

# PROMETHEUS

2021 Edition



The Johns Hopkins University  
Undergraduate Journal of Philosophy

Mackenzi Barrett	On Self-Determination and Coercion: A State's Right to Control its Border	1
Yunlong Cao	A Defense of the Second Analogy	9
Ibrahim Dagher	Expanding William Hasker's Transcendental Refutation of Determinism	15
Nathan Jacob	Phenomenology of Consciousness: Understanding its Functional Role	21
Campbell Rider	Dualism and Organic Form	30
Musa Saleem	A Distinctive Type of Anthropocentrism in the Philosophy of Personhood	37



## *Note From the Prometheus President*

Without a doubt, this past year's crisis has given us a great degree of grief, pain, and suffering. It has been a difficult time for virtually every single one of us, and the crisis is still ongoing to this very day. It is a time like this when philosophers are needed the most, contrary to what people may think. Of course, the work that the epidemiologists, doctors, nurses, and everyone in the healthcare industry do is crucial during a pandemic like this, and by saying that philosophers' work is important during times like these, I am not trying to discount the value of the work of those in the healthcare industry. What I mean to say, instead, is that the fundamental value of philosophy, which is to question everything and examine the nature of things that we take for granted, can help us guide through difficult times such as this pandemic, and the philosophers' works should not be neglected. Especially at a time like this, some people may wash away our works as philosophers and treat them as just some "playing with words" or whatnot. However, by doing so, they are neglecting the true value of philosophy and forgetting that there are many things that we can gain from doing philosophy, both as a philosopher and a non-philosopher. Accepting and understanding the crisis is as important as dealing with, preventing, and stopping the crisis.

I am very proud of the works that were submitted to Prometheus the past year. The quality of works that were submitted was excellent and it gave us, the Prometheus editors, a great degree of hope that the undergraduates from all over the world are still continuing to do philosophy even amidst the chaos that this crisis has caused. It is with great honor and privilege to introduce you to the best works that we have selected to be in this journal. We as editors thought that these articles embody the mission of philosophy as aforementioned the best and articulate the issues in a clear and precise manner. Perhaps they are not necessarily all so related to the issue of the pandemic, but the real mission of philosophy, examining the fundamental nature and questioning everything, does not necessarily have to be always related to the current events, and we can always learn from many different branches of philosophy. It is our hope as editors that this year's journal will show the degree of hope and excellence that we have seen so far.

Phillip Yoon  
President of Prometheus

## **Prometheus Undergraduate Philosophy Club and Journal 2021**

### **Advisor:**

Simon Brown

### **President:**

Phillip Yoon

### **Board Members:**

Shour Arashanapalli

Lawrence Cao

Damon Choi

Devin Harsch

Joi Haskins

Smitha Mahesh

Christopher Slama

### **Editors:**

Colten Fisher

Jayden Montalvo

Christian Reidy

Maxwell Sahhar

Michelle Schipilliti

Spencer Shattuck

Stefano Tusa

The Johns Hopkins University  
Baltimore, Maryland

Copyright © 2021 Prometheus  
All rights reserved.



## ***On Self-Determination and Coercion: A State's Right to Control Its Borders***

MACKENZI BARRETT

THE UNIVERSITY OF ALABAMA

**Abstract:** Andrew Altman and Christopher Heath Wellman argue that legitimate states have a right to freedom of association as a component of their right to self-determination. According to Altman and Wellman, freedom of association includes the right *not* to associate, which has implications for a state's right to control its borders. Because the members of a state have a right to not associate with whomever they choose, the state has a *prima facie* right to exclude all outsiders. However, Altman and Wellman's argument relies on a conception of self-determination as self-composition, but this is an improper conception of self-determination. The proper conception, non-intervention and self-rule, does not lead to a state's right to exclude all outsiders. Finally, states coerce potential immigrants through border control, and to be legitimate the state must justify its coercion, which contradicts the right to self-determination, or eliminate the coercion. Ultimately, contra Altman and Wellman, the state's right to exclude does not apply to all outsiders but is limited to those whose entry would undermine the state's ability to fulfill the basic rights of those under its control.

### **Introduction**

In "Immigration and Freedom of Association," Christopher Heath Wellman argues that legitimate states have a *prima facie* right to exclude all potential immigrants from their territories. Wellman, in collaboration with Andrew Altman, reaffirms this conclusion in *A Liberal Theory of International Justice*. The argument is based on a right to freedom of association, which they argue legitimate states have as a result of a right to self-determination. The argument proceeds as follows:

- (1) Legitimate states have a right to self-determination.
- (2) Freedom of association is a component of the right to self-determination.
- (3) Therefore, legitimate states have a right to freedom of association.
- (4) Freedom of association includes the right not to associate.

---

<sup>1</sup> Notice that Altman and Wellman distinguish between the constituents of a state and non-constituents of a state, such that the state has to protect the rights of the former and merely respects the

- (5) Therefore, legitimate states have the right to exclude all outsiders from their territories.

To be clear, Altman and Wellman do not assert that the right to freedom of association is absolute. There are circumstances in which other rights can override the right to freedom of association. Thus, the state's right to freedom of association is a *prima facie* right, and consequently, so is the right to exclude outsiders from its territory.

In response to the issue of when a state has a right to control its borders, Altman and Wellman assert that a state has this right whenever another right does not override the state's right to freedom of association. I will argue, however, that their argument is invalid. Specifically, the state's right to self-determination and associated right to freedom of association does not entail the state's *prima facie* right to exclude *all* outsiders from its territory. Rather, the state's *prima facie* right to exclude outsiders from its territory is far more limited than Altman and Wellman conclude. To accomplish this task, I will proceed in the following way. Part I will describe what a state is and what the purpose of the state is. The purpose of the state has implications for the rights associated with the state. Part II will discuss the state's right to self-determination and consider whether Altman and Wellman's conception of self-determination is correct. Part III will introduce coercion to the discussion on border control. Particularly, I will demonstrate that contra Altman and Wellman, states coerce all potential immigrants. Furthermore, the state's coercion of potential immigrants has implications for their legitimacy and right to exclude outsiders from their territories. Ultimately, I will conclude that Altman and Wellman's argument in support of a state's *prima facie* right to exclude all outsiders from its territory fails and that a state's *prima facie* right to exclude is far more limited than they propose.

### **I. On the State**

The notion of legitimacy is a useful starting point for determining what exactly the state's purpose is. Altman and Wellman assert that "political states are legitimate if they adequately protect the human rights of their constituents and respect the rights of all others" (1).<sup>1</sup> Thus, legitimacy is predicated on human, or basic,

rights of the latter. In Part III, I will discuss a problem with their distinction.

rights.<sup>2</sup> Importantly, our notions of legitimacy are often intimately connected to notions of purpose: some action is legitimate if that action is in accordance with the purpose of that action. Altman and Wellman never explicitly discuss the purpose of a state or the connection of purpose and legitimacy. However, they do state that if a state is legitimate, then the state “successfully carries out the ‘requisite political functions’” (Altman and Wellman 3). In other words, if a state is legitimate, the state is fulfilling its purpose. For example, suppose that I am a teacher. In my role as a teacher, my purpose is to teach a particular subject to my students. If I choose to sit at my desk and allow my students to act as they please, I am not fulfilling my purpose as a teacher. My action is not legitimate.<sup>3</sup>

Similarly, when starting with a particular description of legitimacy, we can work backwards to determine purpose. A state is not legitimate if it does not protect basic rights because the state is failing to fulfill its purpose. What, then, is the state’s purpose? The state’s purpose is to protect the basic rights of those under its control.<sup>4</sup> If a state is not protecting the basic rights of those under its control, then the state is not fulfilling its purpose and therefore is not legitimate. Importantly then, fulfilling its purpose is both a necessary and sufficient condition for a state’s legitimacy. Furthermore, I am choosing to focus on a state’s purpose rather than whether it is legitimate for the following reason.

Suppose that I am a parent. As a parent, I have a duty to my child, e.g., to fulfill her basic needs. I have

a right to go beyond that duty (as long as I am not contradicting my duty or harming anyone else in doing so). However, the rights that I have as a parent are restricted to my child; they are the result of the relationship between me and my child. I do not have any rights against other parents or other children beyond those that are necessary to fulfill my duty because they are outside of the relationship. For example, I have a right to defend my child against other people because I am protecting her; this is a duty that arises from our relationship. Similarly, there is a specific relationship between a state and the individuals under its control that is relevantly different from any relationship, if one exists, that a state has with other states or individuals in other states. Furthermore, the state exists only because of its relationship to those under its control (just as a parent exists only because she has a child). A state must, at minimum, fulfill its duty to those under its control, and it has a right to defend those individuals from others that attempt to interfere with that duty.

Thus, the rights that a state has are those that are relevant to its relationship with those under its control. It has rights to fulfill that duty and in certain cases to go beyond that duty, but the rights it has against those outside of the relationship are restricted to rights that arise from the duties of the relationship.<sup>5</sup> In other words, when evaluating whether or not a state has a particular right against outsiders, we are concerned with the proposed right as it relates to the state’s ability to fulfill its purpose—its duty.

<sup>2</sup> There is, of course, a question of what these basic rights are. Following Joseph Carens, basic rights “correspond to conditions in whose absence human beings cannot live decent lives, no matter what particular values and plans of life they choose to pursue” (Carens 195). Carens’s definition alone does not provide a detailed list of basic rights, and according to this definition, people could reasonably argue over the inclusion and exclusion of various rights. Perhaps at minimum we would all agree that rights to life, basic liberties, and security should be included on the list. Moreover, determining which basic rights should be on the list is not critical to my argument.

<sup>3</sup> This connection between legitimacy and purpose has a strange implication. Suppose I am a thief. My purpose is to steal things, and any action that I take in accordance with that action is legitimate. Some might protest that my thieving actions are *not* legitimate, however, because my purpose itself is not justified. I agree that my purpose may be unjustified, but my thieving actions can still be legitimate *with respect to my purpose as a thief*. Claiming that my thieving actions are illegitimate requires looking beyond my purpose as a thief. For example, as a human being, my purpose may involve not harming others. Because my purpose as a thief contradicts my greater purpose as a human being, thievery is

not justified, and my thieving actions are illegitimate *with respect to my purpose as a human being*.

<sup>4</sup> “Under its control” does not contradict Altman and Wellman’s description of a state’s constituents, as constituents are under the state’s control. However, it is more general, and my reason for making the statement more general will become obvious in Part III.

<sup>5</sup> Some might argue that states have more rights concerning outsiders beyond those rights that arise from the state’s relationship with those under its control. For example, in much the same way that parents may have a right to help needy children who are not their own, states may have a right to help needy foreigners. While I agree that parents have a right to help needy children who are not their own, this right is not because they are *parents* but because parents are necessarily *persons*. Any rights that a parent has as a parent are restricted to rights arising from the relationship between parent and children. Similarly, any rights that a state has as a state are restricted to rights arising from the relationship between state and constituent. While it is obvious that parents are also persons and thus have rights as a result of their personhood, it is less clear what rights a state might have not as a result of its statehood but from some other source. Without any indication of what this source might be, I will continue to evaluate what rights a state has by considering its status as a state.

Therefore, because a state can go beyond its purpose and still be legitimate, what is relevant for determining a state's rights against outsiders is its purpose and not its legitimacy. Furthermore, the purpose of the state has important implications for the state's right to control its borders.

## II. The State's Right to Self-determination

Like Altman and Wellman, I will assume that states have a right to self-determination.<sup>6</sup> Even assuming that states have a right to self-determination, a question still remains regarding what exactly a state's right to self-determination entails. Traditionally, a state's right to self-determination is understood as a negative duty on other states.<sup>7</sup> More specifically, other states are bound by a duty of non-intervention as long as the state is legitimate. This duty of non-intervention is paired with a right to self-rule. This pairing is relatively intuitive: if a state has a right to rule *itself*, other states are prohibited from ruling that state. Thus, this conception of self-determination has two aspects: non-intervention and self-rule. Furthermore, this conception of self-determination protects the legitimate state from actions like forced mergers and annexations. Yet, this conception is not what is meant by a state's right to self-determination in Altman and Wellman's argument. This conception concerns interactions between entities with the intention to exercise the monopoly of physical force within a territory. For the most part, these sorts of entities are states, such that the conception concerns interactions between states and not between a state and *individuals* from other states, i.e., potential immigrants. If one entity is legitimate and is already exercising that monopoly, that entity has a right to continue exercising that monopoly (self-rule) and a right against other entities to not attempt to acquire that monopoly (non-intervention). Thus, if a right to freedom of association were the result of a right to self-determination understood as non-intervention and self-rule, the resulting right to not associate would merely apply to associations between these certain entities, which are usually states, and Altman and Wellman's conclusion that states have a right to exclude all outsiders would not follow.

Instead, Altman and Wellman understand self-determination as a right to self-composition: "having

control over what the 'self' is" (Altman and Wellman 163). This conception of self-determination concerns a group's, specifically the state's, *membership*, such that the group has a right to determine *who* its members are. In this way, the corresponding right to freedom of association is attached to the state's *members* rather than the state as a whole. The current members of the state get to determine whether to allow new members into the state, i.e., whether to associate with new members. Obviously, current members can refuse to associate with new members as family, friends, etc., i.e., personal or intimate associations. However, Altman and Wellman are concerned with a different type of association, a political association. Thus, Altman and Wellman initially assert that "a state's freedom of association entitles it to exclude all foreigners from its political community," i.e., refuse a political association (164). This assertion is not equivalent to their conclusion, wherein a state is entitled to exclude all foreigners from its *territory*. Denying entry into the political community concerns preventing a forced political association. It does not, however, concern entry into the state's territory. For even if a foreigner is allowed into the territory, that does not mean that any member of the state has to associate, even politically, with her because the state can simply deny her entry to the political community. This would mean that she is unable to participate in the politics of the country, e.g., voting, which is sufficient to prevent a forced political association. Denying her from the territory is not necessary to prevent a forced political association.

Altman and Wellman never explicitly state the reasoning for jumping from a state's right to exclude foreigners from its political community to its right to exclude foreigners from its territory. Our only insight into this jump occurs after they explain their argument: states "may not treat the [outsiders] they do admit as a class of permanent political subordinates," i.e., keep them from being citizens forever (Altman and Wellman 171). In other words, for their conclusion to follow from this conception of self-determination, Altman and Wellman are implicitly assuming that if a foreigner resides within a territory for a significant period of time, that foreigner eventually is entitled to join the political community. For members' freedom of association to function as defined, the state must be

<sup>6</sup> Altman and Wellman do not provide a conclusive argument for a state's right to self-determination. Instead, they appeal to our intuitions. For example, without a right to self-determination, we are

unable to account for our condemnatory responses to actions like forced mergers and annexations. See Altman and Wellman 161.

<sup>7</sup> See, e.g., Lægaard 658 and Walzer 88.

able to exclude foreigners from its territory. Otherwise, its current members would be forced into political associations with foreigners once those foreigners have been present in the territory for some specified period of time because they would no longer be able to deny those foreigners entry into the political community. Moreover, once they become citizens, assuming that citizens have some control over policy decisions, they have the potential to impact the future of the state. As David Miller puts it, new members gaining control could result in original members losing control, such that original members “have reason to prevent the newcomers from taking up residence in the first place” (265). However, having a reason for something does not necessarily give rise to a right to that something. I may have a reason to steal a laptop, but that does not mean that I have a *right* to steal that laptop. Does the original members’ desire to exercise complete control over the state’s future result in a right to control its composition?

Although self-determination as self-composition allows Altman and Wellman to draw their conclusion, this conception is not necessarily the right conception, and if it is the wrong conception, then Altman and Wellman’s conclusion does not follow. Let us consider some reasons against the self-composition conception. First, I would like to point to some unsavory implications of this conception. For Altman and Wellman, freedom of association includes a right not only to associate but also “the right *not* to associate, and even, in many cases, the right to dissociate” (Altman and Wellman 159). Consequently, it seems as though freedom of association, as applied to current members of a state, could allow for forced emigration if there were current members that wanted to end association with another current member. Additionally, allowing the entry of immigrants is not the only, or most prevalent, way that membership of a state changes; membership also changes through reproduction of current members. A right to freedom of association so construed could imply the right of a state’s members to implement strict population control policies on its own members.<sup>8</sup> This is because current members have a right to refuse political associations with individuals that are not currently members. Individuals that are not yet born are not currently members, so under Altman and Wellman’s view, the state could implement various policies such as

controlling which members of the state are able to reproduce. For example, the state could institute a vast regime of eugenics.

Altman and Wellman only specifically address the former concern, but their response also applies to the latter: “states have special responsibilities not to make unreasonable demands upon their citizens...As a result, our view does not imply that legitimate states may kick insiders out” (184). In other words, forced emigration (except perhaps under extreme circumstances) and population control are unreasonable demands. I think that most would agree with Altman and Wellman that these are unreasonable demands. Regardless, even if Altman and Wellman’s defense saves their view from these implications, their defense of their view against these concerns still seems unsatisfactory. Although I never explicitly established any of the basic rights that a state must protect, it seems reasonable to suggest that rights to live in one’s own state and reproduction are among those basic rights. That the conception of self-determination as self-composition even *implies* that a state can infringe on the basic rights of those under its control is deeply problematic for Altman and Wellman’s view. This is because they claim that individuals’ “rights...provide the normative considerations that ground a legitimate state’s own right to self-determination” (Altman and Wellman 5). While under the conception of self-determination as non-intervention and self-rule, the state only has a right to self-determination when the state is protecting those rights, this is not the case for the self-composition conception. The former conception has none of these implications because it is directly connected to the state’s legitimacy, but the latter conception does not have this same connection. As Sune Lægaard points out: “the fact that a right to control composition violates human rights unless constrained shows that not only is such a right *not derivable* from a concern with human rights; it *directly contradicts* human rights” (662). Lægaard’s point is that if legitimacy—protecting human rights—is what grounds a state’s right to self-determination, it seems counterintuitive to suggest that the proper conception of self-determination is one that contradicts human rights unless constrained. However, the other conception of self-determination, non-intervention and self-rule, does not require constraints. It is fully compatible with Altman and Wellman’s assertion that

---

<sup>8</sup> Sune Lægaard also addresses this concern in “Territorial Rights, Political Association, and Immigration,” 661.

individuals' rights *ground* the state's right to self-determination.

