The Ethics within Metaphysics

by

Michele Odisseas Impagnatiello

DENS Philosophy, École Normale Supérieure, 2018
M.A. Philosophy, Université Paris-Sorbonne, 2017
B.A. Philosophy, Università di Bologna, 2015

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2024

Authored by: Michele Odisseas Impagnatiello
Department of Linguistics and Philosophy
August 30, 2024

Certified by: Caspar Hare
Professor of Philosophy, Thesis Supervisor

Accepted by: Bradford Skow
Laurence S. Rockefeller Professor of Philosophy
Chair of the Committee on Graduate Students

The Ethics within Metaphysics

by

Michele Odisseas Impagnatiello

Submitted to the Department of Linguistics and Philosophy
on August 30, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

This dissertation consists of three chapters at the intersection of ethics and metaphysics.

In the first chapter, I put forward a new theory of personal identity, give arguments for it, and defend it from objections. In the first part, I argue that the two most prominent theories of personal identity, the psychological theory and the physical theory, do not satisfy some constraints on any acceptable theory: that personal identity be all-or-nothing, determinate, principled, and substantive. I then put forward a new theory, the phenomenal theory, on which personal identity is determined by the uninterrupted continuity of a stream of consciousness. I argue that this theory does satisfy all the desiderata, and is as such a better theory. In the second part, I argue that the phenomenal theory also solves the problem of fission cases, because there are no cases of phenomenal fission. In the third and last part, I consider the objection that, on the phenomenal theory, we do not survive interruptions of consciousness such as sleep; I argue that this objection doesn't succeed in refuting the theory.

In the second chapter, I generalize a debate about laws of nature to the domains of metaphysics and ethics. Patterns in the natural world lead us to the postulation of laws. A metaphysical dispute arises as to whether these laws are mere summaries of the mosaic (as the Humean would have it), or whether they govern the mosaic (as the Anti-Humean would have it). In this paper, I first argue that similarly, patterns in the metaphysical and ethical facts should lead us to the postulation of metaphysical and ethical laws, which are the proper subject of metaphysical and ethical inquiry. Then, I argue that the Humean/Anti-Humean debate also arises when it comes to metaphysical and ethical laws. Finally, I argue in favor of the Anti-Humean conception of metaphysical and ethical laws, both adapting standard arguments used in the debates about laws of nature, and with new arguments specific to metaphysics and ethics.

In the third chapter, I investigate conflicts between ethics and metaphysics. Sometimes, a metaphysical theory has revisionary ethical consequences: for example, some have thought that modal realism entails that there are no moral obligations. In these cases, one may be tempted to reject the metaphysical theory on the grounds that it conflicts with commonsensical ethics. This is an ethics-to-metaphysics inference. My claim is that this inference is in general irrational, and that the fact that a metaphysical theory has highly revisionary ethical consequences is no reason at all to reject the theory. I argue for this claim on the basis of general epistemic principles about the transmission of justification, and what makes for a good argument. Furthermore, I argue that my account can explain why a certain narrow class of ethics-to-metaphysics inferences are rational.

Thesis supervisor: Caspar Hare
Title: Professor of Philosophy

Acknowledgments

Writing the acknowledgments for this dissertation is a tricky matter. If what I argue in chapter 1 is right, we all die when we go to sleep, and a new person comes into existence in our body when we wake up. So, I didn't write this dissertation. Who wrote it? No one alive now. In fact, no single person can be singled out as the author; different people wrote different parts of it. Some wrote a few pages, some just a few words. Some merely corrected typos. I, however, wrote nothing but these acknowledgments; and so, I have no one to thank.

Bracketing the question of who I should thank, there is still the question: who should be thanked? Sadly, the thesis defended in chapter 1 bites again. No one currently alive should be thanked. Everyone who helped those past authors of this dissertation, whether through mentorship or friendship, is also dead. This is sad, to be sure; but also very impractical. The list of people to be thanked would contain thousands and thousands of entries; furthermore, I wouldn't know how to individuate the different people that inhabited the same body throughout the years.

Perhaps it's time to put philosophy aside. Let us pretend, for the sake of the acknowledgments, that the thesis defended in chapter 1 is false, and that things are as we naively think they are with respect to facts about personal identity.

So, let me first thank my committee: Caspar Hare, the chair, for his guidance; Jack Spencer for his open door; Roger White for his silences. Most of my philosophical development is due to them. Thank you. I would also like to thank all the faculty members who have endured me in their classes, especially Alex Byrne and Brad Skow who taught me prosem (beyond the aforethanked Caspar and Roger), and Agustín Rayo for our philosophical walks.

With respect to philosophy, this dissertation has also benefited greatly from long conversations with David Balcarras, David Builes, Thomas Byrne, and Jonathan Fiat. However, I should really thank the whole department. Everyone has heard, willingly or unwillingly, my ideas, and has had very useful things to say about them, whether at WIP or MATTI or at a reading group, or just in the lounge or in someone's office. At this point, the MIT Philosophy department feels like my

intellectual home.

With respect to my personal life, my greatest friends during grad school have been the three *B*s: David Balcarras, Allison Balin, and David Builes. My time at MIT is memorable thanks to you. To them, I should add a fourth *B*: Thomas Byrne. Many others have been my friends in different ways: Tyler Brooke-Wilson, Felipe Doria, Kevin Dorst, Bennett Eckert, Jonathan Fiat, Félix-Antoine Gelineau, Jessica Heine, Josh Hunt, Abe Mathew, Josh Pearson, Ryan Ravanpak, Bess Rothman, Pauliina Rumm, Haley Schilling, Abi Thwaites, Eliot Watkins, and Katie Zhou. Thank you all.

I would also like to thank my past mentors: Paul Égré and Uriah Kriegel, my professors in Paris who helped me apply to grad school and led me to MIT; and Paolo Leonardi, my professor in Bologna who first put me in touch with analytic philosophy.

Finally, thanks to my mother, my father, and my brother, for their support throughout the years.

Contents

Chapter 1

The Phenomenal Theory of Personal Identity

1.1 A New Theory of Personal Identity

1.1.1 Introduction: the Problem of Personal Identity

The question of personal identity is the following. Consider person A, who exists at a certain time t_1; and person B, who exists at a later time t_2. Under what conditions is A the same person as B? That is the question. It can be made vivid if considered in the first person: what are the criteria for *my* persistence through time, and *my* continued existence?

Two answers dominate the contemporary debate. On the one hand, there are *psychological* theories, which say that personal identity holds in virtue of some psychological connection between one's earlier and later mental states; on the other hand, there are *physical* theories, which say that personal identity holds in virtue of some physical connection between one's earlier and later bodily state.

More precisely, the psychological theory is stated as follows:

> *Psychological theory of personal identity*: suppose A is a person who exists at t_1, and B is a person who exists at t_2. A is B iff B's temporary mental state at t_2 is psychologically continuous with A's temporary

mental state at t_1.

Psychological continuity between mental states consists of overlapping chains of direct psychological connections. In turn, there is direct psychological connection between A's mental states at t_1 and B's mental states at t_2 if B's mental states at t_2 are appropriately caused by A's mental states at t_1, and are relevantly similar with respect to beliefs, desires, intentions, memories, character traits, and so on. On this view, Joe today is the same person as Joe yesterday in virtue of the fact that Joe today remembers being Joe yesterday, and has the same, or similar, beliefs, desires, and so on; and Joe today is the same person as Joe as a kid in virtue of the fact that Joe today remembers being (and has similar beliefs as, etc.) Joe yesterday, who in turn remembers being (etc.) Joe the day before yesterday, and so on until we reach Joe as a kid. The psychological theory is the most popular view of personal identity.[1] Different versions of the theory place greater emphasis on some mental states (such as memory) rather than others.

The other great theory is the physical theory. It can be stated as follows:

> *Physical theory of personal identity*: suppose A is a person who exists at t_1, and B is a person who exists at t_2. A is B iff the temporary state of B's body at t_2 is physically continuous with the temporary state of A's body at t_1.

The physical theory is really a family of theories, which differ in how they understand "physically continuous", namely on what is the physical thing that determines personal identity. The main candidates are the continuity of the *brain*, or of the entire *body*, or the continuity of the same *organism*. On this view, Joe today is the same person as Joe yesterday (and Joe as a kid) in virtue of the fact that Joe has the same brain (or the same body, or is the same living organism), as Joe yesterday; psychological factors are irrelevant to personal identity. The physical theory is the main rival of the psychological approach.[2]

[1] Defenders of the psychological theory include Lewis (1976), Parfit (1984), Shoemaker (1984). Its contemporary form descends from Locke.

[2] Proponents of physical views (in one form or another) include Williams (1970), van Inwagen (1990), Thomson (1997), Olson (1997).

So these are the two great theories we have. They both fare pretty well, at first glance. They both have some initial plausibility: it seems intuitively plausible that mental factors, and bodily factors, play a role in our identity through time. They both agree with many ordinary personal identity judgments we make about ourselves and others. The psychological and physical theory come apart in certain exotic cases, such as Star Trek-like teleportation, brain transplants, body switching, total psychological wipeout, or the like. So, which theory is the correct one?

1.1.2 Desiderata for a Theory of Personal Identity

Neither. Both those theories are flawed in a fundamental way. They're flawed because they cannot give us *what we want* from a theory of personal identity. To see what I mean, we need to consider two sorts of problematic cases for those theories: the borderline cases of personal identity, and the hard cases of personal identity.

Consider, first, the *borderline cases of personal identity*. The psychological and physical theories analyze personal identity in terms of a relation that comes in degrees. At one extreme, the relation clearly holds; at the other extreme, the relation clearly doesn't hold. In the middle, there are going to be borderline cases. This is analogous to what happens with vague predicates in general. The predicate "tall", for example, clearly applies to people who are above 6 feet, and clearly doesn't apply to people who are below 5 feet. But 5 feet 9 is a borderline case for the predicate.

The psychological theory has borderline cases. It makes use of the notion of direct psychological connection. This consists in the similarity of a sufficient part of one's memories, desires, beliefs, and so on (plus the right kind of causal connection between such mental states). But how much similarity is enough? Say that A and B share the entirety of their memories, et cetera. Then it's clear that there is psychological connectedness between them. Say that A and B share no memories, et cetera. Then it's clear that there is no psychological connectedness. But what if A and B share half of their memories? Or one-quarter, or three-quarters? There are going to be borderline cases somewhere in the middle.

The physical theory too has borderline cases. Take the brain continuity version, for example. How much of the brain is needed for identity to hold? If a certain process keeps the entire brain intact, then clearly that's enough for identity. If, instead, most of the brain is destroyed except for a handful of brain cells, and the rest of the brain reconstructed from a different biological origin, then that's not enough for identity. What if half of the brain is preserved? Or one quarter, or three-quarters? Again, there are going to be borderline cases.

Secondly, there are the *hard cases of personal identity*. A "hard case" is a scenario where a theory of personal identity is apparently incapable of giving a satisfactory response.

Take the psychological theory. The theory requires, for personal identity, a chain of direct psychological connections. Direct psychological connections require similarity of psychology, as well as the later psychology being caused in "the right way" from the earlier psychology. But the theory doesn't tell us much about what counts as similarity of psychology, or about the kind of causal connections there need be between the relevant psychological states. Consider, then, the following cases involving person A:

- A loses the entirety of their memories, but retains the entirety of their beliefs and desires and character traits.

- A completely changes their personality and their desires, while keeping their memories.

- A's entire bodily state, down to the elementary particles, is scanned by a machine. A's body is destroyed, and a perfect duplicate is created using the scan. The duplicate has A's psychology.

- The machine scans A's brain to obtain an extremely accurate description of A's psychology. A's body is destroyed; we take a new body and put A's psychology onto the new body's brain.

- As above, but we don't use a machine to scan A's brain. Rather, we do extensive testing and questioning to get an accurate understanding of A's psychology.

- The machine scans A's psychology. We print down the results in English, in a piece of paper. Years later, someone finds it. They don't understand English perfectly; they translate it to their language, mostly correctly, and put that psychology into a new body.

- We know that A has a perfect psychological duplicate. We destroy A's body; then we scan A's duplicate's psychology instead of A's, and put that psychology into a new body.

In each of these cases, we can ask whether A is identical to the person after the process; that is, we can ask whether A survives those processes. And in all these cases, it's just unclear how to apply the theory to obtain an answer.

A similar situation arises with the physical theory, according to which what matters is the continuity of the body. Consider now the following cases about person A:

- A's body is frozen. After a million years, it's thawed.

- A's body is frozen, then cut into big chunks, which are then reassembled and thawed.

- A's body breaks down into cells and tissues, which are carefully reconstructed by a surgeon to remake the body, putting every cell and piece of tissue exactly where it was.

- As above, but A's body breaks down into elementary particles before reconstruction.

- As above, but qualitatively identical particles are switched in the reconstruction.

- A's body smoothly morphs into a dog, before morphing back to how it was.

- A's body smoothly morphs into a rock, before morphing back to how it was.

- A's body is mostly destroyed, except for a few cells that manage to replicate and reconstitute the body.

As before, we are unable to apply the physical theory to find out what happens to A.

In looking at the hard cases and at the borderline cases, I feel a sense of vertigo. Intuition doesn't help determine an answer. But the theories, as stated, don't help either. The two great theories are formulated in vague terms, and the hard cases are precisely designed to exploit such vagueness. The question is what the two theories could possibly say about those cases; the problem is, I will argue, that anything they could say is deeply unappealing.

