a person walking a dog on a leash for example. Due to that presupposed coupling, "...*how people interact* does not in itself become an explicit topic for investigation" (Di Paolo, Rohde & De Jaegher, 2010, p.61). The idea that an interaction process could become a measurable system is foundational in understanding how participatory sense-making (PSM), an enactive approach to embodied social cognition, goes about understanding and utilizing the individual in terms of social situations. De Jaegher and Di Paolo find that in attempting to understand the meaning that stems from sociality, not only the embodiment of the interactors, but the interaction process which occurs between them, both must be considered as relevant areas of focus.

ENACTIVISM, PARTICIPATORY SENSE-MAKING, AND EMPIRICAL POTENTIAL

The enactive account of intersubjectivity brings new light to participatory and non-individualistic processes in social cognition (Di Paolo & Thompson, 2014, p.75). Enactive accounts consider concepts which allow meaning to be drawn not only from individuals within social situations, but the interaction processes themselves. Intersubjectivity here can be best understood as the psychological and interactional relations between two or more individuals. To reiterate, traditional theories focus on the inferences that an observer can make

using theory or simulation about the intentions, feelings, and desires of another person based on their external behavior. An embodied account considers more thoroughly an individual's bodily interactions with the world. That said, in applying an enactive approach, the domain of social interaction is made explicit. In their paper "An Inter-Enactive Approach to Agency: Participatory Sense-Making, Dynamics, and Sociality," Torrance and Froese consider the key attributes of an enactive approach to cognition and agency. Enactivism, to the authors, addresses the question: "What is it to be an (cognizing, conscious) agent?" (Torrance & Froese, 2011, p.22). Torrance puts forward a five-fold response to enactivism's foundational question. To be a conscious agent is to have the following attributes:

> …(a) to be a biologically autonomous (autopoietic) organism – a precarious, far-from-equilibrium, self-maintaining dynamic system; (b) with a nervous system that works as an organizationally closed network, whose function is to generate significance or meaning, rather than to act via a set of continually updated internal representations of the external world; (c) the agent's sense-making arises in virtue of its dynamic sensorimotor coupling with its environment, such that (d) a world of significances is 'enacted' or 'brought forth' by a process whereby the enacted world and the organism mutually codetermine each other; and (e) the experiential awareness of that organism arises from its lived embodiment in the world. (Torrance & Froese, 2011, p.22).

As PSM is laid out, the complex terminology in the above excerpt will fall into place. The most important concepts in the authors' understanding of enactive social cognition, with respect to this discussion, are autonomy and sense-making. These two concepts serve to shift how the individual is commonly understood in social cognition. PSM, as proposed by De Jaegher and Di Paolo in their work "Participatory sense-making: An enactive approach to social cognition," provides a starting point in coming to terms with that shift. The theory is held up by two key pillars. The subjects involved in the interaction process, and the definition and operationalization of the interaction process itself.

In PSM, the individual is a sense-maker, or a cognizing agent. Sense-making as a concept is understood to be the relational process between an autonomous self-organizing agent and their world. Sense-making implies active engagement.

This will be important in considering the shortcomings of the theory later in the paper. According to De Jaegher and Di Paolo, sense-making "...is an intentional activity that can become expressive in social situations through embodied action" (De Jaegher and Di Paolo, 2008, p.41). The concept of sense-making contrasts the commonly held view that organisms passively receive information from their environment and then process it into internal representations which are given significance only after further processing (De Jaegher and Di Paolo, 2008). Importantly, a sense-maker's self-organization entails particular needs and constraints which produce a perspective on the world. That perspective, according to De Jaegher and Di Paolo, stems from the thought that sense-making "...grounds a relational and affect-laden process of regulated exchanges between an organism and its environment…" (De Jaegher and Di Paolo, 2008, p.35). Those regulated exchanges give rise to normative conditions attached to a sense-maker's experience of the world at multiple levels of its identity. Meaning can then be drawn from the normative conditions which are attached to what is relevant to the needs and constraints of the self-organizing system. The idea of self-organization can be understood more simply in terms of existence. Meaning here can be pulled from any level of the identity of the existing system. From a metabolic perspective, meaning might stem from the nutrients a system requires to perpetuate its state of being. On a more conceptual level, meaning might stem from things that matter to the system existentially and are relevant to how it organizes itself conceptually. Autonomy is integral here, and an explanation of an autonomous system, as it's understood through an enactive lens, is necessary in drawing parallels between individuals as autonomous sense-makers and interactions in themselves being understood as autonomous sense-making systems.

To De Jaegher and Di Paolo, an autonomous system is "...a system composed of several processes that actively generate and sustain an identity under precarious circumstances" (De Jaegher and Di Paolo, 2008, p.35). The concept of operational closure will be necessary in considering identity generation with respect to interaction. Operational closure is the idea "...that among the enabling conditions for any constituent process in the system there will always be one or more other processes in the system" (De Jaegher and Di Paolo, 2008, p.35). In other words, with respect to each process in a given network of processes, there will always be another process by which they are conditioned. Precarious conditions point to the idea that without the organization of the system as a network of processes,

isolated component processes would tend to run down or extinguish (De Jaegher and Di Paolo, 2008). In terms of a sense-making thing, precariousness generates meaning and necessitates the normativity connected to its existential regulation. This idea links back to the needs and constraints relevant to a sense-maker at multiple levels of its identity. An individual sense-maker then, is uncontroversially autonomous.

According to De Jaegher, Di Paolo and Gallagher, "Autonomy can happen on different levels (metabolic, neural, cognitive and social) and different timescales, and autonomous agents can interact at various levels" (2010, p.443). It will be shown that the precarious, operationally closed conditionality of interaction processes can fit neatly into the conceptualization of autonomy as put forward above. Interaction processes often parallel and incorporate the needs and constraints attached to individual sense-makers at multiple levels of their identities. Meaning, therefore, can be drawn from interaction by considering more explicitly the interaction process as a sense-making thing with needs and constraints of its own. The relational processes that occur between two or more individuals when they encounter each other is the next step in coming to terms with how PSM conceptualizes the individual and the interaction process.

Sense-making is an embodied process of active regulation of the relation between an agent and its world. Through relational patterns of coordination and breakdown, this sense-making process can be shared between individual interactors (Di Paolo and Thompson, 2014, p.75). These patterns give meaning and perspective to interaction processes in a way that parallels how meaning arises from the needs and constraints of an individual sense-making system. In this way, encounters between two or more sense-makers can, in some circumstances, take on a life of their own and in turn become sense-making processes in themselves. Important here is the idea that the actions or intentions of the agents involved may sometimes fall short in determining the outcome of an encounter (Di Paolo and Thompson, 2014, p.75). In some cases, an interaction may emerge and keep existing against the intentions of the individual interactors involved. Take the narrow hallway thought experiment as an example. When walking down a hallway, sometimes a person will attempt, multiple times, to shift out of the way of a person coming in the other direction, only to unintentionally remain in the way. To De Jaegher and Di Paolo, this illustrates a way in which the relational patterns of coordination and breakdown in interactions can modulate, enable

and constrain individual sense-making processes and even supplement or replace aspects of individual cognitive functions (2007, p.491). In other words, an interaction process which could count as an autonomous system is one that would inherently incorporate individual sense-makers into the overarching process. The dynamic nature of the patterns of coordination and breakdown attached to social interactions, too, exemplify the various levels at which an autonomous system can be considered as such. Given the operationally closed and precarious nature of some interaction processes, one can be justifiably considered to be an autonomous, sense-making system.

To clarify, when two sense-makers (self-organizing systems) interact, another self-organizing system emerges between them which itself possesses the necessary properties to be considered an autonomous system. This emergent self-organization is the interaction process. According to Di Paolo and Thompson, this "…shared form of sense-making is what is meant by 'participatory sense-making'" (2014, p.75). The individual here can be understood in terms of their own sense-making attributes as they are affected by the coordination dynamics described above, as well as the joint sense-making processes which often co-opt those individual processes. The individual is a sense-maker with normative needs and constraints which apply to each level of their identity. The interaction process is a combination of two or more sense-making things which itself has needs and constraints (patterns of coordination and breakdown) which include those of the individual sense-making things involved. In a given social interaction, the agents involved sustain the encounter, while the encounter itself influences the agents and places them into the role of interactors (De Jaegher and Di Paolo, 2007, p.492). The dynamic organization of these joint sense-making processes is significant due to the empirical potentialities attached.

Using dynamical systems tools, an interaction process can be made explicitly measurable. According to Thelen, a dynamical systems model can explain the full range of an individual's behaviors without "…invoking constructs of 'object representation,' or other knowledge structures" (Shapiro, 2019, p.61). Due to the patterns of coordination and breakdown which mirror and exemplify the functions of an autonomous system, the various facets of an interaction process can be reliably accounted for. Here, one's individual autonomy is a necessary precondition for the autonomy of the system. Dynamical systems can be used to show not only

an interaction process, but the blurring of the individual within the process when it is implemented to its fullest extent.

De Jaegher and Di Paolo take this idea farther than Gallagher and other embodied theorists are willing to. The two theorists have consistently alluded to the idea that such models could nullify the relevance of individuals' internal states, as PSM would inherently provide meaningfully measurable representations of those functions. Gallagher, on the other hand, still points to the relevance of the individual as an entity less wholly integrated into the interaction process. The individual, to Gallagher, has a personal and social narrative experience of the world which informs and adds to interaction processes. While that narrative is necessarily present in De Jaegher and Di Paolo's view, it is simply seen as a nondescript portion of the overarching interaction process, particularly when that process takes on a life of its own (as in the hallway example). The inner lives of individuals, through the lens of PSM, do not matter when taking seriously the extent to the potential meaning that could be drawn from interaction processes. A disagreement is clearly present here in terms of how these embodied theorists understand the individual. The conversation takes an interesting turn, however, when De Jaegher, Di Paolo and Gallagher write a paper in which they appear to make concessions to one another regarding the ways they conceive of the individual and interaction in social cognition.

SUBJECTIVITY CAPTURED

It has been shown that De Jaegher and Di Paolo put forward a view of social cognition that focuses on the operationalization of the interaction process as an autonomous and measurable joint sense-making venture. In applying the concept of coordination to the shifting and emerging levels of autonomous identity present in sense-making things, the authors successfully embody social interaction (De Jaegher and Di Paolo, 2008). Through this enactive conception of embodiment, PSM pulls away from both traditional approaches, as well as more recent embodied approaches with respect to the use and role of the individual.

In a paper published by De Jaegher, Di Paolo and Gallagher, titled "Can social interaction constitute social cognition?" the three theorists put forward a carefully constructed opinion. That opinion, however, is one that the theorists seem hesitant to agree upon. The authors state, "Our proposal is…that the role

of interactive and individual elements in social cognition must be systematically re-evaluated" (De Jaegher, Di Paolo & Gallagher, 2010, p.441). In achieving this goal, the authors make concessions to one another which fail to remain present in their later pieces. Throughout the paper, a division between Gallagher's more standard understanding of embodiment and portions of PSM becomes apparent. Though the authors offer a synthesized view, lines can still be drawn between their theories which show a failure to sufficiently recognize implications attached to certain aspects of the individual in less obviously social situations.

To accentuate the division present in the language used throughout the collaborative paper, take another glance at Gallagher's individual. This time focusing on his emphasis on subjective narrative competency, as offered in his book, "The Phenomenological Mind." Narrative competency here can be understood in terms of the continuity attached to an individual's experience as an actor in the world. To Gallagher, narrative competency counts in understanding and operating in interaction. He claims that the pervasiveness of narratives in most cultures, from nursery rhymes to performances of theater and film, expose individuals to a variety of characters, situations, and reasons to act in certain ways (Gallagher, 2012). A narrative then, is an unavoidably relevant portion of an individual's worldly and cognitive experience. Moreover, he claims that a combination of one's cultural and personal narratives, "...provide the background knowledge that allows us to implicitly frame the actions of others in understandable narratives, providing a fallible and revisable sense of what the other is up to" (Gallagher, 2012, p.226). In other words, an individual's understanding and experience of the world, including both interactions with others, as well as passive interaction processes (like watching a movie), are inextricably linked to the cultural and personal narratives attached to that individual's embodied experience. Gallagher's conception of narrative competency applies an embodied twist on concepts fundamental to methodological individualist theories.

Gallagher's disposition becomes apparent in the collaborative piece written with De Jaegher and Di Paolo. The authors make a claim about the inherently sophisticated nature of cognitive processes. They state "…interactive processes are not automatic and higher cognitive processes such as reflection, imagination and self-monitoring can influence them" (De Jaegher, Di Paolo & Gallagher, 2010, p.443). The emphasis placed here upon those higher-level cognitive processes leans towards Gallagher's understanding of the ways in which individuals operate

both implicitly and explicitly in terms of their own narrative competencies and experiences. In this paper, De Jaegher and Di Paolo concede the idea that some individual processes may be left out of their otherwise broadly sweeping enactive account. Cases that involve watching a movie, interacting with a social robot, and giving presentations to a virtual audience exemplify just a few of the issues relevant to both ends of the present social cognition conversation.

Weak concessions lead directly to relevant contradictions. The authors collectively state, "…if we take seriously the idea that interaction can enable and constitute social cognition, we can conceive of interaction dynamics as… delivering the necessary cognitive performance" (De Jaegher, Di Paolo & Gallagher, 2010, p.445). To the authors, in cases of social interaction, there would be no need to duplicate cognitive effects by considering individual mechanisms. This is no longer Gallagher's voice. In fact, he would likely take issue with this devaluation of higher cognitive processes due to the significance he places upon narrative. To Gallagher, narrative matters in terms of both the individual's situatedness, as well as the situation itself.

To PSM theorists, in many cases of social interaction the individual would become a non-distinct portion of an operationally closed system. To De Jaegher and Di Paolo, the individual sense-makers within an interaction process can still affect and be affected by the system as well as processes outside of the system. Whereas Gallagher would presuppose a somewhat dichotomous relationship between individuals and interactions, De Jaegher and Di Paolo attempt to "… supersede such a dichotomy…" using their enactive standpoint (De Jaegher and Di Paolo, 2013, p.2). At its most pure, the way De Jaegher and Di Paolo conceive of an interaction process blurs the line between an individual sense-maker and the participatory joint sense-making venture which accounts for the individual mechanisms present in the subsymbolic system. According to Di Paolo and Thompson in a more recent paper, taking a non-individualistic enactive perspective "…does not imply positing either the individual or the interactive levels as fundamental, but rather understanding the mutually enabling relations between the two levels" (Di Paolo and Thompson, 2014, p.75). In other words, studying social situations requires that the complex relations between individual and collective levels be understood as mutually determining concepts.

To clarify, in the paper "Can social interaction constitute social cognition?" De Jaegher, Di Paolo and Gallagher make concessions to one another in an

unproductive manner. With respect to both of their non-traditional accounts of social cognition, the authors weaken their standpoints. Gallagher weakens IT by accepting that interaction processes can, and in some cases do take into account the whole of individual cognitive functions, including their subjective narrative competencies and experiences. He accepts that in some cases that dichotomy can be made irrelevant on account of the meaning that can be drawn from the overarching interaction process using dynamical systems tools. This is significant in that Gallagher's concession warps his conception of the individual in social cognition. In his work before and after this paper was published, Gallagher consistently leans more heavily upon the subjective intricacies attached to individuals in social situations. The PSM theorists differently weaken their view by conceding that there are social situations which are difficult to explain within their framework. While the paper served to synthesize some of the new thinking present in the social cognition conversation, certain complications arise from the concessions made.

There are some cases of interaction which are disputable. Rather than two or more people interacting with one another in an autonomously organized encounter, these cases involve more of an observational stance. "Such situations are social in an obvious sense and have measurable cognitive effects, but do not involve interactions" (De Jaegher, Di Paolo & Gallagher, 2010, p.443). Cases such as these, especially in light of the more moderate conception of PSM pitched in the collaborative paper, call into question the efficacy of the whole approach. In attempting to blur the line between an individual sense-maker and the sense-making interaction process, PSM theorists leave out explanations about how their model can be applied to passive social instances. For example, PSM fails to reconcile how an individual sense-maker might go about understanding a performance, or an interaction with a non-autonomous social robot. Moreover, traditional views of social cognition can explain these situations by applying foundational aspects of the theories to the passive instance. A person could draw meaningful understanding out of an interaction with a social robot, for example, by using their own internal states to theorize or simulate things about the robot based on its external behavior. In IT, these passive situations could likely be explained with respect to Gallagher's emphasis upon individuals as one traceable piece of embodied socialization. Unfortunately, given the concessions to PSM which Gallagher works with, IT is pulled in two opposing directions. Questions

then arise which remain unanswered. How can more observational forms of social understanding be taken into consideration using an embodied or enactive approach? How do these theories reconcile interaction that is less obviously social? Moreover, is interaction still present in these observational situations? Though these questions haven't been wholly answered, in more recent work, De Jaegher and Di Paolo, as well as Gallagher have each pulled back from the middle ground which they together constructed.