Additionally, recall that in Part I, I said that what rights a state has is connected to what duties a state has as a result of its relationship with those under its control. Thus, which conception of self-determination that a state has as a right is dependent on which conception is necessary for the state to protect the basic rights of those under its control. Non-intervention and self-rule are obviously necessary for that *particular* state to fulfill its purpose. The right to self-determination as self-composition is not necessary. A particular state can fulfill its purpose no matter *who* is part of the state. Furthermore, remember that Altman and Wellman motivated their claim for a state's right to self-determination with an intuitive appeal: without a right to self-determination, we cannot explain what is wrong with forced mergers and annexations. Yet, the conception of self-determination that is needed to prevent forced mergers and annexations is not self-composition but rather non-intervention and self-rule. The reasons that they provide us for accepting that a state has a right to self-determination does not provide a reason for accepting the self-composition conception of self-determination.

All of the above considerations demonstrate that the proper conception of self-determination is non-intervention and self-rule, not self-composition. As I mentioned earlier, however, Altman and Wellman's argument requires a conception of self-determination as self-composition in order to draw their conclusion. Self-determination as non-intervention and self-rule does not suffice as this conception applies to interactions between states only. Consequently, a state's right to self-determination, properly defined as non-intervention and self-rule, does not result in a right to freedom of association that applies to the members of a state rather than the state as a whole. Ultimately then, a state's right to self-determination does not entail that legitimate states have even a *prima facie* right to exclude *all* outsiders from their territories.

### III. Coercion

In case the arguments above have not been sufficiently convincing of the state's limited right to exclude, I turn now to the issue of coercion. Recall that Altman and Wellman base the legitimacy of a state on its protection

of the human rights of its constituents and respect for all others'. The state's coercive power motivates this condition: "If a state adequately protects and respects human rights...the state is doing the job that it needs to do in order to justify its coercive power and thereby be legitimate" (Altman and Wellman 3). I take no issue with the connection that Altman and Wellman have made between a state's coercive power and its legitimacy. As Altman and Wellman also point out, individuals in a state do not actually consent to the state's coercion, i.e., the coercion is nonconsensual: "universal political consent is a fiction" (Altman and Wellman 183).<sup>9</sup> In the absence of consensual coercion, it is right to assume that the state's coercive power requires justification. My issue with their justification is their distinction between a state's constituents and all others.

To start, it is unclear what exactly Altman and Wellman mean by "constituents." I am going to assume that by constituents, Altman and Wellman were referring to those that are subject to a state's coercive power. It was with this in mind that I defined a state's purpose as protecting the rights of those *under its control*. Thus, Altman and Wellman are asserting that the legitimate state *protects* the rights of those within its territory and merely *respects* the rights of those outside of its territory. Presumably the difference between protecting and respecting is the following: protection is active while respect is not. A state protects rights both when it does not infringe on those rights *and* prohibits others from doing so, but a state merely respects rights when it does only the former. Under this distinction, it is obvious why a legitimate state is under no obligation to intervene in an illegitimate state's affairs.

On its face, this distinction seems uncontroversial. However, the distinction becomes problematic when applied to immigration. According to Altman and Wellman, "states do not nonconsensually coerce foreigners" (184). Further, because states do not nonconsensually coerce foreigners, states are allowed to refuse admission to foreigners. When they say that states do not nonconsensually coerce foreigners, it is unclear whether they mean that states consensually coerce foreigners or that states do not coerce foreigners, either consensually or not. I am going to

<sup>9</sup>One might wonder how coercion can be consensual in the first place. In other words, it seems as though all coercion is nonconsensual, so saying "nonconsensual coercion" is redundant. I agree that this seems strange. Perhaps what Altman and Wellman

intend by referring to nonconsensual coercion when it comes to states is that individuals do not consent to the state's coercive power. Regardless, I will continue to use their terminology.

assume that they meant the latter. Altman and Wellman do not explicitly state what they mean by coercion, so I will offer a plausible definition of my own to determine whether or not states coerce foreigners. I define coercion as the following:

Coercion occurs when one agent successfully executes a set of actions such that another agent acts in accordance with the first agent's intentions and contrary to her own intentions.

On this definition of coercion, it is clear that the state coerces those *within* its territory. Now, I want to consider this definition as applied to immigration; i.e., Does border control coerce foreigners or not? Since border control is a function of the state, if border control does coerce foreigners, then Altman and Wellman are wrong to assume that states do not nonconsensually coerce foreigners.

The issue at hand is whether or not would-be immigrants are coerced by a state's immigration policy. Suppose that a state has completely closed borders. In other words, its intentions are to keep all individuals outside of the state out. Obviously, the state wants all would-be immigrants to stay out. Furthermore, in order to enforce its intentions, the state builds a wall, impossible to breach, and has armed guards patrol the wall as extra protection. The armed guards are authorized to use any means necessary to keep outsiders out. Thus, any would-be immigrant that attempts to enter the state will be prevented from doing so. In other words, the state's enforcement of its intentions makes it such that would-be immigrants act in accordance with the state's intentions because they never enter. Furthermore, we can assume that there are some people who would attempt to immigrate to the state if they knew it was possible, but knowing of the state's enforcement techniques—knowing that it is impossible—, they do not even attempt to immigrate. They too act in accordance with the state's intentions. Ultimately then, following the definition above, border control coerces all potential immigrants, both those that do attempt to enter the state and those that do not even bother. This example was of completely closed borders, but even states with only partially closed

borders coerce all potential immigrants, including those to whom it eventually grants entry.<sup>10</sup> Since border control coerces all potential immigrants, and border control is a function of the state, the state coerces all potential immigrants. Contra Altman and Wellman then, the state coerces foreigners.

Recall that coercion requires justification, according to Altman and Wellman, when the coercion is nonconsensual. Altman and Wellman assert that nonconsensual coercion is not justified if states "...did not supply these extremely important benefits [i.e., protection of human rights]" (184). Altman and Wellman clearly only intend that a state must supply these extremely important benefits to those already in the state's territory for its nonconsensual coercion to be justified since a state's legitimacy depends on protecting the human rights of its *constituents* and not outsiders. This is because Altman and Wellman assume that states do *not* nonconsensually coerce outsiders. Presumably, if this were not the case, i.e., if Altman and Wellman instead assume that states do nonconsensually coerce outsiders, then Altman and Wellman would instead require a state to protect the human rights of outsiders also. In other words, a state's legitimacy would require the state to protect both the rights of its constituents and all others that it nonconsensually coerces, and not merely respect the rights of all others.<sup>11</sup>

If a state does not protect the human rights of all those it nonconsensually coerces, then its use of nonconsensual coercion is not justified, and it is not legitimate. However, requiring a state to protect the rights of all others contradicts other states' right to self-determination. This requirement goes beyond simply implying that states *must* intervene in the affairs of illegitimate states, which is perhaps not that unpalatable. To protect rights, a state must not infringe on those rights, and it must prohibit others from doing so. States can do this in a variety of ways, which is why the right to self-determination is important. However, if states are also required to protect the rights of those outside of their territory to fulfill its purpose, their coercive power must extend outside of their territory.

<sup>10</sup> It may seem strange to think that the state coerces even successful immigrants. States with partially closed borders usually have lengthy processes for immigrants to obtain entry. In the United States, for example, the State Department is still processing visa applications from the 1990s. All of these processes require potential immigrants to comply or else they risk rejection.

<sup>11</sup> Altman and Wellman may reply that states coerce those within its territory and outsiders in relevantly different ways, such that a

different justification is required for the coercion of outsiders. However, I am not convinced that there is a relevant difference between the coercion used on insiders and outsiders. What it means to be a state is that the state has a monopoly on the use of physical force within a territory. Thus, the coercive power that a state has is based on physical force. On insiders and outsiders alike, the state uses this physical force, whether it be through prisons for insiders or armed guards for outsiders.

The state must use its coercive power to enact and enforce laws on those outside of its territory. If states' coercive power is not contained within a territory, then no state has a monopoly of physical force within its own territory. Other states' laws and coercive powers apply too. This contradicts not only the very definition of a state but also the right to self-determination of states because it destroys the right to self-rule.<sup>12</sup>

To save the right to self-determination, which Altman and Wellman no doubt would prefer, we must eliminate the nonconsensual coercion of foreigners. If states no longer nonconsensually coerce potential immigrants, or foreigners more generally, they no longer must justify such coercion (because it does not exist). Since justification is no longer required, states do not have to protect the human rights of foreigners. Thus, as Altman and Wellman initially supposed, a legitimate state protects the rights of its constituents (i.e., those under its control—those it *does* nonconsensually coerce) and respects the rights of all others. Fortunately, eliminating the nonconsensual coercion of foreigners is not a difficult task. Since border control is a function of the state and is coercive, eliminating border control will also eliminate the state's nonconsensual coercion of foreigners.<sup>13</sup> Furthermore, because any form of border control, i.e., any control that is more than no control, constitutes coercion, eliminating border control does not mean *reducing* border control. It means completely open borders, but as Carens points out, "Having borders that are open is not the same as having no borders" (231). Because borders define a state's territory, and states are territorial entities, states do require borders. Open borders simply mean that a state cannot deny entry to a potential immigrant. There may be an exception to this, however, if, as in the above sections, granting entry to a particular immigrant would undermine the state's ability to fulfill its purpose. In such circumstances, the state's nonconsensual coercion of that

particular immigrant, through denying entry to the territory, may be justified. In the absence of such circumstances, though, nonconsensual coercion of potential immigrants is not justified. To be legitimate, a state must eliminate border control and have open borders.<sup>14</sup> Consequently, the connection between coercion and legitimacy entails that legitimate states do not have a right to control their borders.

### Conclusion

Altman and Wellman argue that as a result of a legitimate state's right to self-determination and corresponding right to freedom of association, a state ultimately has a right to exclude all potential immigrants from its territory. This argument is predicated on a conception of self-determination as self-composition. In Part II, I argued self-composition was an incorrect conception of self-determination. Instead, the proper conception is non-intervention and self-rule, which does not give rise to a *prima facie* right to exclude outsiders. Finally, in Part III, I argued that the state's use of border control constitutes coercion of outsiders. Altman and Wellman contend that states' use of nonconsensual coercion requires justification, and I showed that states fail Altman and Wellman's justification conditions when it comes to border control. However, fulfilling these conditions would contradict state self-determination rights. Thus, to be legitimate and maintain a right to self-determination, states must eliminate border control and open their borders. Through my discussions of self-determination and coercion, I have refuted Altman and Wellman's conclusion—that states have a right to exclude *all* potential immigrants from their territories. Specifically, the right that a state has to exclude potential immigrants from its territory is very limited and only applies when to admit a particular immigrant would prevent a state from fulfilling its purpose.<sup>15</sup>

<sup>12</sup> Since self-rule and non-intervention are corollaries, it ultimately also destroys the legitimate state's right to non-intervention. While we may not have an issue with intervening in an illegitimate state's affairs, most do when it comes to a legitimate state.

<sup>13</sup> This is assuming that border control is the only way in which a state nonconsensually coerces foreigners. I am not denying the possibility that there are other ways in which a state nonconsensually coerces foreigners. If these other ways do indeed exist, then in order for the state to be legitimate and maintain its right to self-determination, the state must eliminate these other ways as well.

<sup>14</sup> To eliminate nonconsensual coercion of foreigners and thus be legitimate, instead of opening their borders, states could also

alternatively switch to consensual coercion of foreigners. This would require all foreigners, specifically potential immigrants, to consent to the state's coercive power over them, i.e., border control. Again, consensual coercion seems to be oxymoronic, but if my interpretation of Altman and Wellman is correct, then consensual coercion simply means that individuals consent to the state's coercive power in general rather than specific acts of coercion. For this alternative option, see e.g., Abizadeh 37-65. However, it seems unlikely that potential immigrants, those that want to enter the state, would ever consent to the state's border control.

<sup>15</sup> I would like to thank Dr. Rekha Nath for her help throughout the process of drafting and completing this paper.

**Works Cited**

- Abizadeh, Arash. "Democratic Theory and Border Coercion: No Right to Unilaterally Control Your Own Borders," *Political Theory*, vol. 36, no. 1, 2008, pp. 37-65.
- Altman, Andrew and Christopher Heath Wellman. *A Liberal Theory of International Justice*. Oxford: Oxford University Press, 2009.
- Carens, Joseph. *The Ethics of Immigration*. Oxford: Oxford University Press, 2013.
- Lægaard, Sune. "Territorial Rights, Political Association, and Immigration." *Journal of Moral Philosophy*, vol. 10, 2013, pp. 645-670.
- Miller, David. "Territorial Rights: Concept and Justification." *Political Studies*, vol. 60, 2012, pp. 252-268.
- Walzer, Michael. *Just and Unjust Wars*. New York: Basic Books, 2006.
- Wellman, Christopher Heath. "Immigration and Freedom of Association." *Ethics*, vol. 119, 2008, pp. 109-141.

## *A Defense of the Second Analogy*

CAO YUNLONG

THE JOHNS HOPKINS UNIVERSITY

**Abstract:** In his book, *The Bounds of Sense*, P. F. Strawson commented that Immanuel Kant’s argument in the second analogy “proceeds by a *non sequitur* of numbing grossness,” causing a fair amount of debates. Kant’s task in the second analogy is to argue that every event has a cause. Strawson criticizes Kant by claiming that in his argument, Kant not only changes the content of necessity but also shifts a conceptual necessity to a causal one. In this paper, I defend Kant’s second analogy against Strawson’s objection by arguing that Strawson misinterprets Kant’s strategy.

**Keywords:** Immanuel Kant, P. F. Strawson, The Second Analogy, Causality

### Introduction

In his book, *The Bounds of Sense*, P. F. Strawson commented that Immanuel Kant’s argument in the second analogy “proceeds by a *non sequitur* of numbing grossness,”<sup>1</sup> causing a fair amount of debates. In this paper, I argue that Kant’s argument can be defended against Strawson’s objection because Strawson misinterprets Kant’s strategy. First, I will demonstrate Kant’s account of events; then, I evaluate Strawson’s argument; after that, I reconstruct Kant’s arguments that every event presupposes a cause; finally, I will briefly mention some philosophers’ comments on Strawson’s objection. In this paper, I use the following notations: capitalized letters, A and B, to represent the empirical objective states of an event; capitalized letters with p in subscript, A<sub>p</sub> and B<sub>p</sub>, to represent the perception or apprehension of an objective state of an event; “–” between letters to represent an event composed by two elements; pre-A to represent the presupposition from which the event A–B follows.

#### I. Kant’s Account of Events and Objects

The underlying question in Kant’s second analogy is: how are objective experiences of events possible? I argue that Kant has three requirements: 1) an event must be composed of two objective states; 2) an event must not follow from an empty time; 3) the two objective states must follow according to a rule of

apprehension in the manifold, which means the order of the objective states is determined objectively. In this section, I discuss them in order.

Kant first defines an event as “a perception that follows upon another perception.”<sup>2</sup> An event must be composed of two parts, A and B. We cannot perceive an event unless something contained in state B is not part of state A. This is intuitive: I cannot perceive the event—e.g., the lights turned on—unless I perceive it is bright now and dark seconds ago. If it is always bright, I would never perceive an event. I use the notation A–B to represent the objective experience of an event, such that A is the preceding, and B the succeeding, objective state of appearances.

Kant then claims that an event should not follow from empty time. He writes that “for an event which should follow upon an empty time, that is, a coming to be preceded by no state of things, is as little capable of being apprehended as empty time itself.”<sup>3</sup> To perceive an event without a preceding state implies that we can perceive empty time before perceiving such an event. Since Kant has already proved that empty time is not perceivable, it follows that there is always a pre-A preceding A–B. This requirement of an event is an essential step in Kant’s argument for all events must have their causes.

The third requirement, events must have a necessary objective order, takes a few steps to reason. First, to distinguish one event from all our apprehensions, i.e., pre-conceptual “sensory data,” we need a concept of such an event, similar to how we need concepts to distinguish objects from apprehensions. For example, I must have the concept of a table to experience the table in front of me; similarly, I also need the concept of “opening doors” to experience any specific door-opening event. Without such a concept, I cannot pick out relevant apprehensions from the manifold. For example, I might perceive a door opening and a person walking in at the same time. I need my concepts to distinguish which parts of my apprehension belong to which event; without the ability to distinguish, I would have no objective experience: I might perceive A<sub>p</sub>–B<sub>p</sub> as arbitrary apprehensions, but never A–B as objective events. Therefore, we must have concepts of events.

A question that follows naturally from the third requirement is: what are some necessary rules of apprehension in the concepts of events? Strawson used

<sup>1</sup> Strawson, 136.

<sup>2</sup> Kant, A192/B237.

<sup>3</sup> Ibid.

the term “order-indifference”<sup>4</sup> to mark the difference between the concepts of events and concepts of objects. Kant thinks that the concepts of an object or co-existing object parts have order-indifference. In contrast, the concepts of an event lack order-indifference: i.e., there is a necessary order of  $A_p-B_p$  ( $A_p-B_p$ -irreversibility). Kant used examples of a house and a ship to demonstrate.

My objective experience of a house contains a necessary rule of apprehension; however, the rules do not require the order of apprehension.<sup>5</sup> My apprehension of a house can either start with the roof or the basement: the rules of connection in the manifold guarantee that I can experience the house with order indifference. In contrast, the rule for an objective event requires a determined order. If a ship moves downstream, I always perceive it first in a position higher up, then lower down.<sup>6</sup> In the event of  $A-B$ , I always perceive  $A_p-B_p$ -irreversibility.  $A_p-B_p$  alone is not enough for an objective experience of an event. I cannot perceive my objective concept  $A-B$  as sometimes  $A_p-B_p$  and sometimes  $B_p-A_p$ . One may argue that sometimes one does perceive the representation  $B_p-A_p$ : e.g., one can first see the ship in a position lower down, then see it higher up; she can reverse the engine. However, this kind of objection bears no significance. Kant does not need to argue that some representations never precede others; he only argues that in cases of events, the order of perceptions is necessary. One sure can perceive  $B_p-A_p$ ; however, that would be a perception of event  $B-A$  instead of  $A-B$ .

Moreover, when one perceives the event  $B-A$ , she must perceive  $B_p-A_p$ -irreversibility. If I first perceive the room as dark, then the room as bright, I perceived the event of lights-turning-on; if I first perceive the room as bright, then the room as dark, I perceive the event of lights-turning-off. There is a clear difference between the two. Thus, we can conclude that for any event  $A-B$ , my perception of an event follows a necessary order,  $A_p-B_p$ -irreversibility.