The first thing they could say is to claim *indeterminacy*. One might think that in the borderline cases and in the hard cases, there is *no fact of the matter* about whether there is the required psychological or physical continuity; and so, that there is no fact of the matter about personal identity. This doesn't seem too implausible. After all, in the various hard cases, the subject undergoes some very complex psychological and physical manipulations. Maybe the relevant concepts just fail to deliver a verdict in such a case.

A second, similar, answer is the appeal to *degrees*. One can say that personal identity itself comes in degrees, and that what happens in the cases above is that there is a middling degree of personal identity. This fits neatly with the borderline cases, where one could put forward a direct connection between the degree of psychological or physical continuity and the degree of personal identity. For example, if B has 75% of A's psychology, then B and A would be the same person to degree 75%. In more sophisticated ways, degrees could be assigned to the various hard cases, as the strength of the psychological or physical connection increases or wanes.

Indeterminacy and degrees of personal identity, however, are unacceptable. Not merely because identity in general seems to be incapable of any form of vagueness or degrees. The idea that personal identity in particular can be indeterminate, or a matter of degree, seems nonsensical. I can't conceive of myself existing only to a certain degree, or of there being no fact of the matter about whether I will exist. In fact, it's hard to even understand what it means, for a subject, to exist indeterminately or only to a certain degree. Rather, either I will clearly be around, or I will clearly not; it's as simple as that. As Lund writes: "I am unable to imagine being involved in circumstances under which I am neither fully admitted nor fully excluded. The experiences occurring under these circumstances would have to be something for me if I am to be involved in them at all, but the suggestion that they would be something for me even though it is indeterminate as to whether I am having them seems simply unintelligible when one takes the first-person perspective and reflects upon what it is to have experience" (2005, p. 229).

Furthermore, degrees and indeterminacy seem incompatible with the possibility of practical rationality. In looking at a certain prospect, I must consider what will happen to *me* and whether *I* will be around. But if I'm considering a situation containing a person such that there is no fact of the matter about whether that person is me (or in which it's a matter of degree), the question of what to do doesn't make sense anymore. In order to evaluate a situation, I must project myself in it as present or absent; if there is no fact of the matter, I cannot even begin to deliberate. In the words of Williams, "To be told that a future situation is a borderline one for its being myself that is hurt, that it is conceptually undecidable whether it will be me or not, is something which, it seems, I can do nothing with; because, in particular, it seems to have no comprehensible representation in my expectations and the emotions that go with them" (1970, p.174-5).[3]

A third response to the problematic cases is to say that there *is* a sharp answer to each of them. For example, one could say that exactly 73% of one's psychology or brains is what's needed for personal identity; that we survive freezing, and disassembly and reconstruction, only if the pieces are bigger than some threshold. In general, that we survive such-and-such weird psychological or physical manipulations, but not those others. The theories can correspondingly be formulated more precisely, to entail these verdicts.

The problem is that these responses, and the corresponding precisified theories, are intolerably *arbitrary*. Why is 73% the sharp threshold, rather than 72%? Why

[3] Parfit (1984), who defends the psychological criterion, responds by building a time-slice ethics: what really matters isn't personal identity, but just the degree to which a future person is psychologically connected to me; egoistic concerns must then be discounted by that degree.

do we survive disassembly and reconstruction only if the pieces are cut that finely? There is no explanation as to why. After all, the pairs of relations "connected to degree 72%" and "connected to degree 73%", or "connected without cutting into pieces smaller than cells" and "connected without cutting into pieces smaller than subcellular organs" are so similar, that it's hard to see what could possibly explain why one of them is identity-making while the other is not. But it couldn't be a brute fact, either. It can't be that 73% is the magic number that determines personal identity, and that that's all there is to say about it. Nothing seems to "click", in reality, when we reach such threshold, and nothing special seems to happen. But personal identity is something special.

At this point, one might think that the correct diagnosis of the problematic case is a completely different one. In each of those cases, it's really a matter of *stipulation* or *convention* whether A is identical to B. Just as we stipulate that 18 is the age of adulthood, knowing full well that nothing special happens the moment one becomes 18, so we can stipulate that personal identity requires, say, at least 73% of psychological connection, or that it holds across certain complex manipulations and not across others. We need to make a decision, so to speak, which in general is informed by pragmatic considerations. In this sense, the question of whether we "really" are identical to a certain future person is *empty*. Parfit, who held this view, gave the example of the identity of a club through time. Say a certain club dissolves; then, years later, the members decide to reconstitute a club with the same name and rules. Is it the same club or not? Parfit says that we can stipulate either answer, and that the question is empty. Similarly, the identity question in borderline and hard cases is empty.

But this line of thinking is profoundly misguided. It might be plausible in the case of clubs, but it's not in the case of the self. When it comes to the club, once I learn all the subvenient facts (about the members of the club and its rules at each time, etc.), I don't feel like there is any substantive question left. But in the case of myself, when I consider one of the problematic cases above, even if I know all the subvenient facts about people at times and their physical and psychological connections to me now, I still wonder whether I will survive or not. Here I am; I'm wondering what will happen to me. I'm definitely *not* wondering anything about conventions, or about whether I will "count" as the same person or not. I'm happy to have a court make a stipulative ruling about whether the club is the same club or not; but I couldn't care less about what a court stipulates about my identity through time, and I don't see how that could have any relevance to the question I'm interested in.

We've exhausted the possible answers to the problematic cases that the psychological and the physical theories could give. In so doing, we have seen the deep flaw behind those theories. What we've found out is that there are some desiderata on a theory of personal identity. When we look inward, at what we are *qua* subjects, we find that there are some constraints on our nature that any acceptable theory must satisfy. They are the following:

1. *Determinacy*: there is always a fact of the matter about personal identity;

2. *All-or-nothing*: personal identity doesn't come in degrees;

3. *Principledness*: personal identity facts aren't arbitrary;

4. *Substantiveness*: personal identity facts aren't conventional.

These desiderata can be summarized as the need for "metaphysical respectability". They capture the way we think we are, as subjects. The psychological and physical theory, however, cannot satisfy them all; that is the lesson of the borderline and hard cases. Those two theories, then, cannot be right.

1.1.3 The Phenomenal Continuity Theory

But not all is lost. There is a different theory of personal identity that we can put on the table. A theory that has received little attention in the debate, but which has the potential to solve all of the above problems. I'm talking about the *phenomenal theory of personal identity*.

The theory says that personal identity is a matter of a specific sort of mental continuity, namely *phenomenal continuity*. Phenomenal continuity is a relation between conscious experiences. When we consider our conscious experiences, we find that each experience "flows" into the next one, smoothly merging into it, and forming a long stream of consciousness. As William James (1904) aptly wrote: "My experiences [...] pass into mine, and yours pass into yours in a way in which yours and mine never pass into one another". Two experiences are phenomenally continuous if they belong to the same continuous stream of consciousness.

In its simplest form, the phenomenal theory says that phenomenal continuity is the relation that determines personal identity, as follows:

> *Phenomenal theory of personal identity*: suppose A is a person who exists at t_1, and B is a person who exists at t_2. A is B iff B's experience at t_2 is phenomenally continuous with A's experience at t_1.

Equivalently, A is B iff B's experience at t_2 and A's experience at t_1 belong to the same stream of consciousness. The phenomenal theory, then, is a theory that puts large emphasis on the continuity of the *first-person perspective*. The question to ask is: what is going to happen to this stream of consciousness, where is it going to flow?

Let me say a few words to defend the initial plausibility of this theory. There seems to be a clear connection between selfhood and conscious experience, and as such the theory seems to be onto something. And intuitively, phenomenal continuity is sufficient for personal identity. When I look at my phenomenal stream, it seems undeniable that where it goes, I go. As Dainton & Bayne (2005) argued, it seems that phenomenal continuity trumps both biological and psychological continuity. Imagine that right now, as you are reading, without interruptions in consciousness you leave your body behind, float a bit around the air, and then enter a new body. It seems obvious that you survived this process, but your body is not the same one as before.[4] The conclusion is that phenomenal continuity trumps bodily continuity.

Similarly, phenomenal continuity trumps psychological continuity. Imagine that, without any interruption in consciousness, all your memories and desires

[4] Indeed, this thought experiment is a better version of Locke's famous thought experiment, in which we imagine a prince waking up in the body of a pauper and vice-versa. There is room, in that case, for the claim that the prince either died or was brainwashed. But in this case, if we imagine the prince's consciousness continuously "jumping" into the body of the pauper, the verdict is clear that the prince survived and has merely switched bodies.

and character traits are removed; then again it seems you've survived this process, yet you've left all your psychology behind. This is might seem less intuitive; but imagine that you're focusing on a certain particular thing, say a painful experience or the sight of a landscape, and while you focus on that very experience, your memories etc. get erased. Now it seems clear that you were the same thing who began thinking about the pain and who is there now; what happened is that you were brainwashed in the while. Therefore, phenomenal continuity trumps psychological continuity.

This argument can be strengthened following Duncan (2020), who argued that phenomenal continuity gives us *cartesian certainty* of our identity through time. I know that I exist, given that I have some experience; but an experience takes time. I can then know that I exist over the positive duration of this experience, and that I am the same person throughout. We can't doubt, in a cartesian way, that the person who is experiencing this extended pain is one and the same: just consider an extended pain and wonder if it could really be that the person who experienced it a millisecond ago is different from you. What happens is that over short intervals we can directly perceive our own persistence and be certain of it.

While I'm sympathetic to these arguments, they only establish that phenomenal continuity is *sufficient* for personal identity. And these authors do not claim that phenomenal continuity is necessary for personal identity, because that would entail that we do not survive interruptions of consciousness such as sleep. I, however, do claim that phenomenal continuity is necessary for personal identity (I will discuss the issue of sleep later in the paper).

The argument I want to give is a different one. My argument in favor of the phenomenal theory is that it satisfies the desiderata of the last section: it is a theory on which personal identity is determinate, all-or-nothing, principled and substantive. If so, then the theory succeeds where the psychological and physical theories have failed. To give the argument, we first need to formulate the theory more precisely, and give a more precise account of the stream of consciousness. What is a stream of consciousness, and how is it built out of individual experiences? In what follows, I mostly adapt the model developed by Foster (1991) and Dainton (2008).

The first thing to note is that the focus is on our phenomenally conscious experiences: e.g. an individual pain, or an experience of seeing a red wall. There is something it's like to feel pain, and to see a red wall. In this sense, the intrinsic and qualitative aspect of an experience is completely characterized by the way it feels, i.e. by its phenomenal character.

The first notion we need is that of *phenomenal connectedness*. This is the relation that holds between experiences which are experienced together. To illustrate it, consider these two scenarios. In scenario A, there are two people in a room: one is in pain, the other is looking at a red wall. In scenario B, there is just one person in the room: it's in pain while looking at a red wall. Let's suppose that the pain in A is qualitatively identical to the pain in B, and the same for the red experience. The experiential description of scenario A is this: there's a pain, and there's a red experience. What about the experiential description of scenario B? It's still true that there's a pain and a red experience. But that doesn't capture all: there's a difference with respect to scenario A. In scenario B, it seems that the pain and the red experience are somehow "sticking together", while in scenario A they are separate.

In which sense are the two experiences in scenario B "sticking together"? One could say many things: for example, one could say that in scenario B (unlike scenario A), the pain and the red experience belong to the same person; or are realized by the same brain; or are caused by interactions of the world with the same body; or are spatially colocated. But this isn't getting at the core of the sticking. There is a purely *phenomenal* difference between the two scenarios: *what it's like* to be the person in scenario B is different from what it's like to be either person in scenario A. If you were one of the persons involved, you could easily tell whether you're in scenario A or B, without having to look at brains or bodies or anything else, but just by focusing on your phenomenology.

There is something it's like to experience pain and redness together. This can be seen from the fact that when we experience pain and redness together we don't just experience the two individual experiences: there is an additional experience of the difference and contrast between the two, and an experience of the fusion of the two experiences. In sum, in scenario B there is an experiential relation holding between the pain and the red experience: "being experienced together", or *co-consciousness*, or phenomenal connectedness. Phenomenal connectedness typically holds between one's visual, and auditory, and tactile, etc., experiences, which gives them the feeling of unity. I'm not giving an account of what that is or how it works (which is the question of the unity of consciousness[5]); I'm just pointing out that it's a manifest feature of our conscious experience, and something we're all familiar with.

Phenomenal connectedness also holds diachronically. Consider scenario X: a person hears, in quick succession, the musical tones C-D. One might think that the experiential description is simply this: at t_1, C is experienced; at t_2, D is experienced. This, however, doesn't capture what's going on. This same description applies to a different scenario Y: in which one person hears C at t_1, and a second person hears D at t_2. But these two scenarios differ experientially. Imagine for example what it's like to be the person in scenario X. You hear C-D in quick succession. You don't just hear C and then hear D; you also hear C flowing into D. The content of the experience is "C followed by D". In scenario Y, instead, someone hears C, and someone else hears D. In scenario Y there is just a *succession of experiences*; is scenario X there is also an *experience of succession*. There's a purely phenomenal difference between those two scenarios.