CONCLUSIONS

In theories attempting to explain how humans go about understanding others, as well as understanding with others, serious shortcomings exist with respect to the role and use of the individual. Each theory takes the individual seriously. In TT and ST, the individual is an observer who must use internal theories or simulations in understanding other people and situations. In IT, the individual's embodied experience, as well as their narrative competency are integral to social cognition. IT attempts to consider both interactive and individual levels of socialization. That said, the individual is an acting agent in the world, whose high-level concepts complicate and inform interaction processes. PSM attempts to supersede the implied dichotomy between an individual and the interaction process itself by using an enactive approach. PSM takes an interaction process to be a sense-making thing that enacts its world in a way that at once parallels and also incorporates individual sense-making processes. Using the measurability attached to and the subsequent meaning that can be drawn from the way PSM conceives of interaction, the pieces which make one up become a non-distinct portion of the process as a whole. Meaning can then be drawn from the autonomous nature of the process through the patterns of coordination and breakdown which are present at every conceptual level of the operationally closed system. Individual cognition is present but seen as a portion of the essentially cognizing interaction process.

Though separately these theories each put forward claims about the individual, not one considers the totality of social cognition as a subject. TT and ST fail to recognize the significance of the interaction process and draw hard lines between fundamentally interconnected concepts. IT also fails to take seriously the versatility and measurability of interaction processes. Moreover, it places too

much weight upon the inner lives of individuals in considering social interactions. PSM fails in considering social situations which lack obvious interaction. It takes individual mechanisms to be a portion of the social system, paying less attention to the subjectivity of the individuals involved, and placing emphasis instead upon the subsymbolic system as a whole. This conception of the individual does not well enough consider passive, observational social situations like watching a movie, interacting with a social robot, or giving a presentation to a virtual audience. Regardless, advancement is clear. Moving forward, the role and use of the individual within social cognition should be re-examined once again. Because such a fundamental piece of the discipline is used so inconsistently, progress in solving the problem of social cognition will continue to be a tedious, but worthwhile trek.

REFERENCES

Apperly, Ian A. 2008. "Beyond simulation–theory and theory–theory: why social cognitive neuroscience should use its own concepts to study "Theory of Mind"." *Cognition* 107 (1): 266-283.

De Jaegher, Hanne, and Ezequiel Alejandro Di Paolo. 2013. "Enactivism is not interactionism." *Frontiers in Human Neuroscience* 6 (345): 1-2.

De Jaegher, Hanne, and Ezequiel Di Paolo. 2007. "Participatory sense-making." *Phenomenology and the cognitive sciences* 6 (4): 485-507.

De Jaegher, Hanne, Ezequiel Di Paolo, and Shaun Gallagher. 2010. "Can social interaction constitute social cognition?." *Trends in cognitive sciences* 14 (10): 441-447.

De Jaegher, Hanne, and Ezequiel Di Paolo. 2008. "Making sense in participation: An enactive approach to social cognition." *Emerging communication* 10: 33-47.

Di Paolo, Ezequiel, and Evan Thompson. 2014. "The enactive approach." *The Routledge handbook of embodied cognition*: 86-96.

Di Paolo, Ezequiel, Marieke Rohde, and Hanne De Jaegher. 2010. "Horizons for the enactive mind: Values, social interaction, and play." *Enaction: Towards a new paradigm for cognitive science.*

Gallagher, Shaun. 2005. *How the Body Shapes the Mind*. New York: Oxford University Press.

Gallagher, Shaun, and Dan Zahavi. 2012. *The Phenomenological Mind*. New York: Routledge.

Gallagher, Shaun. 2001. "The practice of mind. Theory, simulation or primary interaction?" *Journal of Consciousness Studies* 8 (5-6): 83-108.

Gallagher, Shaun. 2009. "Two problems of intersubjectivity." *Journal of Consciousness Studies* 16 (6-7): 289-308.

Heath, Joseph. 2015. "Methodological Individualism." *The Stanford Encyclopedia of Philosophy*.

Shapiro, Lawrence. 2019. *Embodied Cognition*. New York: Routledge.

Torrance, Steve, and Tom Froese. 2011. "An inter-enactive approach to agency: participatory sense-making, dynamics, and sociality." *Humana. Mente* 15: 21-53.

compos mentis

Against Virtual Realism

Coner Segren
University of Michigan-Flint

ABSTRACT

As virtual reality technology grows more sophisticated, questions about how we define reality versus unreality become more imperative. David Chalmers tackles this question in his 2017 paper "The Virtual and the Real," where he contrasts two views which he calls virtual realism and virtual irrealism. In his paper Chalmers wants to show that even with our imperfect and temporary VR, the view that virtual objects, events, etc., are real and can be thought of as digital objects. He defines VR as immersive, interactive, and computer-generated. VR proper is said to have all these traits, but some things can be called VR with only one or two of the three. One example Chalmers cites is World of Warcraft. WOW is a virtual world, an interactive computer-generated environment that we seem to inhabit. In these environments we would find virtual objects. The virtual realist would hold that virtual objects really exist, events that happen in virtual reality really happen, experiences we have in virtual reality are not illusory, and that these experiences in virtual reality are valuable just like non-virtual experiences. Virtual irrealism argues the antithesis of each of these qualities. Chalmers wants to defend the view of the virtual realist. In this paper, I will argue against Chalmers' theory of virtual realism. I will present three objections based on what I see as a mistake Chalmers makes regarding how we experience VR and also how the user interacts with VR.

As virtual reality technology grows more sophisticated, questions about how we define reality versus unreality become more imperative. David Chalmers tackles this question in his 2017 paper "The Virtual and the Real," where he contrasts two views which he calls *virtual realism* and *virtual irrealism*. Chalmers defines virtual reality as an "immersive, interactive, computer-generated environment." (Chalmers 2017, 3)

In his paper Chalmers wants to show that even with our imperfect and temporary VR, the view that virtual objects, events, etc., are real and can be thought of as digital objects. He defines VR as immersive, interactive, and computer-generated. VR proper is said to have all these traits, but some things can be called VR with only one or two of the three. One example Chalmers cites is World of Warcraft. WOW is a virtual world, an interactive computer-generated environment that we seem to inhabit. However, we cannot say it is immersive, since the borders of the frame are visible. It is clear we are playing a game. In these environments we would find virtual objects. The virtual realist would hold that virtual objects really exist, events that happen in virtual reality really happen, experiences we have in virtual reality are not illusory, and that these experiences in virtual reality are valuable just like non-virtual experiences. Virtual irrealism argues the antithesis of each of these qualities. Chalmers wants to defend the view of the virtual realist. In Chalmers view, if tomorrow we were to find out we lived in a Matrix, "instead of saying there are no tables, we should say instead that tables are digital (computational) objects made of bits." (Chalmers 2017, 2) In this paper, I will argue against Chalmers' theory of virtual realism. I will present three objections based on what I see as a mistake Chalmers makes regarding how we experience VR and also how the user interacts with VR.

1. CHALMERS' ARGUMENT FOR VIRTUAL REALISM

Chalmers begins by laying out the definitions he will use throughout the paper. It is helpful here to define these things for this paper, as well. Chalmers points out the "Virtual X" used to be defined as something like "as if X but not X." More recent definitions take virtual to mean "a computer-based version of X." In Chalmers view, a virtual sword is simply a computer-based version of a sword, and, just a non-virtual sword is made of atoms at its core level, so a virtual sword is made of "bits."

Chalmers also defends virtual objects and experiences from charges of fictionalism. Fictionalism is the view held by virtual irrealists who believe that virtual worlds, and by extension all objects and events within, are fictional, akin to Middle Earth in the works of Tolkien, which has no geographic location outside of Tolkien's books and all characters and events within are purely products of imagination. It likewise follows from this view that all virtual objects within a fictional virtual world are fictional objects. Chalmers, however, wants to say that even in these worlds there are real objects. Virtual objects are objects within virtual worlds which we perceive and that we interact with as part of the world. These are things like virtual bodies, virtual swords, virtual trees, etc. These are the things that cause our perceptions in the virtual world and digital objects have those causal powers in non-virtual reality by virtue of existing on real computers.

Take a virtual sword as an example. A user may see a virtual sword, they may see a glint of sunlight shine off the blade, the intricate carving on the handle. Chalmers would say the sword has real causal power because it is causing you to see all of these things, to *experience* the sight of a sword. Because virtual objects have this causal power, we can think of them as *digital* objects, as real things that are not fictional and are made of computational bits in the same way an object in the non-VR world may be made of atoms.

Chalmers also argues against "virtual illusionism," wherein the perception of visual worlds is considered illusory. Specifically, the premise he wants to argue against is the premise that "We perceive virtual objects as having the ordinary (non-virtual) colors, locations, and shapes that a corresponding nonvirtual object has." (Chalmers 2017, 15) Also, when laying out the premises for an argument in favor of virtual illusionism, another of the premises he uses is "If one perceives an object as having properties that it does not have, the perception is illusory." This is important because Chalmers writes that this premise "can be regarded as a definition of 'illusion.'" (Chalmers 2017, 15) This is the only definition of illusion he provides, and so it is the one I will use for the sake of my argument. The reason we may consider VR worlds as illusory is because we perceive virtual objects as non-virtual ones. To this point, Chalmers says that there are naïve and sophisticated users of VR technology, wherein naïve users may have false beliefs such as if they are in a VR world, they may believe they are in a non-VR world, in non-virtual space and interacting with non-virtual objects.

As for experiences, Chalmers argues that VR experiences have value just as non-virtual experience does. For Chalmers, the more sophisticated VR users become, the more immersed into the world they are, and are better able to apprehend objects and experiences as being "about as valuable as non-virtual experiences of a non-digital world." (Chalmers 2017, 2) This is because in VR, users can interact and direct the course of the game. Things are not preprogrammed in advance and they can use the game however they want. Their achievements are real achievements, because they came about through their effort. They can also do things in VR, like make friends or write a book or fall in love, that have value as real experiences. In Chalmers' view these are plausibly real actions accomplished through a virtual body.

One last clarification before moving onto the body. I should say that the goal of this essay is to argue Chalmers' claims and definitions in the context of what he calls, "temporary and imperfect virtual realities that are possible with current VR technology." (Chalmers 2017, 2) It is often a fool's errand to speculate on potential future developments and I will not be arguing for the possibility or impossibility of VR that is capable of completely supplanting our existing reality in the future.

2. OBJECTIONS

2.1 The Frame Objection

Chalmers wants to say virtual things are real things, even though they may not be the same things in the real world. Virtual space is still space. His theory is neutral on whether a virtual X is the same as an X in physical space. Chalmers simply wants to define something virtual as being "a computer-based version" of the thing in question. I would dispute Chalmers argument from the standpoint of perceptions.

Chalmers argues that the virtual objects we perceive in a virtual world are the causal basis of perceptions. The causal basis of our perceptions are digital objects, therefore virtual objects are digital objects. And since digital objects are real, virtual objects must be real in this sense. However, I would say that the objects we perceive are not the causal basis of our perceptions. Take the case of a photograph of Albert Einstein. Chalmers would say that when you look at a picture of Einstein you see Einstein, because Einstein is the causal basis of your

perception because our perception depends on Einstein's features at the moment he was photographed and we perceive him *via* the photograph. However, to say this is to ignore the medium which mediates the real Einstein and our perception of him. In truth, the causal basis of your perception is the *picture of* Einstein. The basis for your perception of Einstein is a photo of Einstein.

Chalmers describes the photo as being the "causal basis of our experience, and the features of our experience depend systematically on the features" the man had "when he was filmed." (Chalmers 2017, 9) However, this photo is only a representation of Einstein, lacking many key features Einstein would have had when the picture was taken like consciousness and three-dimensionality, things like his wit, the sound of his laugh, his particular odor. The Einstein of the photo is not real, at least not anymore, it is only a representation. The photograph only is the causal basis of perception.

In VR, your perception is caused by the VR glasses themselves. If you were to use an Oculus Rift, which is an immersive VR headset which covers your eyes and offers handsets allowing you to manipulate the VR, you might feel totally immersed in the reality. If you find a red apple the VR simulation, Chalmers would say the apple is the cause of your "red experience" and that it would be wrong to say that the apple is not really red. Rather, it is *virtually* red, and since it is the cause of a red experience, we can simply say it is red. However, the apple is not the cause of your red-experience, but instead it is the headset itself. VR may seem totally immersive, but the level of immersion in VR is still limited by a frame in the way a picture is, it just so happens in the case of VR that the frame is the boundary between the real and digital world itself.

VR is also similar to a photograph because it is created as a representation of the world. The trees or apples or landscapes of a VR world may be exaggerated, but they nonetheless all carry elements borrowed from the world we live in, i.e. in a fantasy world where the trees have blue leaves, the tree is still meant to be recognized as a tree. Even Chalmers' view that virtual objects are made of computational bits does not rescue VR from the status of representation. Instead of thinking of these bits as akin to the atoms that make up matter, it is better to think of them in the same way we think of the individual brush strokes of a painting. Or to carry on with the camera metaphor, we can think of them in the same way we think of traces of light on film brought out by photochemical compounds.

2.2 The Problem of Illusion

Chalmers holds that VR are not illusory. In VR worlds, things are presented to us as being nonvirtual. If we pick up our virtual sword and closely examine it it will seem to have all the properties of a real sword. This is by design. Chalmers lists two ways VR may (mistakenly he will argue) be perceived as illusory. One is by a false belief wherein the VR world is presented as non-virtual. The second is via perceptual illusion, which Chalmers defines as "a case where an objects[sic] looks a certain way, when it is not that way." (Chalmers 2017, 16) Chalmers believes that there are naïve versus sophisticated users of VR, and this may play a part in whether they have a false belief about VR or whether they experience a perceptual illusion. Naïve users suffering false beliefs about VR may persist in thinking that the environment the are interacting with is physical, non-virtual space. A similar thing happens with perception.

To help illustrate this, Chalmers uses the example of a rear-view mirror in a car. A naïve user with no background using mirrors may look in the rear-view and plausibly experience an illusion by seeing what looks to be cars on the far side of the mirror in front of the car. What the naïve user doesn't know and has no context for, is that the cars are really on the near side and behind the car. The sophisticated mirror user will have an entirely different experience. They will *from the instant they gaze at the mirror* experience the cars as being behind them.

However, mirrors are not obviously analogous to VR experience. Mirrors are designed only to reflect the world; they are not painstakingly crafted to be a simulation of a world. Every open-world video game that Chalmers cites, such as *World of Warcraft*, as well as every VR simulation in general, is made with intention that it *be deceptive* by design which we can tell simply by the fact that creators want it to be as immersive as possible. Any creator of a video game is going to boast about the painstaking detail with which the VR world they created was rendered, and it is done with the intention to immerse the users and deceives them into thinking the VR is as real as possible. In this way, VR and video games are indeed their own artform, and are thus more analogous to art. Specifically, the medium I would liken VR to the most is painting. This is because painting like VR are ultimately representations of the world. Even fantastical VR worlds, there are still recognizable elements drawn from our own world to help orient viewers, and

like a still-life painting, it is created *to leave the impression* of seeing the thing it is representing.

Because of this, it is more appropriate to stay with the original definition of "virtual" which Chalmers argues against in his paper. Remember, Chalmers wants to define virtual as "a computer-generated version of x" as opposed to the older definition, "as if x but not." But the older definition is more accurate because a virtual apple is made by VR designers to appear as if an apple but it is not really an apple. And Chalmers may respond by stressing that it is a *computer-generated* apple made of computational bits. But we know VR is illusory because it was designed to be so. It was designed to fool the user because the end goal of VR is to immerse the user as much as possible into the world. In this way, VR should be judged as an artform or a different form of media than as its own virtual space. Most people would probably agree on the aesthetic merits of VR. After all, much detail and careful design work was put into the creation of the digital world. And as I said before, the goal is much the same as any other artform: to fool the user as to the nature of the world to totally immerse you in it. Even a very dissimilar art form such as literature shares that goal. Readers of a book like *The Lord of Rings* often report feeling "as if they were there" in Middle Earth.

Looking at similar art forms makes this point even more strongly. The details of a still-life painting, such as in Caravaggio's *Basket of Fruit*, which contains the illusion of texture and shadow and depth, are similar to techniques in VR. The art of film is even more immersive since it adds motion and sound and characters. VR adds the appearance of three-dimensionality and the stimulation of more senses. Like other art forms, VR also takes elements from the real world to help draw users in. Even the most fantastical games still have familiar landscapes or human characters. VR should therefore be judged as an artform the *represents* the world in an illusory way. In the same way a painting is representational, we would not call a painting of an apple a paint-generated apple and imply it is its own qualitatively different entity with the medium producing it. A painting is only meaningful in the totality of all its brushstrokes. Similarly, lines of code create the appearance of an apple.