Lastly, why must  $A_p-B_p$ -irreversibility come from the rules under the concepts  $A-B$ ? Kant claims: “we must derive the subjective succession of apprehension from the objective succession of appearances.”<sup>7</sup> Since we need to have the objective

experience, the necessity of our apprehensions must not come from our understanding; otherwise, the order of apprehension would be arbitrary. Kant writes, “the objective succession will therefore consist in that order of the manifold of appearance.”<sup>8</sup> The objective order must be given to me in the manifold already. Like how I perceive an object, my apprehension must follow the specific rules of the manifold, according to which I cannot help but experience an object in a specific way. For example, I cannot resist experiencing a table upside-down if the rule in the manifold makes me do so; similarly, I cannot help but experience a door opening instead of closing if the rules in the manifold require it. Kant summarizes by saying that “this is another way of saying that I cannot arrange the apprehension otherwise than in this very succession.”<sup>9</sup>

It is worth noting that Kant has not yet accomplished that an event  $A-B$  follows from a cause necessarily. Instead, Kant proves that one can experience the event only in the way the event happens: what I can experience is determined already in the manifold because my sensibility is merely passive and receptive. I will discuss Kant’s argument that all events are caused in section three. Briefly, Kant points out three requirements for the experience of an event. 1) An event must have two parts, A and B; 2) an event must follow from some other objective state, pre-A; 3) the order of perception of an event is always necessary, such that  $B_p$  follows from  $A_p$  according to a rule in the manifold. I now reconstruct Strawson’s notorious *non sequitur* criticism on Kant’s Second Analogy.

## II. Strawson’s *Non Sequitur* Objection

Strawson claims that Kant “not only shifts the application of the word ‘necessary,’ but also changes its sense substituting one type of necessity for another.”<sup>10</sup> He reconstructs Kant’s premises as it is necessary that if  $A-B$ , then  $A_p-B_p$ -irreversibility; he writes,

It is conceptually necessary, given that what is observed is, in fact, a change from A to B, and that there is no such difference in the causal conditions of the perception of these two states as to introduce a differential time-lag into the perception of A, that the observer’s perceptions should have the order:

<sup>4</sup> Strawson, 132.

<sup>5</sup> Kant, A190-191/B235-236.

<sup>6</sup> Ibid., A192/B237.

<sup>7</sup> Ibid., A193/B238.

<sup>8</sup> Ibid.

<sup>9</sup> Ibid.

<sup>10</sup> Strawson, 137.

perception of A, perception of B – and not the reverse order.<sup>11</sup>

Strawson thinks that  $A_p$ 's dependence on A is the same as  $B_p$ 's dependence on B. James Van Cleve, in his reconstruction of Strawson, called this idea “perceptual isomorphism.”<sup>12</sup> Given that A–B and perceptual isomorphism, it conceptually necessarily follows that  $A_p$ – $B_p$ -irreversibility. I think Kant would agree with this premise: one's apprehension follows the order of her objective experience. For example, when I experience a door opening (A–B), I first have an apprehension of the door being closed ( $A_p$ ), then have an apprehension of the door being open ( $B_p$ ). Strawson thinks it is only a conceptual necessity that  $A_p$ – $B_p$ -irreversibility follows from A–B because it is not the case that A–B caused the order of their apprehension to be necessary; the necessity follows from the concept of perceptual isomorphism.

Strawson then reads Kant's conclusion as a causal necessity, as Kant describes in his thesis: “everything that happens, that is, begins to be, presupposes something upon which it follows according to a rule.”<sup>13</sup> It is worth noting that Kant is not trying to argue A–B-necessity, objective state A causally necessitating objective state B. As Strawson points out that “Kant does not say that to conceive the order of perceptions of A and B as necessary is equivalent to conceiving A as causally necessitating B. He says that it is equivalent to conceiving the change from A to B as causally necessitated by some unspecified antecedent conditions.”<sup>14</sup> In other words, Kant concludes that pre-A–(A–B)-necessity, for any event there presupposes a condition from which the event necessarily follows; instead of A–B-necessity, A causally necessitating B.

Strawson thinks Kant cannot conclude that pre-A–(A–B)-necessity with the premise  $A_p$ – $B_p$ -irreversibility. Kant derived his conclusion by shifting the application of “necessity” from 1) necessarily if A–B then  $A_p$ – $B_p$  to 2) A–B necessarily follows from pre-A. Simultaneously, the sense of “necessity” shifted from a conceptual necessity to a causal necessity. I agree that if this is Kant's strategy, he would be guilty of *non sequitur*; however, I disagree that Kant's argument proceeds in this direction. I argue Kant did not achieve pre-A–(A–B)-necessity through the necessary order of one's perception of an event;

instead, Kant's argument for the cause of events follows from the mere possibility of an objective experience of events.

### III. Kant's Arguments for the Causal Principle

In this section, I reconstruct Kant's arguments that all events are caused basing on the mere possibility of objective experience. In previous sections, I have shown that Kant requires necessary relations between the two perceptions of an event: i.e.,  $A_p$ – $B_p$ -irreversibility if A–B are events. Kant's next step is to argue that any event A–B presupposes something, pre-A, upon which it follows according to a rule: i.e., pre-A–(A–B)-necessity. Kant's argument is based on that certain rules are required for objectivity. He says, “I render my subjective synthesis of apprehension objective only by reference to a rule in accordance with which the appearances in their succession, that is, as they happen, are determined by the preceding state. The experience of an event [i.e., of anything as happening] is itself possible only on this assumption.”<sup>15</sup> I now demonstrate Kant's arguments.

Kant argues that an event needs a determinate position in time to be objective.<sup>16</sup> I do not need to know the precise time when an event happens; instead, Kant only claims that I have to put the event into the temporal framework to experience it: e.g., I have to know the door opens after I push the door and before the dog walks out. I have to provide every event with a determinate position in time to have experience. This is similar to how we must provide a determinate spot in space to experience an object. I cannot objectively experience an a-spatial table; similarly, I cannot say I experience an event without knowing its temporal position. The next question for Kant is, from what authority can events obtain a determinate temporal position?

Recall from the last section that all events presuppose something preceding them, pre-A because we cannot perceive empty time. Kant argues that I have to presuppose that any event, A–B, necessarily follows from something preceding, pre-A, to satisfy the determined time relation. Moreover, since this time relation is objectively determined, there must be a governing rule in the manifold; I cannot provide such a rule myself, which would be arbitrary. Kant argues that “[an event] can acquire this determinate position

<sup>11</sup> Ibid.

<sup>12</sup> Van Cleve, 82.

<sup>13</sup> Kant, A189.

<sup>14</sup> Strawson, 137.

<sup>15</sup> Kant, A195/B240.

<sup>16</sup> Ibid., A198/B243.

in this relation of time only in so far as something is presupposed in the preceding state upon which it follows invariably, that is, in accordance with a rule.”<sup>17</sup> In other words, Kant argues that since all events have determined time relation, which can only derive from rules in the concept of objects, and that time only flows in one direction, we conclude that an event can only acquire its determinate time relation from its preceding cause.

Back to Strawson’s objection, The Kantian concept of events has two necessities: (1) the necessary irreversible order of perception given any event and (2) the necessary temporal determination of an event given its preceding cause. I have shown that Kant’s argument that every event has a cause does not follow from (1), which Strawson claims he does. Instead, Kant treats the entire event, A–B, as a whole and argues the event can only be experienced if it has a determined time relation, which should follow upon a rule. In our case, the rule for the ordering is causation. Strawson’s objection rests only on the assumption that Kant argues for objective causation based on subjective apprehensions; however, this is not the case. Therefore, Strawson’s objection fails. Next, I wish to defend Kant’s view against three possible objections.

One possible objection Kant addressed is that events might acquire their determinate temporal relation directly from their relation to absolute time. If each event is marked with a specific time in the manifold, events will not follow from each other necessarily. For example, it could be the case that a rule in the manifold says that “Kant pushes the door at a second before five” and that “the door opens at five o’clock,” instead of “the door opens after Kant pushes it.” Since we have no access to the rules in the manifolds, how can we rule out the former option and claim that events must acquire their temporal relation through each other?

Kant argued that “since absolute time is not an object of perception, this determination of position cannot be derived from the relation of appearances to it. On the contrary, the appearances must determine for one another their position in time and make their time-order a necessary order.”<sup>18</sup> Appearances can only acquire determined relations from other appearances. Time is only the form of sensibility, which has no determinate markers to bear relations to appearances: the rules can only say A–B follows pre-A, but never

A–B happens at five o’clock because the absolute five o’clock is not an object of perception. Therefore, events must acquire their determined relations among each other according to the rules.

Another possible objection Kant addressed is that there might be simultaneous causes.<sup>19</sup> For example, a lead ball is causing a hollow on a cushion. The ball must be on the cushion at the same time when the hollow exists; if I remove the ball, the hollow will disappear. Kant agrees that there could be simultaneous causation; however, he also pointed out, “it is the order of time, not the lapse of time, with which we have to reckon.”<sup>20</sup> The order of time still exists even when there is no lapse in time. Let us scrutinize the event: a hollow appears on a cushion: the first perception is a flat cushion, A; the second perception is a cushion with a hollow, B; the condition is the placing of the lead ball, pre-A. We can see that the posit of pre-A will necessarily lead to the event, A–B, where A–B would not lead to pre-A. Therefore, we can determine that the event A–B shall follow pre-A in the order of time. The necessary temporal relation is determined by order of time, not the lapse of time.

Here is another possible objection: why must it be the case that what determines the temporal relation between events corresponds to what we perceive as causal relations? We perceive so many events at any given time, even if there must be some pre-A; how do we know it picks out the one we want, i.e., the one of causation? Take the door-opening event for example: I pushed the door, then the door opens. The moment right before I open the door, I happen to turn on the lights as well. In this case, we have two pre-A candidates for the event A–B of door-opening: (1) turning on the lights and (2) pushing the door. (2) causally determines A–B while (1) is just a random event; however, they both can determine A–B in temporal order. How can we guarantee that the rules in the manifold are the causal relation?

I think Kant would argue that, if it is the case that we arrange door-opening after lights-turning-on, then we would think that lights-turning-on *causes* door-opening. However, we do arrange door-opening after door-pushing, and we know this because we perceive door-pushing as causing door-opening. The rules already determine the causal relations in the manifold. We are in no position to alter; otherwise, our experience is not *objective*. If it is really the case that

<sup>17</sup> Ibid.

<sup>18</sup> Ibid., A200/B245.

<sup>19</sup> Ibid., A202-203/B247-248.

<sup>20</sup> Ibid., A203/B248.

lights-turning-on causes door opening, potentially how alien determines event order, then they will not feel weird about this determination. An alien philosopher might ask why it is not the case that door-pushing causes door-opening; a similar response is available. Objective experience requires rules coming from the manifold. Kant does not have to worry if the determined temporal order mismatches with causation because it is causation.

From the above argument, we can get two consequences, as Kant claimed: “in the first place, I cannot reverse the series, placing that which happens prior to that upon which it follows. And secondly, if the state which precedes is posited, this determinate event follows inevitably and necessary.”<sup>21</sup> I cannot reverse the objective series because of the governing rules in the manifolds; events inevitably follow from their conditions. If the necessary time relation fails, no appearances would be possible at all. Kant says that “the principle of sufficient reason is thus the ground of possible experiences, that is, of objective knowledge of appearances in respect of their relation in the order of time.”<sup>22</sup>

In brief, Kant first argues that for any event A–B, one necessarily presupposes something, pre-A, before A–B. Also, for both pre-A and A–B to be objective, i.e., not merely subjective apprehensions, they need to have temporal positions. Their relationship is determined in the objects, according to rules in the manifolds. Finally, Kant shows that pre-A and A–B’s temporal position can only derive from relating to each other rather than relating to absolute time. Kant concludes that to experience any event A–B, one needs to presuppose from something Pre-A, from which A–B necessarily follows according to a rule.

I wish to clarify that the causation defended in the second analogy might not be as strong as our ordinary conception of causation: it is only a necessity for our objective and conceptual experiences of events. In our everyday life, we may think causation is a real relation between real events in the world. For example, my throwing of rock actually causes a window to break. However, Kant would argue that the causation is only necessary because my perception of the rock-throwing precedes my perception of the window breaking. Although this may sound unsatisfactory, it is not because the Kantian strategy is inadequate. Instead,

it is because we ask too much of necessity in our everyday life. We cannot directly access the objects in themselves; thus, we cannot attribute causation to them.

#### IV. Commentators’ Replies to Strawson

Henry Allison argues that Strawson’s *non sequitur* objection assumes a “transcendentally realistic standpoint.”<sup>23</sup> He claims that Strawson “treats Kant as if he were an empirical idealist, concerned to ground a conclusion regarding the causal relations of ontologically distinct things and events on a feature of our perceptions (their irreversibility).”<sup>24</sup> Strawson’s objection would be sound if Kant shifts the causal necessity of objective experience to the causal necessity of objects in themselves; however, we are not even entitled to argue for order in the noumenal world. At the beginning of his argument, Kant claimed that “how things may be in themselves, apart from the representations through which they affect us, is entirely outside our sphere of knowledge.”<sup>25</sup> One has to pay attention to the requirement that objects in the phenomenal realm need to be constructed by us according to the rules.

Lewis Beck also points out that Kant does not draw the inference Strawson objects against, namely from  $A_p-B_p$ -irreversibility to A–B-irreversibility, which would be a *non sequitur*. Beck claims that the former is only one evidence for me to know an event A–B, where the latter is a rule which is also necessary for me to regard A–B as a necessary event<sup>26</sup>, which I think Kant derived independently from the former. The apprehension’s irreversibility is necessary under the concept of any event, while the causal necessity rests on the objective temporal order of appearances. The rule and the evidence together enable us to experience events objectively.

Michael Rosen points out that Strawson is wrong by assuming series of events and asking whether those events are objective; however, for Kant, one only has series of apprehensions that need to be synthesized into objective events. Kant’s project is to determine the rules, according to which we can match our subjective apprehensions to the objective experience. He writes that “the constraint upon those rules is that they must be such as to allow us to order what is given to us into a seamless and coherent

<sup>21</sup> Ibid., A198/B243-244.

<sup>22</sup> Ibid., A201/B246.

<sup>23</sup> Allison, 255.

<sup>24</sup> Ibid.

<sup>25</sup> Kant, A191/B235.

<sup>26</sup> Beck, 389.

reality.”<sup>27</sup> Without such rules, e.g., causation, the objective experience is impossible.

### Conclusion

Commentators disagree on whether Kant argues that objective state B follows from objective state A according to a rule, or he argues that the event A–B follows from a presupposed pre-A according to a rule. I think this distinction is subtle: whether causation resides between objective states or between events. Both objective states and events belong to our objective experience, where they necessarily require a relative temporal position. Kant’s argument flows on both reads of his causality, where Strawson’s objection flows on neither.

Kant argued that under the concept of any event, the perceptions must acquire necessary orders. Strawson reads this necessity as the reason Kant uses to prove that every event has a cause, on which He charged Kant of shifting both the application and the sense of necessity in his argument. However, I argue that Kant’s argument does not advance in the direction Strawson attributed to him; Kant achieved the causal necessity based on the relative temporal order between event, which is required for our objective experience; instead of the conceptually necessary perceptual isomorphism, on which Strawson based his attack. I think Strawson mistakenly read Kant’s argument of pre-A–(A–B)-necessity as following from A<sub>p</sub>–B<sub>p</sub>-irreversibility; however, Kant’s arguments are based on the necessity of our experience of events. Therefore, Kant’s argument can be defended against Strawson’s objection.<sup>28</sup>

### Works Cited

- Allison, Henry. *Kant’s Transcendental Idealism*. New Haven and London: Yale University Press, 2004.
- Beck, Lewis. “Is there A Non Sequitur in Kant’s Proof of the Causal Principle?” *Kant-Studien*, vol. 67, no. 3, 1976.
- Kant, Immanuel. *Critique of Pure Reason*. Translated by Norman Kemp Smith. Reissued edition. Hampshire: Palgrave Macmillan, 2007.
- Rosen, Michael. 1996. “The Second Analogy: No Non Sequitire – But does Kant prove too much”. <https://scholar.harvard.edu/michaelrosen/publications/second-analogy-no-non-sequitire-does-kant-prove-too-much>
- Strawson, Peter. *The Bounds of Sense*. London and New York: Routledge, 1966.
- Van Cleve, James. “Four Recent Interpretations of Kant’s Second Analogy.” *Kant-Studien*, vol. 64, no. 1, 1973.

---

<sup>27</sup> Rosen, 18.

<sup>28</sup> I wish to thank Prof. Eckart Förster for his teaching and the audience at the 2021 EMU Undergraduate Conference in Philosophy for their helpful comments.

## *Expanding William Hasker’s Transcendental Refutation of Determinism*

IBRAHIM DAGHER

THE UNIVERSITY OF CALIFORNIA, DAVIS

**Abstract:** This paper is an evaluation and expansion of William Hasker’s transcendental argument against determinism. Hasker’s argument attempts to show that determinism is logically incompatible with rationality and justified belief. Hasker claims this argument to be conclusive given two independent qualifications: first that the argument only applies to a specific form of determinism, and second that the argument rests on a specific conception of rationality. My aim in this paper will be to modify and expand Hasker’s argument such that it (1) applies to the more general deterministic thesis (as defined below), and (2), rests on a modified epistemic conception of rationally unaffirmable beliefs. I will attempt to do (2) by modeling Hasker’s argument after one of Alvin Plantinga’s epistemological arguments against naturalism, and hope to show that such an argumentative model is superior for the conclusion Hasker wishes to bring out. I will begin by explicating Hasker’s argument, and then I will move to the modification and expansion of the argument.

### I. Hasker’s Argument

William Hasker’s argument, entitled ‘The Transcendental Refutation of Determinism,’ is an argument against determinism which seeks to demonstrate a logical incompatibility between determinism and justified belief.

The argument is best summarized by James N. Jordan, who writes:

Suppose that our assessments of arguments are the results of sufficient causal conditions whose complete statement involves no reference to “rational insight into a nexus between premises and conclusion,” no mention of judgment “in accordance with objective laws or principles.” Would it not be merely fortuitous if our assessments were in accord with what “objective laws or principles” prescribe as conditions for sound argumentation?...Of course, one still might occasionally believe what is true, but this would always be the outcome of happy circumstances, never of reasoned investigation. And if this is true of our rational assessment of any argument, it is true of our attempts to determine the strengths and weaknesses of any argument for [determinism]. If the latter is true, any argument for it is self-defeating, for it entails that

no argument can be known to be sound (Jordan 1969).

Hasker’s argument, then, is not one against the *truth* of the deterministic thesis. Rather, the transcendental argument aims to establish that, if determinism were true, no one could ever *know* that determinism (or any other thesis for that matter) were true, because all of our beliefs, even if true, are not the result of reasoned judgement, but rather fully the result of antecedent physical causes. Thus, determinism is self-defeating, and so rationally we cannot affirm it (although it may end up being true). Hasker goes about establishing this claim by first noting two definitions necessary for the argument to work.