In scenario X, C and D are phenomenally connected: you experience them together, albeit in a succession. This is an instance of the general phenomenon of the *specious present*. In our experiences, we aren't merely aware of the momentary, durationless present moment: rather, we can be directly aware of an extended, if brief, temporal interval. This is manifested in our direct awareness of change and persistence. In the example above, we're directly aware of the transition from C to D. But the phenomenon is widespread in our experience. In looking at a moving car, we're not merely experiencing that the car is in position p at time t; we're also directly perceiving the motion of the car (contrast this with a case in which we're shown, separated by one second each, a series of frames showing the car at distinct positions: then we would be aware of the car's motion only indirectly). In experiencing a long pain, we're directly perceiving its duration and change through

[5] See Bayne & Chalmers (2003), and Bayne (2010) for a book length treatment.

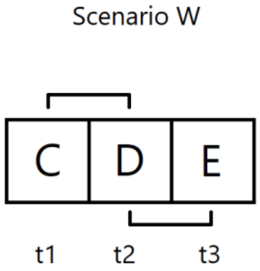
time. This is a phenomenological datum.

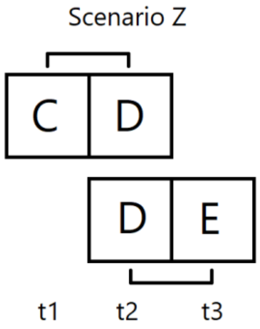
There are several accounts of the specious present. The account I favor says this: that we can experience all at once an extended group of momentary experiences spread out over an interval. As James said, "We do not first feel one end and then feel the other after it, and from the perception of the succession infer an interval of time between, but we seem to feel the interval of time as a whole, with its two ends embedded in it." (1890: 610). If we hear C and D in quick succession, we experience a whole interval containing both C and D, and that D follows C. It's a diachronic case of phenomenal connectedness. This is the extensional account of the specious present.[6]

Diachronic phenomenal connectedness and the specious present is short lived, maybe around a second or so; but as long as it's of any positive duration, that will do for us. In virtue of its being extended, it can be used to construct longer streams of consciousness. Consider scenario W: the experience of hearing the succession of three notes C-D-E. The two experiences C and D are phenomenally connected in a single specious present. D is similarly phenomenally connected to E in a single specious present. There is then an experiential chain from C to E, and we can say that C and E are part of the same continuous stream of consciousness.

Crucially, for the chain to be a genuine one, the specious presents must overlap and share a component experience. To illustrate this requirement, consider scenario Z: a person hears C at t_1 and D at t_2; a different person hears D at t_2 and E at t_3. In this case there is no experiential chain from C to E. There is an experience of C that is connected to an experience of D, and there is an experience of D that is connected to an experience of E. But the two experiences of D are distinct things: they are qualitatively identical, and happen at the same time, but they are two different token experiences. In Z, there are two distinct streams of consciousness. In contrast, in scenario W, C is connected to D, and it is the *very same* token experience D that is connected to E; so there is only one stream. The following experiential diagrams illustrate the difference:

[6] See Dainton (2017) for more on the extensional model, and for rival models.





In the diagrams, adjacent experiences (and adjacent experiences only) are experienced together within one single specious present. In case W, the two specious presents [C-D] and [D-E] overlap, sharing the experience D. When this happens, we say that C and E are phenomenally continuous, and we say that C-D-E form a stream of consciousness. In general, two experiences are *phenomenally continuous* if there is a sequence of phenomenally connected experiences (i.e. of overlapping specious presents) which leads from one to the other. A *stream of consciousness* is a maximal sum of phenomenally connected experiences. This is the "overlap model" of the stream of consciousness, developed and defended by Foster (1991) and Dainton (2008). In our experience, we have long streams of consciousness built in this way. Say we hear a musical scale, then see a red wall, then feel pain, then feel cold (a nice, good life). Then our stream is something like this: (earlier stuff)-C-D-E-F-G-A-B-redness-pain-cold-(later stuff), each element phenomenally connected to its neighbors. This representation is an obvious oversimplification: but it will do for our purposes.[7]

[7] One way in which it's a simplification is that, since I recognize synchronic connectedness (e.g. pain and redness at the same time), the stream involves several experiences at a single time all connected to several others at different times. Another simplification is the usage of discrete time, and well-spaced apart specious presents which overlap in a single experience. It's more likely that specious presents overlap continuously, so that phenomenal connectedness doesn't merely hold between successive experiences, but between all experiences across an interval. Finally, there are open issues about the topology of time and of experiences: if, say, time is continuous and the shortest experiences are truly durationless, then the model would need to be adjusted accordingly.

1.1.4 How to Get What We Want

Now that we have a clearer understanding of the stream of consciousness, I can give my argument that the phenomenal theory satisfies the desiderata. The desiderata are the following:

1. *Determinacy*: there is always a fact of the matter about personal identity;
2. *All-or-nothing*: personal identity doesn't come in degrees;
3. *Principledness*: personal identity facts aren't arbitrary;
4. *Substantiveness*: personal identity facts aren't conventional.

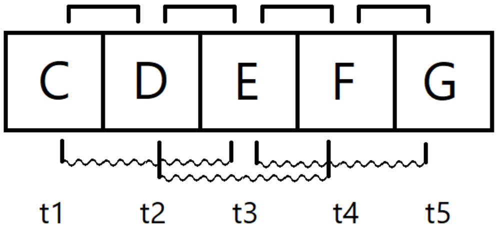
The phenomenal theory analyzes personal identity in terms of phenomenal *continuity*, which in turns consists of overlapping chains of phenomenal connections. I will then argue for the following: that the relation of phenomenal *connectedness* is well behaved so as to satisfy the desiderata; that ensures that phenomenal continuity, and so personal identity, satisfies them as well.

First, *synchronic* phenomenal connectedness satisfies the desiderata. If it didn't, then there would be borderline cases or hard cases of connectedness. Borderline cases of connectedness, however, are impossible. It can't be vague whether two simultaneous experiences are experienced together or not. To see this, consider these two scenarios. In scenario A, there are two simultaneous but disconnected experiences, one of seeing red and one of feeling pain. In scenario B, those two experiences are phenomenally connected. Could there be a sorites series going from one scenario to the other? From the third personal point of view, it would seem so. We can imagine that in scenario A there are two independent brains, and in scenario B just one brain; we can then build a sorites series of cases in which we being with one brain and, through a series of very small particle changes, we end up with two distinct brains. We would judge that somewhere in the middle, it is vague whether the two experiences of red and pain are experienced together.

But try considering the situation from the first person. Imagine you are the person in scenario B, feeling red and pain; then imagine going through the sorites series. I maintain that when we try doing this exercise, we realize there can't be a borderline case of the red and the pain being experienced together. What would what be like? I can certainly imagine an experience of redness together with an experience of pain which becomes fainter and fainter, until it disappears; but this is not what we're required to imagine. This would be a case in which there is non-vague phenomenal connection between red and an ever fainter pain. In the sorites series, we need to keep fixed the phenomenal character of the relevant experiences. What we're required to imagine is an experience of red and an experience of a vivid (and extremely intense, say) pain, but such that it's not clear whether the pain is co-conscious with the red or not. But that cannot be conceived. The phenomenal character of my overall experience is clearly determinately in a certain way. It either includes the intense pain, or it doesn't. I may not be paying attention to such experience (though it would be hard, if it is of intense pain); but it's either there, or not there.

If this is right, then, synchronic phenomenal connectedness doesn't have borderline cases. Indeed, this thought experiment showed how synchronic phenomenal connectedness satisfies the four desiderata. There is always a fact of the matter about whether two simultaneous experiences are phenomenally connected; and it is a relation that doesn't come in degrees. When it comes to principledness, we can notice that once the vagueness problem is done away, there is no room for arbitrariness: phenomenal connectedness is a special relation, metaphysically distinguished, and not a member of a spectrum of similar relations. When two experiences are linked by connectedness, it is a special link in reality. And it is evident that it's also a substantive question, and not one up to stipulation, whether two experiences are so connected: again, just imagine being the person experiencing the red and the intense pain.

Let's now turn to *diachronic* phenomenal connectedness. Diachronic phenomenal connectedness is the experiencing together, within one specious present, of two experiences at different times, e.g. a quick succession of notes C-D. Most of the above discussion carries over, with one proviso. Here, one might think that vagueness does creep in, because it might be vague how long exactly one specious present lasts. If we hear C-D a millisecond apart, they're clearly connected; if C-D are ten seconds apart, they're clearly disconnected. If C-D are, say, 1 second apart, then it's not clear whether they're connected or not, and this looks like a borderline case. A reply is that even if it looks vague, it's not, and our uncertainty about the 1-second case is just an artifact of our limited powers of discrimination. However, I needn't say that. I can agree that, strictly speaking, the relation of diachronic phenomenal connectedness admits of borderline cases. This, however, doesn't mean that phenomenal *continuity* is vague. As long as experiences which are close enough in time (e.g. one millisecond apart) are such that phenomenal connectedness between them is sharp, then there is no problem. To see why, consider the following experiential diagram:



We can assume that it's vague whether C and E are phenomenally *connected*, similarly for the pairs D-F and E-G. But it needn't be vague whether C and G and phenomenally *continuous*. As long as adjacent experiences are phenomenally connected, then C and G are phenomenally continuous. If C and D are determinately connected, and so are the couples D-E, E-F and F-G, then these couples, given their overlap, form a determinate chain from C to G. The determinacy of the shorter interval washes away the indeterminacy of the longer interval.

So all I need is just the claim that there is *some* positive duration such that diachronic phenomenal connectedness, for experiences which are that duration apart, isn't vague. And this, I think, it's clear. Considering two experiences which are extremely close to each other, say one millisecond apart (in the extreme, experiences which are in the most immediate succession), a borderline case of connectedness is inconceivable. Consider a single subject experiencing an atrocious pain, and a millisecond later seeing a red wall; then a subject experiencing an atrocious pain, and, a millisecond later, a different subject seeing a red wall. The difference between these two scenarios can't be made fuzzy, for the same reasons discussed above in the synchronic case.

To make things precise, we can then define *strong diachronic connectedness* as connectedness between very close, or adjacent, experiences, and define psychological continuity in terms of overlapping chains of strong connectedness. Strong connectedness then satisfies our desiderata. Therefore, phenomenal continuity, being built out of chains of strong connectedness, satisfies the desiderata too. And finally, personal identity, which consists in phenomenal continuity, will satisfies the desiderata too.

Could there be different hard cases of phenomenal connectedness? Once the relation is all-or-nothing, and as simple as stated, I don't see any hard cases. The most I can do is imagine *weird* cases. A succession of phenomenal experiences, for instance, consisting of a very fast alternation of extreme pain and extreme pleasure, coupled with random and fast changing visual experiences and sounds, and immediate memory loss and random false memories implanted every second. But if I imagine being such person, as long as I'm phenomenally continuous throughout, there is no obstacle in saying that I survive that; I just have very weird experiences.

And the hard cases discussed earlier now become easy cases. In all those cases, if phenomenal continuity holds, then there is identity; if it doesn't hold, then there is no identity. From the lens of the phenomenal theory, none of these cases will be truly puzzling. If someone is frozen, or cut into pieces and then reassembled, or has their psychology juggled between several transferring machines, we just need to look at whether phenomenal continuity holds; it's an empirical question (the answer for all these cases seems to be "no", though one can imagine them with the continuity preserved). It might be a hard empirical question, and so they might be hard cases epistemically; but they're not hard cases fundamentally.

Indeed, the phenomenal theory makes questions of personal identity pretty easy. Consider reality, and the myriad of experiences it contains. Take any two of them. Are they phenomenally continuous or not? Can you find a chain of connected experiences from the first to the second? The answer will always be either a clear "yes" or a clear "no". Of course, the answer may be hard to know by just looking at physical phenomena, which is the evidence we have when wondering about the persistence of other people. But if you had the experiential map of the world in your hands, determining personal identity facts would be an easy exercise; much reminiscent of the game of connect the dots.

The argument is concluded. To sum up, the phenomenal theory satisfies the desiderata, because its fundamental linking component, phenomenal connectedness, does. This means that the phenomenal theory succeeds in doing what the two standard theories have failed: it provides us with an account of our persistence that satisfies our most deeply held intuitions about ourselves. This, I claim, is a strong reason to prefer the phenomenal theory over the two more popular counterparts.

1.2 Fission: Impossible

1.2.1 Fission Cases

In addition to not satisfying the four desiderata, the psychological and physical theories face another problem: the problem of *fission*. Consider the following scenario:

> *Brain bisection*. Adam is about to undergo a special brain operation. His brain is going to be split in half. Each half is, by itself, sufficient to sustain his entire mental life. Each brain half is then transplanted into a new body. At the end of the procedure, then, there are two people, each of which has half of Adam's brain, and each of which has all of Adam's memories, desires, and so on.

Let's call "Lefty" the person who, at the end of the process, has the left half of the brain; and "Righty" the person who has the right half of the brain. Prima facie, it looks like Lefty and Righty both have a claim to being Adam. This is true both on the psychological and on the physical theory (at least, in the brain version). Lefty and Righty, after all, both have the memories and beliefs, etc., of Adam. In addition, they each have half of Adam's brain, and half of a brain is normally sufficient for survival. It can't, however, be the case that they're both Adam, since one person can't be two people.