2.3 The Problem of Value

Chalmers argues that virtual worlds and virtual experiences have value. Chalmers defends the value of virtual worlds by arguing against Nozick's parable of the Experience Machine, in which Nozick gives three reasons against plugging into a VR machine. We will focus on the first reason which Chalmers summarizes as "We want to *do* things, and not just have the experience of doing them." (Chalmers 2017, 25) Chalmers believes this doesn't apply to VR, however. "In virtual reality environments, users make real choices, they really do things, and they are genuine sorts of people. Even in limited existing environments such as Second Life, a user can genuinely write a novel, or make a friend, or read a book." (Chalmers 2017, 25)

Chalmers is not wrong to say these acts done in VR can have a kind of value. Some may even find them highly rewarding. However, experiences in VR cannot have equal value to experiences in non-VR or even be said to be real at all because they lack one key feature of real experiences, something which makes them necessarily incomplete, which is the feature of scarcity. It is true that human beings want to do things, and that there are things they can do in VR which will bring them satisfaction. However, it is also true that humans *need* to do things which VR cannot provide. One unique thing about humans is that they are a synthesis of the mental and physical. VR seems to certainly satisfy the mental needs, such as the need for distraction, creativity, and storytelling. But an immersive VR is incompatible with the physical aspect of humans. No matter how immersive the world, a person will also need to break away to eat, since they could not sustain themselves on virtual apples. The experience of eating a real apple will always be more meaningful, since not only will the apple eventually rot, but the labor to grow and harvest the apple is much greater than that needed to create an apple in digital space.

3. RESPONSES & REPLIES

3.1 The Frame Objection

Frame Objection: Chalmers argues that seeing in VR is more reminiscent of ordinary seeing than something like a photo or a film. Chalmers supports this in

three ways. First, that when plugged into a VR experience one needn't have a sense of seeing a screen, and might not see the screen at all. Second, VR provides immersive, three-dimensional perceptual experience and this can be seen from a perspective. Third, one can move around, interact with the world, and potentially alter its course in a VR environment. To Chalmers, the immersive and interactive quality of VR makes it qualitatively different than the aforementioned other mediums and thus more realistic.

In response to Frame Objection, I think it is helpful to point out a certain caveat to the immersive quality of VR, which is that it can only ever be immersive *to a certain extent*. This means that there is a limit to the immersive quality of VR because, unlike real life or non-virtual life, VR can be *unplugged* from. There is always an implicit border between the virtual world and the non-virtual world. This border may vary between the naïve and sophisticated users. For instance, the naïve user may see a virtual object in the distance and reach for it only to end up grabbing air. For the sophisticated user, the border may only exist in the feel of the VR goggles on the face or when the goggles are removed. But the point is that there is a line of demarcation that distinguishes the virtual and the real, and while one may not see the border of the screen it is always implicitly there.

Given Chalmers fascination with *The Matrix*, he may reply that we cannot be sure the reality we inhabit when we take the VR goggles off is real-life, that we may be totally sophisticated, immersed users in a Matrix we are unaware off. While we can never say with all certainty that this is the case, until such a thing can be shown the question is irrelevant. If we think of different reality's, both virtual and non-, as Russian nesting dolls, we should say that the outermost shell in which we have no awareness of a reality outside as our baseline reality, and this is what we should consider real.

3.2 The Problem of Illusion

In response to the illusion problem, Chalmers would say that naïve users are more likely to interpret virtual worlds as fictional. This applies to virtual experiences: how you experience the VR world becomes a matter of sophistication and comfort with the virtual world.

Chalmers cites the use of mirrors as a case for the naïve vs. sophisticated users. The process of looking in a mirror and knowing that what you see in the

mirror is behind you not in front, is not a very intuitive one. Yet, we are able to, without hesitation, interpret the car we see in our rearview mirror as being behind us. This is a case of a sophisticated mirror user. Cognitive penetration, which is the influence of cognition on perception. What one knows or believes influences their experience of the world. If a person has extensive background knowledge of mirrors or has a strong belief about how mirrors operate, it will condition their use of mirrors. Chalmers presents a hypothetical case of naïve versus a sophisticated mirror user "in which a subject sees a chair in a mirror, where in one case the subject believes a mirror is present and in the other subject believes a window is present. The two subjects may have quite different visual experiences: the chairs[sic] appears to be on the near side of the glass for one subject, and on the far side for another. This suggests a direct dependence of perceptual appearance on belief." (Chalmers 2017, 18)

Responding to Chalmers, we must note the way he distinguishes between naïve and sophisticated users. The distinction for Chalmers lies in the users' perceptual orientation of the virtual world. Naïve users experience the world as being illusory, they are fooled into believing virtual objects are non-virtual objects. The sophisticated user on the other hand is more experienced and is better able to interpret the world as virtual. To quote, "A naïve user who does not know they are using virtual reality will undergo the illusion that certain objects are present in physical space in front of them. After they learn they are using virtual reality, the perceptual illusion may persist for a period, but they will not be fooled into believing that the objects are present. After some time, a sophisticated user will become familiar with VR, and they will act in ways that turn on interpreting themselves to be in VR." (Chalmers 2017, 19) On this we can probably all agree. But where I would push back against Chalmers is in his assertion that the sophisticated user is more immersed in the world by virtue of experiencing it *as a virtual world*, and not a fictional one. In fact, the more sophisticated user is more acutely aware of VR's artificiality. Videos of people playing games reveal players who revel in exploring glitches in the game or utilizing cheat codes to gain unrealistic advantages or in general push the limits of the games' worlds. Because they are better aware of the heightened and fictitious nature of the game, these sophisticated viewers take pleasure in manipulating the mechanics of the game to their own ends.

To even better illustrate this, let us look at an example from *The Matrix*. At the end of the film, the main character Neo dies in confrontation with the

film's antagonists, only to be reborn. When he is revived, however, it is with the additional power to see the Matrix as it truly is, which is structures of code. The layer of artifice is stripped away, and Neo sees that in fact the virtual objects he had been perceiving are not truly there. This gives him the ability to manipulate reality and gain enhanced abilities like inhuman speed and dexterity and even flight. Neo in effect becomes the ultimate sophisticated user of the matrix, and the effect is not to be more immersed in the VR, but rather to see through it and its underlying artifice in ways that allow him to take advantage of it.

3.3 The Value Objection

Chalmers wants to say virtual reality has value. To illustrate the value of VR, he responds to three criticisms by Robert Nozick. We will focus on the first two because Chalmers cites them as the more serious objections. To reiterate, the first objection is that people want to do things, not just simulate themselves doing things. Nozick's objection specifically is that in the experience machine, your physical body would simply be in a vat, while scientists would use machines to, "stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book." (Nozick 1974, 44-45) The second is that life in the experience machine is entirely preprogrammed, that you select the experiences you want and they play out on a track for you.

Chalmers thinks neither of these apply to VR. If you make a friend in VR, you genuinely make a friend. You can genuinely read and write a book. And VR as we know it allows you to collaborate and make choices when building your life in the virtual world. In more modern video games and VR machines, nothing is guaranteed, and the user has to act in the right way to accomplish an objective in the game.

However, Chalmers response to the first criticism mistakes the medium for the experience. Let's take the example of writing a book, and imagine two hypothetical video games, of which the objective of both is to write a 200-page novel. Either a) you play as a virtual avatar. The virtual avatar sits down and writes a book, and you watch as a passive observer while the player sits passively watching on their monitor. Maybe you have to make your avatar eat and drink or do things to keep them alive, but the content of the novel is purely generated by the program. Or b) you are the one actually producing the content of the novel. The game is you

typing into your computer and watching your avatar type the things you type. Maybe you make your avatar mimic you as you get up to eat or drink or smoke.

In the first game, the content of the novel would be preprogrammed and the user couldn't say they had actually written a novel because they had put no effort into it. Therefore, it would be difficult to say they'd achieved anything or had a valuable experience. The second case does have value as an experience. However, the fact of it taking place through a VR simulation is not what inscribes value into the event and to say that the value of the experience stemmed wholly or in part from the medium in which it took place would be false. In that scenario, where you had written a 200-page novel as part of a game, it would be strange to say you'd played a game where you wrote a novel. You would simply say you'd written a novel. It is the same with making a friend through VR. The value of the experience, making a friend, is independent of the means through which it happened. Before VR and the internet, people had pen pals, where they had the experience of making a friend without the friend being in physical proximity to the person. In this scenario, the act of letter writing is not in itself valuable, it is valuable as a means to an end. The same is true of experiences in VR.

In conclusion, the main problem with Chalmers theory of the virtual is that it assumes a more idealized and perfect form of VR than what currently exists. As I have stated, the intent of the paper is not to argue the possibility of such an apparatus, but to say that VR as it exists is incapable of meeting the qualities described by Chalmers. While for some the allure of VR may seem more appealing than life outside, the technology is not at place where VR can reasonably even be called real.

REFERENCES

Chalmers, David. 2017. "The Virtual and the Real." *Disputatio* 9 (46): 309-352.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York City: Basic Books.

compos mentis

Enaction, Anātman, & Episodicity: The Self in Principle & Practice

Jack Swick
Villanova University

ABSTRACT

In this paper I consider various accounts of selfhood as articulated in certain traditions of Buddhism, phenomenology, and enactivism. First, I present arguments from the Abhidharma movement of early Buddhism against the existence of an ontologically independent self. Then, I discuss enactivism. I contend that the existence of autopoietic systems contradicts Abhidharmic mereological reductionism, thus undermining early Buddhist arguments for nonself (*anātman/anattā*). Next, I draw on Dignāga's principle of self-awareness (*svasaṃvedana*), and examine phenomenological and enactivist theories of selfhood. I endorse a notion of prereflective self-awareness that constitutes a thin self/subject. Finally, I situate this account of selfhood within Galen Strawson's discussion of episodic lived experience. I put forth the ethical Episodicity thesis, which maintains that the episodically-lived life is desirable because it 1) aligns with the actual nature of selfhood and 2) can reduce suffering, thus preserving the Buddha's insight that attachment to a false notion of self promotes dissatisfaction and unhappiness.

Changing it rests.
—Heraclitus

I. ANĀTMAN

Early Buddhists of the Abhidharma movement argue against the existence of an ontologically independent self. That is, they argue against some unchanging entity that gives each person an individual identity. This entity would hypothetically persist across time and physical change. This might be conceived of as a soul. From the time one is born until death, one's mind and body are under incessant transformation. The atoms of which one's body is constituted are in constant flux. We finish each year with almost none of the same physical material as when we brought in the new year. Similarly, the beliefs, desires, motivations, and even personality traits that characterize our mental lives are subject to change over time. So what accounts for the seemingly persistent sense of identity that sets each one of us apart as individuals? Here one might posit a self that maintains our identity throughout the lifetime. It is this kind of permanent self that early Buddhists reject.

The Buddha holds that we suffer because we are ignorant of impermanence (*anitya*), the true nature of suffering, and the reality of nonself (*anātman/anattā*). Understanding nonself is vital in Buddhist soteriology: *saṃsāra* (the cycle of birth, death, and rebirth characterized by suffering) continues as long as we misunderstand the self (Siderits 2007, 32). The doctrine of nonself maintains that there is no part of a person that accounts for permanent identity over time. In short, people are empty of selves (33). If there were an independent entity that accounts for individual identity, it would probably be a part of a person, since, after all, it is what makes that person *them*. So nonself arguments analyze the constituents of the human person in order to see if we can find anything we might call a 'self.' According to this view, humans consists of five *skandhas*, which include: *rūpa:* anything physical (i.e. your body), *feeling*: positive, negative, and neutral sensations (here *feeling* is a technical term that doesn't denote emotions, but instead refers to responses of varying valence to changing phenomena), *perception:* mental events involving sensing characteristics of an object (e.g. seeing the color blue), *volition:* mental forces responsible for activity (both physical and psychological) such as hunger and wakefulness, and *consciousness*:

awareness of physical and psychological states. Collectively the *skandhas* are referred to as *nāma-rūpa*, which translates as name and (physical) form. *Rūpa* — physical matter — can be seen, while the latter four *skandhas* can only be named, since they are not immediately observable by others (*ibid.* 35-37).

Early Buddhism mounts two arguments for nonself: the argument from impermanence, and the argument from control. The argument from impermanence holds that *nāma-rūpa* cannot constitute the self, since the *skandhas* are impermanent. The argument proceeds as follows:

1. *Rūpa* is impermanent

2. Sensation is impermanent

3. Perception is impermanent

4. Volition is impermanent

5. Consciousness is impermanent

6. If there were a self it would be permanent

IP. [There is no more to the person than the five *skandhas*]

C. Therefore there is no self. (Siderits 2007, 39)

Physical matter is under constant change: atoms move and are replaced all the time. So *rūpa* cannot be permanent. Sensations arise and pass away, e.g. one's back can feel fine one instant, and then one can be gripped with pain, and then be fine again later. Perceptions come and go depending on what is going on in one's environment — they depend completely on external phenomena. Volitional desires emerge only in response to specific situations (e.g. pulling one's hand back from a hot stove) and then pass away. Once away from the stove, that volition passes, and others will soon take its place, and so on. So therefore sensation, perception, and volition are impermanent. What about consciousness? Consciousness here is merely the awareness of psychophysical events. This can't be permanent since it depends on being aware (you can't be asleep or unconscious). Would we say that we're a different person each morning following a deep sleep? Would getting surgery mean the pre-surgery 'you' was extinguished and a new and numerically different 'you' came about? Few would grant these

points. Therefore consciousness is impermanent as well. Since there is no more to a person than the *skandhas,* and the *skandhas* are impermanent, we see that there is no permanent self (Siderits 2007, 39-46).

Here one might object that the *skandhas* are not an exhaustive list of what constitutes the human person. Could there be more to a person than *nāma-rūpa*? Here we might posit a permanent entity 'hidden' above or among the *skandhas.* This entity might be called a soul, and would be what gives each person her individual identity. The early Buddhist response to this appeals to the 'principle of lightness,' which holds that we should choose the 'lighter' or more parsimonious of competing theories in order to arrive at the best explanation. We should do this since otherwise we would posit and believe in things that are unobservable and for which have no evidence. If we're trying to explain something, why respond by positing an entity which itself requires further explanation? If we were to posit an unseen self, we would need further explanation and evidence. We can instead turn to what is observable and needs no further explanation (Siderits 2007, 43-46). Siderits writes that the "Principle of Lightness says we should resort to positing unobservable entities only when the world tells us we have no alternative" (45). And in the case of the *skandhas*, the principle of lightness applies, leaving us with no hidden permanent entity.

The control argument begins with the assumption that, if there were a self, it would be the thing from which executive control emanated; it would be the source of autonomy. The self would be able to respond to feelings and change them when unpleasant. The self would be in control. The control argument appeals to the anti-reflexivity principle, which states that a thing cannot operate on itself. The classic metaphor for this is a knife: a knife can cut other things, but it cannot cut itself. But what about a doctor that operates on herself? Here again, the doctor is not truly operating on herself, but is operating on a *part* of herself, say, her foot. She would likely need her hands to do this. Her feet are not operating on themselves, her hands are. Thus, the anti-reflexivity principle holds. The control argument goes as follows:

1. 'I' consist only of *skandhas.*

2. I can change the *skandhas.*

3. An entity cannot operate on itself.

4.   A self would be the part of the person that performs executive functions.

C.   There is no self.

Since the *skandhas* are subject to executive control, they cannot be the source of control (anti-reflexivity). If there were a self, it would be the part of the person that is in control. Yet we consist only of *skandhas*, so there is no abiding self (Siderits 2007, 46-49).

The Abhidharma movement of early Buddhism lays a foundation for arguments for nonself in mereological reductionism/nihilism. On this view, wholes are not ultimately real. What is real are the most fundamental parts that constitute the whole. Anything that can be reduced to smaller entities is not ultimately 'real,' but rather is conveniently labelled as a whole in order to make communication and interaction easier. The only real things are impartite entities. The classic example is a chariot. A chariot is not really a whole thing, since it can be broken down into its subcomponents. It has wheels, a carriage, shaft, axles, and other parts that, when assembled in a particular way, are referred to as 'chariot.' We designate this assembly as its own entity, but the whole 'chariot' depends on its parts for existence, so it is neither fundamental nor independently existant. Thus a chariot is not ultimately *real*. We don't call chariots the collection of chariot-parts; we call them chariots since it is easier to communicate that way. It is in our interest to refer to chariots as independent wholes.

Mereological reductionism is justifiable because our interests do not determine reality. Just because something is useful or convenient for us to believe doesn't necessitate that it is ultimately real. (Siderits 2007, 54-56). The same applies to living beings. Since what we call a 'person' can be analytically reduced to the *skandhas*, that person is not an 'entity' existing independent of its constituents. Though we refer to people as independent selves, selves do not really 'exist' independently of their parts. This is why we can't label each collection of changing *skandhas* as a self.

For the sake of communication and getting by in daily life, early Buddhists advance the concepts of convenient designation and conventional truth. For the purposes of day-to-day life, we have to use words such as "I." "I" is a convenient designator or helpful fiction we use to refer to the collections of *skandhas*. 'Persons' are not the wholes we refer to them as, but are instead their

fundamental parts. But for daily life, we must refer to wholes as wholes and not as a fiction superimposed on a bunch of impartite components (Siderits 2007, 49). Here we see a distinction between conventional and ultimate truth. Something is conventionally true if and only if it is commonsensical and leads to successful practice. Something is ultimately true if and only if it corresponds to reality as it really is and does not assume the reality of any conceptual fictions (56). It is unhelpful to think of chariots as not real. Sometimes it is helpful to think of oneself and others as whole persons. But suffering arises from attachment to the self, which for Abhidharma Buddhists is illusory. So understanding that what we refer to as 'selves' are really impermanent packets of *skandhas* is necessary for *nirvāṇa* (*ibid.* 56-64).