First, Hasker notes that the determinism targeted by his argument is a specific conception of determinism, namely the thesis that:

Every event, including in particular every event in the life of each person, has a set of temporally antecedent physical conditions which are sufficient for its occurrence (Hasker 1973).

Second, the argument rests on a key epistemic principle (let us call it H):

(H): For a person to be justified in accepting a conclusion, she must be aware of reasons for the conclusion and her reasoning process must be guided by rational insight on the basis of principles of sound inference (Hasker 1973).

Given these qualifications, Hasker’s argument follows rather naturally. For, under such a determinism, a person’s acceptance of any conclusion has a sufficient explanation in a series of physical causes that determine their belief. And, if this series of physical causes is sufficient for explaining a person’s belief, then it follows that said person’s mental states are not *necessary* for their holding the belief. However, given H, this person cannot be said to be justified in accepting any conclusion.

The function of Hasker’s argument, at heart, is to distinguish between what can be called the “cause” of an agent’s belief and the “reason” for the belief itself. We can define “belief causation” as the reasons for which an agent holds a belief (that is to say, the actual *cause* of the agent’s belief) and “belief reasoning” to be the actual reasons a belief is true. Given our key epistemic principle H (that a person must be aware of the reasons for her belief and that those reasons must be sound), it follows that a necessary condition for an agent being justified in a belief is a correlation between

the agent’s belief causation and belief reasoning. And since belief causations, under determinism, are always the combination of the laws of nature acting upon an agent’s neurological physical state, it seems that agents in a deterministic world are never justified in *any* belief they hold. Their beliefs, rather than being the results of a careful consideration of logical principles and relevant data, are the results of a rationally indifferent neuro-chemical causal history. To put it in the words of James N. Jordan, all of our beliefs would be “the results of sufficient causal conditions whose complete statement involves no reference to ‘rational insight into a nexus between premises and conclusion,’” and it would be “merely fortuitous if our assessments were in accord with what ‘objective laws or principles’ prescribe as conditions for sound argumentation” if determinism were true (Jordan 1969). Indeed, given that we accept that justified beliefs require our awareness of sound rational insight, it seems we are compelled to say no such scenario is possible under determinism.

We can roughly reconstruct Hasker’s argument as follows:

1. For a person to be justified in accepting a conclusion, she must be aware of reasons for the conclusion and her reasoning process must be guided by rational insight on the basis of principles of sound inference.
2. If determinism is true, no person is guided by rational insight on the basis of principles of sound inference in believing anything.
3. [Therefore, from 1 and 2] if determinism is true no person would be justified in accepting any conclusion.
4. So, determinism is rationally unaffirmable.

Clearly, both of Hasker’s qualifications (that the argument refers to a specific type of determinism, and that it rests on the assumption of H as true) do the majority of the work for Hasker’s argument. In one sense, the argument is limited, as it targets only a specific type of determinism (albeit a popularly motivated one) which states that every event has a temporal *physical* antecedent sufficient for its occurrence. Let us refer to this sort of determinism as ‘physicalist determinism’. As Hasker notes, not all advocates of determinism would agree with physicalist determinism. Many would argue for what I shall call ‘general determinism’, the broader thesis that every event has a temporal antecedent sufficient for its occurrence. Note that physicalist determinism specifies that all antecedents which determine any event are physical. General

determinism, on the other hand, leaves open the question of the nature of any event’s determining antecedents.

As Hasker concedes, such a determinism would not be affected by his argument, because general determinism is fully compatible with the epistemic principle H. Under general determinism, it may be that an agent’s mental states, or more broadly that the laws of logic and inference carry *causal ability* such that they determine an agent’s beliefs. Or, to make use of our prior terminology, general determinists can maintain that a potential form of belief causation *just is* belief reasoning. Hasker mentions Brand Blanshard, among other determinists, as advocates for such a position, writing that:

Suppose someone were to hold that the laws of logic, and the principles of good reasoning generally, hold an irresistible attraction for the human mind—a sort of epistemic counterpart to the theologian’s “irresistible grace”—such that we just cannot help but reason correctly whenever the opportunity presents itself (Hasker 1973).

In such a case, premise (2) of Hasker’s argument no longer applies, as in fact agents *can* be guided by rational insight on the basis of principles of sound inference (perhaps even more so than in a world with free will). It is this limit that I mean to expand in this paper. I contend that the transcendental argument can in fact apply to general determinism, given that we alter the epistemic principle H. Rather than targeting rationally justified belief, I will focus on rationally unaffirmable beliefs to show that any sort of determinism is ultimately irrational.

## II. On Rationally Unaffirmable Beliefs

To take on such a task, we must first say more about rationally unaffirmable beliefs. More specifically, we must define what precisely makes a belief rationally unaffirmable. Generally, a belief that cannot be rationally affirmed has been conceived of as a belief which, when affirmed, undermines the rational faculties of the agent affirming the belief. Consider the explanation and example offered by Peter van Inwagen of a rationally unaffirmable belief:

Let us use ‘anti-rationalism’ as a name for the thesis that none of our beliefs is based on reasoning—the thesis that reasoning plays no role in the explanation of why anyone holds any belief. Anti-rationalism is a position that cannot be *rationally* accepted. It cannot be rationally accepted because, as one might say, it undermines a claim that anyone

might make on its behalf to be a thesis that any person can rationally believe (van Inwagen 2013).

Clearly, ‘anti-rationalism’ is a rationally unaffirmable belief because my affirming anti-rationalism would mean my affirmation that none of my beliefs are rationally accepted, *including my belief in anti-rationalism*. To put it more rigorously, anti-rationalism provides a *rebutting* defeater of any belief I hold, including anti-rationalism itself. This is to say that in affirming anti-rationalism, I not only remove any reason to think any of my beliefs are true, but I have direct reason to think my beliefs are *not* true, including my belief in anti-rationalism itself. And so anti-rationalism cannot be rationally affirmed. As van Inwagen suggests, this connection between rebutting defeater and rationally unaffirmable belief can more clearly be seen if one considers the counterfactual case. Suppose we wanted to construct a rational case or argument for one’s affirmation of a thesis. If said thesis is one which, in our affirmation of it, rebuttingly defeats our beliefs, or our reasons to take our beliefs as rational, then our task of constructing a rational case on the thesis’ behalf is not even possible in theory, because, as van Inwagen puts it, in our construction of such a case we construct the very propositions which make it so that no such case is tenable. If it is not possible to build a rational case for a thesis, this is just to say that it is rationally unaffirmable. It is worth noting that anti-rationalism may end up being true—but given that one cannot rationally affirm it, the rational course of action would be to reject belief in anti-rationalism. The condition can be stated like this then: a belief X is rationally unaffirmable if affirming X provides a rebutting defeater of any belief I hold.

Yet it seems the principle of what makes a belief rationally unaffirmable can be expanded to not only include beliefs that rebut an agent’s confidence in her rational faculties, but beliefs that undercut it. In other words, insofar as a belief of mine removes any reason I have in thinking that my beliefs are true, then it is rationally unaffirmable. Consider the thesis of ‘quasi-rationalism’, which states that we have no good reason to think anyone ever uses reason when acquiring any belief. This thesis, too, seems rationally unaffirmable. For, if I attempt to affirm quasi-rationalism, I will have just conceded that there is no good reason for me to think I affirmed the belief rationally, and as such, I have no good reason to think it is true. Again, this fact becomes obvious after considering the process of constructing a rational case for a belief. If a

belief is one which, by constructing a rational case on its behalf, an agent acquires an undercutting defeater of the belief itself, then again we ought to think no rational case for the belief is even possible. The principle, then, can be reformed to state: a belief X is rationally unaffirmable if affirming X provides an undercutting defeater of any belief one might hold. The proper course of action, of course, when met with a belief that is rationally unaffirmable is to reject that belief.

To clarify this concept, I will model my argument after Alvin Plantinga’s evolutionary argument against naturalism, itself an epistemological argument against the thesis of evolutionary naturalism (for a rich analysis of the argument, see Plantinga 1993). Plantinga’s argument is centered around a consideration of the conditional probability statement  $P(R \mid (N\&E))$ , where R is the statement that our cognitive faculties are generally reliable (by this Plantinga means that our faculties are capable of reaching truth in elementary arithmetic, logic, inference, and basic observation (see Plantinga 1993)), N is the statement that metaphysical naturalism is true, and E is the statement that our cognitive faculties are the results of biological evolution. Plantinga notes the following with respect to  $P(R \mid N\&E)$ :

You might reasonably hold, therefore, that the right course here is simple agnosticism: one just does not know (and has no good way of finding out) what  $P(R \mid (N\&E))$  might be...What would be the right attitude to take to R? Well, if we have no further information, then wouldn't the right attitude here be agnosticism, withholding belief? The person convinced of N&E who is agnostic about  $P(R \mid (N\&E))$  ...has a defeater for any belief B he holds. Now the next thing to note is that B might be N&E itself; our devotee of N&E has an undercutting defeater for N&E...the rational course would be to reject belief in N&E (Plantinga 1993).

Plantinga’s epistemological analysis is identical with our principle above: a belief X is rationally unaffirmable if affirming X provides an undercutting defeater of any belief one might hold. This principle is particularly insightful as we turn back to our original task: expanding Hasker’s argument to include general determinism. Following Plantinga’s argumentative structure, it becomes clear that rather than constructing an epistemic principle for rationally justified beliefs (which in itself is a controversial principle), we should rather ask the following question: does an affirmation of general determinism leave us agnostic about R, and thus provide

an undercutting defeater of our beliefs? In other words, what is the  $P(R \mid GD)$ , where  $GD$  is the statement that general determinism is true? If general determinism does indeed undercut our beliefs and force us to at least be agnostic with respect to  $P(R \mid GD)$ , as I shall argue it does, then the rational course of action is to reject belief in general determinism.

### III. Rachel’s Plight and Expanding Hasker’s Argument

To argue for the proposition that general determinism provides an undercutting defeater of our beliefs, let us begin by supposing that the world we live in is such that general determinism is true. Suppose further that, indeed, some agents are predisposed to act rationally and are determined to be guided by rational insight and sound inference when forming beliefs (agents whose belief causations just *are* belief reasonings). These we can call “rationally determined agents”. There are also (in virtue of the obvious variety of beliefs and epistemic methodologies) agents predisposed to reason *irrationally* and to hold beliefs for reasons *other* than for which they are true or false beliefs (agents whose belief causations are *in no way* belief reasonings). These we can call “irrationally determined agents.”

Now consider this skeptical scenario. Given that this is in fact the state of affairs, how do I know if any of *my* beliefs are rationally produced and correctly reasoned through? In other words—how do I know that I am not in fact one of those “irrationally determined agents” which we have all agreed indubitably exist? Surely every agent is determined to hold and form their beliefs for either correct or incorrect reasons. In terms of our prior terminology, every agent will have belief causations which either do or do not correlate with the proper belief reasonings. And surely every agent would be convinced of precisely those reasons as those correct in producing their beliefs, regardless of if in fact they are. In other words, every agent will be convinced of the correlation between their belief causations and belief reasonings, irrespective of if in fact there is such a correlation. So the question persists: how do I know that any of my beliefs are rationally produced? How do I know there is in fact any correlation between my belief causations and belief reasonings? Any—rational—answer given here would have to *assume* that I am, in the first place, such an agent.

To better understand the implications of this argument, consider the following thought experiment. Imagine an agent—Rachel—with a special type of

consciousness, such that she can only perceive that which is around her and perform motor skills (move her limbs, walk, etc.) and she cannot perform any sort of reasoning or abstract intellectual activity. Imagine now that Rachel is given a very difficult mathematical problem, requiring multiple rigorous mathematical steps, and has only one solution. In front of Rachel is a box which contains two helmets, physically indistinguishable from one another. Each helmet, when Rachel puts it on, will take complete control of her mind and force Rachel to write down a series of mathematical steps and a solution to the problem, all to Rachel’s cognizance. However, only one of the two helmets in the box will force Rachel to produce the correct solution with correct mathematical axioms and reasoning utilized in the steps. The other helmet will produce an incorrect solution via incorrect steps. In all cases, though, Rachel will be convinced her work and solution was correct. Curious, Rachel picks a helmet at random, and solves away. The central question can be formulated as follows: can Rachel, with any confidence, affirm the proposition that she has correctly solved the math problem?

It seems that, at the very least, Rachel would be forced to reserve her judgement about whether she in fact solved the problem correctly. Perhaps she did, perhaps she did not: the proper course of action would be to be agnostic about the proposition. Rachel has an undercutting defeater of her mathematical faculties’ efficacy, despite the fact that her belief causation may in fact be perfectly in line with the correct belief reasonings.

Rachel’s plight, I argue, is precisely the same as general determinism’s. Given a deterministic world, we all live with helmets on our heads, aimlessly ‘reasoning’ with all conviction and no warrant, robotically moving along our predetermined lines of reasoning in the same respect that Rachel moves her pen to solve the mathematical problem. Some of us move to the hum of the ‘correct’ helmet, while others to the ‘incorrect’ helmet, and others to anything in between. Whether our predetermined belief causations in fact correlate with the proper belief reasonings is as insignificant as the question of if Rachel’s work correlates with the correct mathematical answer.

What seems to be at the core of the skeptical reasoning here is some principle of indifference. That is to say that although it is metaphysically *tenable* that agents are determined to be guided by rational insight and sound inference in the production of their beliefs, there just is no independent reason to think that they

are. It seems, at least *prima facie*, equally plausible that the  $P(R | GD)$  is low compared to high. What, then, is the probability that our cognitive faculties are generally reliable given general determinism (what is  $P(R | GD)$ )? It seems, just as in the case of  $P(R | E\&N)$ , the proper course of action is to reserve judgement about the probability of R given general determinism. Especially if one is willing to grant that the truth of general determinism is indifferent, or inconsequential, to the  $P(R)$ . And it seems that thought experiments such as Rachel’s Plight provide ample evidence of the fact that a deterministic world, even one where it is metaphysically possible to be guided by rational insight and sound inference, is at minimum inconsequential, if not damaging, to the  $P(R)$ . If this quintessential point is agreed to, we now have a rational argument against general determinism. For if we are at least agnostic about this probability, if we agree that we ought to reserve judgement about R given the truth of general determinism, it follows we will be agnostic about R, meaning anyone who affirms general determinism now has an undercutting defeater of any belief he might hold—including his belief in general determinism. And so, in the absence of any defeaters for this defeater, the rational course of action is to reject belief in general determinism.

Hasker’s argument, in effect, has been expanded to general determinism in the following way:

1. A belief X is rationally unaffirmable if affirming X provides an undercutting defeater of any belief one might hold.
2. If general determinism is true we ought to reserve judgement about the reliability of our cognitive faculties.
3. If a belief causes us to reserve judgement about the reliability of our cognitive faculties, then the belief provides an undercutting defeater of any belief we might hold.
4. Therefore, general determinism is rationally unaffirmable.

Premises (1) and (3) are the results of Plantinga’s aforementioned analysis, and premise (2) is the central premise which I have argued for, making use of Rachel’s Plight as a skeptical scenario to warrant our reservation about the reliability of our cognitive faculties under general determinism.

It should be noted that, especially in the space for which I have argued here, the discussion concerning rationality and general determinism does not stop here. The critical phrasing in our conclusion is that *in the absence of any independent defeaters of this defeater*, we ought to reject belief in general determinism. It very well may be argued that various defeaters of this undercutting defeater exist (perhaps if theism is true, or if evolutionary mechanisms following a deterministic path are highly conducive to R), which make it such that we ought not reserve judgement concerning the  $P(R | GD)$ . If such contentions are advanced, then an insightful discussion will begin concerning the success of such defeaters in ‘saving’ rational belief in general determinism. Similarly, there may be additional contentions (both empirical and philosophical) to motivate the aforementioned principle of indifference, or even  $P(R | GD)$  being low (such as Plantinga’s points about the nature of evolutionary naturalism). My purpose is to show that such a discussion must be had, for at least *prima facie* it seems issues concerning rationality are relevant in our evaluation of the deterministic thesis.

It is also worth adding that this new argumentative structure that I have offered, modeled after Plantinga’s argument, also expands the epistemological playing field around Hasker’s argument. As has been discussed elsewhere (see Hookway 1989), Hasker’s key epistemic principle H centers around an internalist view of justification, and can be readily denied by externalists. By shifting the argument to concern the probabilistic relationship of the reliability of our cognitive faculties given general determinism, the argument against the rationality of determinism no longer relies on a somewhat controversial conception of rational justification. The expanded argument now applies to externalists who would be willing to accept premise (2).<sup>1</sup>

---

<sup>1</sup> I would like to deeply thank Dr. Hannah Tierney and Dr. David Glick for their many helpful comments on earlier drafts of this manuscript, which made this publication possible.

**Works Cited**

- Hasker, William. “The transcendental refutation of determinism.” *Southern Journal of Philosophy*, vol. 11, no. 3, 1973, pp. 175-183.
- Hookway, Christopher. “The epicurean argument: Determinism and scepticism.” *Inquiry: An Interdisciplinary Journal of Philosophy*, vol. 32, no. 1, 1989, pp. 79-94.
- Jordan, James N. “Determinism's Dilemma.” *Review of Metaphysics*, vol. 23, no. 1, 1969, pp. 48-66.
- Plantinga, Alvin. *Warrant and Proper Function*. Oxford: Oxford University Press, 1993.
- van Inwagen, Peter. “C. S. Lewis’s Argument Against Naturalism.” *Res Philosophica*, vol. 90, no. 1, 2013, pp. 113-124.

## ***Phenomenology of Consciousness: Understanding its Functional Role***

NATHAN JACOB

THE EMORY UNIVERSITY

**Abstract:** As intimately as we know conscious experience, it is difficult to explain how it fits within our understanding of the natural world. The hard problem of consciousness, posited by David Chalmers, addresses the problem of conscious experience: how and why do physical processes give rise to qualia, or phenomenal experiences? According to Chalmers, a solution to this problem “would involve an account of the relation between physical processes and consciousness, explaining on the basis of natural principles how and why it is that physical processes are associated with states of experience” (Chalmers, 248; sec. 2). In relation to this problem, Joseph Levine argues that there is an explanatory gap between our understanding of the physical world and consciousness (which will be discussed further in *Objections and Responses*). The aims of this paper are threefold. I will examine the two main metaphysical theories, materialism and dualism, that address the hard problem of consciousness and explanatory gap in different ways. I will then put forward my own phenomenological position that establishes the presence of a functional role of consciousness. Finally, I will discuss other metaphysical theories, such as mysterianism and panpsychism, which address the remaining unanswered questions and could be incorporated into my position.