Or consider the following case:

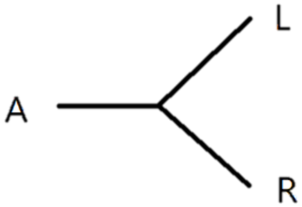
> *Double Teletransportation*. Adam's body is scanned by a machine, and then destroyed. The machine then uses the scan to create two perfect duplicates of Adam, Lefty and Righty, both of which claim to be Adam and remember Adam's life.

Since both Lefty and Righty's psychological states are appropriately connected to those of Adam, they both have a claim to be Adam, at least according to the psychological theory.

These two cases are fission cases. In general, a fission case for a theory of personal identity is a case in which:

- It appears that there is a single person A at time t_1;
- It appears that there are two distinct people L and R at a later time t_2;
- The relation which, according to the theory, makes for personal identity, holds both between A and L, and between A and R.[8]

A fission case can be represented with the following image:



The brain bisection scenario is a fission case for both the psychological and the physical theory (in its brain version). Double teletransportation is a fission case for the psychological theory.[9]

Fission cases are a big problem. The problem is that there's no good thing to say about personal identity facts in those scenarios. The question is: what happens

[8] To avoid counting time travel cases as fission cases, another condition must be added: that the relation doesn't hold between L and R, or that neither L nor R is a later stage of the other.

[9] For a fission case for the physical theory in its organism version, we can imagine a scenario in which someone smoothly divides in two like an amoeba. A bit fanciful, to be sure; but metaphysically possible.

to Adam? And what are the identity facts involving Adam, Lefty, and Righty? There are six possible answers. All of them are unsatisfying. Let's go through them.

1. *Double existence*: $A = L$ and $A = R$. Adam is both Lefty and Righty. This is what we get if we straightforwardly apply the psychological or physical theory as stated above: Adam is, after all, psychologically and physically continuous with both Lefty and Righty. Given the transitivity of identity, this entails that Lefty is Righty; which seems absurd, given that we seem to have two distinct people in front of us, which can play tennis against each other, and which will then go on (we may presume) to live far different lives in distant lands. So this first view isn't very plausible.

2. *Bigger person*: $A = L + R$. Adam is the fusion of Lefty and Righty. On this view, after fission Adam is the object composed of Lefty and of Righty. So Adam has four arms, four legs, two heads, two… et cetera. Adam is some sort of strange creature; and Lefty and Righty are merely parts of a bigger person. This seems absurd as well. It doesn't seem plausible that two perfectly functioning, independent people can compose another person. In addition, there are two distinct spheres of consciousness, two points of view: one for Lefty and one for Righty. What is the point of view of Adam? How could one person have two completely disjoint points of view?

3. *Asymmetry*: $A = L$ and $A \neq R$, or vice versa. Adam is identical to exactly one among Lefty and Righty. The problem of this view is that it's intolerably arbitrary. Given the symmetry of a fission case, why would Adam be identical to Lefty (say), rather than Righty? What could make it the case? And how would we possibly know? Without further explanation, this seems like magic.

4. *Fission is death*: $A \neq L$ and $A \neq R$. Adam is neither Lefty, nor Righty: Adam died. This is the view of Parfit (1984). To get this result, one must supplement the psychological/physical theory with an extra condition: a "no branching" clause, which says that psychological/physical continuity makes for personal identity only if it doesn't take a branching form, that is, if it doesn't hold two-to-one.

This view, too, is quite implausible. If the fission process only led to the existence of Lefty (if only one clone is created; or if only half the brain is implanted into a new body, the other half being discarded), then Adam would survive as Lefty. How is it, then, that when, in addition to Lefty, Righty is around, Adam dies? How could *more* of Adam make him die? How could a double survival result in death? Another objection is that personal identity is, intuitively, an *intrinsic* matter: whether a certain process guarantees survival shouldn't depend on what goes on elsewhere.[10] But this view violates this requirement. For whether Adam survives with one half of the brain depends on whether the other half is merely destroyed or put into a new functioning body.

5. *Colocation*: 'A' is ambiguous. $A_1 = L$ and $A_2 = R$. There is no single person named 'Adam'. On this view, not only there are two people after the fission (Lefty and Righty), but there were two people all along. Before the fission process, Lefty and Righty are coincident people: they occupy the same place and share the same body. After fission, they become separated. Lefty and Righty are two four-dimensional worms that overlap before fission, sharing some temporal parts, much like two conjoined twins can share a torso, or two roads can coincide before bifurcating. This is the view of Lewis (1976).

While this view is theoretically elegant, it has its own costs. For one thing, it requires the acceptance of the theory of temporal parts, which is quite contentious; ideally, that metaphysical issue should be independent of the specifics of personal identity. Furthermore, the idea that before fission there are two people is extremely unintuitive. There is just one body, and just one brain, and just one group of mental states: how could there be two people? And it becomes unbearable in the first person. Am I one person or two people? It seems obvious that I am one person only. But if I will undergo fission tomorrow, then *already now* I am two people, or the word "I" fails to refer. Really, then, I don't know how many people I am, or whether "I" refers to anything. And, to make things worse, if there is no fact of the matter about whether I will undergo fission – suppose it will depend on the outcome of a genuinely chancy coin flip – then there is no fact of the matter as to how many people I am. All of these are very unpalatable consequences.

6. *Indeterminacy*. The last answer is that there is no answer. On this view,

[10] Noonan (2019) calls this the "only-*x*-and-*y* principle", that says whether *x* is identical to *y* can only depend on intrinsic facts about *x* and *y* and the relationships between them.

there is simply *no fact of the matter* about the identity facts involving Adam, Lefty, and Righty. It's not correct to say that Adam is Lefty, and it's not correct to say that Adam is Righty. The propositions "Adam is Lefty" and "Adam is Righty" are neither true nor false. Johnston (1989) has this view. This is a not a very satisfying answer; it's more like giving up than actually providing an answer. But secondly, it's unclear whether we can make sense of it. It entails that, in a fission case, there is no fact of the matter as to whether I will survive. But it's hard to make sense of the idea that it's going to be indeterminate whether I will survive.

This is it. In summary, all the responses to fission are not so good. Fission is the thorn of the side of personal identity; it's the hardest of hard cases. But at this point, a natural and mildly comforting thought is that while this is a problem, it's not a problem for anyone in particular. After all, both the psychological and physical theories have fission cases. One might think that the solution to the problem of fission, whatever that is, is orthogonal to the issue of which criterion of personal identity is the correct one.

I will argue that this is false. The phenomenal theory of personal identity has the resources to solve the problem of fission in a satisfactory way, once and for all. It does because on the phenomenal theory – so I'll argue – there are no fission cases. And so, we have a reason to like that theory better than the psychological and physical theories.

In the next section I give my argument that genuine fission is impossible on the phenomenal theory; in the following section I consider what really happens in cases that look like fission cases, including cases that might seem like fission cases for the phenomenal theory.

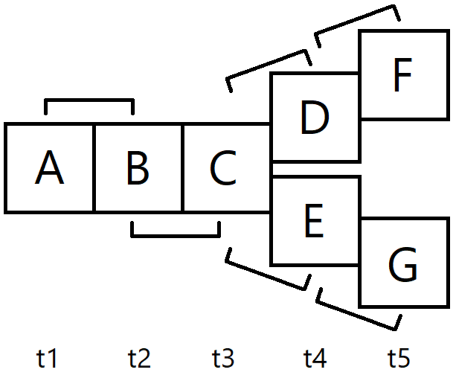
1.2.2 Why Phenomenal Fission Is Impossible

I claim that on the phenomenal theory of personal identity, fission is impossible. In this respect, the phenomenal theory is unlike the physical or the psychological theory. Before going on to argue for this claim, two clarifications are needed about what I mean by it.

Firstly, when I say that fission is impossible, I don't mean that the cases described as "fission" in the previous section are impossible: I'm not claiming that it's impossible to divide a brain in two and put it into two new bodies, or that it's impossible to create two perfect duplicates of somebody. Rather, my claim is that there are no fission cases *for the phenomenal theory*. What this means is that it's impossible for the relation of phenomenal continuity to have a branching structure, or to hold two-to-one.

Secondly, it's in a way easy to make fission impossible *by fiat*, by adding a "non-branching" condition to whatever continuity relation one adopts as the ground of personal identity. Clearly, the relation of being psychologically, physically or phenomenally continuous to only one future person can be had to only one future person. But this is not a genuine sense in which fission is impossible. Rather, this is an implicit commitment to the acceptance of the "fission is death" response to fission, which is (for reasons discussed above) undesirable.

My claim is, then, that phenomenal fission is impossible. The relation of phenomenal continuity, without a non-branching condition, cannot hold in a branching way; in other words, that streams of consciousness can't split. A case of phenomenal fission would be a case in which a continuous stream of consciousness branches off into two, as follows:



I claim that a pattern of experiences like this is metaphysically impossible. I'm swimming against the tide: little has been written about phenomenal fission, but the few who have mostly favor the idea that it is possible.[11]

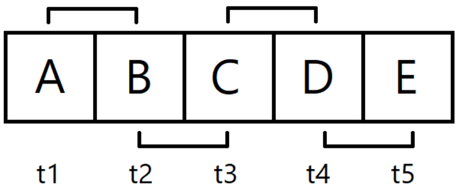
[11] Parfit (1984, p.278), Shoemaker (1984, p.148), Unger (1990, p.191) and Dainton (2008, p.373) say it's possible; Bayne (2010, p. 284) is agnostic; Duncan (2020, p.2048) speculates that it's impossible.

Before giving my argument for why phenomenal fission is impossible, here's a preliminary thought. There's already something suspicious about phenomenal fission. In a certain sense, phenomenal fission is *inconceivable*. One can't imagine what it would be like to smoothly fission into two distinct streams of consciousness. Try it. Imagine experiencing a pure white visual field. And now imagine that your stream of consciousness is going to be fissioned, with one branch experiencing a red visual field and the other branch experiencing a blue visual field. I don't think you could conceive that. What would that be like? I can only imagine the white visual field morphing into a red visual field or a blue visual field, or into a half-red and half-blue visual field. I cannot imagine the white visual field morphing into a fully red and fully blue visual field. This, by itself, might not mean much. Maybe this failure of conceivability is simply due to the fact that one can't conceive being two distinct subjects at once. I'm inclined to think there's more to it than that. Still, it already points to an asymmetry between phenomenal fission and psychological or physical fission. There is no question that fission is possible for the psychological or the physical theory: the cases described above, of brain bisection or double teletransportation, are clearly conceivable and possible. But when it comes to phenomenal fission, it is *not* clearly possible. Whether phenomenal fission is possible or not is an open question.

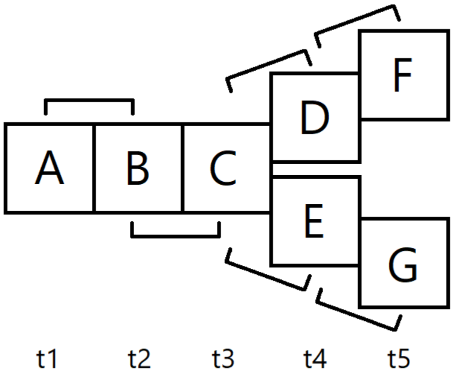
Now, here is my argument that phenomenal fission is impossible:

P1 If phenomenal fission were possible, two distinct subjects could have the *very same* experience;
P2 Distinct subjects can't have the *very same* experience; therefore,
C Phenomenal fission is impossible.

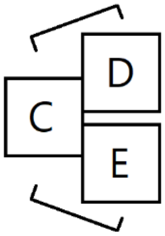
Let's first see why P1 is true. As said earlier, the stream of consciousness is built out of overlapping extended experiences. A standard stream of consciousness looks like this (where A, B, C, etc. are individual experiences):



While in a case of phenomenal fission, we would have a situation that looks like this:



Firstly, we note that after the fission, there are two distinct subjects: Lefty and Righty, each with its own stream of consciousness. Secondly, there will be a specious present which overlaps the fission moment. Here, it's the specious present between times t_3 and t_4. Let's "zoom in" on that interval:



What is going on here? The experience C is part of two specious presents: [C-D] and [C-E]. And those two specious presents belong to two distinct subjects of experience: [C-D] is experienced by Lefty, and [C-E] is experienced by Righty. This means that Lefty and Righty share the experience C: they both experience that same experience C. That is, P1 is true.

I claim that that's impossible: two distinct subjects cannot share an experience. That is P2. It says that if we consider a token experience, for example an individual instance of pain, it can only be experienced by one particular subject. In other words, P2 is the claim that phenomenal experiences cannot be *shared*, and that distinct subjects of experience cannot *overlap*. Unger (1990) calls it the principle of *Privacy of Experience*, and formulates it as follows: “Except for that particular subject himself, nobody else and nothing else can *have* that conscious experience that he has [. . .] nobody else, and nothing else, can be *directly conscious* of the experience of that particular subject” (p.40). Dainton (2011) and Roelofs (2016) call it the *Exclusivity Principle*, and formulate it as follows: “if an experience e_1 belongs to a subject S_1, it belongs ONLY to S_1, so e_1 cannot also (and simultaneously) belong to a distinct subject S_2”.