II. THE AUTOPOIETIC SELF

In *Mind in Life*, Evan Thompson discusses the enactive approach to cognition, also called enactivism. Enactivism is a type of embodied dynamicism, which maintains that cognitive systems are self-organizing systems which emerge from circular, nonlinear causality of continuous sensorimotor interactions between brain, body, and environment. The metaphor for this is not a neural network enclosed by skull and flesh, but instead a mind as an embodied dynamical system in the world. Embodied dynamicism draws from two approaches. The first is the dynamic systems approach to cognition, which maintains that cognition is a temporal phenomenon and must be understood in terms of dynamic systems theory. The second is embodied cognition, which holds that cognition is the use of skillful know-how in situated and embodied action (Thompson 2007, 11).

Enactivism contends that "the human mind emerges from self-organizing processes that tightly interconnect the brain, body, and environment at multiple levels" (Thompson 2007, 37), and is based on two ideas: autonomous or self-determining systems, and emergence, which describes large, collective patterns of functioning. Enactivism unifies several ideas:

1. Living beings are autonomous agents that generate and maintain themselves and thus enact their own cognitive domains.

2. Nervous systems are autonomous agents that actively generate and maintain their own coherent and meaningful patterns of activity in

conformity with their operation as circular and reentrant networks of interacting neurons. Nervous systems create meaning; they don't process information as in the computationalist account.

3. Cognition is the exercise of skillful interaction in situated and embodied action. Cognition emerges from recurrent sensorimotor loops of perception and action.

4. A being's environment is not represented internally in its mind, but instead is a relational domain enacted by that being's agency and means of interacting with the environment.

5. Experience (sentience) is not epiphenomenal, but rather is central to understanding minds.

Enactivism thus offers a new way of understanding cognition not as the doings of a neuronal computer situated within the skull, but as the activities of organisms situated within meaning-imbued environments, and the interactions of brain and body with the world (13).

Dynamic interaction between organism and environment begins at the "lowest" level of organism. Thompson's philosophy of the organism depends on the notion of autopoiesis, which contradicts Abhidharmic mereological reductionism, and thus undermines arguments for nonself. Autopoiesis (literally "self-making") is a property of living systems, such as cells, that permits those systems to maintain and renew themselves by creating and regulating their boundaries (e.g. a cell wall), and by regulating its composition by means of metabolism and other processes. Autopoietic systems are a form of emergent dynamical systems, such as tornadoes, but are different from other dynamical systems because they regulate and maintain themselves. In an autopoietic system, the whole depends on its parts, but the existence of the parts depends on the whole as well. For instance, a cell is not reducible to its subcomponents, since its organelles cannot exist except within a functioning cell. We can see how autopoietic systems undermine mereological reductionism. Abhidharmic arguments in favor of mereological reductionism presuppose that wholes are able to be reduced to their constituents. Living things such as cells *cannot* be reduced to their parts; the whole is something different than its parts. The groundwork of Abhidharmic mereological reductionism is thus in trouble.

Thompson's ontology of the organism establishes notions of identity and minimal selfhood. Autopoietic systems are emergently but immanently purposeful; meaning that purposiveness is constitutive of the system, and is not determined from the outside. This consists of two "modes": identity and sense-making. *Identity* holds that autopoietic systems create and maintain an identity amidst change. *Sense-making* posits that autopoietic systems interact with and make sense of their environments in order to remain viable; enaction endows the otherwise insignificant physicochemical environment with significance and valence, creating an *Umwelt* or environment (Thompson 2007, 146-147). Thompson lays out two propositions that complement each other and describe two aspects of the autopoietic process:

1. An organism is fundamentally a self-affirming, identity-producing process based on autopoiesis.

2. A self-affirming identity establishes logically and operationally the reference point or perspective for sense-making and a domain of interactions. (147)

Organisms thus create for themselves an identity. This identity is not independent of an organism's environment, but instead is established *in relation* to the environment. The organism's sense of meaning — what it takes as good or bad, what is useful or not, etc. — is not 'built into' the environment, but is endowed onto the world by that specific organism by means of ongoing interaction.

An organism's dynamic identity and sense-making gives it a sort of minimal selfhood. By autopoietically differentiating themselves from their environments, organisms maintain their identities. Thompson maintains that "a living cell stands out from a chemical background as a closed network of self-producing processes that actively regulates its encounters with the environment" (Thompson 2007, 149). The difference between a cell and the chemical soup in which it is situated is its active maintenance of its boundary and composition. The autopoietic minimal self is not the *what*, but the *how*. Numeric identity/ontological independence refers to what something *is*, especially in contradistinction to other things. The self cannot be the things of which it is constituted, since these are impermanent, as the Buddhists noted. So the self must be the *way in which* matter is organized

— the pattern that endures amidst perpetual material change. Thompson puts it well:

> An organism is a material being, and its reality at any given moment coincides completely with its material constitution. Yet its identity cannot be based on the constancy of matter because its material composition is completely renewed…Only at the level of form or pattern can we find constancy in the flux. (150-151)

> Metabolism is the constant regeneration of an island of form amidst a sea of matter and energy. Metabolism establishes a self with an internal identity marked off from the outside world and whose being is its own doing…An organism must subordinate every change it undergoes to the maintenance of its identity and regulate itself and its interactions according to the internal norms of its activity. Life is thus a self-affirming process that brings forth or enacts its own identity and makes sense of the world from the perspective of that identity. The organism's 'concern,' its 'natural purpose,' is to keep on going, to continue living, to affirm and reaffirm itself in the face of imminent non-being. Incessant material turnover and exchange with the environment is both a reason for this concern and the only way to meet it. (*ibid.* 152-153)

Thus an organism's identity is not established independent of its environment, but instead it is in relation and interaction with the world that organisms create and maintain identity and meaning. The cell paradigm of selfhood (i.e. enactive or autopoietic identity) is therefore a 'verbal' conception of selfhood, in which 'self' is more verb than noun. Autopoietic systems such as cells are first-order autopoietic systems, whereas multicellular organisms, like human beings, are second-order autopoietic systems. In either case, the system establishes for itself its own identity, its own 'self' (*ibid.* 105). This selfhood is "minimal autopoietic selfhood," which does not imply consciousness or "phenomenal selfhood," which requires a nervous system (*ibid.* 162). So Thompson here is not claiming that individual cells are sentient, nor is he endorsing panpsychism. Instead, he is establishing that autopoietic systems have distinct, irreducible identities that persist through time.

Though the existence of autopoietic systems contradicts mereological reductionism, it is clear that there is still no need to indulge the idea of an eternal and ontologically independent self. Thompson's notion of minimal autopoietic selfhood is quite different from the self with which early Buddhists were concerned. Now that we have arrived at a basic conception of selfhood that is more than mereological reductionist nonself but much less than an eternal soul, let us build upon this 'middle ground' between the two extremes.

III. THE MUTE SELF

Before discussing alternative theories of selfhood, we must explore the notion of self-awareness as articulated by the Buddhist logician and scholar Dignāga (c. 480-540 ce). Dignāga's concept of self-awareness (*svasaṃvedana*) holds that all mental states are intrinsically self-aware. This awareness is not the result of another mental state, but instead is an intrinsic property of mental states, such as memories, sense-perception, et cetera (Kellner 2010, 204). Self-awareness is not when one is consciously aware of something or of oneself, rather, it is prior to conscious recognition of an object of cognition. Self-awareness describes how one is aware of both an object of perception and also the awareness of *being* aware. For instance, when one sees the color blue, one is aware of blueness, but one is also aware of *seeing* the blue. Self-awareness is "an immediate, non-conceptual mode of awareness that provides access to how mental content (including feelings, etc.) presents itself subjectively." In this way, self-awareness is an explanation of the *how* of mentality, not a descriptor of *what* constitutes mentality (227-228). Self-awareness is not reflective or introspective, since both of these require a higher-order mental state (*ibid.* 215).

Why can't each mental state be made aware by another mental state? Dignāga's proof of self-awareness addresses this objection (Kellner 2010, 213). It must be the case that either 1) self-awareness is an intrinsic property of cognition/cognitions are self-experiencing, or 2) cognitions are made aware by separate cognitions. Option 2) implies an infinite regress of cognitions, since perception *A* would be cognized by cognition *B*, which would be cognized by cognition *C*, which would be cognized by cognition *D*, and so on *ad infinitum*. Dignāga chooses 1) over 2) as the best explanation, because the regress implied by 2) would itself need further explanation, while self-awareness by Dignāga's account would

explain experience and why experience seems first-personal. Self-awareness is prior to becoming consciously or reflectively aware of experience. Self-awareness is logically independent of intentional self-awareness (in which one is consciously aware of being aware of something), and intentional awareness itself may require self-awareness (206).

As a Buddhist, self-awareness for Dignāga does not constitute a self. Like Dignāga, Dan Zahavi maintains that consciousness is self-aware, yet for Zahavi this constitutes a thin or minimal self (Zahavi 2011, Zahavi 2017). According to this view, phenomenal consciousness or experience is intrinsically first-personal, even when one cannot linguistically articulate it (such as in the case of infants and nonhuman animals). Hence the label 'prereflective,' since one need not have any advanced cognitive or linguistic abilities to be a self. Selfhood is prior to reflection, that is, prior to thoughts and words. Sentience is "self-disclosing" or "self-revealing" (Zahavi 2017, 198). The first-personal character of consciousness is what differentiates sentience from non-sentience, and one's own experience from the experiences of others. Zahavi puts it nicely:

> Some might object that there is no property common to all my experiences, no stamp or label that clearly identifies them as mine. But this objection is misplaced in that it looks for the commonality in the wrong place. The for-me-ness or mineness in question is not a quality like scarlet, sour, or soft. It doesn't refer to a specific experiential content, to a specific *what*, nor does it refer to the diachronic or synchronic sum of each content, or to some other relation that might obtain between the contents in question. Rather, it refers to the distinct givenness or *how* of experience. It refers to the first-personal presence of experience. It refers to the fact that the experiences I am living through are given differently (but not necessarily better) to me than to anyone else. It could consequently be claimed that anybody who denies the for-me-ness or mineness of experience simply fails to recognize an essential constitutive aspect of experience. Such a denial would be tantamount to a denial of the first-person perspective. It would entail the view that my own mind is either not given to me at all — I would be mind- or self-blind — or

> present to me in exactly the same way as the minds of others."
> (2011, 59, emphasis original)

This says nothing about the contents of consciousness. Instead, the prereflective self is the reflexive *what-it-is-like-for-me-ness* of consciousness (2017, 194). Here Zahavi is not positing a new and permanent entity, nor is he arguing for anything more than what a materialist account of the human person would grant. Instead, Zahavi submits that the self is "the very subjectivity of experience, and is not taken to be something that exists independently of, or in separation from, the experiential flow" (2001, 60). The very nature of consciousness makes a subject out of each sentient organism.

If the requirement for selfhood is only to be conscious, it isn't much to be a self. Hence the alternative terms 'thin self' and 'minimal self.' The prereflective self is a 'thin' or 'minimal' self in the sense that it is not a rich enough concept to contain every sense of the word 'self,' such as a social self, but rather is prior to *all* uses of 'self' (Zahavi 2011, 67). In this way, the prereflective self predicates any interpersonal or narrative sense of self, but does not contradict it (2017, 194-195). Zahavi's notion of prereflective selfhood is useful because it allows for more highly-elaborated conceptions of self. Any account of self, however, will necessarily presuppose the prereflective self.

In *Engaging Buddhism*, Jay Garfield challenges prereflective selfhood. Garfield identifies several tautologies in contemporary phenomenological accounts of self. Consider the claim that self-awareness is the "'first-person givenness or manifestation of experiential life.'" Garfield responds that this is tautological: "How else could I know my own life?" (Garfield 2015, 163). Or examine the claim that "'if there is no awareness of the experience, the object does not appear at all.'" This again is a tautology: this basically says "if there is no appearance of the object, there is no appearance of the object" (165). Once more, Garfield writes that "Kriegel claims that consciousness consists in a kind of penumbral halo around every experience, whether perceptual or cognitive, that reveals it as *mine*...If the argument is meant to show that when I have an experience, it is *mine*, the claim is an empty tautology" (*ibid.* 166, emphasis original). Garfield is arguing that invoking 'experience' explains nothing, and that claiming consciousness is simultaneously aware of its content and its own awareness is to make an empty and false claim.

Garfield here does not consider why these claims are ostensibly tautological. The arguments for prereflective selfhood sound this way for a reason: they are describing what it is to be conscious, something with which all of us are intimately acquainted. To lay out the properties of *A* is not to say that "*A* is *A*," but rather to make a descriptive claim about how *A* is. These 'tautological' claims seem this way because to say that consciousness has first-personal character is seemingly to state the obvious. Garfield's accusations of a dearth of scientific or philosophical evidence (Garfield 2015, 166) miss the point: does the fact that consciousness is first-personal need any further defense? At the level of prereflective selfhood, we reach a certain epistemic bedrock.

Thompson defends the reflexive prereflective self by reconstructing a classic memory argument and addressing its objections. The argument goes as follows:

1. When one remembers (say) yesterday's vivid blue sky, one remembers not simply the blue sky, but also seeing the blue sky. In other words, one remembers not just the object seen, but also the visual experience of seeing. Thus the memory comprises both the objective side of the perception (the object seen) and the subjective side of the perception (the seeing). (Phenomenological claim)

2. Thus no additional cognition is necessary in order to recall the subjective side of the original experience. (Phenomenological claim)

3. To remember something one must have experienced it. (Conceptual claim)

4. The causal basis for features of the present memory is corresponding features of the past experience. (Causal claim)

5. So the past visual perception must have included an experience of the seeing, along with the object seen. In other words, the perception must have included an awareness of itself as a visual perception, which is to say that it must have been reflexively self-aware. (Conclusion) (Thompson 2011, 162)

Thompson contends that this argument is an appeal to the best explanation (163). One may object to premises 1 and 4, and argue that the experience of seeing blue is only inferred after the fact upon visiting the memory of seeing blue. One has the memory of seeing blue, and infers that one was aware of seeing blue at the time of the original perception. Thompson submits that this objection is false because it gets the nature of memory wrong.

Memory entails an intrinsic "character of pastness" that imbues it with a feeling of being a former experience. Perception is presentational, while memory is re-presentational (Thompson 2011, 164). When we remember an object, it is "re-presented" to us in consciousness in the present. Yet, though we are conscious of a memory in the *present*, the object of memory retains an historical tone. The experience is given to us as having already happened. The objection fails to explain why memories retain this character of pastness, that is, why they 'feel' different from perceptions of presented objects and imagined futures. The Husserlian account of memory as presented by Thompson contends that every memory contains in it not only the object of memory, but also the implicit awareness of experiencing that object (164-166). This implicit "character of pastness" emerges because of the nature of time-consciousness: the "now-phase" of consciousness retains the "just-past" phase, and is retentionally self-aware, allowing one to be aware of objects over time. The previous *conscious experience* of perceiving the object, as well as the *object* of consciousness itself, is re-presented in consciousness. Each conscious moment retains an impression of the moment that just passed, giving consciousness a temporal character. Thus one is implicitly aware of having experienced something in the past, allowing one to nonreflectively differentiate between a presented perception and a re-presented object of memory (*ibid.* 166-167). Thompson argues that this implicit awareness is only possible by means of the reflexively self-aware nature of consciousness (*ibid.* 166-167). Again, consciousness is simultaneously aware of its object and of experiencing the object, which includes the "pastness" of experience reconstructed in memory. This, Thompson contends, is a better account of memory than that presented by objectors.

As a result of its intrinsically first-personal, reflexive nature, consciousness creates a subject, which, Thompson argues, constitutes a prereflective self. From the standpoint of phenomenology, we need not posit an enduring self independent of psychophysical events; rather, consciousness is first-personal,

which constitutes a 'thin self,' which is by definition prior to any reflective sense of "I" (Thompson 2011, 168). This constitutes a "self-as-subject" (172). Here Thompson and Zahavi are in agreement. Any sense of an enduring self emerges *within* consciousness, that is, it is *constituted* by consciousness. There emerges a "self-as-object" of consciousness, just as there are other objects of consciousness (*ibid.* 172-173). The "self-as-object" emerges only later, and is predicated on the existence of a "self-as-subject." The reflective sense of self emerges from its reflexive first-personal stream of psychophysical events, and is "fundamentally I-making (*ahaṃkāra*)." Subjectivity implies a prereflective self, and later feelings of 'I-ness' emerge, giving oneself the impression that one has a self or ego which persists through time (*ibid.* 173).