In this paper, I will start with a brief overview of two metaphysical theories, materialism and dualism, that address the hard problem of consciousness and the explanatory gap. After discussing objections and the unanswered questions associated with these theories, I will provide my own phenomenological position on consciousness. I take the position that consciousness, namely qualia, has a functional role that involves its interaction with physical properties. My general point is this. Certain mental diseases, such as pseudocyesis, cannot be treated or stimulated by changing physical processes, but rather through manipulation of qualia (conscious experiences). This manipulation is associated with measurable changes in physical symptoms, suggesting that change at the conscious level has an influence on those physical properties. I acknowledge that physical processes are connected to and influence consciousness: any conscious state is associated with a

given cortical state (associated cerebral cortex activity). Nonetheless, treatment strategy used for mental diseases like pseudocyesis demonstrates that change at the conscious level can be more effective in treatments for physical symptoms, which can be explained through the functional role of consciousness. Consciousness itself is ambiguous and has many interpretations. For the sake of clarity, I will consider consciousness as synonymous with experience or quale, similar to Chalmers’ interpretation. After addressing anticipated objections and limitations to this position, I will discuss metaphysical theories (mysterianism and panpsychism) that are capable of incorporating my position while addressing its limitations.

### **I. Consciousness According to Materialism and Dualism**

There are countless metaphysical theories within philosophy of mind, so for the sake of concision I will only discuss the encompassing ones: materialism and dualism.

Materialism, or physicalism, employs a reductionist view and argues against the existence of the explanatory gap. Materialists believe that consciousness should be reduced to entirely physical explanations. If consciousness is understood only as an assortment of brain processes, then there would not be an explanatory gap between the physical world and consciousness. Patricia Churchland expands upon this idea of “mind-brain reduction” by arguing that folk psychological concepts, or common phenomenological experiences (like anger or joy), should be explained through brain states (Churchland 23). Since consciousness is a specific type of brain state, folk psychological concepts that arise from consciousness should be explained through brain states. Paul Churchland takes a more radical materialist view through eliminative materialism; he argues that folk psychological explanations constitute “a radically false theory... [which] will eventually be displaced, rather than smoothly reduced, by completed neuroscience” (Churchland 568). According to Churchland, folk psychology provides ambiguous and inadequate explanations of human behavior when compared to explanations of brain processes. Hence, it is misleading and should be replaced with the more reliable and more developed brain-based explanations.

Dualists generally support the existence of the explanatory gap and employ a non-reductionist view in

describing consciousness. Dualism posits that consciousness must be non-physical and thus cannot be reduced to physical processes. Hence, the explanatory gap exists and presents the problem as to how physical processes influence non-physical conscious experiences, and vice versa. In the *Meditations*, René Descartes argues that the mind and body are distinct because the mind is non-extended and indivisible in nature while the body is extended and divisible in nature. He equates consciousness entirely with the mind, believing that it should be treated as a separate entity from the body. Property dualism emerged as a new type of dualism that attempts to preserve the mental and physical distinction in a manner more in accordance with the contemporary understanding of the material world. It posits that the world is made up of only physical substances, commonly understood as a fundamental kind of entity that can bear properties. Property dualists believe that all physical substances are made up of two distinct types of properties: mental and physical properties. Hence, it logically follows that they consider consciousness as an aspect of its otherwise entirely physical substance, the brain, and is influenced by its physical properties. Epiphenomenalism is a kind of supervenience property dualism that is used to explain consciousness and qualia. In arguing that qualia are left out of the reductionist view, Frank Jackson argues that qualia can be best explained through epiphenomenalism. Epiphenomenalism refers to the view that mental events are caused by physical events in the brain, but mental events have no effects on physical events. According to this view, mental processes (like consciousness) are inert byproducts of physical processes. Jackson argues that qualia are best explained through epiphenomenalism because they have no effect on the physical world. The absence of qualia would make no difference to the physical world, so therefore they have no effect on the physical world. However, qualia are caused by something physical so they must be a byproduct of physical processes.

## II. Objections and Unanswered Questions of These Theories

For the sake of brevity, I will omit discussing the myriad of counterarguments to specific theories within materialism and dualism. Rather, I will provide my own objections to these metaphysical theories and address the remaining unanswered questions.

I will start by providing an epistemological argument as to how materialism does not clearly refute the idea of an explanatory gap, as it claims it does.

Let's assume that the theory of materialism is correct: everything can be exhaustively described and explained through physical processes/explanations. This would mean that conscious experience can be reduced to physical processes and explanations. If this theory successfully argues against the existence of an explanatory gap, then it is necessary to state that the explanation of physical processes should entail and explain the entirety of conscious experience (otherwise, one could argue that there still exists a gap in using physical processes to explain consciousness). However, this doesn't necessarily hold true. For example, take the statement: activation of the amygdala leads to anger. This is a true statement; activation of the amygdala is in fact a scientifically proven and biological basis of anger. However, one cannot imagine or experience anger based on the objective account of associated brain states. In other words, the meaning of a statement about brain states does not express the meaning of a statement about conscious states. We can go deeper in scientific explanation. The DARPP-32 gene and change in left amygdala volume leads to anger (Reuter et al., 2009). Again, this is a scientifically proven statement, but it does not entail the conscious experience of anger. Joseph Levine provides a similar objection in relation to the statement: pain is the firing of C-fibers. There is nothing about the firing of C-fibers that "makes it naturally 'fit' the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties" (Levine, 356). This leaves the connection between C-fiber firing and the experience of pain as a mystery. The existence of the explanatory gap serves as an objection to the main claim of materialism, but materialism does not sufficiently or logically disprove the explanatory gap. Physical properties certainly influence consciousness, but this doesn't logically imply that consciousness can be reduced to physical properties. An anticipated objection to this point could be: "a" can be "b" without understanding why it is in fact "b". This statement is not applicable to this refutation because the second part of the objection, "without understanding why it is fact 'b'", denotes the existence of an explanatory gap, which materialism claims to reject entirely.

Dualism avoids this logical fallacy by supporting the existence of the explanatory gap and distinguishing consciousness from physical processes. There are many arguments against the various forms of dualism, but the main one is that dualism fails to address how physical processes can interact with non-

physical consciousness. If dualists assert that consciousness is distinct from physical reality but acknowledge the explanatory gap between the two, they must be able to explain the gap or provide reasoning as to why it cannot be explained. Dualists have tried to answer this question in a variety of ways but have fallen short of an entire and complete explanation. Descartes proposed that consciousness interacts with physical processes via the pineal gland, but this claim has been empirically disproved by modern science. Modern property dualists look to epiphenomenalism, which proposes that qualia are inert. In other words, physical processes affect qualia, but qualia don't affect physical processes (they are functionally inert). This still fails to explain how physical processes lead to conscious experiences.

### III. The Functional Role of Consciousness

Before explaining my position on consciousness, it is important to clearly state its limitations based on its nature. As this is a phenomenological position, it seeks to describe specific phenomena through direct investigation of qualia. While the position itself may not address the problem of consciousness or the explanatory gap, it can incorporate different metaphysical theories that do address this problem (which will be addressed in *The Hard Problem of Consciousness*). Altogether, it aims to provide more clarity on the metaphysical reality of consciousness by carefully investigating qualia.

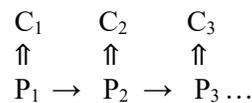
I will begin by introducing an empirical study about the cerebral cortex and consciousness, as well as a corresponding schema (schema 1) that demonstrates the interaction between physical and cortical states associated with consciousness in the study. I will refer to the cortical states associated with consciousness simply as conscious states/properties going forward. As this schema represents a simple scenario with basic interactions, it will serve as a baseline schema and establish a premise that will be utilized when constructing further schemas throughout the discussion of my position. I will then introduce a different empirical study about pseudocyesis, a mental disorder, and explain how its treatment strategy highlights the functional role of consciousness. I will propose a schema (schema 2) that represents the interactions between physical and conscious properties in this study. Finally, I will present a new revision of schema 1 (schema 3) that incorporates the functional role of consciousness.

It is proven that the cerebral cortex has an influence on conscious activity. When specific layers of

the cerebral cortex are stimulated, there is a change in reported conscious experiences/associated neural processing as well as activation of other neurons in the thalamus (Redinbaugh et al., 2020).

Schema 1 will represent this straightforward scenario, with P representing a physical property/properties and C representing the complex cortical state associated with consciousness. According to this study, stimulation of the cerebral cortex ( $P_1$ ) can change conscious experience ( $C_1$ ) and/or stimulate other neurons in the thalamus ( $P_2$ ):

#### Schema 1 – Cerebral Cortex and Consciousness



This schema establishes an important premise that will be used in the next schemas. Every conscious experience must have a biological, physical basis that influences it. In other words, every conscious property “C” on the schema must be connected to a physical property “P” (it cannot stand alone).

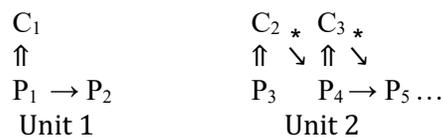
Now, let's consider studies that demonstrate the functional role of consciousness. By functional role, I mean the ability of cortical states associated with consciousness to affect brain and bodily states. These studies focus on the treatment of mental health disorders. In order to avoid ambiguity in interpretation, it is important to focus on a study that undisputedly demonstrates that a new conscious experience ultimately leads to a change in physical properties. The primary study used to demonstrate my position involves the treatment strategy of pseudocyesis. Pseudocyesis is a psychosomatic condition, in which females have a false belief of pregnancy but display real pregnancy symptoms. According to biopsychosocial studies of pseudocyesis, this disease occurs when a woman has an intense belief that she is pregnant. This is accompanied with physiological manifestations of pregnancy, including “irregular menstruation, abdominal distention, the subjective sensing of fetal movements” and more (Azizi et al., 2017). Since there are biological, psychological, and social factors that are demonstrated to underlie this disease, it is impractical to argue that it is only the patient's intense, conscious desire that leads to the physical symptoms displayed. However, we reach a different conclusion when considering the treatment strategy. Published literature from over 1000

studies on pseudocyesis have come to the same conclusion about the best and most successful treatment strategy for this disease (Azizi et al., 2017). “Confrontation with reality and a true diagnosis, together with a supportive therapeutic approach, in most patients lead to disappearance of the symptoms of false pregnancy, and in other patients the symptoms evanesce during a six-month period” (Koić et al., 2002). Many patients affected by this disease saw a reduction of physical pregnancy symptoms after they were made aware that they were not actually pregnant, most commonly done through presentation of an ultrasound (Azizi et al., 2017). Treatment at the level of consciousness and belief, rather than at a biological level, is more successful for patients afflicted with pseudocyesis. In other words, it is the introduction of a new, complex cortical state associated with consciousness that results in the reduction of physical symptoms, not a physical property like drugs or medicine. While drugs like antidepressants or antipsychotics have been used occasionally in treatment, it was only demonstrated to be effective in combination with the treatment strategies mentioned earlier (Azizi et al., 2017). The leading treatment for this mental disease remains a psychotherapeutic one. This study indicates that once new information was understood consciously, there was a resulting change in physical symptoms. This is what I define as the functional role of consciousness: consciousness is affected by physical processes, but its cortical states also function to affect other physical processes. This functional role could also explain the progression towards establishing evidence-based psychological treatments (EBPTs) as the primary treatment for mental disorders. There is a growing list of EBPTs that are linked to significant improvements in physical symptoms associated with a variety of mental diseases, like post-traumatic stress disorder, depression, and much more. In all of these studies, it was found that introduction of new, conscious experiences is associated with cortical states that led to a change in physical symptoms, rather than a manipulation of physical properties (Harvey et al. 2015). However, the functional role of consciousness is most clearly demonstrated in the treatment of pseudocyesis, in which the supporting science has clearly demonstrated the efficacy of treating patients by carefully introducing a new, conscious experience.

To simplify my position, I will propose a schema (schema 2) that addresses the entire case of pseudocyesis and the functional role of consciousness. Like schema 1, physical properties are represented by

P and conscious properties are represented by C. Let’s consider the case of a female patient who suffers from pseudocyesis. She experiences an intense desire to get pregnant ( $C_1$ ), and this experience of motivation/desire is known to arise from biological factors like dopamine and associated biological activity ( $P_1$ ). She subsequently starts to experience pregnancy symptoms ( $P_2$ ), but at this point it is inconclusive if these symptoms resulted from the desire to get pregnant or its underlying biological factors. Since the functional role of consciousness cannot be clearly defined at this point, let’s assume that the pregnancy symptoms arise from the rush of dopamine and other biological factors. The interactions in this situation are contained within unit 1, which is essentially the same as schema 1 since the functional role of consciousness is yet to be definitively established. Let’s move on to the treatment strategy, which introduces a new unit of interactions (unit 2). The patient is made aware that she is not pregnant, which constitutes a new conscious experience ( $C_2$ ) that is influenced by certain physical properties ( $P_3$ ). In this psychological treatment strategy, the introduction of a new, conscious experience and its associated cortical state is the *sole* variable that results in a reduction in physical pregnancy symptoms ( $P_4$ ), since the existing physical properties ( $P_3$ ) are not being manipulated directly in treatment. This relationship demonstrates and establishes the functional role of consciousness (indicated by “\*” in schema 2). Proceeding onwards, the reduction in pregnancy symptoms can induce downstream changes in physical hormone levels ( $P_5$ ) or lead to a new, conscious experience ( $C_3$ ). Since it has been established that the cortical states of conscious experiences have a functional role, it is certainly possible that the new conscious experience ( $C_3$ ) could also lead to the changes in physical hormone levels ( $P_5$ ). However, this position does not yield insight into whether  $C_3$  or  $P_4$  have a greater influence when it comes to inducing  $P_5$ ; rather, it simply introduces a new interaction between conscious and physical properties. Schema 2 represents these interactions, specific to pseudocyesis:

#### Schema 2 – Pseudocyesis

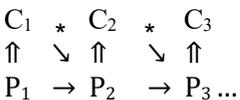


“\*” represents the functional role of consciousness

This schema clearly illustrates why the treatment strategy of pseudocyesis demonstrates the functional role of consciousness. By establishing that changing physical properties cannot result in a reduction of pregnancy symptoms (no arrow between  $P_3$  and  $P_4$ ) but introducing conscious experience can (arrow between  $C_2$  and  $P_4$ ), this study eliminates ambiguity in determining what could have led to this physical change.

To create a simple revision of schema 1 and 2 that most accurately incorporates the functional role of consciousness, let's assume that all of the conscious experiences are related to previous changes in physical symptoms (if this were not the case, there would simply be the formation of a new unit with the same schema structure). The conscious and physical interactions are represented through schema 3:

#### Schema 3 – The Functional Role of Consciousness



A clear example that demonstrates schema 3 is the treatment strategy of a mental disease: it involves EBPTs (conscious), biological treatments (physical), or a combination of both to see a resulting change in physical symptoms.

A limitation to these schemas come from the simplicity of design. To explain my position with the most clarity, the schemas are designed in a linear fashion to most clearly represent the sequential explanation of pseudocyesis. However, when considering the real-world application of consciousness and its function, conscious experiences can influence a variety of physical properties, not just one. In schema 3,  $C_1$  is not only limited to affecting  $P_2$ ; it can impact a large variety of physical properties and processes. In fact,  $P$  can refer to a single physical property or a variety of related physical properties (similar to how  $P_4$  represents the reduction of physical pregnancy symptoms). The schemas and its symbols were simplified in order to explain the interactions in pseudocyesis as clearly as possible.

#### IV. Objections and Responses

In order to provide more clarity and defend the interactions proposed by my position and schema, I will address various anticipated objections to the proposition that consciousness has a functional role.

Materialists and epiphenomenalists would likely argue that the subjective experience isn't causing the physical symptoms of pseudocyesis to go away. Rather, it is entirely a change in the activity of the cerebral cortex or another physical process that causes it to go away. Hence, schema 1 should represent a more accurate understanding of consciousness because it is only the physical change that leads to a reduction in symptoms (materialists would likely revise schema 1 to remove  $C$  states altogether). It is important to note that in this position, I am not arguing that the subjective experience is what is directly causing the physical symptoms to disappear. Rather, it is the physical, cortical states associated with consciousness that leads to the disappearance of physical symptoms. However, since qualia is being manipulated rather than cortical states, it is impractical to reduce qualia to these cortical states. In other words, it is important to consider the two as associated rather than equal, since treatment for pseudocyesis involves change at the level of qualia, rather than at the physical level through altering cortical states. If qualia or consciousness could be reduced to their cortical states, scientists or doctors should be able to induce changes to cerebral cortex activity to see at least some of the physical symptoms go away. However, various attempts to create physical treatments that implement physical changes to the brain or body have not seen stand-alone success in reducing or completely removing these physical symptoms (Azizi et al., 2017). Rather, affecting the patient at the level of belief (telling patients, showing ultrasounds, or other imaging techniques) is associated with cortical states that lead to more efficient treatment. One can argue that scientists are beginning to understand the complex cortical activity associated with the reduction in physical symptoms, and that once this is understood then a physical treatment can be created. While the associated cortical activity could be complex, the same could be stated for pregnant women whose symptoms are uniformly and successfully treated via medication. While the cortical activity is complex and different among these women, doctors are able to reliably treat all of their pregnancy symptoms based on administering medication that predictably targets associated parts of the body and brain. If these same symptoms are present in patients with pseudocyesis, it is likely that there are similar effects happening in related parts of the body. However, administration of the same medication does not result in its predicted effect and the symptoms persist. For materialists to say that promising medication can be introduced if the cortical activity is better

known is introducing an impractical argument, because this was not the main factor considered in creating most successful medications. Pregnant women with a variety of cortical activities saw similar results in a reduction of symptoms due to medication, but this does not hold true for pseudocyesis. The only plausible treatment remains that change should be induced at the level of belief, which is associated with the necessary cortical and physical state to support a reduction in physical symptoms.