This principle has been taken to be very plausible. James (1890), for example, writes that “No thought even comes into direct *sight* of a thought in another personal consciousness than its own. Absolute insulation, irreducible pluralism, is the law [. . .] Neither contemporaneity, nor proximity in space, nor similarity of quality and content are able to fuse thoughts together which are sundered by this barrier of belonging to different personal minds.” (p. 226). Dainton (2011) writes “It is difficult to see how your current experiences and mine could possibly be parts of the consciousness of a larger and more encompassing mind” (p. 246).[12]

And indeed, I find it highly intuitive; it’s hard to conceive of it being false. Suppose for example that I’m having an experience of pain. It’s *my* pain. What would it even mean for someone else to feel *this very* pain? I cannot understand that. Of course, I can easily conceive someone else having a qualitatively identical pain, if their body undergoes the same processes as mine. But still, they would have *their* pain and I would have *mine*. I conclude that the principle is very plausible.

So that’s the argument. Phenomenal fission is impossible because it would involve distinct subjects sharing the same experience. Before moving on, it is instructive to see where this argument would fail if applied to other kinds of fission. After all, some things *do* fission. Amoebas, for example, when they reproduce; or bricks when we cut them in half; or roads when they bifurcate. What if we tried to give a parody of this argument to show that, say, an amoeba or a brick or a road cannot fission?

The parody argument would be as follows. If an amoeba could fission, then distinct amoebas would share a part (P1*). But distinct amoebas cannot share parts (P2*); therefore, amoebas can’t fission. This argument breaks down at both

[12] See, however, Roelofs (2016) for a critique of this principle and a defense of a restricted version.

premises. In the original argument, P1 is motivated by the overlap model of the stream of consciousness and the notion of specious present. And it's not very plausible that ordinary things are like a stream of consciousness in that respect. The specious present is peculiar to experiences; arguably, there is no specious present for amoebas or other physical objects. So already P1* in the parody argument is dubious, unlike P1 in my argument. Even if P1* were true, however, P2* is definitely false. P2* says that two distinct amoebas or bricks or roads cannot share parts. Well, this is false. Distinct amoebas, bricks and roads *can* share parts. It just isn't plausible to say that such objects can't overlap. However, it *is* plausible to say that distinct subjects of experience can't overlap. This is the crucial disanalogy between bricks and selves: bricks can overlap, while selves can't. The argument, then, works in the case of streams of consciousness because of those two specific features of them: they're built of overlapping specious presents, and their subjects can't overlap.

1.2.3 Fission Cases, and Split-Brains

Alright. So, fission is impossible. A worry that one might have is that this can't be right, because fission cases like those described earlier are clearly possible. I've already answered this worry: the question isn't whether those cases are possible (they are); the question is whether the phenomenal theory admits of fission cases. And I've argued that it doesn't. So, if the phenomenal theory of personal identity is true, those cases aren't genuine fission cases. Rather, those are merely cases that *look like* fission cases.

O.K.; but what happens to the person in those cases? Well, what happens in those cases is going to be determined by the phenomenal theory of personal identity. In general, all we need to do is look at the experiences involved, and "connect the dots" to make streams of consciousness, and see where each stream goes. It's an empirical question, really; and there is no general answer, because it depends on the details of each case. Let me illustrate.

Consider first the case of double teletransportation. In this case, Adam's body is destroyed, and two perfect duplicates are created *ex nihilo*. Plausibly, Adam's stream of consciousness came to a halt when his body was destroyed, and neither of the two duplicates' streams of consciousness is phenomenally continuous with Adam's stream. The phenomenal theory will then straightforwardly say that Adam died.

Consider now a different fission case: delayed teletransportation. Adam's body is *not* destroyed; but a perfect duplicate, let's call him Schmadam, is created alongside him. In this case, Adam's stream of consciousness keeps on going uninterrupted, while Schmadam's stream comes out of nowhere and is not phenomenally continuous with Adam's stream. The phenomenal theory will say that Adam survives the creation of Schmadam, while Schmadam is not identical to any person who existed before his creation.

What about the case of brain bisection, where Adam's brain is removed from his body, cut in half, and the two halves put into two new bodies? Plausibly, such an operation would involve an interruption of consciousness. As in the case of double teletransportation, Adam dies, and neither of the two resulting people is Adam.

One could build more fission cases, but a similar story will be told. In general, in all cases that look like fission the phenomenal theory delivers a straightforward verdict. One might worry, however, that we could construct cases that *are* problematic for the phenomenal theory. In particular, that we could construct cases where the most natural thing to say is that phenomenal fission occurred. For example, one could take the brain bisection case, and argue that if we perform the operation while the person is conscious, then plausibly we would have one stream of consciousness splitting into two. However, that's not so simple. Brain bisections have never actually happened. I can concede that they are possible; and it is clear, from their possibility, that psychological and physical fission is possible. But we cannot infer from their possibility that phenomenal fission is possible, because we don't know whether it's possible to perform a brain bisection while the person is fully conscious; this is because we don't really know yet the physical bases of consciousness, and how they would interact with the bisection and the subsequent double transplant. And one can't just postulate that in such a case phenomenal fission occurs. Rather, what the objector needs is a real-life case where phenomenal fission seems to occur.

Maybe there *is* a real-life case of phenomenal fission. This would be the case of *split-brain syndrome*. In a split-brain operation, the two hemispheres of the brain are separated by severing the corpus callosum (this is sometimes done as a treatment for severe epilepsy). In some carefully controlled environments, people who undergo this operation exhibit some peculiar behaviors. For example, a patient can be presented with the word “key” in their left visual field, and the word “ring” in their right visual field. The patient will then verbally report only seeing the word “ring”, yet with their left hand they will pick up a key and reject a ring. This might suggest that there are actually *two* conscious agents housed in the brain, one for each hemisphere.[13] A natural thought is then that a split-brain operation *is* a case of phenomenal fission: not only possible, but actual. Right? Not so fast. In order for this to be true, what needs to be the case is that (a) split-brain patients have two streams of consciousness; and (b) severing the corpus callosum smoothly generates those two streams of consciousness from the original one. I argue that both these conditions are dubious.

Firstly, (a) is doubtful. It’s actually not clear at all that a split-brain harbors two distinct streams of consciousness: in a recent review of the empirical literature, de Haan, Corballis, Hillyard *et al.* (2020) write “Clearly, the central question, whether each hemisphere supports an independent conscious agent, is not settled yet” (p. 229). Indeed, there are several ways of understanding the situation as one in which there is only one stream of consciousness. For example, on one view there is one stream of consciousness supported by one hemisphere, while the other one is a non-conscious “zombie” hemisphere. On another view, there is one stream of consciousness which “switches” periodically between the two hemispheres. On a third view, there is one stream of consciousness over both hemispheres; however, a failure of cognitive access to the experiences results in the strange behavioral patterns.[14] If any of these views is the correct one clearly the split-brain syndrome is no evidence of phenomenal fission.

[13] Nagel (1971) popularized the split-brain phenomenon among philosophers. See Schechter (2018) for a book-length treatment.

[14] See Bayne (2010) for the second view, and Bayne & Chalmers (2003) for the third view.

Even conceding that there are two streams of consciousness, and so two subjects of experience, this by itself doesn't establish phenomenal fission, because (b) must be true as well. That two people are housed in a single body is, *per se*, not a problem, and not evidence of phenomenal fission: what we need is that those two streams are both phenomenally continuous with the pre-operation stream. And this is doubtful as well. Firstly, since the split-brain operation is performed under general anesthesia, (b) is false in actual cases. None of the two resulting streams of consciousness is phenomenally continuous with the pre-operation stream. Well, one might then say, what if the operation were done without general anesthesia? Wouldn't (b) be true then? My reply is that we can't really know this *a priori*. Maybe severing the corpus callosum necessarily renders the patient unconscious for a while. Or maybe, if the person is conscious throughout the whole operation, the original stream of consciousness will be continuous with the one supported by just one hemisphere, while the other hemisphere will start its own stream of consciousness anew. Given that there are asymmetries in the brain, this isn't implausible. And so, even if there are two streams of consciousness as a result of split-brain syndrome, it is doubtful that the operation preserved phenomenal continuity between both of them and the original stream.

In conclusion, the split-brain syndrome doesn't constitute good empirical evidence that phenomenal fission is possible. This being said, I grant there *could* be empirical cases which would give grounds to think that. Suppose for example, that we found ourselves in a scenario in which the brain is perfectly symmetrical, and it is split in half perfectly symmetrically, all the while the processes responsible for consciousness and for phenomenal continuity continue uninterrupted during the splitting, and such that after the splitting the two halves are conscious. If we found ourselves in such a situation, we would have good empirical evidence for the claim that phenomenal fission is possible, and therefore to conclude against my arguments. But we definitely *do not* find ourselves in such a situation. The brain is not a perfectly symmetrical organ, and we have not perfectly isolated the neural bases of consciousness and of phenomenal continuity.

This concludes my discussion of fission. If all of this is right, then phenomenal fission is impossible. If the phenomenal theory of personal identity is the correct one, then it follows that fission is impossible.[15] The phenomenal theory then has a distinctive advantage over the rival psychological and physical theories. This is because those rival theories will face fission cases and will have to say something unpalatable or absurd. The phenomenal theory, instead, doesn't face fission cases, and so won't have to say something unpalatable.

This is reason to favor the phenomenal theory over its rivals. One might have thought that the problem of fission, formidable as it is, really cuts across theories of personal identity. If I'm right, that's not true: the problem of fission is really an argument in favor of the phenomenal theory, which is the only theory with the resources to solve it in a satisfactory manner.

1.3 The Problem of Sleep

1.3.1 Sleep and Interruptions of Consciousness

It's time to address what might be the reason why the phenomenal theory is not that popular. The theory says that personal identity is exactly preserved by the continuity of a stream of consciousness. This means that when a stream of consciousness is interrupted, the person stops existing; when a stream of consciousness begins, this is always a new person.

The problem is that interruptions of consciousness are commonplace. When we go to sleep, our stream of consciousness is interrupted, and a new stream begins after waking up. The same happens under general anesthesia, or when someone knocks us out with a bat. It seems that we don't cease to exist after such interruptions; but the theory says that we do. The theory has the very radical consequence that our lives are very short lived: around sixteen hours or so, on average. This seems wrong; it seems that sleep isn't death. So we have a problem; and this is the problem of sleep, even if in its full generality it concerns

[15] The argument of this part can equally well apply to the case of *fusion*. Fusion cases are the opposite of fission cases: they are cases in which (what appears to be) two distinct people "merge" into (what appears to be) one single person. Fusion cases pose the same problem as fission cases. The argument I put forward can also be used to show that phenomenal fusion is impossible: much like a case of fission, a case of fusion would involve two distinct subjects sharing the same experience.

all interruptions in consciousness. My discussion will revolve, for simplicity's sake, around sleep.

This is the major and strongest objection to the Phenomenal Theory. The few others philosophers who are sympathetic to phenomenal continuity are convinced by the objection, and abandon the theory as formulated; they provide a more complex theory to "bridge together" different streams of consciousness as part of the same person.[16] I cannot, however, follow any such strategy. This is because building any such bridge nullifies the advantages of the phenomenal theory, that is, the satisfaction of the desiderata. If, for example, one says that different streams of consciousness are tied by underlying continuity of brain or psychology, then the borderline cases and the hard cases arise again. Similarly, the argument that fission is impossible only works if we assume that the subject is always conscious; while they're unconscious, we can clearly split a brain in half and put the halves into two bodies.

This means that I have to accept that yes, strictly speaking, we don't survive sleep. I need to say something, then, about this awkward consequence of the theory.

It is certainly an unintuitive consequence of a theory that we do not survive sleep.[17] In general, when a philosophical theory contradicts common sense, a standard move is to say that the issue is now that of the theory vs common sense. That the question is whether such violation of common sense is a good price to pay or not for the theory's attractive features. In this case, I could claim that yes, we don't survive sleep. However, we get a theory of the self which satisfies all of our deep desiderata. And that I think the price is right. But other people will think the price is not right. Stalemate and deadlock ensue.

[16] Foster (1991), Dainton & Bayne (2005), Dainton (2008), Bayne (2010), Duncan (2020) are all examples of phenomenal theories which build a bridge between different phenomenal streams.

[17] Although Strawson (1997) has defended the even more radical view that we last for just a few seconds.

1.3.2 Why the Problem of Sleep Is No Problem at All

I think we can do better. I think the objection can be answered properly. My claim is that it is *not* a bad consequence that we don't survive sleep. The problem of sleep *is not actually a problem*. It's not a reason against the phenomenal theory that it entails that sleep is death.

How am I going to argue for this claim? We need to look at why the problem of sleep is supposed to be a problem. Well, simply put, it just doesn't sound right that we don't survive sleep. And I agree: it doesn't sound right. But what is it about it that doesn't sound right? In other words, what are the *reasons* we have to disbelieve it? In yet other words, how would we *justify* the claim that we don't die during sleep?