The nature of time-consciousness influences the lived experience of one's self. As noted above, time-consciousness "comprises both awareness of external things and their temporal characters, and awareness of experience itself as temporal and as unified across time" (Thompson 2007, 318). Experience includes not only the objects of consciousness, such as changing perceptions and emotions, but also the "character of pastness" that gives one the impression of enduring through time. Though the stream of consciousness is in perpetual flux, we do not experience life as a series of instantaneous moments. Rather, our experience is constituted by a temporal character, giving rise to the impression that the *I* that exists in this moment was the same *I* that existed in the past. The present is not experienced like a "knife-edge," but as a "duration block." This duration block is an intentional object of time-consciousness, and is constituted by three intentional acts: *primal impression*: the "now-phase" of an experience; *retention*: the "just-now phase," directed towards the moment that just slipped away; and *protention*: the future-oriented phase, which anticipates something coming next (318-319). These three 'acts' occur together, and collectively make up the duration block that marks our moment-to-moment experience (The words *intentional* and *act* might be misleading here. *Intentional* refers not to an intention or something of which we are consciously aware, but instead is used in its phenomenological meaning, referring to how consciousness 'aims toward' or 'intends' something [*ibid.* 22]. *Act* here doesn't refer to anything done voluntarily or consciously, but is rather something that 'is done.'). To use Thompson's example, consider listening to a melody. At any moment of listening, one experiences the note or notes being played *at that instant*. At the same time, one is co-aware of a note having

just 'slipped-away.' One *retains* this implicit knowledge of having just heard a note. One is also simultaneously anticipating a new note to arise to follow the one being played right now. This threefold character of experience — retention, primal impression, and protention — makes up the duration block. If it were not for retaining and anticipating coming experiences, we wouldn't experience a melody as a coherent unit. Instead, we would only hear a series of distinct and unrelated sounds.

The way we experience time gives rise to the impression of enduring through time. Thompson writes that "the unified operation of protention, primal impression, and retention underlies our experience of the present moment as having temporal width" (Thompson 2007, 319). Each moment we experience what is directly at hand, we retain what just happened (the object of experience as well as the experiencing of it), and expect something new to come soon. This continues as long as one is conscious. Consciousness is thus horizontally unified, and is related to itself, since the present moment of consciousness is implicitly aware that one was conscious in the past (322).

For Thompson, the duration block of time-consciousness *is* the prereflective self (Thompson 2007, 322-328). Continuing with the example of hearing a melody, Thompson asserts that "the threefold structure of time-consciousness entails prereflective self-consciousness. At the same time one is aware of the melody, one is implicitly co-aware of one's ongoing experience of that melody, thanks to the threefold temporal structure of one's experience" (322). The very nature of consciousness creates a subject, a prereflective self. Time-consciousness — the experience of the duration block — is the bedrock of all consciousness; it is presupposed by all other conscious experiences. It is not constituted by time but rather is *constitutive* of time (*ibid.* 323-325). Thompson follows Zahavi in equating inner time-consciousness and prereflective self-awareness. There is no transitive/object-directed experiential awareness, but instead:

> there is only experience of temporal objects and events in the world, as well as the prereflective and intransitive self-awareness of those very experiences. When we listen to a melody we hear the melody (transitive consciousness), but we also subjectively live through our listening (intransitive consciousness). The listening has a subjective character that makes it immediately manifest, without observation or inference, as one's own experience...The

> subjectivity of the experience consists essentially in its being intransitively and nonreflectively self-aware. Or rather it consists in its being prereflectively self-aware, for it can come to be reflected upon but is necessarily prior to any such reflection." (*ibid.* 327)

To sum it up: consciousness is prereflectively, intransitively, and reflexively self-aware. It is prereflective since one need not express it, and we can assume that non-speaking beings like infants and sentient nonhuman animals are subjects as well. Consciousness is fundamentally intransitive, since the fundamental subjective nature of consciousness is not 'directed at' any object. Subjectivity is prior to all experiences, so it doesn't matter what is *in* consciousness. Subjectivity abides as long as one is conscious. And relatedly, consciousness is reflexively self-aware because it is simply a property of consciousness to give the impression of first-personal subjectivity. That is, consciousness makes a subject. We might refer to this notion as minimal phenomenal selfhood. I find this theory of selfhood, as articulated by Thompson and Zahavi and supported by Dignāga's argument for self-awareness, to be not only convincing, but also existentially interesting, as we will see in the next section.

IV. THE LIVED SELF

In 'Against Narrativity,' Galen Strawson argues against two popular claims: the *psychological Narrativity thesis*, and the *ethical Narrativity thesis.* The psychological Narrativity thesis is a descriptive theory that claims humans experience their lives narratively (i.e. like a story) (Strawson 2004, 428). This thesis contends that humans are natural story-tellers, and we all tell ourselves stories about our own lives, with each of us the protagonist of one's own story. This is a descriptive claim, so it doesn't say whether our self-concerned narrative predilection is good or bad — it just *is*. The ethical Narrativity thesis, on the other hand, is a normative theory that maintains that humans *should* view their lives narratively, because a rich personhood necessitates a story-like outlook on one's own life (428). On this view, self-narration is necessary for leading a healthy and moral life.

Strawson submits that both of these claims are false. Strawson draws a distinction between two types of lived experience: diachronic and episodic self-experience. In diachronic self-experience, "one naturally figures oneself,

considered as a self, as something that was there in the (further) past and will be there in the (further) future" (Strawson 2004, 430). In this type of self-experience, one remembers one's own past, is aware of the present, and can expect a future. One views oneself as persisting as the same person through time. In episodic self-experience, on the other hand, "one does not figure oneself, considered as a self, as something that was there in the (further) past and will be there in the (further) future" (430). Life is a series of 'episodes.' Although one has memories of being a person in the past, and can expect to exist in the future, one does not feel like the same 'person' or 'self' that was there in the past, and doesn't expect to be the same 'person' there in the future. Diachronic self-experience typically involves narrativity, while episodicity implies a non-narrative outlook (*ibid.* 430-432).

In order to understand what episodic self-experience is like, Strawson expounds two ways to think about one's 'self.' One can consider oneself as a human being, an organism (i.e. a second-order autopoietic system/a minimal autopoietic self). One is, of course, the same organism from the beginning of one's life until death. Each human being has memories, experiences, desires, personality traits, can think about existing in the future, etc. Strawson does not deny this. Being the same organism makes memories and expectations possible. But one can also think about oneself as a "mental entity," as an experiencer or locus of consciousness (i.e. phenomenal selfhood, but a richer notion than Zahavi's prereflective self). Strawson refers to this experiencing 'thing' as I*, me*, my*, and so on (Strawson 2004, 429-430). I exist as a human being, and did in the past. But only I* exist now, as a subject. Strawson maintains that "I'm well aware that my past is mine in so far as I am a human being, and I fully accept that there's a sense in which it has special relevance to me* now, including special emotional and moral relevance. At the same time I have no sense that I* was there in the past, and think it obvious that I* was not there, as a matter of metaphysical fact" (434). One's past as an organism does shape what it is to be me* right now, but that does not mean that I* existed in the past. One can expect to exist and be conscious in the future, but one cannot expect I* to exist in the future.

While it is true that some people experience life narratively, the psychological Narrativity thesis is false because not *all* humans experience life that way. Strawson writes that he himself, like many others, experiences life episodically, thus contradicting the psychological Narrativity thesis (Strawson 2004, 433-434). Strawson also argues that the ethical Narrativity thesis is false because it

is possible to live a healthy, fulfilling, moral, and emotionally rich life without thinking of oneself narratively (432-433). The episodic life is in no way deprived of meaning or morality, and the narrative life is no more desirable than episodic self-experience. In short, narrative self-experience is not the only, nor the best or healthiest, way to experience one's self.

In *Waking, Dreaming, Being*, Thompson considers how one should think about the self as a living person. Thompson contends that, though we are routinely deceived about the nature of selfhood, the "mineness" of conscious experience is not a delusion (Thompson 2015, 359). Prereflective minimal selfhood (which gives rise to the feeling of "mineness"), as we have seen above, implies a subject and an agent, but not a substantially existent (ontologically independent) ego. The minimal self distinguishes 'my' experiences from everyone else's experiences (361). "I" is a label given to individuated streams of conscious experience. Using the word "I" doesn't imply the existence of an enduring ego. Instead, saying "I" is a performative utterance: "I" appropriates experience as one's own in contradistinction to the experiences of others (*ibid.* 362-363). The fact that phenomenal experience is available to you and only you, makes you *you*; it makes you a subject and permits you to say "I."

Drawing on Candrakīrti, Thompson posits that the self is "the dependently arisen and constructed appearance of an independent subject of experience and action" (Thompson 2015, 365-366). The metaphor for the self is an image in a mirror: the mistake is not in taking the mirror to be *real* — since it exists in some capacity — but rather in taking the image in the mirror to exist in its own right. The self is a construction, not an illusion. Cutting through the illusion of an independent and eternal self can still be done by means of contemplative practices and analytical insight (365).

So how should we think about our 'selves?' Thompson argues that wisdom does not entail annihilating all sense of self, but instead, it includes "knowing how to inhabit that activity ("I-making") without being taken in by the appearance of there being an independent self that's performing the activity and controlling what happens" (Thompson 2015, 366). That is, it is wise to understand that what we call "I" is a label placed on the stream of consciousness, which is predicated on the thin self. Yet where does the thin or minimal self actually get us? Given the importance of interaction in enactivism, and the importance of sociality for human well-being, how do we square such a sparse account of selfhood with

the need to act as enduring entities? Another way of putting it is this: don't we need some richer notion of selfhood that includes a sense of narrativity in order to meaningfully engage with others? Thompson's (and my) answer to this is "yes," but in a qualified way. Memory and prospection are necessary for constructing a narrative of self. Memory endows us with an autobiographical history, and prospection allows us to imagine ourselves in the future (348). These allow one to feel like the same person throughout the 'story' of one's life.

Yet one need not *identify* with one's memories or prospective thoughts — one can instead take them for what they really are: they are just *thoughts* arising and passing in a moment full of other arising and passing phenomena, such as sounds and feelings (Thompson 2015, 349-350). Being lost in thought generates so much of the suffering individuals put themselves through: rumination over past embarrassments and failures, worries over the uncertain future, etc. These thoughts are harmful insofar as one takes them to define who one is. But one can choose not to identify with one's thoughts, and instead recognize them simply as occurrences in experience. One can *feel* the difference between identifying as the "I" of those thoughts, and identifying that a thought is passing by within one's greater experiential field (350). Here Thompson distinguishes between "narrative focus," in which one identifies with descriptions of oneself, and "experiential focus," in which one can observe one's phenomenological experience from moment to moment without identification or judgement (*ibid.* 354). One can train to be more experientially focused by means of meditation and other contemplative activities. After reviewing neuroscientific studies of mindfulness practices, Thompson asserts:

> it's easier to disengage from narrative forms of self-identification
> when we have the kind of training in present-centered awareness
> that mindfulness practices provide. Although we need narrative
> thinking to understand ourselves as individuals with personal
> histories and plans for the future, and as members of traditions
> and communities, we can easily get stuck in worrisome rumination
> about our past and future selves, or become attached to some
> mental representation of ourselves. Individuals with mindfulness
> training seem better able to adopt an experiential focus and
> avoid getting stuck in the narrative focus. In other words, they
> seem able to move flexibly between narrative thinking about

themselves and present-centered, embodied awareness, and imaging their brains accordingly brings to light the distinct neural systems supporting these two kinds of self-experience. (*ibid.* 355)

Accordingly, episodic lived experience is not a theoretical goal or something only certain people are born to do: it can be learned and practiced. There is a place for narrativity in one's life. We have to indulge some illusions in order to get by. However, exactly how much narrativity is necessary for a good life? Thompson doesn't say, and there may not be just one answer. Maybe it's up to each of us to discern that relationship for ourselves. Perhaps wisdom entails contemplative practices such as mindfulness in conjunction with deep reflection about how much narrativity one really needs in one's life.

Strawson's notion of episodicity squares nicely with Thompson's advocacy of experiential focus and mindfulness. To abet episodic self-understanding, one might employ mindfulness practices. Given the amount of suffering brought on by narrative conceptions of self, I see little reason to entertain the ethical Narrativity thesis. I would like to put forward, to parallel Strawson, the *ethical Episodicity thesis.* One should live/view one's life episodically because 1) it is closer to the truth about selves (the minimal self is a constantly fluctuating and egoless *process* of subjectivity), and 2) I think the Buddha is right in his assessment of the self: seeing the self for what it is reduces suffering. I see the success of mindfulness as support for the ethical Episodicity thesis. Through some forms of meditation, one can discover for oneself the centerlessness or egolessness of consciousness. One can reap real benefits from meditation. Episodicity and mindfulness allow one to cultivate for oneself a mature relationship with one's thoughts, memories, goals, and anything else we would usually ascribe to an enduring self. In this way, we can better understand ourselves as organisms and as mental entities. In doing so, hopefully we can save ourselves from some unnecessary suffering.

REFERENCES

Garfield, Jay L. 2015. *Engaging Buddhism: Why It Matters to Philosophy.* New York: Oxford University Press.

Kellner, Birgit. 2010. "Self-Awareness (*svasaṃvedana*) in Dignāga's *Pramāṇasamuccaya* and *-vṛtti*: A Close Reading." *Journal of Indian Philosophy*, 38: 203-231.

Siderits, Mark. 2007. *Buddhism as Philosophy: An Introduction*. Indianapolis: Hackett Publishing Company, Inc.

Strawson, Galen. 2004. "Against Narrativity." *Ratio* XVII: 428-450.

Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge: The Belknap Press of Harvard University Press.

Thompson, Evan. 2011. "Self-No-Self? Memory and Reflexive Awareness." *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*. Edited by Mark Siderits, Evan Thompson, and Dan Zahavi. New York: Oxford University Press.

Thompson, Evan. 2015. *Waking, Dreaming, Being: Self and Consciousness in Neuroscience, Meditation, and Philosophy*. New York: Columbia University Press.

Zahavi, Dan. 2011. "The Experiential Self: Objections and Clarifications." *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*. Edited by Mark Siderits, Evan Thompson, and Dan Zahavi. New York: Oxford University Press.

Zahavi, Dan. 2017. "Thin, Thinner, Thinnest: Defining the Minimal Self." *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. Edited by Christoph Durt. Thomas Fuchs, and Christian Tewes. Cambridge: MIT Press.

Aesthetic Flow: The Implications and Neural Correlates of a Goal-Directed Aesthetic Experience

Jack P. Vavrinchik
Elmhurst College

ABSTRACT

This paper offers an argument for the existence of a type of goal-directed aesthetic experience. By weaving together recent research in the fields of philosophy, psychology, and neuroscience, I show how an artists in flow experience—a necessarily goal-directed mental state—can nevertheless have aesthetic experiences (Csikszentmihalyi 1991; Csikszentmihalyi 1996; Dietrich 2004; Ulrich et al. 2014). Recent approaches to understanding aesthetic experiences from the perspective of neuroscience have suggested that a non-goal-directed (or disinterested, in Kantian terms) mental state is needed to achieve an aesthetic experience (Kant 2009; Brincker 2014). Contrary to this, I will argue that aesthetic experiences can occur in the context of goal-directed states.

INTRODUCTION

Simply put, flow state is the feeling of being "in the zone."[1] Experts in nearly any skill-based activity describe experiences in which they felt completely immersed in the task at hand, with a sense of control over their actions without "thinking" about them. Examples of flow vary from music ensembles executing difficult music with extreme precision for extended periods of time, to assembly line workers being able to construct their units in record time. Flow state allows for explicit thought in relation the agent's actions to be bypassed, removing the risk of failure due to not being able to "think" fast enough (Dietrich 2003, 746-761; Csikszentmihalyi 1991, 36-49).[2]

The connection between flow state and the aesthetic experience lies within the creative state of the performer. A number of psychological studies have illuminated how musicians and songwriters have aesthetic reactions to their work while in flow. In a study of musical therapy techniques, a survey of highly trained musicians revealed that they felt a "higher desire to experience and express feelings through music [during flow]" (Woody and McPherson 2011, 405). Subjects who composed original songs during flow reported meaningful emotional responses to their art in the moment of creation as a result of the experience. These studies went on to suggest that subjects who are able to achieve deeper flow states while playing music or writing a song will have stronger and more meaningful experiences while creating (Baker and Macdonald 2013, 131-140; Baker, MacDonald, and Pollard 2018, 17-19). Woody and McPherson's 2011 study focused on the group flow state that occurs between musicians in an ensemble setting. During this type of flow experience, musicians were quoted explaining their emotional reactions to the structural components of the music they just played as "lovemaking," being "high," or in "ecstasy" (Woody and McPherson 2011, 405). The emotion and meaning experienced by these artists during the moment of creation is clear evidence of an aesthetic response during flow. Their aesthetic states appeared to be a response to their own and other's creations, as well as perceptions of their external environment. This evidence demonstrates that some flow states can constitute aesthetic experiences. And, if flow is a necessarily goal-directed mental state, then it must also be true that some aesthetic experiences are goal-directed

---

1. Throughout this paper, I use flow, flow state, and flow experience interchangeably.

2. *Agent* and *Performer* will also be used synonymously throughout this paper.

mental states (Woody and McPherson 2011; Baker and Macdonald 2013; Baker, MacDonald, and Pollard 2018).