Another objection could propose that there is a connection between conscious properties, like  $C_1$  and  $C_2$ , that is not accounted for in schema 3. To dualists, it would logically make sense that non-physical conscious states interact with other non-physical conscious states, and this interaction would not need to account for the explanatory gap. To this objection, I will explain why it is impossible to clearly establish a connection between conscious states. Let's say a scientist wants to propose an experiment to prove that  $C_1$  influences  $C_2$ . To do this, the scientist would need to point to a scenario where a change in conscious states ( $C_1$  to  $C_2$ ) occurs without a change in physical states (otherwise, one could just say the physical state is what caused the change in conscious state). This would logically and clearly show that there is an interaction between conscious states. The theoretical schema to prove would be as follows:

$$\begin{array}{c} C_1 \rightarrow C_2 \\ \uparrow \\ P_1 \end{array}$$

The scientist would quickly realize that this scenario is impossible based on modern scientific understanding. Every conscious experience has to have a physiological basis, and countless experiments have proven that stimulations to the cerebral cortex result in changes to conscious experiences. Without a functioning cerebral cortex or underlying physical processes, qualia would not exist. Given this scientifically proven premise, it is impossible to reach the conclusion that conscious experiences can change without any change in physical properties because it implies that there exists a conscious experience without a physiological basis. In other words, it implies that there exists a conscious property (in the case of the schema,  $C_2$ ) that is in no way influenced by a physical property ( $P_1$  or any other  $P$ ). This is why, in the case of schema 2 and 3, connecting conscious states would be illogical. Pseudocyesis has never isolated changes between conscious

states that don't involve any change in physical states, so implying that this is the case in the schema would be an extrapolation.

A different objection could accept the idea that consciousness has a functional role but question its implications. If consciousness functions to influence a variety of physical properties, then in relation to pseudocyesis, it should logically follow that anyone with an intense desire to get pregnant could develop pregnancy symptoms. It is not being argued that consciousness directly causes all of the changes seen in developing pseudocyesis. There are many physical properties, like the brain and its associated neurotransmitters, that are also responsible for onset of pregnancy symptoms. This is supported by the schema, which shows that both  $C_1$  and  $P_1$  can affect  $P_2$ , for example. Based on this analysis, it cannot be said that anything with a consciousness and intense desire to get pregnant would see false pregnancy symptoms. A man, for example, would be less likely to see pregnancy symptoms because he has very different physical properties (different hormone balance, lack of uterus, etc.) that may not be sufficient in physically manifesting pregnancy symptoms. However, there is still a chance of this happening: there are many recorded cases of expectant dads who have experienced pregnancy symptoms. Sympathetic pregnancy is a condition that is similar to pseudocyesis, but it occurs in men. Men with pregnant wives often experience similar pregnancy symptoms, such as nausea, food cravings, etc. However, this is different from pseudocyesis because the men know they are not pregnant. There is no known cause, but symptoms have been associated with elevated cortisol, adrenaline, and norepinephrine levels (which are hormones related to stress). Similarly to pseudocyesis, there is no physical cure for this strange condition. However, when the men's partner is not pregnant anymore, physical symptoms naturally go away. It is harder to see a clear distinction as to how conscious awareness is responsible for some of these physical changes rather than other physical processes in the body, but it is certainly an interesting case that could further suggest involvement of conscious awareness on physical symptoms. Even though men rarely experience pregnancy symptoms because of their different physical properties/composition, conscious awareness and its interaction with physical properties could explain why 80-90% of expectant dads suddenly experience these symptoms during the time they are aware of their wife's pregnancy (Hall-Flavin, 2019).

## V. The Hard Problem of Consciousness

This position does not address how consciousness associated with certain cortical states; it rather shows that there is an association present. As a result, it doesn't inherently address the cause of the explanatory gap or answer the hard problem of consciousness. While it doesn't quite answer the hard problem, this position provides more clarity into the metaphysical nature of consciousness, and this clarity can direct us to different metaphysical theories that could be incorporated in the position in order directly address the hard problem and explanatory gap. I will conclude this paper by discussing some of these theories, specifically mysterianism and panpsychism, in more detail.

Mysterianism, a view pioneered by Colin McGinn, argues that the hard problem of consciousness cannot be solved due to the principle of cognitive closure. According to McGinn, this hard problem arises because "we are cut off by our very cognitive constitution from achieving conception of that natural property of the brain (or of consciousness) that accounts for the psychophysical link" (McGinn 395). This is a sort of causal nexus that we cannot comprehend because of the way we develop and form concepts. To McGinn, a mind is cognitively closed to a brain property if its concept-forming procedures cannot extend to understand the property. Through his argument of cognitive closure, McGinn argues that we are cognitively closed to a brain property (P) that naturalistically explains consciousness. According to McGinn, all of our concept-forming procedures, which includes introspection, perception, and inference, are cognitively closed to P. He posits that the scientific explanation of the hard problem of consciousness is found in P, but only a mind that can form concepts without perception or introspection could understand P and solve this problem. This theory does not attribute the explanatory gap to something supernatural; rather, it can only be understood through a scientific explanation that is cognitively closed to us.

Based on this theory, it is likely that the cognitively closed property P explains how consciousness functions to interact with physical properties. As McGinn suggests, a mind that can form concepts without cognitively closed faculties of understanding could solve this problem of interaction, but this is impossible. Mysterianism's unique argument avoids the objections and unanswered questions faced by materialists and nonmaterialists alike. It doesn't reduce consciousness

to physical processes and acknowledges an explanatory gap, but rather than leaving the explanatory gap as an unanswered question, it provides a careful argument to show that the explanatory gap is explained by science that is inaccessible. Accordingly, this theory would suggest that the function of consciousness is unexplainable due to its inaccessibility. While an unsatisfying answer, it is helpful to understand the metaphysical status of consciousness through its complex and possibly incomprehensible nature rather than an easier but inaccurate explanation.

Panpsychism is a meta-theory that is commonly seen as an intermediate between materialism and dualism. According to this theory, consciousness is an inherent and unique property of matter that operates on a continuum. For instance, an electron would have the most simple form of experience while human brains have complex forms of experience. This conclusion is reached by four premises:

1. All living organisms are material systems entirely composed of matter, which is all that exists.
2. Mental states are not physical properties of the organism, nor implied by them. Its association with subjectivity, as opposed to objective reality concerned with physics, makes it irreducible to physical properties.
3. Mental states necessarily exist as properties of an organism because there is no soul, nor can they be properties of nothing at all.
4. There are no emergent properties in complex systems. As a non-emergent property, mental states are derived from properties of the organism's constituents.

From these premises, Nagel proposes that mental states exist within a physical world as a unique property of matter. Further, matter "must have properties that imply the appearance of different mental phenomena when the matter is combined in different ways... this would amount to a kind of mental chemistry" (Nagel 182). Nagel posits that different combinations of matter lead to different mental capacities through a sort of "mental chemistry". This implies a bottom-up approach to human consciousness, which is described as a higher-level mental capacity that arises from lower-level mentality (like mental particles).

The conclusion provides a general explanation of how consciousness arises, and supporters of panpsychism argue that it solves the hard problem of consciousness because it makes consciousness a fundamental feature of reality. Additionally, the mental

chemistry model proposed by panpsychism provides a possible explanation for the function of consciousness. According to the model, consciousness would arise from a variation in combination of mental particles. As levels of consciousness change (according to the continuum of consciousness proposed), there are different mental particle combinations present, which implies that the inherent nature of matter in the system is different. It is up for debate if this change in inherent nature influences the physical nature of matter, but the functional role of consciousness suggests that it does and could fit into this metaphysical theory.

I discuss these metaphysical theories to illustrate the direction that must be taken in pursuing a deeper and clearer understanding of consciousness. Metaphysical theories that 1) can incorporate the functional role of consciousness and 2) seek to fill the explanatory gap/address the hard problem of consciousness are ones that are more sound in metaphysical understanding. There are likely many counterarguments that mysterianism and panpsychism have yet to address, but these theories guide us closer to the metaphysical reality of consciousness.

## VI. Conclusion

The hard problem of consciousness and explanatory gap remain as the main barriers to understanding the true metaphysical nature of consciousness. I discussed the main metaphysical theories of consciousness, materialism and dualism, that address the explanatory gap and hard problem of consciousness. I then provided my objections and commentary on the unanswered questions posed by both of these theories. I propose the phenomenological position that consciousness and its associated cortical states possess a functional role that allows for interaction with physical properties. This position is based on schema derived from studies on pseudocyesis, a mental health disorder whose treatment strategies effectively establish the functional role of consciousness. Based on this finding, I further propose a revised schema that incorporates the function of consciousness and provides a new understanding into the metaphysical nature of consciousness. I addressed multiple anticipated objections to the proposed position and schema. Finally, upon acknowledging that there are still objections and unanswered questions related to my position, I discussed promising theories

that can incorporate this phenomenological position and seek to answer these remaining questions.<sup>1</sup>

## Works Cited

- Azizi, Marzieh, and Forouzan Elyasi. "Biopsychosocial view to pseudocyesis: A narrative review." *International journal of reproductive biomedicine*, vol. 15, no. 9, 2017.
- Cook, Gareth. "Does Consciousness Pervade the Universe?" *Scientific American*, 2020. <https://www.scientificamerican.com/article/does-consciousness-pervade-the-universe/>
- Churchland, Patricia. "Can Neurobiology Teach us Anything About Consciousness?" *Proceedings and Addresses of the American Philosophical Association*, vol. 67, no. 4, 1994, pp. 23-40.
- Churchland, Paul M. "Eliminative Materialism and the Propositional Attitudes." In *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. Oxford: Oxford University Press, 2002.
- Descartes, René. "Meditations on First Philosophy". In *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. Oxford: Oxford University Press, 2002.
- Hall-Flavin, Daniel. "What can you tell me about couvade? Can men really experience sympathetic pregnancy symptoms?" Mayo Clinic, 2019.
- Harvey, Allison G, and Nicole B Gumport. "Evidence-based psychological treatments for mental disorders: modifiable barriers to access and possible solutions." *Behaviour Research and Therapy*, vol. 68, 2015.
- Jackson, Frank. "Epiphenomenal Qualia." In *Philosophy of Mind: Classical and Contemporary*

---

<sup>1</sup> I would like to express my special thanks of gratitude to Dr. Jessica Wahman for helpful comments and thoughts on this paper.

*Readings*. Edited by David Chalmers. Oxford: Oxford University Press, 2002.

Koić, Elvira et al. “Pseudocyesis and Couvade syndrome.” *HRCAK* 2002, pp. 1031-1047.  
<https://hrcak.srce.hr/19614>.

McGinn, Colin. “Can We Solve the Mind-Body Problem?” In *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. Oxford: Oxford University Press, 2002.

Nagel, Thomas. “Panpsychism.” In *Mortal Questions*. Cambridge: Cambridge University Press, 1979.

Place, Ullin. “Is Consciousness a Brain Process?” In *Philosophy of Mind: Classical and Contemporary Readings*. Edited by David Chalmers. Oxford: Oxford University Press, 2002.

Redinbaugh, Michelle J. et al. “Thalamus Modulates Consciousness via Layer-Specific Control of Cortex.” *Neuron* vol. 106, no. 1, 2020.

Reuter, Martin et al. “The biological basis of anger: associations with the gene coding for DARPP-32 (PPP1R1B) and with amygdala volume.” *Behavioural Brain Research*, vol. 202, no. 2, 2009.

## Dualism and Organic Form

CAMPBELL RIDER

THE UNIVERSITY OF MELBOURNE

**Abstract:** This paper explores a tradition in the philosophy of biology critical of physicalism’s ability to explain organic form, and examines the resonances between this tradition and the anti-physicalist dualism of David Chalmers and Frank Jackson. I argue that opponents of physical reductionism in biology — whom I call “teleomechanists” — couch their position in terms suggestive of the same metaphysical commitments defended by Chalmers and Jackson with respect to the nature of consciousness. This similarity invites speculation as to whether their arguments — the celebrated “zombie” argument and “knowledge” argument, respectively — can be adapted for a defense of dualism about biological form. I conclude that although these arguments have limited success when applied to the question of life, the reason for this failure is instructive. It illuminates the difference of philosophical orientation between the teleomechanists and Chalmers and Jackson: the former group begin from a phenomenological position that is difficult to reconcile with the latter’s more metaphysical concerns.

### I. Chalmers and Jackson’s Dismissal of the Life Question

In this paper I compare David Chalmers and Frank Jackson’s “dualist” positions on consciousness with a field of philosophical inquiry into the life sciences I term “teleomechanism.” I adopt this term in order to capture the spirit of a view that, although fully accepting of a naturalistic understanding of life and organism, nonetheless insists that the notion of teleology is required to fully comprehend biological function. Among the thinkers I place in this camp are Immanuel Kant, Hans Jonas, Francisco Varela and Evan Thompson.

The teleomechanists defend a position on biological life that closely resembles David Chalmers’ position on consciousness in his seminal work *The Conscious Mind*. Like Chalmers, they do not believe in a supernatural explanation for consciousness or life, but they nonetheless believe that there is something fundamentally deficient about the physicalist or mechanist model: a deficiency we can describe as a failure of physical entailment to fully explain a phenomenon of which we have direct evidence. The difference is that Chalmers is satisfied that the phenomenon of biological life is quite sufficiently explained by

microphysical description, and it is only consciousness that stands alone as an anomalous phenomenon. Therefore, despite their resemblances, David Chalmers argues that there can be no biological analogue for the anomaly of consciousness:

“Emergent [biological] properties of this sort are not analogous to consciousness. What is interesting about these cases is that the relevant properties are not obvious consequences of low-level laws; but they are still logically supervenient on low-level facts” (Chalmers 129).

Likewise, Frank Jackson rounds off his thesis that all the facts contained in our folk vocabulary are *a priori* deducible from a complete microphysical description of the world with the example of an amoeba:

“It is implausible that there are facts about very simple organisms that cannot be deduced *a priori* from enough information about their physical nature and how they interact with their environments, physically described. The physical story about amoebae and their interactions with their environments is the whole story about amoebae” (Jackson, 83).

However, the point made by teleomechanists like Evan Thompson is that although the materialist account of organic systems is perfectly intelligible and has profound explanatory value, it is, nonetheless, lacking. Thompson claims that this is what is at issue when it comes to the life sciences and the failure of the self-organising features of living organisms to be entailed by a complete physical description of the world. As though replying directly to Chalmers and Jackson, Thompson writes:

“These concepts and the biological accounts in which they figure aren’t derivable, even in principle, from some observer-independent, non-indexical, objective, physico-functional description (according to the physicalist myth of science). As Jonas puts it, no disembodied and purely intellectual mind, like Laplace’s divine mathematician, would be able to comprehend the form of the organism simply from a complete knowledge of the microphysical state of things” (“Life and Mind” 393).

This position can be traced back to Immanuel Kant’s *Critique of Judgment*. There, Kant argues that mechanical explanations of biological phenomena fail to explain the *appearance* of teleological organisation characteristic of organic life. He does not believe this is an extrinsic, divinely-ordained purposiveness. He

describes organic entities as formally possessing “purposiveness without purpose” — a teleological form which necessarily enters into our conceptualisation of life, despite our inability to locate it in the world described by empirical science (Teufel 4).

Kant’s antinomy is, of course, a question we can couch in terms of Jackson’s “location problem”: Kant is doubtful of the extent to which we can locate (and thus reduce) this mystifying quality of living things, the appearance of teleological purposiveness that to us appears inseparable from what it means to be organic, in purely physical terms. In this respect, Kant is doing Jacksonian “serious metaphysics” (Jackson *From Metaphysics to Ethics* 3). He has identified an explanatory gap in materialist reduction, one that is more or less equivalent to the denial of physical entailment quoted above from Thompson.

There is thus a committed intellectual tradition in the history of the philosophy of the life sciences that denies the sufficiency of physicalism to describe biological phenomena. The question, then, is whether Chalmers and Jackson are wrong to discount organic matter as an example of an explanatory gap in materialism; alternatively, we shall see whether recent teleomechanists exaggerate the anti-physicalist implications of their theorisations of life.

## II. The Concept of Life

In this section, I introduce the core tenets of the “teleomechanist” theory with which we will be dealing. Exemplary of this tradition is the philosopher Francisco Varela, who describes his work as a biologist to have led “to the realization that what was missing [in biological understanding] was precisely the fact that life and cognition are actively done by an agent, an autonomous being who does not suffer passive world encounter, but fashions a world of meaning from within” (Weber and Varela, 115). Varela’s subsequent account of biological teleology — the theory of “autopoiesis” — makes the radical move of reintroducing subjectivity into the life sciences. He proposes the emergence of an autonomy — even at the cellular level — whose ontological foundation ensures that the relation between world and organism is mediated by valences according to the ends for which that organism strives. This is not to impute conscious selection of ends to all living beings — it is merely the supposition that life distinguishes itself from matter by its self-generation of a world, in the sense that there are things that matter for it, things “for-which” and “towards-which” it acts.

There are a number of more concrete conditions for life prescribed by the resulting theory of *autopoiesis* (“self-production”). Among them is the requirement that the “subject-pole” of a lifeform be determined by active maintenance of a boundary separating the inside of the organism from its surroundings. The distinctive feature of this organic interiority is that it is defined by “operational closure”: the internal parts of the organism reciprocally influence and determine each other self-referentially — lifeforms ‘bootstrap’ themselves into existence (Weber and Varela 120). Contrast this with a machine, which must be powered, programmed, designed and constructed from without. Autopoietic organisms also undergo continual metabolic exchange of matter with the outside world. Like the ship of Theseus, a living organism is constantly destroying and rebuilding itself through consumption and excretion. What defines the continuity of the living thing despite this dynamism is the stability of the “subject-pole” which retroactively emerges as a constant locus of all the organism’s actions as it strives to maintain itself in the face of its innate biological impermanence:

“organisms are not only self-regulating but built from cells that materially establish themselves... Bluntly stated self-production is already and inevitably a self-affirmation that shows the organism as involved in the fundamental purpose of maintaining its identity” (Weber and Varela 116).

This “striving” from within a self-created world of meaning is where the teleology (purposiveness) comes in. These are *natural* purposes, however, not divine ones: they are *immanent* to the organism and emerge naturally from the functional unity constituted by its life-processes.

## III. Zombie Life

Chalmers’ celebrated “zombie” argument for property dualism about consciousness begins from the position of physical entailment or logical supervenience he shares with Jackson: if physicalism is true, then a minimal physical duplicate of our world will be a *complete* duplicate, containing everything that exists in this world. He argues, however, that we can imagine a minimal duplicate of the world where human organisms are *functionally* equivalent to us, but entirely lack consciousness or qualia — they will not possess a subjective first-person perspective. For the simple fact that these “zombies” are logically conceivable, physicalism fails to logically entail every feature of the

*actual* world (Chalmers 93-9). A physical duplicate of this world might satisfactorily replicate the outward appearance of human consciousness, but fail to actually contain subjective interiority. Consciousness, Chalmers concludes, must be a metaphysical property distinct from physical matter.