I see two main families of reasons behind such reticence. There are theoretical reasons, and moral reasons. I shall examine each in turn, and argue that in each case, those reasons either aren't there, or don't actually have any dialectical force. If I succeed, then the objection will have been neatly answered, and the major obstacle removed, and the path definitively cleared for the Phenomenal theory to be, truly, phenomenal.[18]

1.3.3 Reason #1: Theory

If I were to ask you, or any other sensible person for that matter, why it is that we survive sleep, I think I would receive a response like the following: look, it's the same living organism that exists before and after sleep. Or maybe something like: look, when we wake up, we remember being the person going to sleep, have the same beliefs as that person, and so on. In other words, the most natural justification is to point out that there is physical and psychological continuity between the

[18] One should mention that a different answer is possible. Maybe there is no real problem of sleep because, in spite of appearances, we actually are continuously conscious during sleep, with a "dim" level of consciousness of which we aren't aware. Such a view has been held by, among others, Descartes, Spinoza, Leibniz, and Husserl. This isn't, however, a satisfactory answer, for two reasons: firstly, it turns on a highly uncertain empirical claim about our consciousness during sleep; and secondly, even if such empirical claim held, we certainly think we can survive "super sleep", where it is stipulated that during super sleep we *do* have an interruption of the stream of consciousness.

person before and after sleep.

This justification is then theory-driven: it derives from a tacit adherence to a physical, or psychological, theory of personal identity. What this entails, however, is that it cannot be used as an objection against the phenomenal theory.

Why is that? Well, it's weird to say that the phenomenal theory is false because it gives a false verdict in the case of sleep, and then say that that verdict is false because a rival theory says so. What's at stake is precisely what theory of personal identity is the right one. We aren't entitled to assume a rival theory and use that theory to assess a particular case, and then use that case against the theory under consideration: to do so is question begging (To illustrate: I could equally well say that the psychological theory is false because it gives the false verdict in the case of sleep! And justify that by saying that the phenomenal theory says that we die during sleep).

Still, something seems fishy. Inasmuch as we are justified in believing the psychological or the physical theory, we are justified in believing their consequences, including their verdict in the case of sleep; and if this verdict is inconsistent with the phenomenal theory, we thereby gain justification against the phenomenal theory. Isn't this right? So what of the question begging charge of the previous paragraph?

What we need to get clear on is the nature of the justification that we gain against the phenomenal theory, and whether it is or not a direct objection. I have no issue admitting that we *do* have justification in the psychological and the physical theory. They're good theories, and they have good arguments on their side. And this justification *does* give us justification for thinking that we survive sleep. And so, we *do* gain justification against the phenomenal theory. But what kind of justification is it? It's not that there is any problem with the phenomenal theory; it's just that there are other theories for which we have some (maybe a little, or maybe a lot) justification. So the objection isn't "the phenomenal theory says something strange about sleep"; it's just "what about these other theories?". But this objection has nothing to do with sleep. We already knew that there are other theories out there; no extra reason against the phenomenal theory is gained by pointing out that those theories give a different verdict than the phenomenal one in the case of sleep. So we needn't really address the problem; the only reply to this "objection" is just to wait till the end of the day, "when all is said and done", and compare the rival theories as a whole and see which one is better. But sleep isn't the issue.

In conclusion, then, in order for the problem of sleep to be a genuine problem, the objector needs *independent*, not theory-derived justification that we survive sleep. If we did, then the fact that we survive sleep could be treated as a data point, and a reason to favor theories that entail it.

1.3.4 Reason #2: Intuition

Such independent justification doesn't seem hard to find. It's a Moorean fact that we survive sleep. Isn't it pretheoretically obvious that we survive sleep? Well, yes, yes, yes, yes, but say more. Surely, we have an intuition to that effect. An extremely strong intuition, in fact; an intuition that I reckon every human being has. And as we know, intuitions pull a lot of weight in philosophy. Isn't that enough? What am I going to say about that?

(first of all, the intuition must be a "pure" intuition about the case. If the intuition is directed toward a certain theory, like the intuition that we're animals, then this is just liket he case discussed...

I'm *not* going to say that I don't have the intuition, because I do have it. And I'm *not* going to say that sometimes intuitions need to be given up in favor of theory; while that's true, "biting the bullet" would be conceding that the problem of sleep *is* a problem, and I want to argue that it's not. Rather, what I want to say is that if we take a closer look at the intuition, we can see that it's not a good intuition. It's not truth conducive; it can be debunked, explained away. And so it can be safely cast aside.

To see why the intuition isn't a good one, let's first assume that the objector is a physical theorist (the arguments would go through even if they're a psychological theorist). Then, let's consider the following fanciful scenario:

> *The Teleportation World.* The world is very much like our own. The only difference is that every night, when people are asleep, they un-

> dergo “teleportation in place”. A machine scans everyone’s body, records the body’s exact physical state, and destroys it; then it creates a perfect duplicate of the destroyed body, and places it asleep where the previous body was. This has always happened and will always happen to everyone. Everyone is aware of this.

Outside of what happens in people’s beds at night, things are mostly the same. People carry on their lives pretty much like we do. But what happens to them at night? The physical theorist thinks that the people in the Teleportation World die every night. The body, brain and life are destroyed every night, after all. But what would the inhabitants of that world say? I think they would beg to differ, and that they would have a very strong pre-theoretical intuition that they survive sleep, just as strong as our intuition that we survive sleep. After all, that’s the way it’s always been: people’s bodies are destroyed and duplicated at night. A physical theorist would be ridiculed by them, because for them, the notion that we die every night is absurd.

The physical theorist has to say that their intuition is just wrong. Fair enough; but where does their intuition come from? The correct story, I believe, is along the following lines. As William James said, “Each of us, when he awakens says, Here’s the same old self again, just as he says, Here’s the same old bed, the same old room, the same old world” (James 1890/1950: vol I: 317). The people in the Teleportation World self-identify themselves every morning, and build a person awakening after awakening: their memories, beliefs and desires form a coherent chain; they have an established practice of person-construction.

This is what they do every day; and it is this practice that gives them the intuition that they survive sleep. But, hold up… that’s also what *we* are doing. The source of *our* pre-theoretical intuition that we survive sleep is the same. It is a certain practice of self-identification and person-building that, as the Teleportation World shows, is independent of our actual survival. It should be clear then that the intuition isn’t truth-tracking, it doesn’t have any weight, and that it can be discounted. Pretheoretically, then, it is an entirely open question whether we survive sleep or not. The phenomenal theorist then can provide an error theory to explain what went wrong.[19]

The Teleportation World case allows us to attack the intuition by showing it to be, in a sense, too strong, and to be present in cases where, by the lights of the very opponents of the phenomenal theory, we don't survive sleep. But the intuition can also be attacked from the opposite angle. Consider the following case:

> *The world that never sleeps.* The world is very much like our own. However, no one ever goes to sleep (or faints, or undergoes general anesthesia). People enjoy an uninterrupted stream of consciousness from biological birth to death.

What would the people in this world think of sleep, and interruptions of consciousness? I think they would pretheoretically think of them as death. After all, they always had a continuous stream of consciousness, and existence always was for them conscious existence. We can imagine giving these people a "preview" of what interruptions of consciousness are like. We can gradually put them into a state of dim consciousness, making their experiences less and less intense, as in right before sleep, right before the lights go off, and then bring them back to full consciousness. They would be scared; they would be afraid of the gradual dimming in consciousness; they may well think that if consciousness went all out, it would be death. If we told them not to worry, that a new stream of consciousness will begin in the same body, with the same memories and all, they would think of it in the way a physical theorist thinks of teleportation: namely as death, followed by cloning.

[19] One might worry that the Teleportation World case is only effective against the survival intuition of the physical theorist, because a psychological theorist would be happy to say that the people in the Teleportation World *do* survive sleep. True; but for them, a different scenario can be put forward. Consider *The Coincidence World*: The world is very much like our own. The only difference is that whenever someone goes to sleep, their body is annihilated by some process; Then, without any reason whatsoever, by sheer coincidence, a duplicate body appears in the bed where the previous body was before being annihilated. There is no causal relation whatsoever between the destruction of one's body and the appearance of a duplicate. It's just a cosmic coincidence, which has happened every night to every person since the dawn of time. Everyone is aware of this. The inhabitants of the Coincidence World die every night, even by the psychological theorist's lights. All my arguments can be put in terms of this scenario.

This, I think, would be their intuition.[20] What this shows is that the intuition that we survive sleep is also pretty weak, and could easily not have been there; indeed, in this scenario people's intuitions accord with the phenomenal theory of personal identity. This is yet more evidence that our intuitions about whether we can survive interruptions of consciousness aren't really connected to the truth, but are the product of habit and mostly reflect the way things have always been in a certain community.

All in all, then, our intuition that we survive sleep is just the product of habit; and from the from the point of view of a physical or psychological theorist, it is both too strong and too weak. Too strong because it would easily be present in cases where (by my opponent's lights) we do die every night; too weak because it would easily not be present in cases where (by my opponent's lights) we don't die every night. And so, much as it seems obvious that we survive sleep, this doesn't actually provide us with justification that we do.

1.3.5 Reason #3: Semantics

Intuitions aside, there's other things that the commonsensical chap could say. They could argue that there is something deeply confused about the idea that we don't survive sleep. The idea is that it is *semantically* or *conceptually necessary*, or maybe *analytic*, that we survive sleep. On this view, it *doesn't make sense* to say that we don't survive sleep. If one says so, they are deeply mistaken – conceptually confused, more than wrong. The basic idea is that the meaning of the word "person", or maybe the nature of the concept "person", automatically guarantees that we survive sleep. It just *can't turn out* that we don't. It's as if one's theory of bachelorhood entailed that some bachelor are married.

One way to argue for this claim is to say that meaning is determined by use, and that we clearly use the word "person" to refer to extended entities spanning across interruptions of consciousness and who live for a long time. Another way is

[20] This isn't to say that their intuition would be unshakable. If things started to change and people started going to sleep and waking up, and claiming to be the same person they were before, then after a while people would arguably stop thinking of sleep as death, and they would start viewing it as we do. And who knows, maybe that's what would happen to a population of physical theorists if their world gradually evolved into the Teleportation World.

to say that there is a principle of charity in semantics, such that we couldn't be *that* wrong about the reference of "person" (but we would be that wrong if we didn't survive sleep). Yet another way is to say that the meaning of "person" is given by some more-or-less complicated theoretical role, part of which are commonsensical facts including the truisms that we can survive sleep and we can live for many years.

Such a line of thinking is certainly attractive at first glance. It rests, however, on a mistake of semantics — or rather, or metasemantics.

To see why it fails, let's borrow the Teleportation World from the previous section. This is a world, remember, in which everyone's body is destroyed at night, and a perfect duplicate body created in its place. Now, let's suppose that we discover that *our* world is the Teleportation World. For example, we're shown tapes showing people's bodies blowing up every night, and being replaced with a duplicate. Now, what should we think? In particular, what should a physical theorist think? It seems that a physical theorist should simply think that we just discovered something pretty surprising (and disturbing), namely that people die when they go to sleep, and that no one has ever lived for more than a day. But the semantic argument deployed above would say something different. After all, the meaning of "person" guarantees that we survive sleep and that we live for a long time. So really, if we follow the semantic argument we should conclude that the physical theory can't be right; and that holding onto the physical theory after we discover that we are in the Teleportation World would be a semantic and conceptual confusion.

But this verdict can't be right. While I don't like the physical theory of personal identity, it cannot be falsified by tapes showing us that we blow up every night. It is clear that the physical theorist should *not* give up his theory upon discovering that we're in the Teleportation World. Rather, they should hold on to it and say that we're radically wrong about our survival and about how long we've been around. What this means, then, is that the concept of "person" doesn't guarantee that we live long lives and survive sleep; that it's a perfectly legitimate possibility that we don't; and indeed, it could turn out that we don't. Who knows, maybe tomorrow we'll find such tapes. Not that I hope so. But at any rate, we got an argument against the semantic strategy.

The semantic strategy can be refined, however, to counter this argument. If we learn that we're in the Teleportation World then we learn something quite unexpected about some underlying facts about the world. Namely, we learn that people's bodies blow up at night, while we all believed that they didn't. The semantic argument can be patched by saying that what's semantically/conceptually necessary/analytic isn't that we survive sleep *no matter what*; but that, *if the underlying facts are such-and-such*, then we survive sleep. In other words, the idea would be that our pattern of usage of "person", or the nature of the concept "person", is such that it applies to extended creatures that survive sleep and are long lived, as long as certain conditions are satisfied. These conditions include that people don't blow up overnight, et cetera. This patching does avoid the argument above, because those conditions are satisfied in the actual world (I presume), but would not be satisfied if we learned we were in the Teleportation World.

The patched version of the semantic strategy runs into a different problem, however. Let's look again at the Teleportation World, this time not from the perspective of us discovering that we live in that world, but from the perspective of the people who have always known that they blow up at night (and, we can suppose, who use teleportation all the time to travel fast). They don't seem to care. The physical theorist thinks that they're wrong, and that they die every night. But the patched semantic strategy disagrees. After all, *they* use the word "person" to refer to extended entities, and they are aware of all the underlying facts, so they couldn't be that wrong about their survival. So, people survive being blown up at night. So, the semantic strategy entails that we can disprove the physical theory just by pointing to the linguistic behavior of the people in this other world. Which clearly can't be right.

The only reply available to the semantic strategist is to say that the people in the Teleportation World actually have a different concept of person than we do. When we say "person", we mean roughly "long living things, conditional on underlying facts such as not blowing up every night". When they say "person", they mean roughly "long living things, conditional on underlying facts such as possibly blowing up every night". On this view, we and them are just talking about different things. The problem is that this view deflates the question of personal identity. Do we survive sleep? According to our concept, we do; according to their concept, we don't; and that's the whole story. If we were to go to the Teleportation World and argue with them about survival, we would be talking past each other. But intuitively, we would *not* be talking past each other. Intuitively, we are arguing about something substantial and important. We would feel bad for them, because they are so misguided. We would say that the people in the Teleportation World have the same concept of person; it's just that they have radically views about the survival of people.