In a 2015 article, Maria Brincker developed a theory that proposes a neurological explanation of Kant's disinterested aesthetic experience. To have an aesthetic experience, Kant argues, we need to get our minds into a state of "free play," where our imagination can take its course due to scaled back cognition. Brincker argues that the disinterested states described by Kant are necessarily non-goal-directed. So, we must get our mind out of a practical, goal-directed state before we can have a disinterested brain state; only then do we have the chance to have an aesthetic experience. Though disinterest seems to be a part of how our brains approach the aesthetic state, it does not account for the contexts of all of our aesthetic experiences. As a goal-directed state in which we execute practical thinking and have the opportunity for aesthetic experiences, flow provides a counterexample to Brincker's claim (Csikszentmihalyi 1991, 39-40; Dietrich 2004, 746; Kant 2009, 198; Brincker 2014, 5). The aesthetic moments one can experience during flow state calls for a reassessment of the breadth of Brincker's argument.

FLOW STATE

For a person to get into flow, the defining characteristics of the mental state, first identified by psychologist Mihaly Csikszentmihalyi, must be fulfilled (Csikszentmihalyi 1991, 39-41, 66-67):

1. The skill/challenge balance.
2. Goal-orientation.
3. Immediate feedback to one's actions.
4. Absorption.

Once fully immersed in the experience of their task, the performer begins to undergo the characteristic phenomenological experiences of flow:

1. A mitigation of self-consciousness.
2. Subsequent eliminations of processing the experience of time, worry of failure, and distractions.
3. Autotelic feelings related to the performed task.

TRANSIENT HYPOFRONTALITY THEORY

Flow experience involves a continuous interaction between perception, lower level information processing, and the skills responsible for the execution of the task. These functions span what some neuroscientists call the "implicit" and "explicit" systems of the brain. The implicit system, which is located primarily within the basal ganglia and the larger networks it belongs to, is responsible for skill-based knowledge and efficiency, while the explicit system is responsible for executive function, self-consciousness, and cognitive flexibility. Dietrich's proposed explanation of flow experience is that the brain enters a state of "transient hypofrontality," temporarily blocking the self-conscious and self-analytical functions of the explicit system. Transient hypofrontality theory has been confirmed by observations of decreased regional cerebral blood flow in the dorsolateral prefrontal cortex (DLPFC) and the medial prefrontal cortex (MPFC)--areas of our brains responsible for self-conscious and -analytical processing--during flow conditions (Dietrich 2004, 746, 748-750; Ulrich et al. 2014, 199).

Transient hypofrontality does, however, allow for the cognitive flexibility and "sustained and direct attention" from the DLPFC to be activated. The resulting state is a trade-off between the efficiency of the implicit system and the flexibility of the explicit system. When acting on a task during flow, the implicit system's routine skills are merged with the explicit systems novel approach to a task. In music, for example, the jazz musician's implicit skills are informed by the technique she has developed practicing the transcriptions of "the greats" she's played hundreds of times. Their practiced musical vocabulary forms a bank of skills to draw from and is supported by the basal ganglia and procedural memory. The cognitive flexibility afforded by the explicit system allows for novel approaches to the note "choices" around those licks while improvising over the song's chord changes. These functions achieved by brain networks provide a sense of constant interaction between processing of incoming information and the agent's output in the task at hand, equating to a high level of focus, and the feeling of being "in the zone" (Csikszentmihalyi 1991, 41; Dietrich 2004, 746, 748-750; Ulrich et al. 2014, 199).

THE SKILL/CHALLENGE BALANCE

Flow experience requires a happy medium between the challenge of the task and the agent's skill set. If this isn't satisfied, the ability to get into, or maintain a flow state can be interrupted. If a melody is perceived to be too difficult, and broke our musician out of flow, they would need to engage their executive functions, i.e. their explicit thinking, for however brief a time, to try to re-enter flow. At the other end of this continuum, are those tasks which are not challenging enough and do not present the need for novel and semi-demanding action. Because of the lack of perceived challenge, the explicit and implicit systems don't achieve transient hypofrontal interaction, which can lead to daydreaming or a similar state (Dietrich 2004, 757). Furthermore, boredom is a highly self-conscious state in which the person knows that "what *I am* doing right now is not stimulating." This is not to say that ostensibly boring tasks themselves cannot be used to begin to enter into a flow experience, but rather an example of a self-conscious state that has the ability to end flow conditions (Csikszentmihalyi 1991, 39). In a 2014 study, Ulrich et al. were able to observe the skill/challenge condition during induced flow experiences. They found that neural activity in the inferior frontal gyrus (IFG), an area aligned with adaptation to task goals, peaked when presented with the optimal level of difficulty during flow (Ulrich et al. 2014, 199).

ATTENTIONAL REQUIREMENTS

Focused attention must be sustained until complete absorption is achieved in the task at hand in order to have a flow experience. In this way, the process of concentration involved in attaining flow is very similar to the process used to enter a meditative state. This highlights the idea that the attentional systems are still involved in flow, but their function and experience are highly specific and specialized. Therefore, understanding the agent's concentration as *absorption* in their task, rather than sustained attention, is much more useful. This description gets at to the unique relationship between information processing and output, which includes the processing of multiple perceptions at the same time without self-conscious and executive interference. We can thus identify the state of absorption during flow as deeply related to transient hypofrontality, which is also responsible for the loss of the sense of self and time due to the shutdown of

nearly all of the MPFC and DLPFC (Morita et al. 2008; Dietrich 2003, 243, 244; Csikszentmihalyi 1991, 66-67).

GOAL ORIENTATION

A series of nested goal-directed mental states are necessary to achieve flow experience. An agent's actions must have clearly defined goals every step of the way, which are often combined with larger overarching goals. For example, while playing a song, a musician has the small, clearly defined goals of playing the right note after each note for the duration of the song. Each note gets built into a melodic line, which, when compounded, expresses the entirety of the song's form, further expanding the breadth of their smaller goals. All the while, she has the overarching goal of playing the song to completion (Csikszentmihalyi 1991, 39-40; Dietrich 2004, 757; Ulrich et al. 2014, 3530-3532, 3540-3543).[3]

AUTOTELIC ACTIVITY

According to Csikszentmihalyi, activity which results in flow states results in the intrinsic, self-contained benefit of the task. This concept in flow is called the autotelic experience, meaning that the activity is an end in itself. By playing the saxophone with their group for the sake of creating and interacting with other musicians and audience members, the musician is focused on the benefits only found by the very of performing the activity. Conversely, if the saxophone player was playing only for the prospect of acquiring prestige, fame, and money, they would be focused on the "exotelic" results of the action. Therefore, Csikszentmihalyi argues, even though the product and execution of both situations are completely the same-- playing the song successfully--the saxophone player who is going into the activity focused on exotelic results does not achieve the intrinsic benefit that is attributed to flow experiences. The problem with Csikszentmihalyi's analysis of the exotelic approach is that it disregards the agent's ability to achieve flow experiences based purely on the skill based, goal-oriented, and attentional aspects of flow. Positive feelings in response to flow states have been reinforced by neuroscientific analysis;

---

3.   These studies found that the goal-directed determinant of flow correlated with increased regional cerebral blood flow to the putamen, which is involved in guiding and coding goal-directed action.

however, these results were independent of what each subject's autotelic and exotelic attitude on the tasks performed. Therefore, it appears that the autotelic aspect of flow may not be as necessary as previously outlined by Csikszentmihalyi (Csikszentmihalyi 1991, 39-40, 66-67, 111-113; Dietrich 2004, 757-758; Dietrich 2003, 243, 244; Ulrich et al. 2014, 194, 198, 200).

CHARACTERISTIC MENTAL STATES

The sharp decrease in self-conscious processing during flow can be attributed to the low levels of neural activity in the MPFC and DLPFC. Since the mind does not have the ability to have additional self-conscious information brought into the flow experience without exiting the flow state, distractions, the fear of failure, the self-critic, and extraneous content, are eliminated from perception. Flow, however, is not unbreakable and can be interrupted by external stimuli. If an audience member is acting belligerent, yelling, or throwing tomatoes at the performer, this would surely be enough to break the musician out of their flow experience. Tomato throwing aside, low to mid-level distractions, like chatter throughout the audience or a loud conversation, would not be enough to break a seasoned performer's flow experience. Distortion of time during flow experience is a hallmark of the flow state condition. Our perception of time is handled by the DLPFC. When the mind enters transient hypofrontality during flow, the portions responsible for temporal perception are shutdown. This is what accounts for musicians who often recall playing or practicing for hours on end without any clue to how long (or short) of a time they've been playing (Ulrich et al. 2014, 195, 200; Dietrich 2004, 756-758).

AESTHETIC MOMENTS DURING FLOW STATE

The individual musician's need to listen to themselves or the rest of the ensemble that they're playing with gives them the ability to have aesthetic responses to both their own sounds and the sounds of the group. Aesthetic experiences constituted by flow are non-cognitive due to the all but complete shut-down of the MPFC and DLPFC. So, to the musician with their group, the aesthetic flow perception would come across as the feeling of just "beautiful," or "groovin'," rather than the propositional thought of "this melody sounds beautiful," or "the drummer is groovin'." The first example makes internal, non-

thought-like reference to the external sounds the group is creating. Due to the high level of relevant information processing in flow, the first example is an aesthetic perception, with no *explicit* reference to the self or other individuals involved. Conversely, the second example is cognitive, as the musician makes a reference to herself in relation to events in the external world, as well as a reference to other individuals (Woody and McPherson 2011, 405).

DISINTERESTED AESTHETIC EXPERIENCES
VS. AESTHETIC FLOW STATES

Disinterest is a concept proposed by Immanuel Kant in his 1790 work, *Critique of Power and Judgment.* Kant argues that a level of psychological distance is required during the aesthetic experience to allow our imagination to interpret the art as an art object, as well as have "free play" in our minds (Kant 2009, xxix, 102, 198). Using a neuroscientific approach, Maria Brincker's aesthetic stance hypothesis plays on these themes formulated by Kant. Brincker claims that a non-goal-oriented mental state is the necessary link that allows us to access "free play," in turn allowing "deeper subjective involvement," or "freedom of imagination" (Brincker 2014, 5).

Brincker connects her neurological analysis of disinterestedness with correlating brain states, claiming that deactivation of executive function causes the brain to enter a non-practical mode, while an active DMN is the connection to the "free play of imagination" (Brincker 2014, 25). At first glance, this claim appears to be true. The executive system is responsible for a majority of what we qualify as practical, goal-oriented function. Examples of these actions could be writing a paper, organizing, prioritizing, maintaining focus, and other tasks which require critical thinking. DMN activity is used when we are daydreaming, doing mental time travel, aren't partaking in any particular task, or thinking of ourselves and our relation to others. All of these activities can be easily correlated with freedom of imagination.

Brincker's application of neuroscience to her hypothesis sought to outline the brain conditions of being an aesthetic beholder and show how those conditions were in stark contrast to a goal-directed mental state. Studies on the neural landscape of deep aesthetic experiences continue show increased DMN activity (Reybrouck et al. 2018; Vessel et al. 2012, 1-17; Vessel et. al. 2013, 258-275).

However, shutdown of the DMN during flow states display that there is not a necessary correlation between activity in the default system and the aesthetic experience (Ulrich et al. 2014, 195-200). The previously mentioned musicians and songwriters also showed that they were able to have meaningful aesthetic moments during flow--a state of non-activity in the DMN. The possibility for aesthetic perception during flow shows how the conditions of being an aesthetic perceiver are not necessarily in contrast to goal-oriented attitudes. Transient hypofrontality interaction during flow bypasses executive function within a goal-directed state, demonstrating how our minds are capable of creating the constitutive states for an aesthetic experience even while we may hold a more "practical attitude" (Dietrich 2004, 756-758; Brincker 2014, 21).

CONCLUSION

The possibility for aesthetic perceptions during flow shows that there is no mutually opposing dynamic between goal-directedness and aesthetic perception. In fact, flow experience integrates our ability to access the complex emotional processing of an aesthetic experience while maintaining a practical, task-oriented attitude. In contrast to Brincker's claims, aesthetic flow illuminates how a-typical brain systems can constitute an aesthetic experience depending on the mental state of the agent. In the future, comparing the brain systems active during aesthetic flow with the systems normally active during an aesthetic experience could delineate a framework for where the aesthetic experience exists in the brain as a whole. It is my hope that this paper can be used as a call for further interdisciplinary engagement into flow state's effects on the aesthetic experience, as well as further inquiries into the combination of philosophy and neuroscience as a whole.

REFERENCES

Baker, Felicity A., and Raymond A. R. Macdonald. 2013. "Flow, Identity, Achievement, Satisfaction and Ownership during Therapeutic Songwriting Experiences with University Students and Retirees." *Musicae Scientiae* 17 (2): 131-46. doi:10.1177/1029864913476287.

Baker, Felicity Anne, Raymond Afredo Rossi MacDonald, and Mark Clement Pollard. 2018. "Reliability and Validity of the Meaningfulness of Songwriting Scale with University Students Taking a Popular Songwriting Class." *Arts & Health: International Journal for Research, Policy & Practice* 10 (1): 17–28. do i:10.1080/17553015.2016.1236281.

Brincker, Maria. 2014. "The Aesthetic Stance: On the Conditions and Consequences of Becoming a Beholder." In *Aesthetics and the Embodied Mind: Beyond Art Theory and the Cartesian Mind-Body Dichotomy*, 117-38. Springer Netherlands.

Csikszentmihalyi, Mihaly. 1991. *Flow: The Psychology of Optimal Experience*, New York, NY: Harper Perennial.

Csikszentmihalyi, Mihaly. 1996. *Creativity: Flow and the Discovery of Invention,* New York, NY: Harper Perennial.

de Manzano, Örjan, Simon Cervenka, Aurelija Jucaite, Oscar Hellenäs, Lars Farde, and Fredrik Ullén. 2013. "Individual differences in the proneness to have flow experiences are linked to dopamine D2-receptor availability in the dorsal striatum." *Neuroimage* 67: 1-6.

Dietrich, Arne. 2003. "Functional neuroanatomy of altered states of consciousness: The transient hypofrontality hypothesis." *Consciousness and Cognition 12:* 231-256.

Dietrich, Arne. 2004. "Neurocognitive mechanisms Underlying the Experience of Flow." *Consciousness and Cognition* 13: 746-761.

Hori, Yukiko, Takafumi Minamimoto, and Minoru Kimura. 2009. "Neuronal encoding of reward value and direction of actions in the primate putamen." *Journal of Neurophysiology* 102 (6): 3530-3543.

Kant, Immanuel. 2009. *Critique of Power and Judgment.* Edited by Paul Guyer. Translated by Eric Matthews. Cambridge: Cambridge University Press.

Morita, Tomoyo, Shoji Itakura, Daisuke N. Saito, Satoshi Nakashita, Tokiko Harada, Takanori Kochiyama, and Norihiro Sadato. 2008. "The Role of the Right Prefrontal Cortex in Self-evaluation of the Face: A Functional Magnetic Resonance Imaging Study." *Journal of Cognitive Neuroscience* 20 (2): 342-55. doi:10.1162/jocn.2008.20.2.342.

Reybrouck, Mark, Peter Vuust, and Elvira Brattico. 2018. "Brain Connectivity Networks and the Aesthetic Experience of Music." *Brain Sciences (2076-3425)* 8 (6): 107. doi:10.3390/brainsci8060107.

Vessel Edward, Gabrielle Starr, and Rubin Nava. 2012. "The Brain on Art: Intense Aesthetic Experience Activates the Default Mode Network. *Frontiers in Human Neuroscience* 6: 1-17.

Vessel Edward, Gabrielle Starr, and Rubin Nava. 2013. "Art Reaches Within: Aesthetic Experience, The Self, and the Default Mode Network." *Frontiers in Human Neuroscience* 7: 258-275.

Ulrich, Martin, J. Keller, K. Hoenig, C. Waller, and G. Gron. 2014. "Neural Correlates of Experimentally induced Flow Experiences," *NeuroImage* 86: 194-202.

Woody, Robert H, and Gary E McPherson. 2011. "Emotion and Motivation in the Lives of Performers." In *Handbook of Music and Emotion: Theory, Research, Application.* Oxford: Oxford University Press.

compos mentis

On Epistemic Effects of Cartesian Skepticism

Xisheng Wang
Michigan State University

ABSTRACT

The doubt brought by the hypothesis of Cartesian skepticism is one of the most vital and difficult challenges in the field of philosophy. The reason why it is vital is that it challenges people's everyday knowledge claims about the external world. It is difficult to solve because the reasoning involved in Cartesian skepticism relies on some principles that we use in our daily life. In this paper, I will introduce three important philosophers, G.E. Moore, Robert Nozick, and Fred Dretske and their theories, which are believed by many people to have successfully answer Cartesian skepticism. Nevertheless, in my opinion, these theories, in fact, did not really entirely quell the doubts brought by the Cartesian skepticism. The objective of this article is to analyze the shortcomings of each theory in answering the doubts brought by Cartesian skepticism. Finally, I will elaborate on why Cartesian skepticism is unavoidable and unanswerable under the current conceptions of knowledge and I will offer some advice regarding what strategy we should choose in the future in order to solve the problem of Cartesian skepticism.