Can we postulate a “zombie” world in which biological organisms possess all the physical features they have in this world, and yet fail to *live* in the relevant sense?

Block and Stalnaker arrive almost by accident at a case for “zombie” life. To support their proposal that “it is doubtful that fulfilling any set of functions is conceptually sufficient for life,” they proceed to describe a vehicle such as a van that possesses all the *functional* elements we associate with living creatures: the van moves, it consumes fuel, and creates waste products (Block and Stalnaker 15). To make it analogous to life, all we need to do is add a TV camera, computer screen and self-driving program, as well as “a miniaturized moving van factory” in the rear of the vehicle, which plays the reproductive role (15).

The fact that by functional analysis we would be forced to accept that this van is alive, despite its obvious failure to really meet our folk intuitions about life, accords well with the teleomechanist critique. Without the added notion of natural teleology and its associated concepts, we would not be able to understand what fails to be organic about this van. Firstly, it is not truly autopoietic since it fails to meet the crucial criterion of operational closure: that is, it depends on extrinsic factors to exist and sustain its existence. Someone had to originally forge the metal used to build the van, and had to program the computer that runs it. Next, the van does not exchange materials with its surrounding environment such as to constantly undergo gradual self-replacement of parts, à la ship of Theseus. And finally, it does not mediate a boundary that differentiates its reciprocally determinant interior parts from an external environment.

Block and Stalnaker provide an entry point to something like a zombie argument against physicalist descriptions of life: we can imagine a possible world where counterparts of bacteria, tigers and trees are in fact like the van described above: they have mechanical subunits performing all the ordinary functional roles we associate with lifeforms, such that they are not in fact “living” in the sense understood by teleomechanists; that is to say, they are not autopoietically sustained. Block and Stalnaker gesture to another possible way of conceiving this: an alternative world in

which it is discovered that humans do not in fact digest their own food, instead relying on gut bacteria that digest what we eat and thereby generate byproducts that metabolically perform the same role as food (14-15). What we have here is a possible world where the functional subunits constitutive of living organisms do not contribute to a unified living organism in the way our actual understanding of life prescribes, because the subsystems operate independently from each other and appear to only contribute to the life of the overall organism by sheer chance.

Now, if we look at the notion of life as it is understood by physicalist biologists, and, apparently, theorists like Chalmers and Jackson, then we find that it makes no attempt to explain this self-reflexive, autonomous unity that lifeforms appear to possess. Chalmers is explicitly a functionalist about life: “Once we explain how an organism performs the function of producing another organism, we have explained reproduction, for all it means to reproduce is to perform that function” (44). He would endorse the view that a minimal physical duplicate of this world contains truly living organisms. But it seems conceivable to me that a minimal physical duplicate of this world could be like the scenario described above by Block and Stalnaker — populated by living organisms that fail to possess the characteristics of actual life as our theorists of purposiveness understand it. It seems, then, that Varela and Thompson’s dissatisfaction with biological functionalism and physicalism may lead to a dualism about life: the autopoietic properties of autonomous self-organisation (establishing an interior “world” organised by values relevant to the organism’s self-maintenance) and sustaining a dynamic “subject-pole” are properties over and above the physical properties.

#### IV. The Knowledge Argument

Hans Jonas makes an interesting claim picked up by later thinkers like Varela and Thompson: “life can be known only by life” (Jonas 91). On some level this may appear tautological, but the point is a provocative one: to have a complete picture of what organic life is, notions of self-organisation, purpose, interiority and autonomy are necessary. But these are notions we can only have because of our *own* experience as living beings. The concept of life is something qualitative that cannot be expressed in third-personal vocabulary, because it is an irreducible consequence of having a living perspective.

This phenomenological claim about the necessity of possessing an organic point of view on the

world in order to understand organic life recalls Jackson's celebrated "knowledge argument" concerning Mary, the neuroscientist raised from birth without exposure to the world of colour. Her exhaustive knowledge of brain processes and the neurophysiology of vision does not change the fact that when she encounters colour for the first time she learns something *new*. Her knowledge of the physical facts of experience does not entail knowledge of certain qualitative first-personal aspects of mentality, because they require direct firsthand experience to be properly understood (Chalmers 141-2).

Part of what is compelling about thought experiments involving psychological phenomena is that we can talk about a subject having knowledge and experiences, just not knowledge and experience of a certain kind they have been deprived of. But it is difficult to deprive Mary of the experience of organic life, since she is living herself! The direct evidence of her own status as a living being would undermine the experiment.

Instead, let us borrow from Thompson and run a similar experiment with Laplace's divine mathematician in the Mary-role. This is an artificial intelligence capable of *a priori* deducing complex phenomena from more basic mathematical principles, physical laws and empirical inputs. To make this example analogous to the knowledge argument we will have to construct the experiment so that this (non-living) machine meets what I shall call the 'Mary-condition': it must be sufficiently perceptive so as to be able to interpret the data input and try to anticipate what certain phenomena that it scientifically understands will be *like*; and it must also be sufficiently like Mary that we can construe a moment where the machine *encounters for the first time* a phenomenon it had hitherto only understood in microphysical terms. Therefore, let us imagine this artificial intelligence as a sort of quasi-living brain in a vat, conscious but lacking a body or any life functions. Suppose we programme the machine so that it knows everything about organic life in microphysical terms: we could teach it about the processes of aerobic respiration, DNA replication, natural selection and neurotransmitters. The question is: if we invited the AI to observe an animal in an enclosure, or a bacterium under a microscope, would it 'learn something new'? Chalmers says no:

Laplace's demon... would be able to straightforwardly "read off" all the biological facts, once given all the microphysical facts... Given all that information, it has all the information it needs to

determine which systems are alive, which systems belong to the same species, and so on. As long as it possesses the biological concepts and has a full specification of the microphysical facts, no other information is relevant (35).

(Note Chalmers' point that Laplace's demon, or in our case the Mary-AI, will need to possess the relevant *biological concepts* to translate microphysics into ordinary understanding. The teleomechanist response would be to deny that this would be possible unless the AI were itself living.)

Now we might expect the teleomechanist to say that of course the Mary-AI makes a discovery — it encounters life. It discovers that the entities it had formerly understood as aggregates of biochemical processes are in fact *self-organising* and thus purposive. If the machine experienced what Jonas calls the "ontological surprise" (79) of life, this would surely demonstrate the failure of physical entailment to translate microphysical facts into a "typical macroscopic truth concerning natural phenomena" (Chalmers and Jackson 316). However, there is a catch. If we are committed to the thesis that "life can only be understood by life," then the cohesiveness of the experiment's positive result collapses, since: a) to comprehend the teleological nature of the presented lifeform, the machine would have to already be a living thing, in which case it would not discover something new about life (since it already has direct self-evidence of natural purposes). In this case, we have failed to properly construct the experiment; otherwise b) if Mary-AI was truly not living, then its sense of "ontological surprise" demonstrates that something non-living *can* learn to understand life. In this case, organic form remains a phenomenon insufficiently explained by physicalism (by the failure of the machine to predict the purposiveness of the organism in question) but the claim that "life can only be understood by life" has been falsified. This, however, was the original motivation for denying the entailment from microphysics to biological phenomena. I think it is necessary to conclude that while it is enticing to try and follow through the analogy between Thompson's denial of physical entailment and the knowledge argument, any biologically construed version of the thought experiment is necessarily self-defeating. To understand life, one must already be living, in which case we violate the Mary-condition. And if the being in the Mary-role is *not* living, then the original stipulation that the purposiveness of life is only phenomenally available to living beings has been violated, and the experiment collapses. In the next section,

I wish to argue that the failure of this attempt at a knowledge argument points to a peculiar quality of the idea of natural purposiveness that remains a notable point of contrast between Chalmers and the teleomechanists I have been discussing.

### V. Life Can Be Known Only by Life

According to even the most primitive teleomechanist position defended by Kant, purposiveness is inseparable from our conceptualisation of living phenomena. Someone lacking this concept would, therefore, not be able to discern from the microphysical facts what is living and what is not. Moreover, this concept is not itself derivable from physical phenomena: as Jonas points out, the concept of purposiveness is a concept we have access to exclusively because of our own status as living beings (91). It is a concept derived from qualitatively experiencing our own life and the organic quality of other living things. In this respect, the teleological facet of our concept of life is more than just explanatory, as Chalmers avers.

The purposive features of our concept of life enter into our consciousness through introspection alone and do not depend on any extrinsic physical process. For this reason, the phenomenon of life has a particularly intimate relationship with the immediate experience of being a living thing. This introspection is how we have learnt what it means to be an autonomous, self-organising lifeform. And “what it’s like” to comprehend the natural teleology of a lifeform is just as physically inexplicable as the source of the idea itself, which is generated from within. Now compare what Chalmers has to say about the topic (here he conflates non-materialist theories of life with ‘vitalism’):

“unlike experience, the vital spirit is not something we have independent reason to believe in. Insofar as there was ever any reason to believe in it, it was as an explanatory construct [...] But as an explanatory construct, the vital spirit can be eliminated when we find a better explanation of how the functions are performed.

Conscious experience, by contrast, forces itself on one as an explanandum and cannot be eliminated so easily” (109).

It seems, however, that the purposiveness of a lifeform is in fact an explanandum that does not so easily resist elimination, precisely because it is one of those qualitative aspects of conscious experience that “forces itself” onto our understanding. We *do* have an “independent reason” to believe in natural teleology because

of our own direct experience of being alive. In short, according to the proponents of teleomechanism, to even have a concept of life is to be in a position to know “what it’s like” to be alive — that is, to be alive oneself.

And part of the difficulty we encountered in creating a “knowledge argument” about lifeforms that might serve as a defense of biological dualism is related to this sense in which the qualia of natural purposes are distinct from mere sensory qualia. Mary can be deprived of certain sensory qualia, which she can contemplate but only authentically experience later. But we have such an intimate knowledge of what it means to be alive that only a non-living thing could fulfill the Mary-role, in which case the experiment fails to serve as a meaningful analogue. One way to interpret the failure of the biological knowledge argument is that the perspective of a living organism must be presupposed by any experiment of that kind. Thompson argues that this problem derives from the “ineliminable transcendental character” of the living perspective: “lived experience is always already presupposed by any statement, model, or theory, and the lived body is an a priori invariant of lived experience” (“Life and Mind” 394).

Here, Thompson connects the autopoietic view of organism to a theorisation of consciousness. He advocates for a continuity between life and mind, so that the mental life of a human being expresses the same autopoietic sense-making and self-generation of a phenomenal world instantiated in its most primordial and primitive form by a simple organism. Central to this claim is a rejection of the viability of the “zombie” argument, on the basis that Chalmers’ perspective on the relationship between the physical and mental is too Cartesian; Thompson, by contrast, endorses an *embodied* model of cognition, according to which mentality is always intertwined with a lived body absorbed in the world (*Mind in Life* 230-4). Conversely, the occurrence of any bodily (that is to say, kinaesthetic and proprioceptive) action requires lived subjectivity as its ground. These functions are so essential to the possibility of an organism’s physical functioning and survival that it is inconceivable a “zombie” human would even be able to outwardly reproduce all the behaviours of a really living and conscious organism.

A full treatment of Thompson’s defence of this model of consciousness is beyond the scope of this paper. The important point is that it hints at a crucial difference in orientation between the phenomenological leanings of the teleomechanist view of life and the

more Cartesian commitments of Chalmers and Jackson when it comes to the relationship between the physical and mental. The proponents of teleomechanism I have explored in this paper argue that lifeforms possess an *immanent* purposiveness, that emerges naturally from the distinction between an organism's interior and the exterior to which it is normatively related. So a lifeform's characteristic purposiveness is not a dual property of its physical form. On this interpretation, the biological "zombie" argument I gave earlier would not satisfy the teleomechanists. Thompson would argue that, in fact, lifeforms constructed like the apparently "living" van described in the example above would not even meet the ordinary physicalist criteria for organic functioning: they would not be able to convincingly reproduce even the most basic biological behaviors if they were not autopoietic ("Life and Mind" 396).

This view, I suggest, calls for a reappraisal of the anti-physicalist leanings of the teleomechanists. If their view is not a dualist one, and natural purposes are *immanent* to the lifeform, then this does in fact suggest that the purposiveness of an organism is logically entailed by all the physical facts. A minimal physical duplicate of this world *would* be a world in which lifeforms exhibited autopoiesis. The teleomechanists' gripe with physicalism, then, is perhaps less substantial than Chalmers': the explanatory gap in biology represents a failure to properly interpret the physical facts along phenomenological lines — it does *not* imply the existence of properties over and above the physical ones. This explains why, despite a shared criticism of materialism, Francisco Varela quite explicitly repudiates dualism, associating this with "extra-mechanical entelechy" (Weber and Varela 119). But on this account, it seems as though Thompson has overstated the explanatory gap in the passage I quoted at the beginning of this paper, since the issue is not metaphysical so much as phenomenological. While there is a resonance with Chalmers when it comes to the irreducibility of the phenomenon and concept of life, this pertains to our inability to ever escape the "transcendental" character of the lived perspective.

### Conclusion

In this paper I have discussed the criticisms of physicalist biology defended by Francisco Varela, Hans Jonas, and (in particular) Evan Thompson. My aim has been to explore to what extent their theoretical alternative to biological reductionism can be integrated into the framework advanced by Frank Jackson and David Chalmers. In many respects the autopoietic model of

natural purposes appears dualistic: the purposiveness of life cannot be deduced from an objective, third-personal perspective, nor is it analysable in functional terms. I have considered whether an argument analogous to Chalmers' zombie argument could be constructed, borrowing some promising examples from Block and Stalnaker. The biological version of Jackson's knowledge argument proved more difficult, however, due to the inextricability of knowledge of life from a lived perspective, which must be presupposed in the context of the thought experiment. This difficulty helped clarify the key difference between Chalmers and Thompson and his teleomechanist predecessors: the latter group have a phenomenological orientation to the question of life which leads them to deny that physical facts conceptually entail biological ones. Despite this, they would not in principle doubt that autopoietic life function necessarily supervenes anywhere that the physical structure of an organism is instantiated. Rather, what they dispute is that life can be meaningfully explained outside of the first-personal perspective, a fact which points to the organic foundation of mind.

### Works Cited

- Block, Ned, and Robert Stalnaker. "Conceptual Analysis, Dualism, and the Explanatory Gap." *The Philosophical Review*, vol. 108, no. 1, 1999, pp. 1-46.
- Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.
- Chalmers, David J., and Frank Jackson. "Conceptual Analysis and Reductive Explanation." *The Philosophical Review*, vol. 110, no. 3, 2001, pp. 315-60
- Jackson, Frank. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press, 1998.
- Jonas, Hans. *The Phenomenon of Life: Toward a Philosophical Biology*. Evanston: Northwestern University Press, 2001.

- Teufel, Thomas. “The Impossibility of a ‘Newton of the Blade of Grass’ in Kant’s Teleology.” In *The Life Sciences in Early Modern Philosophy*. Oxford: Oxford University Press, 2014, pp. 48–61.
- Thompson, Evan. “Life and Mind: From Autopoiesis to Neurophenomenology. A Tribute to Francisco Varela.” *Phenomenology and the Cognitive Sciences*, vol. 3, no. 4, Dec. 2004, pp. 381–98.
- . *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge: Belknap Press of Harvard University Press, 2007.
- Weber, Andreas, and Francisco J. Varela. “Life after Kant: Natural Purposes and the Autopoietic Foundations of Biological Individuality.” *Phenomenology and the Cognitive Sciences*, vol. 1, no. 2, 2002, pp. 97–125.

## *A Distinctive Type of Anthropocentrism in the Philosophy of Personhood*

MUSA SALEEM

THE DUKE UNIVERSITY

**Abstract:** Frankfurt's 1971 theory of personhood sought to reserve the label of being a person only for members of those species that displayed a specifically humanlike trait: second-order volitions. His criterion, however, I argue, ends up making it so that any member of any species at all is deemed a person. This is by virtue of a distinctive type of anthropocentrism theories of this kind show: they do not seek to define personhood in a way where it becomes a definitional truth that only humans are persons, but still aim to have a 'humanlike'—but not exclusively human-criterion as their basis. Satisfying this requirement requires defining this 'humanlike' criterion abstractly enough to make its possibility available to other species. However, other species, virtually any at all, can be interpreted to instantiate these abstract properties in each their own way. There can, thus, be no priori-tization of 'humanlike' qualities without a definitional identification of humans with persons: therefore, either we arbitrarily define ourselves to be the only persons, or we define every single member of every species to be a person.

Among the aims of Harry Frankfurt's classic 1971 essay 'Freedom of the Will and the Concept of the Person' is to identify an attribute of human cognition that can function as a criterion for establishing what constitutes personhood. Frankfurt asserts that previous criteria that attempted at such a definition failed to uniquely establish human beings as the only claimants to the title of 'persons', and thus, diluted this 'valuable philosophical term' by allowing species not deserving of it access to it (Frankfurt 1971, p. 5). To remedy this, Frankfurt proposes the notion of 'second-order volitions': human beings are unique in that they are able to reflect on their desires and change them. By this, Frankfurt means not only the ability to change what we desire, but *that* we desire. Human beings can desire to desire in a certain way; I can, for example, want to not want to eat fatty foods so much, or to want to not be so quick to desiring violence in dealing with difficult social situations. Such a 'peculiar character-istic of humans' gives us access to a freedom of will, or at least the possibility of a freedom of will, which Frankfurt

interprets as a freedom from desiring in ways we do so now (p. 6). This ability is what establishes personhood.

What is interesting about Frankfurt's definition is that it sets itself up for it to *potentially* be applicable to other species too. It is not meant to primarily 'distinguish the members of our own species' from those of others, but rather, it is meant to capture:

'those attributes which are the subject of our most humane concern with ourselves and the source of what we regard as most important and most problematical in our lives. What interests us most in the human condition would not interest us less if it were also a feature of the condition of other creatures as well' (p. 6).

That is, the current exclusivity of second-order volitions to human beings is not a necessary one. Another species *can* display such a property; it is simply the fact that these 'animals of various lesser species' that he identifies do not (p. 5). Frankfurt, then, is privileging human beings in his definition of personhood in a very specific way. He is beginning from an attribute displayed only, according to him, by human beings. He is then *abstracting* this attribute from specifically human instances of it by opening up the possibility that other species can show it; it is only a contingent truth that they do not. To perform such an abstraction, Frankfurt must describe his chosen property, second-order volitions, in a way that does not make it, *by definition*, an exclusively human property. By not making exclusivity an element in the definition of this property, Frankfurt allows for the possibility of other species displaying this humanlike—but not exclusively human—property.