This is because the concept "person" is special. It picks out something *natural*, and stable across different possible worlds, and this is one of the reasons it's philosophically interesting. It has a privileged meaning that both us and the people in the Teleportation World share.

This shows that the meaning of "person" doesn't entail being long lived (whether or not conditionally on some underlying physical facts). What is, then, the meaning of "person", and the nature of the concept? The main desideratum is that it be a concept that is in common between us and the people in the Teleportation World. I don't want to be committed to any specific answer, but I'll tentatively put forward two possibilities.

A first hypothesis is that the meaning of "person" is determined by direct reference. Each of us can point inwards, to themselves, and think "I am a person". We are the paradigmatic example of person to ourselves: us, the subject. Then we generalize to other people by saying that in general, people are things like ourselves, like this subject here that we have before our mind. A second hypothesis is that the meaning of "person" is determined by the theoretical role it plays. For reasons given above, it can't be a theoretical role involving surviving sleep. It could play, instead, the theoretical role given to it by ethics when it comes, for example, to praising and blaming and to assigning moral responsibility for one's actions. Either of these accounts can work.

Either hypothesis can explain what's going on. According to the phenomenal theory, we are extremely wrong about the survival of people over time. What's going on can be illustrated using a classic example from Putnam. Consider cats.

What if some crackpot came along and said that cats are actually alien robots? A wild claim, of course. We might be tempted to dismiss it on conceptual grounds, by saying that it's analytically true that cats are animals. But on further reflection, we can't do that: it could turn out that cats are actually alien robots. This would be impossible if part of the meaning of "cat" was that cats are animals. Rather, the meaning of "cat" maybe comes from direct reference – an act of pointing to some furry objects; this leaves open whether those furry objects are animals or alien robots. Similarly, pointing to ourselves leaves open whether we survive sleep. Or maybe it comes from some theoretical role – perhaps "cat" means "those furry things that we keep as pets, etc."; again, this leaves open whether they are animals or alien robots. Similarly, the theoretical role of person can leave open whether we survive sleep.

It's easy to mistakenly think that the meaning of "person" guarantees that we survive sleep. This is because we are so sure that people survive sleep that we might think it's part of the concept. Similarly, we might be tempted to think that the meaning of "cat" guarantees that cats are animals. But once we recognize the semantic nature of the word "person", the semantic strategy becomes unappealing; and the fact that we are extremely wrong about our survival is not puzzling or mysterious anymore.

1.3.6 Reason #4: Tragedy

The claim that we die every night is not just strange and unintuitive in the way many philosophical claims are. It's also deeply disturbing. Dying is bad, after all. The realization that we die every night is the realization of a complete disaster, a catastrophe; it's *very sad*. It's sad on a selfish level, because this is my last day on Earth; and to make it worse, I'm spending it on a philosophy paper. It's also sad on a moral level: everyone is dying all the time! Every day, billions of people die. And, what's worse, everyone is deluded into thinking that they're not.

This is truly a nightmare on all fronts. And so, one might think, it just can't be right.

What should I say? Maybe one could push back against the claim that it's a total tragedy. There's always a silver lining. I'd promised my friend I'd mow his lawn this weekend. But I'll be gone by then, so I won't have to mow it after all. And, no more weddings and lines at the DMV.

I'm joking, of course. Overall, the situation is extremely tragic. But then, what do you want me to say? I sympathize; I really do. But as terrible as the truth might be, this is no reason to reject it. That a theory has tragic consequences is reason to *wish* it not be true; but it's no reason to *believe* it isn't. To do otherwise is wishful thinking, a major epistemological sin.

1.3.7 Reason #5: Ethics

The normative strategy of the previous section can be improved on. One can argue that the claim that sleep is death doesn't just have unsavory ethical consequences; but that it has *false* ethical consequences.

To see how, we start from the fact that personal identity plays an important ethical role. For example, people ought to be blamed for things *they* did at earlier times; or we may have special reasons to care about what will happen to *us*, as opposed to other people. As Locke said, "Personal identity is a forensic notion"; and many have followed him.

If sleep is death, then we are quite mistaken about facts about personal identity; and correspondingly, we are quite mistaken about facts about who should be blamed for what, or what we should care about. For example, suppose that Joe is currently on trial for brutally murdering several innocents a few days ago. Naturally, we think Joe ought to be punished. But if sleep is death, then Joe didn't actually commit any crime. The real murderer went to sleep and thus died; Joe came into existence this morning upon waking up. Hence, Joe is innocent and should not be blamed or punished.

But, one might think, It's obvious that Joe should be punished. That's just a truism; and if a theory entails that Joe should not be punished, then so much the worse for the theory. One, then, can give the following argument:

P1 Joe ought to be punished;

P2 If we don't survive sleep, Joe ought not to be punished; therefore, C We survive sleep.

Now, this looks like a proper argument. It's valid; furthermore, the premises all seem very plausible. Then, it seems like we have good reason to believe that we survive sleep. Many such arguments can be constructed, involving any of the various theoretical roles that personal identity plays (for example: if sleep is death, we don't have to keep the promises made yesterday. But clearly, we should keep those promises; therefore, sleep isn't death). But I'll focus on this argument, as the diagnosis is general.

Yet, this argument is not a good argument. Even if it's valid, it cannot be used to argue for its conclusion. To see why that is, we can borrow yet again the Teleportation World, where everyone blows up at night and a clone is recreated on the spot. Suppose we are physical theorists, and a friend of ours puts forward the possibility that, unbeknownst to us, we are in the Teleportation World. If that were true, then again Joe ought not to be punished. After all, if we all blew up every night, then so did the murderer; Joe is not the same person as the murderer, he came into existence this morning, and is thus innocent.

But clearly, Joe ought to be punished. So we could give the following argument:

P1 Joe ought to be punished;
P2 If we're in the Teleportation World, Joe ought not to be punished; therefore,
C We aren't in the Teleportation World.

This argument, just like the previous one, is valid, and the premises are very plausible. Yet, there's clearly something fishy about using this argument to conclude that we're not in the Teleportation World. Can we really rule out that we blow up every night on the basis that Joe should be punished? It seems like we can't.

This argument exhibits *transmission failure*. Even if it's valid, it fails to transmit justification from the premises to the conclusion. That is, we cannot believe that C on grounds of P1 and P2. Why is that? The issue of transmission failure has received a lot of attention in epistemology, and full theory of transmission failure is outside the scope of this paper.[21] But some kind of circularity seems to be the

[21] Notably, with respect to whether Moore's proof of an external world is an example of trans-

culprit here. We believe that Joe ought to be punished *because* we believe that Joe is the one who committed the crime. The fact that Joe didn't materialize out of nowhere this morning (and so, that we're not in the Teleportation World) is one of the things that determines that he's the one who committed the crime; so, it is epistemically upstream from the fact that he should be punished.

In general, all the facts that determine whether he is the one who committed the crime are epistemically upstream from whether he should be punished. There are many scenarios in which he didn't actually commit the crime. For example, if the police framed him, or if an identical twin committed the crime in his place. Similarly, we cannot argue that he wasn't framed, or that no twin committed the crime in his place on the basis of the fact that he should be punished.

The claim that we survive sleep is just like that: one of the things that feed into the claim that Joe is indeed the one who committed the crime. So, is it similarly epistemically upstream from it. So, we cannot reject it on the basis that Joe ought to be punished. To do so is get things backwards.

You might think that something strange is going on here. Isn't it very plausible, even obvious, that Joe ought to be punished? Yes. But that doesn't mean that we can deduce that Joe survives sleep, just like we can't deduce that he didn't blow up last night or that he wasn't framed, no matter how obvious it is that he ought to be punished. The obviousness that Joe ought to be punished is parasitic on the obviousness that he is the one who committed the crime; correspondingly, on the obviousness that he didn't blow up last night, and, among other claims, the obviousness that we all survive sleep. This just shows something we already know, namely that we all think it's antecedently very unlikely that sleep is death (for any of the apparent reasons I've been talking about), whether rightly or wrongly; but the fact that it has these strange ethical consequences doesn't give us any more reason against it.

This response is enough to put the objection to rest. It does entail, however, that we have to accept a radical shift in our ethical views. Prisoners should be let

mission failure (see e.g. Pryor (2004) and White (2006)). See also Wright (1985) and Davies (1998). For overviews on transmission and transmission failure, see Tucker (2010) and Moretti & Piazza (2023).

free, we don't have to keep promises made yesterday, and so on and so forth. If one absolutely can't live with these consequences, there is something else that one could say; a response of a more conciliary nature, which can preserve commonsensical ethics all the while accepting that sleep is death. A crucial assumption in deriving the strange ethical consequences is that people should be punished for what *they* do; and that they should only keep the promises that *they* made, and so forth. This is a very natural idea, namely the idea that personal identity is what matters. But one could reject this claim. For example, one could say that what matters isn't personal identity but psychological continuity.[22] If psychological continuity is what matters, then P2 in the arguments above is false. Joe should be punished, because he is psychologically continuous with the murderer, regardless of whether he blew up overnight or of whether we survive sleep. Similarly, I should keep promises made yesterday because I'm psychologically continuous with the promise maker. On this view, a biological theorist who learns that we blow up every night shouldn't revise any ethical claim; and neither should the phenomenal theorist who thinks that we don't survive sleep.

So, if you prefer to not change your ethics too much, you can adopt this alternative response to the ethical objection. I won't discuss the merits of this alternative view about what matters, because either way it won't be an objection to the claim that sleep is death. I offered the possibility that identity is not what matters to offer solace to those who don't want to let go of commonsense morality. If you think, instead, that identity is what matters, then my earlier arguments in this section shows that preserving commonsense morality is no reasons to believe that we survive sleep. In conclusion, there is no ethical reason to believe that we survive sleep.

These are all the reasons I could find to justify the claim that we survive sleep. Let's sum up now. We feel a very strong resistance to the claim that we die every time we go to sleep. Where does this resistance come from? It may come from a like of biological or psychological theories of personal identity; but then we must compare the theories on their theoretical merits. It may come from a strongly held intuition; but this is the product of habit, and it isn't conducive to the truth.

[22] Parfit (1984) defended this view, for reasons unrelated to sleep.

It may come from a feeling that it's a conceptual truth that people live a long time; but this rests on a mistaken view of the meaning of "person". It may come from the realization that the world would be a horrible place; but that is wishful thinking. It may come from wanting to preserve commonsense ethics, and a belief that morality would be turned upside down if we did die every night; but ethics shouldn't dictate metaphysics, and at any rate it's possible to revise ethics so as to save all appearances.

And so? On deeper scrutiny, we find that the problem of sleep is no problem at all. The obstacle has been removed in favor of believing the nice, clean theory that is the phenomenal theory. The problem of sleep is laid to rest.

1.4 Conclusion

We've reached the end of the first chapter. I've put forward a radical new theory of personal identity, the phenomenal theory, according to which personal identity is secured by the continuity of consciousness rather than the continuity of any psychological or physical factor. I've argued that this theory is the only one that can satisfy our conception of ourselves as being entities whose persistence conditions are determinate, all-or-nothing, principled and substantive. Furthermore, I've argued that this theory is the only theory that can adequately solve the problem of fission, since there are no genuine fission cases for the phenomenal theory. The theory is radical because it entails that we don't survive interruptions of consciousness such as sleep. Some would consider that a fatal objection; I've argued that, in spite of appearances, this is in fact no reason against the theory. And so, the theory is established.

References

Antony, M. (2006), "Vagueness and the metaphysics of consciousness", *Philosophical Studies*, 128, 515–538.

——— (2008), "Are our concepts conscious state and conscious creature vague?", *Erkenntnis*, 68(2), 239–263.

Bayne, T. (2010), *The unity of consciousness*. Oxford: Oxford University Press.

Bayne, T., & Chalmers, D. (2003), "What is the unity of consciousness?", in A. Cleeremans (Ed.), *The unity of consciousness: Binding, integration, and dissociation* (pp. 23–58). Oxford: Oxford University Press.

Dainton, B. (2008), *The phenomenal self*, Oxford: Oxford University Press.

——— (2011), "Review of Consciousness and its Place in Nature", *Philosophy and Phenomenological Research* 83 (1):238-261.

——— (2017), "Temporal Consciousness", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/consciousness-temporal/>.

Dainton, B. & Bayne, T. (2005), "Consciousness as a guide to personal persistence", *Australasian Journal of Philosophy*, 89(4), 549–571.

de Haan, E.H.F., Corballis, P.M., Hillyard, S.A. et al. (2020), "Split-Brain: What We Know Now and Why This is Important for Understanding Consciousness", *Neuropsychol Rev* 30, 224–233 (2020).

Duncan, M. (2015), "I Think Therefore I Persist", *Australasian Journal of Philosophy* 93 (4):740-756.

——— (2020), "A new argument for the phenomenal approach to personal persistence", *Philosophical Studies* 177 (7):2031-2049.

Foster, J. (1991), *The immaterial self*, London: Routledge.

Gustafsson, J. (2011) "Phenomenal Continuity and the Bridge Problem", Philosophia 39 (2):289-296.