KEYWORDS

Epistemology, Cartesian Skepticism, External World, Epistemic Closure Principle, Tracking Theory, Epistemic Operator, Relevant Alternative Theory, Infallibilism

1. INTRODUCTION

In our daily life, most people usually don't have much doubt about the propositions such as: "I know that I have two hands" or "I know that there is a green tree in front of me." For those people, to be precise, they don't question their "perceptual knowledge" or the veracity of their "perceptual experiences." In other words, they usually believe what they see and what they feel. However, most people also have had the experience of waking up from a dream in a heavy sweat early in the morning and only to realize that what they just "experienced" was only a vivid nightmare. This example clearly illustrates that people can have the exact experience of, say, a green tree in front of them in a dream just as they would experience it in waking life. Now, here is a problem. How do people really know something when they are claiming to know? Or how do they know that they are not just dreaming? This puzzle successfully leads us to today's topic.

The dreaming hypothesis mentioned above is one of the scenarios of Cartesian skepticism's argument, which is our topic today. Under the Cartesian extreme skeptical position, we seemingly don't have any knowledge about the external world since we cannot exclude the possibility that what we take to be veridical experience is just an elaborate dream. Such a skeptical argument is highly destructive because it challenges people's claims to know very ordinary things about their environment. In the book, Critique of Pure Reason, Kant exclaimed that: "It still remains a scandal to philosophy… that the existence of things outside of us… must be accepted merely on faith, and that, if anyone thinks good to doubt their existence, we are unable to counter his doubts by satisfactory proof" (Kant 1929, 34). For many philosophers, like Kant, the conclusion of the Cartesian skepticism is intolerable because we do KNOW that "I have two hands" and "there is a green tree in front of me" and many other perceptual beliefs. As a result, under such an atmosphere, anti-skeptical philosophers came to propose many theories in answer to the challenge of Cartesian skepticism. Some philosophers challenge the Cartesian skeptic by attempting to prove the existence of the external world. Other philosophers challenge the Cartesian skepticism by rejecting the principles relied on in their skeptical reasoning. While these philosophers' theories can seem persuasive, their arguments are not without problems. The objective of this paper is to show the deficiencies of several standard responses to Cartesian skepticism, and the philosophical consequences of these deficiencies. Finally, I will offer some

thoughts about why Cartesian skepticism is unavoidable and unanswerable under the current conceptions of knowledge, and what the strategy should we use in the future in order to solve the doubts brought by Cartesian skepticism.

2. CARTESIAN SKEPTICISM

2.1. Descartes and his Cartesian Skeptical Methodology

René Descartes is one of the most important French philosophers in the 17th century. In his book, Meditations on First Philosophy, he thoroughly introduces what has come to be called "Cartesian skepticism" and the "method of doubt," a form of skeptical reasoning that gives rise to it. After reading the Meditation I, "Concerning Those Things That Can Be Called into Doubt," it's not hard for us to notice that the purpose of meditation is to build an absolutely reliable foundation for knowledge system. To this end, Descartes believes that he must abandon all the opinions and views he held before and start from the ground up. In order to achieve this purpose, Descartes put himself in an extreme skeptical position and tries to raise various challenges[1] to his daily thoughts and beliefs. One of the challenges is the dreaming hypothesis as we learned above. If any belief could survive under such extreme skeptical conditions, this belief must be the foundation of the knowledge system. For Descartes, skepticism serves as a methodological tool to clear out the unreliable beliefs that we took for granted before. The skeptical argument proposed by Descartes is not a pure skepticism, because its purpose is not to doubt for the sake of doubt, but to establish a solid knowledge system. The difference in purpose makes him not a real skeptic. This idea is crucial because when we study other philosophers' theories to challenge the Cartesian skepticism, we have to keep in mind that the rejection of Cartesian skepticism is not the only aim; we have to consider whether these philosophers' theories violate the intention that Descartes purposes the Cartesian skepticism, which is putting thinkers into an extreme environment to establish a solid knowledge system.

---

1. Other challenges include the Cartesian demon hypothesis. Under the Cartesian demon hypothesis, people are living in a world full of illusion created by an evil demon. These illusions can be perceptual, logical, even mathematical. I will mention this hypothesis later in this article.

2.2. G.E. Moore: "I have two hands!"

If people are dreaming, people might not have knowledge about the external world. This is because every sense of the "world" could be merely an illusion in the dream. This reminds us that if we can prove the existence of the external world, we can at least shake ourselves loose Cartesian skepticism. This is the exact strategy that Moore decides to use in his article, "Proof of an External World," where he attempts to prove the existence of the external world by showing that he knows that he has two hands[2]. This is how Moore's proof works:

PI) Here is one hand (making a gesture with left hand)

PII) Here is another hand (making a gesture with right hand)

C) The external world exists

The existence of the two hands intuitively proves that the outside world exists. Moore said confidently that the proof was unquestionable and completely rigorous because this proof meets the three conditions required for a proper proof: (1) the premise is different from the conclusion, (2) the premise is known, (3) the conclusion is derived from the premises.

Let's take a moment to consider each condition: (1) "the premise is different from the conclusion" is a very important precondition for a cogent argument, because if the conclusion is merely a restatement of a premise, then the proof commits the logical fallacy petitio principii, question begging. Moore has succeeded to avoid this. Moore also believes that his demonstration successfully shows (2), "the premise is known", because it's ridiculous for a person to say that "I don't know that there are two hands in front of me," when two hands are presented before him. Finally, (3) is also shown up in Moore's demonstration because the conclusion is indeed a logical consequence of the premises; in other words, the deduction is valid. If Moore has successfully proven that the external world exists, we must be living in a world without the evil demon. Therefore, the argument of Cartesian skepticism is false.

---

2. Hands are somethings that exist independently of the mind, so they are parts of the external world.

2.3. René Descartes: "Are you sure?"

In my opinion, Moore's mistakes can be classified into two kinds: (1) misunderstanding the aim of raising Cartesian skepticism; and, (2) question begging.

First, as I mentioned in section 2.1., the reason why Descartes raises Cartesian skepticism is in order to exclude all possible wrong beliefs and seek the foundation of our knowledge system. The Cartesian skepticism is just a methodology rather than an aim. It's very obvious that, in Moore's demonstration, he does not put himself into an extreme skeptical position in the very first place. Perceptions are questionable under the extreme skeptical position. For instance, how do you know that you are not just dreaming that there are two hands in front of you? Therefore, Moore's demonstration violates the intention of Cartesian skeptical methodology and he is sneaking a doubtful belief, which is "he has two hands", into the field of basic knowledge.

Second, it is true that the premises in his demonstration are different from the conclusion. However, this does not mean that there is no question begging in his proof. This question can be seen in two viewpoints. First, he takes as a premise, a claim, that his audience, in this case, the skeptic would not grant— that here is a hand. For skeptic, maybe it's only an illusion-of-a-hand, or dream-hand, or the appearance-of-a-hand. So, Moore begs the question against the skeptic by assuming something that he needs to prove to the skeptic, namely that here is one hand. Second, "here is one hand" is a belief that needs to be justified by "the external world exists." In Moore's demonstration, Moore claims that there are two hands without justification. The important reason why he cannot offer the justification is that "the external world exists" is Moore planning to prove. Hence, Moore does make the fallacy of question begging during his demonstration.

Based on these two reasons, I suppose that Moore's proof does not really solve the issue brought by Cartesian skepticism; instead, he cleverly avoids the issue.

3. EPISTEMIC CLOSURE PRINCIPLE

3.1. Cartesian Skeptical Syllogism[3]

Our beliefs about the external world come, directly or indirectly, through sensory experience. However, our sensory experience does not seem to be completely reliable. We all have experiences of illusion and hallucinations; even things that seem clear and obvious can, at times, be doubted. These ordinary cases of sensory illusion can make it seems as though it's entirely possible that we are just dreaming or living in a Cartesian demon world. Hilary Putnam in his 1981 book, Reason, Truth, and History, purposed the famous Brain in a Vat[4] (BIV) skeptical scenario. The BIV hypothesis proposes that an evil scientist removes someone, S's, brain and puts it into a vat of nutritious liquid and uses a computer to stimulate the brain to produce sensory experiences qualitatively indistinguishable from those of our ordinary experience. But in fact, everything S feels is just a series of computer signals. The main point of Putnam's skeptical theory is that if we do not know that we are not a BIV, or we cannot rule out the hypothesis, H, that we are, then we don't have the knowledge of the external world. If we simplify the skeptical argument, it can be formulated as follows syllogism:

P1) S doesn't know ~H[5]

P2) If S knows P[6], then S knows ~H

C) Therefore, S doesn't know P

If we put Putnam's skeptical hypothesis into the syllogism, it can be expressed as follows:

P1) We don't know we are not BIV

---

3. Here, when I say Cartesian skeptical "syllogism", I do not really mean the syllogism by logical definition. Cartesian skeptical syllogism is a form a Cartesian skeptical argument that made in three sentences.

4. Interestingly, Putnam himself, just like Descartes, wants to refuse the skepticism, but his theory also accidentally provides a good argument for skepticism

5. "H" represents the skeptical hypothesis. "~" means "negate" or "not"

6. "P" represents the ordinary propositions

P2) If we know we are seated, we know we are not BIV

C) Therefore, we don't know we are seated

If we comprehensively understand the skeptical syllogism, this means that we successfully understand the working principle of the Cartesian skepticism. This reason makes the Cartesian skeptical syllogism play a very important role in the rejection of Cartesian skepticism. In order to reject the skepticism, we can start by considering two approaches. The first approach attempts to prove that the skeptical syllogism is not valid, meaning that: the conclusion is not a consequence of its premises. The second approach seeks to show that one or other of the syllogism's premises is false, so that the argument is unsound.

The first approach is seems to be a nonstarter, because the skeptical syllogism is an instance of modus tollens, a deductively valid form of argument. For this reason, we have to choose the second approach.

3.2. Closure Principle

Since the second approach is the only path that we can choose, now let's start to look at whether the validities (P2) are questionable[7]. If we can question either part of (P2) successfully, including the logic between them, then we can question the validity of (P2); further, we can question the validity of the skeptical syllogism. The validity of (P2) is de facto based on a very important principle, which is called the epistemic closure principle (CP thereinafter). The ordinary applications of closure allow us to infer what we know to be deductive consequences from what we know (Godden 2017, 5). The basic idea of CP is that if S knows that p and knows that p entails q, then S knows that q. It can be formulated as follows:

$$[K(s,p) \text{ \& } K(s,p \rightarrow q)] \rightarrow K(s,q)$$

(P2) is based on the CP because "if we don't know we are not BIV" entails that "we don't know we are seated" because it's possible that we are BIV and "seating" is simply an illusion of our mind. As you may have guessed, if we can reject the CP successfully, for example by proving that "we don't know that we are not BIV",

---

7.   Many philosophers believe that Moore is rejecting the Cartesian skepticism by rejecting the (P1). This is because Moore denies that people do not have the ability to know that they are not BIV.

but "we still know that we are seated" is true, then (P2) can be challenged by the reasoning of contradiction. Since (P2) is questionable, the validity of skeptical syllogism, as well as Cartesian skepticism, can also be rejected meanwhile.

If the properties of CP are just as I interpreted above, it is not much difficulty for most people to reject CP. Here is one of the counterexamples that many people probably think of:

p1) I know that I am reading

p2) I know that I am reading entails that 1+1=2

c) I know that the 1+1=2

The reason why CP can be rejected so easily, in this case, is that there is another significant property of a valid CP that is missing. To be a effective CP, there must be a consequent relationship between "p" and "q." Namely, "p" and "q" cannot be any random proposition. (P2) of skeptical syllogism includes a valid CP because there is a subsequent relationship between the ordinary proposition and skeptical hypothesis. Thus, the above "CP" I give is not a valid CP because there is no consequent relationship between "reading" and "1+1=2."

American philosophers Robert Nozick and Fred Dretske, both reject the Cartesian skepticism by rejecting the CP of (P2). Besides, there is one important point that we have to pay attention to. Nozick and Dretske do not reject the (P2) in its initial format but reject the (P2) in its form of Modus Tollens (contrapositive reasoning). It can be formulated as follows:

P2) If S doesn't know ~H, then S doesn't know P

P2) If we don't know we are not BIV, then we don't know we are seated

The reason why Nozick and Dretske transform (P2) from its initial format into the Modus Tollens is in order to make a better connection between (P1) and (P2). If we try to reject the initial format of (P2) of the skeptical syllogism, it's very hard for us, logically, to understand how it works. This means that in the rest of the paper, when I mention the rejection of (P2), I mean the rejection of transformative (P2).

3.3. Robert Nozick and Tracking Theory

(i) If I am seated, then I know that I am not deceived by the Cartesian demon (ii) Conversely, if I am unable to know that I am not deceived by the Cartesian demon, then I am unable to know that I am seated (Pritchard 2008, 7). This is an analysis given by Duncan Prichard in his paper, "Sensitivity, Safety, and Anti-luck Epistemology." This analysis gives us a closer look at what role the word "know," plays in the (P2) of skeptical syllogism. If people have a different definition for "knowing" or "knowledge;" then they can also have a different version of CP. Since 1963 Edmund Gettier challenged the traditional definition of knowledge, JTB theory, there was not a unified definition of knowledge (Gettier 1963, 1). This gives many philosophers, such as Nozick, a hope to reject the CP by offering their own definition of knowledge. The definition of knowledge by Nozick is as follows:

(1) S has true belief on Q

(2) If Q weren't true, S wouldn't believe it (Sensitivity-based Requirement)

(3) If Q were true, S would believe it (Adherence Requirement)

Nozick's definition of knowledge is also known as tracking theory. In the tracking theory, Nozick makes use of the subjunctive conditional instead of the material conditional.

As I said in section 3.2., Nozick rejects the Epistemic Closure Principle by rejecting the (P2) of skeptical syllogism. The first step that Nozick plans to do is to prove that the first part, "S doesn't know ~H", of (P2) is true. "S knows ~H" is false which means that "S doesn't know ~H" is true. Hence, Nozick only needs to prove that "S knows ~H" is false. The reason that "S knows ~H" is false is that it does not satisfy the sensitivity-based requirement of knowledge. If "S knows ~H" is true, it has to satisfy the sensitivity-based requirement that if "~H" is not true, then "S would not believe ~H" is also true. Please imagine that there are multiple parallel worlds that are existing around the world that we are currently living in. A world that is farther away from the world we lived is a world with more differences with ours. A world that is closer to the world we lived in means that there are more similarities between that world and the world we lived in. In a close possible

world[8] that "~H" is not true, it is a world that "H" is true. "H" is a skeptical hypothesis, such as BIV or Cartesian demon. In other words, in a world that "H" is true means that there is a Cartesian demon in this world. However, in this world, S would still believe that "~H" is true even though "H" is true. Since the sensitive-based requirement does not be satisfied (S still believes that "~H" is true when "H" is true), "S doesn't know ~H" is true. Thus, the first part of (P2) is true.

The second step that Nozick plans to do is to prove the second part of (P2) that is incorrect, which means that "S doesn't know that ~H, BUT S does know P." Please consider the following example:

(i) There is a bunch of fire in front of me, and I believe that there is a bunch of fire in front of me

(ii) If there is not a bunch of fire in front of me, I wouldn't believe it

(iii) If there is a bunch of fire in front of me, I would believe it (the location of the fire is possibly different in a close possible world)

These three conditions, (i) to (iii), are sufficient conditions for the tracking theory, (1) to (3), by Nozick. (i) satisfies condition (1); this is not hard to understand. (ii) satisfies the sensitivity-based requirement because in a close possible world, if there is not a bunch of fire in front of me, I would not believe it (it would be absurd to say that there is a bunch of fire in front of me, if there is no fire). (iii) also satisfies the adherence requirement. In a close possible world, there is a bunch of fire in front of me, but the location of the fire is slightly different than the location in our world. I would believe that there is a bunch of fire in front of me in that world because I see it. Therefore, the second part of (P2) is false, which means that "S does know that P" is true. The acceptance of the first part and the rejection of the second part of (P2) express that CP is incorrect. Therefore, (P2) is incorrect. Since one of the premises of the skeptical syllogism is incorrect, the conclusion of skeptical syllogism should be also incorrect. Finally, we don't have to accept the argument of Cartesian skepticism.

---

8. In metaphysical and epistemological talk, a close possible world is intended to be a world that is particularly similar to the present one, with some details (especially the truth of "H") changed.

3.4. The Issues of Tracking Theory

Since the Cartesian skeptical argument has been challenged by the tracking theory, there are at least two issues that have emerged: (1) Whether Nozick's tracking theory can be precisely applied to every situation in our daily life? (2) Whether Nozick's tracking theory is a really good definition of knowledge without making any contradiction? For the answer to these two questions, my answer is "NO."