I take such an attempt at constructing a definition to be anthropocentric by virtue of its explicitly stated privileging of human experience. However, it is an anthropocentrism of a specific type: it privileges human experience by founding itself on an attribute of it, then makes it so that this attribute is abstract enough for it to be instantiated by other species. I contrast this with what I term to be 'crude anthropocentrism', which I take to mean a privileging of human experience simply by virtue of it being human experience. That is, under crude anthropocentrism, terms such as personhood are defined in such a way that their exclusive instantiation by humans becomes a necessary truth.

In this paper, I will show that Frankfurt's attempt at getting at a criterion for personhood that uniquely identifies something specific about what he terms to be the human condition fails. That is, I argue

that second-order volitions fail to distinctively separate persons considered rightly (i.e. human beings) from what Frankfurt refers to as those not deserving of the label (i.e. ‘animals of various lesser species’ (p. 5)). This is so because of the nature of this kind of anthropocentrism, whose abstract, non-species-specific way of expressing itself allows it to be interpreted in whatever way, and thus, applicable to any species at all. Unless arbitrary restrictions, including those of crude anthropocentrism, are applied, there can, then, be no privileging of human experiences at all.

This paper will be divided into three parts. The first and second focus specifically on Frankfurt’s argument in order to critique it. The third part will allow me to formulate this critique in a way that makes it generalizable to all such arguments.

## I

Frankfurt states his criterion of personhood as follows:

...having second-order volitions, and not having second-order desires generally... [is] essential to being a person (p. 10).

Those who do not fall under this definition are referred to as ‘wantons’—that is, ‘all nonhuman animals that have desires and all very young children’ (p. 11). It is clear then that Frankfurt is making, at the least, an entailment relation between human beings and persons: if *A* is a person, *A* is an adult human being. This holds true insofar as there is no ‘higher’ entity than human beings that is being considered.

Two terms stand out in Frankfurt’s statement: ‘second-order volitions’ and ‘second-order desires’. It is helpful to conceive the former as a subset of the latter, for one can have second-order desires without them being volitions, but such second-order volitions must necessarily be second-order desires. And what are these desires? They contrast with first-order desires. Both kinds, however, are entities expressible in sentences such as ‘*A* wants to *x*’. Though there may be a manifold of meanings associated with the phrase ‘to want’, each use of it can be taken to express a unifying commonality i.e. *A* being in a relation of desire, of whatever kind, towards *x*. What differentiates desires of the first-order from those of the second-order is the referent of *x* in each instance. If the referent of *x* is *not* a desire, then the sentence represents a first-order desire. Likewise, if the referent of *x* is a desire, then the sentence represents a second-order desire.

What this simply means is that first-order desires are those that express wants of activities, while

second-order ones are those that express desires about some desire. If I desire to eat, sleep, read, watch, or dance, for example, I am instantiating a first-order desire. In these cases, I am being led by my want of a thing; I am, however, unable to say anything about how I feel about my want. I can want to smoke cigarettes all day, but I may also be knowledgeable about the health dangers of my continuing to do so. This, in turn, may result in me wanting to quit smoking i.e. in me wanting to *not want* to smoke anymore. The reflection about my first-order desire, then, generates a second-order desire—something which allows me to take a stand on how I feel about my original desire.

The property of instantiating a second-order desire is not sufficient for the determination of personhood, however. One must express a second-order volition: that is, one must express a want for some want to be their *will*. What is a will?

...it is the notion of an effective desire—one that moves (or will or would move) a person all the way to action. (p. 8).

If a will is an effective desire, then a second-order volition is to want one’s want to be their effective desire. Frankfurt gives the example of the ‘unwilling addict’ to clarify what he means (p. 12). Take a person, addicted to a substance, who hates the fact that she is addicted and would like not to be. She, thus, desires to not have a desire to use the substance anymore. However, no matter how hard she tries, she is unable to act on this desire—she, however, *wishes* she could. She wishes that her desire to be free of the desire to take the substance could become her effective desire, and thus, her will. She wishes, therefore, to will differently, even though she is unable at the moment to act out that will. This ability to will differently, that is, to desire to act on a different desire, is what constitutes her free will, and thus, her second-order volition. What matters for second-order volitions, then, is not the ability to act on one’s will—for Frankfurt, it only amounts to a ‘freedom of action’—but to *want* to will differently (p. 15).

Frankfurt asserts that it is only adult human beings that express such a trait. In fact, he takes such a state of affairs to be self-evident. In the next section, I argue that this is not the case by pointing out a counterexample and looking at its implications.

## II

Before we move into the example, we must clarify what exactly it is that we are looking for in it. We are, foremost, looking for an instance of an expression of a

second-order volition by a nonhuman. We know, as stated above, that a second-order volition is the desire that some desire be one's will. But how do we know how an animal wills? A human being, of course, can simply tell us that she is willing so and so. But members of other species do not have this privilege.

Here, one can simply avoid this problem by saying that it is necessarily the case that nonhumans cannot instantiate second-order volitions. But that would be too quick. For the problem here is not about determining whether this instantiation is possible or not, but rather whether it is possible to *test for* the possibility of this instantiation. If such a test is impossible, then the conclusion to derive from it is not that members of other species cannot be persons, but rather that such a judgment of personhood or non-personhood is inapplicable to them. Under such a view, it is necessarily possible, by virtue of the way personhood has been defined, only for humans to be tested for it and deemed persons.

Frankfurt, however, does not want this kind of definition of personhood. He says as follows:

“Our concept of ourselves as persons is not to be understood... as a concept of attributes that are necessarily species-specific. It is conceptually possible that members of novel or even of familiar nonhuman species should be persons” (p. 6).

It is clear that Frankfurt is aiming for a criterion of personhood that makes it so that its applicability or non-applicability is testable in other species. And such a criterion must be so that ‘members of novel species’ can be tested for it: members of species who do not communicate like us i.e. who cannot instantiate language the same way, for example, must also be testable for it. Thus, to make Frankfurt's aims consistent with the theory of personhood he produces, we must locate a criterion that can be tested for any member of any species at all.

Such a criterion must be abstract enough for this aim. For if it were to consist of, for example, concrete terms expressing physical attributes specific to human beings, species not possessing the physical attributes would be simply disqualified. This would, then, be similar to the case of the non-testable criterion: both, via definitions, preclude other species from applicability of the title of ‘persons’. Frankfurt recognizes this and produces notions of second-order desires and volitions, both of which function as abstract psychological attributes rather than as arbitrary physical limitations. The task now, in order to make

Frankfurt's claims consistent, is to interpret both of these attributes in a way where they are testable in all species. Frankfurt was unable to provide a ground for this testability in other species: we will move forward, then, via an example, to search for what would *most reasonably* fit under Frankfurt's definitions of desires and volitions.

Take the case of a common household pet. Your cat enjoys playing with you. That is, your cat has a desire to play with you. Playing can here mean a number of things: receiving belly rubs, fetching toys you throw towards her, and also, unfortunately for you, biting you. Your cat *really* enjoys biting you with its very sharp teeth. You, for obvious reasons, do not like this. So whenever she does bite you during play, you turn away and stop playing in order to discourage her behavior. The cat now faces a dilemma: either bite, and thus, fulfill her desire for biting, but also, simultaneously cause her desire for further play to be unmet, or *curb her desire* for biting it. One may ask: why can't the cat simply stop biting without curbing her desire? This is a valid claim—unless we add the extra premise that the cat *must* curb her desire if it is to enjoy her other desire (for otherwise the discomfort caused by her unmet desire may interfere with her enjoyment). Here, the cat is being forced to confront her desires: she must decide which desire she *desires more*. That is, she must place her desires in something equivalent to the human-specific sentence ‘I desire this desire more than that one’. An evaluation of such a kind fulfills the criterion of what constitutes a second-order desire, for it is a desire about a desire. But is it a second-order volition?

Imagine the case where the cat desires her desire for general play more than the desire to bite. Assume the cat acts on this second-order desire. Although acting on a second-order desire is not a necessary precondition for a volition, one can imagine (and it is not too difficult to imagine) a case where the cat simply wanted this desire to be her will. She may have witnessed her owner being mad at her biting all the time: she, being a loving animal, may then have decided that biting was something she needed to stop doing to endear herself to her owner again. Or biting may have been an instinctual drive: similar to the drug addict who hated her addiction, the cat may have actively been attempting to resist what was so instinctual and immediate to her. In each of these cases, it is reasonable to interpret the cat's behavior as an instantiation of Frankfurt's definition of a second-order volition.

One may anticipate a number of counter-arguments coming here from Frankfurt's camp. Each of these, however, if they are taken to be true, must necessarily make it so that if the member of the other species is to be deprived of personhood, then human beings are also deprived of personhood. I will illustrate *how* this is so in two following claims. A third opposing claim will be discussed in the next section to answer for *why* this is so.

Firstly, one may argue that the cat's curbing of their desire was not a free choice: she was compelled to do so by her owner and the circumstances in which she was placed under. The cat instantiated a second-order desire, but did not instantiate a volition for it did not freely choose to express that second-order desire. That is, the cat would not have willed to reduce her biting had she not been forced to. This, presumably, contrasts with the human being, who, in some way, willed independently of any compulsion directed towards her.

Even if, however, we allow such a modification, it is difficult to determine what would count as being free. For imagine the case of the cigarette smoker again: the cigarette smoker knows that if she continues to fulfil her desire to smoke, she will suffer severe health problems. Less fatal, she will also suffer the negative sanctioning of peers who dislike her smoking. In a way, then, her desire is not free: it is determined by the circumstances *imposed* on her. This is analogous to the pet cat who has circumstances imposed on her by her owner. One may, however, say that the smoker's decision to value the views of her peers and health practitioners is in itself a free choice. Again, however, one may also say that the cat's valuing of her owner's actions is also a free choice. Depending, then, on whether you define second-order volitions in terms of a criterion of 'free choice' or not, you either banish both the cat and the human from the title of personhood, or you end up including both. There is no in-between.

Another option might be to concede to the inclusion of the cat as a person but still maintain that other lower-order animals exist that do not fit this category. Note that this goes against Frankfurt's original definition of 'wantons', of which he included *all* animals and newborn babies. Such a concession, however, simply cannot come to pass, no matter how intuitive it may seem. One could have chosen any animal whatsoever in the place of the cat and posed a scenario where the animal *must* evaluate which desire they desire more. That is, a stimulus of whatever kind may be

posed which forces the animal to decide which desire to fulfil and which one to abandon. We can present a general formula for this, in the form of an experiment, as follows:

A member of Species X is presented a stimulus with effects that are pleasurable. At Phase 1, this member is freely given access to it in order for it to develop an addiction-like property towards its effects. At Phase 2, however, access to the stimulus is given at the price of something necessary for survival—to fulfill the needs of its addiction now means inching closer and closer to death. Here, the member must *evaluate* which desire she desires more—to persist in addiction and die, or to live. And just like the case of the cat whose unmet desire causes her discomfort, an unmet desire of the member caused by its addiction may be debilitating—curbing its desire is a necessity if it wants to live. The member of the species, thus, develops a second-order desire prioritizing one desire over the other. Take the case where it prioritizes its desire to live over the desire of the effects of the stimulus. If the member of the species *wills* this prioritization, that is, actively attempts to fight against its desire for the stimulus in order to live, then it has instantiated a second-order volition.

In this way, the animal undergoing that scenario qualifies itself as a person. We, thus, come to an even stronger conclusion: either all members of all species, including humans, are capable of being persons, or no member of any species, including humans, is a person.

### III

The final opposing claim is less so an argument and more so an appeal to intuition: it simply does not feel right to call, for example, a snake a person. This is, in fact, Frankfurt's motivation behind proposing a theory of personhood that privileges human beings: we have certain qualities, expressible in the 'human condition', that we associate with persons (p. 6). To fail to prioritize them is to 'misappropriate' the idea of personhood (p. 5). Frankfurt felt that his theory could overcome this misappropriation; based on the discussion above, however, it seems that he has failed. But one can still make the argument that there may exist some theory that does do justice to these human attributes, and thus, rightly places them at the center of any criteria of personhood. But just what are these attributes?

A human attribute, at first, can be taken to mean any attribute instantiated by a human being. But a number of different human attributes are also capable

of being instantiated by other species. In fact, an attribute thought to be essential to humans—second-order volitions—is argued here to be one held by all animals. Frankfurt does not want this type of human attribute: he wants something that distinguishes us from other species. What we are after, then, is what *distinctively* makes us human. That is, what makes human beings special compared to members of other species. A human attribute, then, can be taken to mean any attribute uniquely instantiated by human beings.

What is it that is uniquely instantiated by human beings? Here we are at a loss in the sense that we are unable to point to something specific without issue. We can, for example, say something like language—but then our conclusion could simply be refuted by a scientific study that determines that other species, perhaps a number of them in previously unforeseen ways, can use language, in their own ways, too. But though we are unable to get at something specific without issue, we can still point to something that will bring us no issue at all. That is, something which, by definition, is unique to us. Such a thing can only be: whatever we do *in the way we uniquely do it*. A species may be discovered to use language, but it cannot, by definition, be said to use language *in the unique way we do*. And since whatever that is wholly unique to us is something that only we have access to, we know, by sure, that whatever this property concretely is, no other species has access to it.

But if such a property were made to be the criterion for determining personhood, then it would establish complete equivalence between humans and persons. Since no species other than human beings can instantiate a criterion defined to only be expressible by human beings, no species other than humans can bear the title of persons. This, however, goes against the kind of theory of personhood Frankfurt wants: he would like a criterion that is *humanlike*, but not definitionally exclusive to humans. As stated in the last section, Frankfurt does not want a ‘necessarily species-specific’ determination of personhood: he would like for the possibility to remain that it be discoverable in other species. One where personhood is defined exclusively in terms of humans, however, cannot do this. The aim, then, is to have a definition that does justice to this ‘uniquely’ human property, but no longer makes it literally unique: it must be abstract enough so that other species can be tested for it, but ‘unique’ enough that any member of any species at all *cannot* be deemed to have access to it.

Here, before moving forward, it is first of all important to characterize both types of theories as examples of anthropocentrism by virtue of their prioritization of whatever they deem to be human values. We can justify the use of this term via an analogy to another similar term, for example, that of Eurocentrism. A theory that explicitly prioritizes what it deems to be ‘European values’ in setting up criteria of a concept that one seeks to also apply to non-Europeans is being Eurocentric. In the same manner, a concept that is defined by criteria espousing explicitly humanlike values is being anthropocentric. Theories of personhood like that of what Frankfurt wants and the one that definitionally prioritizes humans certainly fit the bill here.

Both, however, prioritize humans in distinct ways. We can use these different ways as determinants of different *kinds* of anthropocentrism. I take definitional prioritization to be ‘crude anthropocentrism’. Here, it is logically impossible for another species to instantiate a concept by virtue of how the concept is defined. I also include another kind of definitional prioritization here: one where the criterion refers to specific physical features that *only* humans (or whatever other chosen species) display. It is similarly impossible for other species to instantiate these properties.

Frankfurt’s anthropocentrism is different in that he does not define personhood in terms of properties exclusively instantiated by humans. Instead he takes something instantiated by humans, then defines it in a way where it is worded *abstractly* enough for other species to show it too. And we have shown that other species do; in fact, we have shown that *all other species* can be taken to show this property. I argue that this is the case for all forms of this kind of anthropocentrism.

Take the case of second-order desires again. Even though the phrase’s precise definition is stated, it consists of abstract terminology that is not directly observable. What I take ‘desiring’ to be, for example, may be very distinct from how someone else may be taking it, *even though both of us may be following the same definition*. We can exemplify with some other criterion for personhood: say that we set up a criterion for a person as ‘whoever that can reflect on their past actions’. The person thinking of this with a uniquely human way of reflection in mind may be satisfied by the thought that this says something specific about the ‘human condition’. Another person, however, can take ‘reflection’ to mean something instantiated by the

smallest forms of life, and *still* be following the definition (one can, for example, take habituation to stimuli by forms of bacteria to count as reflection). If the first person, put off by this, were to now emphasize the uniquely human way of reflection in the definition, they would then be slipping into crude anthropocentrism: a definitional identity of human beings with persons. That is, whenever any ‘distinctively human’ detail is used to expand upon any abstract criterion, it loses its capacity to be applicable to other species.

Why should this be the case? Notice that abstract psychological predicates that we use to describe human beings are equally applicable to all kinds of species: ‘the fox *conceives of* a plan to eat the chick’, ‘the ant *realizes* the mistake it has made’, ‘the bear cub *expresses sadness* at the death of his mother’. No predicate signifies one single process common, in the same way, to each and every species: we may mean something very different when we use it, for example, to describe a human being’s way of expressing sorrow than we may when referring to a bear cub’s way of doing so. But we cannot include this in the criterion *unless* we elaborate on what it is about our way of expressing sorrow that is so distinctively humanlike. But we cannot do this by means of other psychological predicates! For then we would simply be able to re-interpret our new predicate in the same all-species-encompassing way we do so now. The only thing that remains is to either define the specific predicate in a way that necessarily makes it unique to humans (i.e. by simply limiting the domain that the abstraction is applicable to just to humans), or by identifying concrete physical features of behavior that only humans possess. In both ways, it becomes logically impossible for any nonhuman species to hold such a predicate. One, thus, falls into crude anthropocentrism.

It is only, then, crude anthropocentrism that is capable of prioritizing human beings over other species. But this form of prioritization need not tell us anything about the concept which we seek to know more about; it will simply tell us about how we have *defined* the concept, a process which may simply be arbitrary. Any other attempt at prioritizing human beings will fail by virtue of its permissive definition being capable of being interpreted to include any species at all.

### Conclusion

I have argued, first, that Frankfurt’s attempt at characterizing personhood via a criterion reflective of the ‘human condition’ fails because it can be interpreted to

be applicable to any species at all, far removed from whatever Frankfurt takes the human condition to be. I then characterized this conclusion as being a consequence of the theory’s reliance on a specific kind of anthropocentrism that prioritizes human beings *without* definitionally identifying human beings with the concept it seeks to set up criteria for. I then showed that this form of anthropocentrism contrasts with ‘crude anthropocentrism’ which arbitrarily establishes criteria prioritizing human beings. I deemed the latter to be the only way the ‘human condition’, whatever it is taken to mean, can be emphasized in such concepts.

### Works Cited

- Frankfurt, G. H. “Freedom of Will and the Concept of a Person.” *The Journal of Philosophy*, vol. 68, no. 1, 1971, pp. 5-20.