First, Nozick's tracking theory can conduct many contradictive phenomena in our daily life. In other words, this theory cannot be successfully applied to every aspect of our lives. Please compare the following two barn county style cases:

(I) Suppose there is a barn county in which many barn-like structures are scattered. Although they look almost the same, only one of them is a real barn, and the others are fake. Now suppose S drives through this county and just sees the real barn, so S believes there is a barn in the county. S's belief is true, that is, it meets the first requirement of Nozick's tracking theory. However, it violates the sensitivity-based requirement of tracking theory, because in a close possible world, if there is no a real barn, S would still believe that there is a real barn because of the existence of other fake barns. Therefore, in the case (I), S does not know that there is a barn in the field.

(II) Suppose the situation in the barn county is roughly the same as above. The only difference is that the real barn in the county is red and the other fake barns are other colors. Now suppose S drives through the county and S just sees the real red barn, so S believes that there is a red barn in the county. This time, S's belief is true (satisfies the first requirement) but it also meets the sensitivity-based requirement and the adherence requirement of Nozick's tracking theory: in a close possible world, if there is not a red barn in the county (the true barn does not exist), the existence of other fake barns would still make S believe that there are some other barns in the county, but S would not believe that there is a red barn in the county, because the other fake barns were not red. Of course, S also meets the third requirement of the definition: if the location of the red barn in the county changes slightly in a close possible world, S will still believe that there is a red barn in the county.

What is really confusing in these cases is that when you combine these two cases together, they are contradicting. S does not know that there is a barn in the

county, because of the reason shown in case (I); however, meanwhile, S does know that there is a red barn in the county because of the reason shown in case (II). How can a person know that there is a red barn without knowing that there is a barn? This is one of the counterexamples that the tracking theory leads us into an anti-perceptional result. (Kripke 2011, 162-224)

The last, in my perspective, the deadliest flaw of the tracking theory is that we don't know the precise definition of the "close possible world". As I explained in 3.3., the concept of "close possible world" plays a very important role in Nozick's tracking theory. Without understanding this concept, it's impossible for an epistemologist to judge whether someone knows something or does not know something. However, the concept of "close possible world" is very absurd. Nozick does not give us a clear-cut definition of the close possible world. We don't know the boundary between each world and we also don't know how close a parallel world should be to be a close possible world. For instance, in my opinion, between a world with the Cartesian demon and in a world without the Cartesian demon, these two worlds cannot be close possible worlds to each other. Rather, these two worlds are far away from each other because in the world with the Cartesian demon, the external world does not exist, but in the world without the Cartesian demon, the external world does exist. In my opinion, these two worlds should not be close enough to be possible worlds because even though the information that people get in both worlds is the same, but the essence of everything is completely different. In short, if different people may have a different understanding of "close possible world", they will get different outcomes about whether someone knows something. For a definition of knowledge, this instability is fatal.

3.5. Fred Dretske and His Theories of Knowledge

3.5.1. EPISTEMIC OPERATOR

Dretske questions CP by questioning the relationship between the two parts of (P2). In the first step, Dretske challenges the transmissibility of "knowing". In Dretske's point of view, CP can only be applied to a fully penetrating operator. An operator, O, is fully penetrating just in case if P entails Q, then O(P) entails O(Q). These operators are including "it is true that," "it is a fact that," "it's necessary that" and "it is possible that" … (Dretske 1970, 1007). A fully penetrating operator can penetrate to every necessary consequence of P to Q. Semipenetrating operators,

on the other hand, cannot penetrate all its implications to its "receiver" ("Q"). Here is an example of semipenetrating operator:

(R)   (p1)   S regrets P

(p2)   P entails Q

(c)   Therefore, S regrets Q

or with an example, as follows:

(R)   (p1)   S regrets drinking a bottle of tequila

(p2)   Drinking a bottle of tequila entails drinking something

(c)   Therefore, S regrets drinking something

(Godden 2017, 8) "Regretting" is a semipenetrating operator because "regretting" cannot penetrate all receivor's implications to its "regret" receiver. In this case, S regrets that he was drinking a bottle of tequila; however, this is not necessarily expressing that he regrets that he was drinking something. In Dretske's words, these implications that cannot be fully penetrated are called heavyweight implications. Dretske followed by uttering that the epistemic operators, such as "knowing", "believing," are also semipenetrating operators. Here is an example:

(K)   (P①) S knows British PM has COVID-19

(P②) British PM has COVID-19 entails that Boris Johnson has COVID-19

(C) Therefore, S knows Boris Johnson has COVID-19

Does S really know Boris Johnson has COVID-19 in this case? The answer is absolutely "NO"! S only know (C) if S know another premise, (P③), that Boris Johnson is British PM, because that would give S a reason for believing (C). It's only by S's knowing (C) that there is any plausibility to the claim that S's knowing (P③) logically follows from S's knowing that (P①). In this case, the definition for a penetrating operatory does not give me that S knows that (P①) entails (P③). Rather, it only gives me that (P①) does, in fact, entail (P③). This case gives us

a good reason for thinking that "know" is not fully penetrating. Thus, Dretske advocates that there is no such implications about "P", and all implications can be penetrated into a sentence entailed by "P," because some of implications are heavyweight implications. There is a class of heavyweight implication that is highly valued, which is the Modus Tollens of skeptical hypothesis, "~H." Dretske admits that no matter whether by empirical or non-empirical approaches, we cannot know if the Cartesian demon exists. Nevertheless, the ignorance of "~H" does not make us ignorant of "P", which implies "~H."

3.5.2. RELEVANT ALTERNATIVE THEORY

Just like Nozick, Dretske also offers a definition of knowledge. According to Dretske's relevant alternative theory, S knows that P only when:

(P1) P is true

(P2) S believes P

(P3) S can rule out all relevant alternatives to P

According to Cartesian skepticism, the knower has to rule out all alternatives of P in order to have knowledge of P.[9] However, under Dretske's theory, S knows that P does not require that S has ruled out all alternatives of P; it only requires that S has excluded all relevant alternatives of P. Now, there are two questions that need be clarified: (a) What makes an alternative to be relevant alternatives? And (b) How is "ruling out" to be understood? What does it take to "rule out" an alternative?

For question (a), Dretske uses the Gadwall case to explain the relevant alternative. Suppose a birdwatcher sees a bird that looks like a Gadwall on the water. Based on this visual evidence, he believes that what he sees is a Gadwall. Suppose further that this bird is indeed a Gadwall. What this birdwatcher does not know is that the Siberian grebe is no different from Gadwall, except that their belly hair is different in color. The former has red belly hair and the latter does not. Only when the Siberian grebe is flying, are people able to see its red belly hair and distinguish it from Gadwall accordingly. The question now is that does this birdwatcher really know that he is seeing a Gadwall? (Dretske 1981, 368-

---

9. Please remember what I have mentioned in 2.1., for a skeptic, as long as a belief is questionable, such a belief must be treated as wrong. This means that the knower has to rule out all alternatives of propositions.

369) According to Dretske's interpretation of relevant alternatives theory, whether Siberian grebe is a related alternative for Gadwall, it can depend on two factors: (i) Whether it is possible that a Siberian grebe appears in that area; (ii) What is the context, the physical or the intellectual one, that the birdwatcher is standing. For factor (i), if Siberian grebes are appearing in that area (by migrating, smuggling by hunters or by something else), then Siberian grebe is a relevant alternative for Gadwall; otherwise, it is not. For factor (ii), whether Siberian grebes are relevant alternative for Gadwall, it depends on the birdwatcher's context. If birdwatcher is researching in a zoological topic, he might know that is a Gadwall; however, if he is researching in a philosophical topic, he might not know that there is a Gadwall because how does he know that he is not just seeing an illusion made by the Cartesian evil demon?

For question (b), in Dretske's view, in order to exclude an alternative, it requires a conclusive reason (CR thereinafter).[10] There is a similar case given by Dretske in another paper, Epistemic Operators. For a normal zoo, the relevant alternative of the zebras includes mule, elephant, tiger, giraffe… but it does not include mule painted to look like a zebra. If S knows that he is seeing a zebra, he must have CR to exclude that he is not seeing a mule, elephant, tiger, giraffe and all other relevant animals; however, he does not have to have CR to exclude that he is seeing a cleverly-disguised mule. For this reason, it is possible for S to know that he is seeing a zebra without knowing that he is not seeing a cleverly-disguised mule.

In Dretske's point of view, the relevant alternative of "S has two hands" includes, such as, "S is disabled since birth" or "S has experienced a car accident and lost two hands," but "S is a handless BIV" is not included in the relevant alternatives. If S knows that "S has two hands", then S must have CR to exclude its relevant alternative, but S doesn't necessarily have CR to exclude irrelevant alternatives, such as "S is BIV." For this reason, the second part of (P2) cannot be necessarily entailed from the first part of (P2), which means CP is invalid. (P2) of the skeptical syllogism is invalid that leads to the skeptical syllogism which is also invalid. Hence, the argument of Cartesian skepticism is false.

---

10. This is the definition of conclusive reason given by Dresteke:

    (1) S knows that P and he knows this on the basis (simply) of R entails

    (2) R would not be the case unless P was the case

3.6. The Issues of Dretske's Theories of Knowledge

There are two questions that need to be clarified by the relevant alternative theory. They are (a) What makes an alternative to be relevant alternatives? And (b) How is "ruling out" to be understood? What does it take to "rule out" an alternative? In the last section, Dretske detailly answers these two questions. However, here are two issues that also based on these two questions.

First, under what conditions (or contexts) an alternative is a relevant alternative. Please consider the following question: Assuming Judy and Trudy are twins, and Judy lives in the U.S. and Trudy lives in Europe. They are the same regardless of their looks, interests, style of clothing, etc.[11]. Suppose Judy has a very good neighbor Sam and they are familiar with each other, but Sam does not know the existence of Trudy. Under what circumstances we would say that Sam knows that person in front of him is Judy? (Goldman 1976, 778) Now please consider the following three situations:

(I) When we first time look at this question, our intuition tells us that Sam does not know Judy because when Judy and Trudy are standing together, Sam cannot distinguish which person is Judy and which person is Trudy. In this context, Trudy is a relevant alternative of Judy. Therefore, Sam does not know Judy.

(II) Let's further assume that because of Judy and Trudy's family affairs, they cannot leave their place of residence. Under this context, we can intuitively think that Sam should know Judy, because Trudy will never come to the U.S., and Sam can properly identify the person who lives next to his house. Thus, Trudy is no longer a relevant alternative of Judy as Sam knows Judy.

(III) However, if Sam one day travels to Europe, and he runs into Trudy. He probably will say that "Hi Judy, I thought you were in the U.S.!" Obviously, if Sam is capable of traveling to Europe, then Trudy is a relevant alternative of Judy again. Sam does not really know Judy in this case.

A proposition is irrelevant to a certain belief in one case, but it can be relevant to the same belief in another case. Unfortunately, the criteria for determining whether an alternative is a relevant alternative or not is inconclusive. If the factors

---

11. Using the words that Dretske gives to us in the speech, "What we see", in the University of California Berkeley, the properties of Judy and Trudy are the same. https://alchetron.com/Fred-Dretske 13m:05s

that determine relevance are unstable, it also makes people's knowledge relatively unstable.

Second, CR does not seem to be an appropriate condition for the exclusion of alternatives. Consider the following situations: assuming that there are two nasopharyngeal swabs, A and B, for testing COVID-19. If the swab A shows negative, it means that the tested patient is not infected by COVID-19. However, if the swab B shows negative, it means that the patient is not be infected by COVID-19, or it is possible that the tested patient is infected but the concentration of coronavirus is not high enough to make swab B to show positive (this is also called false negative or weakly positive). In this case, S only knows the existence of swab A, but does not know the existence of swab B and the appearances of swab A and swab B are identical. Now let's further assume that a tested patient is not infected by COVID-19, and S wants to detect whether he is infected. In the test center, the test table is filled with nasopharyngeal swabs. S believes they are all swab A. However, in fact, most of them are swab B, and only one is swab A. Luckily, S just takes swab A to test patient and observes that the swab shows negative. Since S believes that he is using the swab A (which is also true), he then believes that the patient is not infected, because the swab shows negative. According to Dretske's definition of CR, S does have CR to rule out that the patient is infected because:

(1) S knows that the patient is infected and S knows this on the basis of swab A shows positive

(2) Swab A will not show positive unless the patient is infected

Since above situation satisfies Dretske's definition of CR, according to Dretske, S knows that the patient is not infected. However, this conclusion is not indisputable, because S's testing includes quite a lot of luck. S does not know that there is another swab, swab B, which can show negative, but the patient is, in fact, infected. This case seems to indicate that the CR for showing negative is not enough to exclude that the patient is not infected by COVID-19. Therefore, CR cannot help S to rule out all relevant alternative of P and the curse of Cartesian skepticism cannot be broken by Dretske's relevant alternative theory.

4. ABSOLUTE INFALLIBILISM LEADS TO CARTESIAN SKEPTICISM

Either from proving the existence of the external world or rejecting the Epistemic Closure Principle; in my opinion, neither of these approaches is feasible as challenges the argument of Cartesian skepticism. It is impossible to prove the existence of the external world through our sensory system because the approach of senses of the external world is exactly the object the skeptics have questioned. It's also problematic and unnecessary to object the Cartesian skepticism by rejecting the CP, because CP is one of the important ways for us to obtain knowledge. The collapse of CP may cause chaos in our daily life. At this moment, I begin to wonder if there is something wrong with our strategy of rejecting skepticism.

Descartes claimed that the only things that we should count as knowledge were things, we could be certain about. This advocation sometimes is also called infallibilism. Infallibilist's view of knowledge holds that knowledge has objectivity, certainty, justification, infallibility, incorrigibility, and indubitability. In Robert Audi's words: "If you know, you can't go wrong." (Audi 2004, 300) That is, knowledge must be a certain belief.

In my opinion, however, the threshold for a belief to qualify as knowledge is too high. In an influential article, "A Defense of Skepticism," published in 1971, Peter Unger raises a similar idea. Unger believes that in terms related to cognition, "certain" is an absolute term, which is a concept without a degree of difference. "Confident," "doubtful," and "uncertain," on the other hand, are relative terms, which with a degree of difference. Here is an example. Unger argues that "flat" is an absolute term. We cannot use the comparative to describe flat like saying "A is flatter than B." In everyday language, when we say, "this plane is flatter than that plane," we are actually expressing that "this plane is closer to flatness than that plane." In Unger's perspective, "flatness" is an absolute concept, it can only be approached, but never be reached. We thought we could find an absolutely flat plane, but every plane is bumpy when it is viewed under a microscope. Although we may not require such strictness for practical purposes, in the study of epistemology we must ensure that the statements made are not false.

Since traditional knowledge definition requires knowledge is something absolutely certain, I believe that knowledge is something not accessible under such a definition. Therefore, Cartesian skepticism is unbreakable without abandoning

infallibilism. For the queries about how should abandon infallibilism and what's the substituted definition for knowledge, they are the topic for another paper.

REFERENCES

Descartes, René. "Concerning Those Things That Can Be Called into Doubt." In *Meditation on First Philosophy*,13.

Dretske, Fred I. 1970. "Epistemic Operators." *The Journal of Philosophy* 67 (24): 1007–1023. *JSTOR*, www.jstor.org/stable/2024710.

Dretske, Fred I. 1981. "The Pragmatic Dimension of Knowledge." *Philosophical Studies* 40: 363-378.

Dretske, Fred I. 1981. *International Journal for Philosophy in the Analytic Tradition* 40 (3): 368-369. *JSTOR*, www.jstor.org/stable/4319495.

Dretske, Fred I. "What We See." The Graduate Council Lecture. The Howison Lectures in Philosophy. Moses Hall, #2390, Berkeley, CA 94720 13 minutes : 5 seconds.

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121–123. *JSTOR*, www.jstor.org/stable/3326922.

Godden, David. 2017. "Scepticism II- Cartesian Scepticism: Method of Doubt · Deductive Closure." Class Handout. Michigan State University. Department of Philosophy. pp 5, 8.

Goldman, Alvin I. 1976. "Discrimination and Perceptual Knowledge." *The Journal of Philosophy* 73 (20): 778. *JSTOR*, www.jstor.org/stable/2025679.

Kant, Immanuel. 1929. *Critique of Pure Reason*. Translated by N.K. Smith, bxxxix, note, 34.

Kripke, Saul. 2011. "Nozick on Knowledge." In *Philosophical Troubles*, 162-224. New York: Oxford University Press.

Moore, G. E. 1993. *G.E. Moore: Selected Writings*. Edited by T. Baldwin, 165-166. London: Routledge.

Pritchard, Duncan. 2008. "Sensitivity, Safety, and Anti-Luck Epistemology." Forthcoming in *The Oxford Handbook of Skepticism*, edited by J. Greco, 7. Oxford University Press.

Putnam, Hilary. 1999. *Reason, Truth, and History*. Cambridge: Cambridge University Press, Chapter 1: 1-21. Reprinted as "*Brains in a Vat*" in DeRose and Warfield (eds.), Chapter 2: 27-42.

Robert, Audi. 2004 *Epistemology: A Contemporary Introduction to the Theory of Knowledge* (second edition). New York: Routledge.

Unger, Peter. 1971. "A Defense of Skepticism." *The Philosophical Review* 80 (2): 198–219. *JSTOR*, www.jstor.org/stable/2184030.

cognethic.org