Hawley, K. (2020), "Temporal Parts", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/ entries/temporal-parts/>.

James, W. (1890), *The principles of psychology*, Chicago: William Benton.

——— (1904), "A World of Pure Experience", *The Journal of Philosophy, Psychology and Scientific Methods* Vol. 1, No. 20, 533-543.

Johnston, M. (1989), "Fission and the Facts," *Philosophical Perspectives*, 3: 369–397.

——— (1989b), "Relativism and the Self", in M. Krausz (ed.), *Relativism: Interpretation and Confrontation*. Notre Dame University Press.

Lewis, D. (1976), "Survival and Identity", in A. Rorty (ed.), *The Identities of Persons*, Berkeley, CA: University of California Press; reprinted in his *Philosophical Papers* vol. I, New York: Oxford University Press, 1983.

Lund (2005), *The Conscious Self.* Amherst, MA: Humanity Books.

Merricks, T. (2001), "Realism about Personal Identity over Time", *Philosophical Perspectives* Vol. 15, Metaphysics (2001), pp. 173-187.

Nagel, T (1971), "Brain bisection and the unity of consciousness", *Synthese* 22, 396–413.

Noonan, H. (2019), *Personal Identity*. Routledge: London.

Olson, E.T. (1997), *The Human Animal: Personal Identity Without Psychology*, New York: Oxford University Press.

Parfit, D. (1971) "Personal Identity", *Philosophical Review* 80: 3–27.

——— (1984), *Reasons and Persons*. Oxford: Clarendon Press.

Roelofs, L. (2016), "The unity of consciousness, within subjects and between subjects", *Philosophical Studies* 173 (12): 3199–3221.

Schechter, E. (2018), *Self-Consciousness and Split Brains: The Minds' I*, Oxford, UK: Oxford University Press.

Shoemaker, S. (1984), "Personal Identity: A Materialist Account", in S. Shoemaker & R. Swinburne (eds), *Personal Identity*. Oxford: Blackwell.

Sider, T. (2001a), *Four Dimensionalism*, Oxford: Oxford University Press.

——— (2001b), "Criteria of Personal Identity and the Limits of Conceptual Analysis", *Philosophical Perspectives* (Volume 15: Metaphysics): 189–209.

Strawson, G. (1997), "The Self", *Journal of Consciousness Studies* 4 (5-6):405-428.

Swinburne, R. (1984), "Personal Identity: The Dualist Theory", in Shoemaker and Swinburne, *Personal Identity*, Oxford: Blackwell.

Thomson, J. J. (1997), "People and Their Bodies", in *Reading Parfit*, J. Dancy (ed.), Oxford: Blackwell.

Unger, P. (1990), *Identity, Consciousness, and Value*, Oxford: Oxford University Press.

van Inwagen, P. (1990), *Material Beings*, Ithaca: Cornell University Press.

Williams, B. (1970), "The Self and the Future", *Philosophical Review*, 79(2): 161–180.

Chapter 2

Laws and Humeanism Generalized

2.1 The Old Stuff: Natural Patterns, Laws, Humeanism and Anti-Humeanism

The natural world isn't random chaos. There are patterns. For example, massive objects attract each other; like charges repel each other; bread nourishes; fire burns; and so on. This is an undeniable empirical fact. So we might as well say:

> **Natural patterns**: there are patterns in the natural facts.

This is what makes science possible. Science, in general, aims at the discovery and systematization of these patterns.

These patterns lead us to the postulation of *laws of nature*. Laws of nature codify and express the patterns that we find in nature. For example, it's a law of nature that massive objects attract each other with such and such force (Newton's law of gravitation); or it's a law of nature that the world's wavefunction evolves according to Schrödinger's equation. At higher levels of the natural world, we can talk of the law of natural selection, or the laws of inheritance, or the laws of supply and demand.

That there are laws of nature is, then, an uncontroversial claim. But what are laws of nature? Neutrally, we can say that laws are regularities of a special kind: they play a significant role in *unification*, *explanation*, *prediction*, and *counterfactual* *reasoning*. Laws *unify* a large number of phenomena: a single law can capture all the gravitational phenomena in the universe. Laws can be used to *predict* the future: for example, we can use the laws of physics to predict the next solar eclipse. Laws can be used to *explain* particular events: for example, we can invoke the laws of mechanics to explain why such-and-such object moved in such-and-such a way. Finally laws can be used to figure out *counterfactuals* about what would happen: we can use the laws of mechanics to evaluate what would happen in a certain hypothetical scenario.

This is, in a way, the "meaning" of law of nature: the theoretical role that something must fill in order to be a law. But what kind of thing fills this theoretical role? That's the metaphysical question of what the *nature* of laws of nature is.

There are two main answers to this question, and a corresponding great divide in the metaphysics of laws. *Humeans* think that the laws of nature are, in an important sense, nothing over and above the particular facts. For the Humean, laws are mere summaries of the regularities we find in the particular facts; the laws just *describe* the facts. *Anti-Humeans*, in contrast, think that laws of nature are, in an important sense, something over and above the particular facts. For the Anti-Humean, the laws push and pull things around; they *govern* the particular facts.

The distinction is intuitive enough, but it's a little handwavy. To get a better understanding of it, we need to clarify the notions of "particular facts" and "over and above".

The particular facts form the so-called "Humean mosaic". It's the collection of all the localized matters of fact in the universe. For Lewis (1986b), these are the facts about the instantiation of intrinsic properties by spacetime points or point-sized objects. This may need tweaking in light of modern physics and entanglement phenomena;[1] but the rough idea is good enough for our purposes. A crucial specification is needed, however: these intrinsic properties need to be non-modal, that is, they cannot include primitive dispositions or causal powers, for otherwise (as will become clear) even dispositionalist theories of laws will count as Humean. Indeed, Humeanism is sometimes understood as the claim

[1] See Maudlin (2007).

that there are no necessary connections in nature. The Humean mosaic, then, is the collection of all instantiations of non-modal intrinsic properties of spacetime points or point-sized objects.

What about the "over and above"? In this context, this is traditionally cashed out with the modal tool of supervenience, as Lewis (1986b) did. Humeanism is then understood as the claim that the laws of nature supervene on the Humean mosaic; and Anti-Humeanism is understood as the claim that the laws of nature don't supervene on the Humean mosaic. That's the standard understanding of the debate.

This traditional definition, however, is not fully satisfying. The locution "nothing over and above" really suggests something thicker, namely some form of metaphysical priority. The Humean spirit is that the laws are somehow posterior to the mosaic and reducible to it, while the Anti-Humean spirit is that they're prior to the mosaic and non-reducible to it. Back in the day, supervenience was often used to get at such thicker metaphysical priority; but in these postmodal days, we can just express it directly as such. Which relation is it? It doesn't matter for my purposes. Take any notion that you think gives hierarchical structure to reality – grounding, reduction, metaphysical explanation, metaphysical dependence, metaphysical priority, building, analysis, truthmaking, whatever really. I will use "determination" as a generic term for such relation of metaphysical priority. Then, we can express the two views as follows:

> **Humeanism**: the laws of nature are determined by the Humean mosaic.
>
> **Anti-Humeanism**: the laws of nature are not determined by the Humean mosaic.

This, I think, is the correct way of understanding the Humean/Anti-Humean distinction.[2] Now that we have the distinction on the table, we can go a bit into the weeds and see some actual Humean and Anti-Humean accounts of what laws are.

[2] In Schaffer (2008), Bhogal (2017), and Emery (2020) one can find such an understanding of the distinction, with grounding as the structural relation.

Let's start with Humeanism. The simplest Humean account is the "Naïve Regularity Theory": laws just *are* the patterns and regularities in the mosaic, understood as true universal generalizations with some bells and whistles. However, the naïve account is aptly named, as it is notoriously plagued by the inability to distinguish true nomic regularities from merely accidental regularities (that all coins in my pocket are pennies; or that all gold spheres are less than one mile in diameter). A better account, and the *de facto* standard Humean theory of laws, is the so-called *Best System Theory* of laws, or BSA, pioneered by Lewis (1973) on the basis of ideas by Mill and Ramsey. According to this theory, laws are the most efficient summaries of the Humean mosaic. More precisely, laws are the generalizations that provide the best balance of informativeness and simplicity with respect to the Humean mosaic. To illustrate: a list of all the facts in the Humean mosaic is very informative, but not simple. The statement "stuff happens" is very simple but quite uninformative. In contrast, generalizations like "all massive bodies attract each other" are both informative and simple; in fact, they achieve the optimal balance. And so, they are the laws of nature.[3]

Now let's turn to Anti-Humeanism. While the Humean has only one option, the Anti-Humean has several options to choose from. Firstly, there is the *Universalist* theory of Dretske (1977), Tooley (1977), and Armstrong (1983), according to which laws are relations between universals. In particular, if it's a law that all Fs are Gs, then it's because a certain relation of necessitation holds between the universals F-ness and G-ness. Secondly, there is the *Primitivist* theory defended by Maudlin (2007).[4] On this theory, the laws of nature are metaphysically fundamental entities that cannot be further analyzed or reduced in any way. Thirdly and finally, there is the *Dispositionalist* theory, defended by Bird (2007), according to which the laws come from the intrinsic nature of physical properties. For example, it is the nature of mass that makes massive objects attract each other, thus giving rise to Newton's law of gravitation. On this view is it the dispositions or powers of things that push and pull things around. These three are the main Anti-Humean alternatives.[5]

[3] Various refinements are needed in order to deal with chance and to avoid artificially simple summaries. See Bhogal (2020) for an overview of Humeanism about laws of nature.

[4] See also Chen & Goldstein (2022).

[5] See Hildebrand (2020) for an overview of Anti-Humeanism about laws of nature.

So these are the views. Humeanism on one side, Anti-Humeanism on the other. Arguments have been given, for and against each side. The question is central in the metaphysics of science. Humeans and Anti-Humeans have very different pictures of reality. Furthermore, one's conception of law is likely to inform one's view of related concepts, such as explanation and modality and causation. Finally, some have argued that whether inductive reasoning is rational depends on whether the laws are Humean or Anti-Humean. Given the importance of inductive reasoning in everyday life as well as science, which side one picks may have radical consequences.

The aim of this paper is to generalize this distinction beyond the domain of natural phenomena to encompass metaphysical and ethical facts. In §2, I argue that there are laws in metaphysics and ethics just like there are laws of nature, and that such laws are the proper subject of metaphysical and ethical inquiry; in §3, I argue that we can generalize the Humean/Anti-Humean distinction to metaphysical and ethical laws; in §4, I argue in favor of Anti-Humeanism about metaphysical and ethical laws, by first adapting the standard arguments in favor of Anti-Humeanism about laws of nature, and by providing some new arguments.

2.2 Patterns and Laws in Metaphysics and Ethics

The issues put forward so far are restricted to natural or scientific facts: the facts in the Humean mosaic involve properties like mass or charge, and the laws in question are about those properties. But reality as a whole is more than just these natural facts. For example, there are *metaphysical facts*: facts about what exists, about which objects compose which other objects, about which event is a cause of which event, about which mental states are realized by which physical states, and so on. And there are *ethical facts*: facts about which actions are right or wrong, about which outcomes are good and bad, about which reasons there are, in a given situation, to do a certain action, about who deserves blame and praise, and so on.

Just like we find patterns in the natural facts, so we find patterns in the metaphysical and ethical facts (from now on, "M/E" will be used to abbreviate "metaphysical and ethical"). For example, when it comes to metaphysical facts, let's look at the *compositional* facts. For any collection of objects, there is a fact about whether those objects compose an object or not. These are the "particular", "localized" compositional facts; they form the compositional mosaic.

And here's the striking fact: there are patterns in the compositional facts. In this room, there are particles arranged chair-wise, and they compose a chair. In the next room over, there are different particles arranged chair-wise, and they also compose a chair. In fact, *everywhere* and *everytime* there are particles arranged chair-wise, there's a chair. There is a clear pattern here. Metaphysicians, of course, don't agree on what are the compositional facts; some deny that particles arranged chair-wise compose a chair. So they don't agree on what are the facts in the compositional mosaic; however, they all agree that there are patterns in the compositional facts. *Nihilists* about composition say that the particles never compose an object; that's a pattern. *Universalists* about compositions say that any group of particles compose an object; that's a pattern. *Restrictivists* believe that only certain pluralities compose an object. They also believe in patterns, depending on which version of Restrictivism they adhere to. For example, Van Inwagen (1990) argued that the only composite objects are living entities; that's a pattern.

All the respectable views, then, say that there are patterns in the compositional facts. It didn't have to be so. It's easy to cook up a view on which there are no patterns of composition. One could believe that in this room, particles arranged chair-wise compose a chair, while in the next room, particles identically arranged don't compose anything. Or one could believe that there is an object composed of this chair, the moon, and the Eiffel tower, but that there is no object composed of this table, the sun, and the Taj Mahal. One could have such a view; but no one does.

So, all metaphysicians believe that there is a pattern in the compositional facts; and in fact, one of the most important questions in the metaphysics of composition is that of finding out *which* is the pattern that obtains. This is the so-called "special composition question".

The same is true throughout all metaphysics. Let's look at the *temporal existence facts*. These are the facts about whether an object exists simpliciter, given that it exists at times other than the present. For example, facts about whether Socrates or lunar outposts exist. Is there a pattern in such facts? Yes. *Eternalists* think that all past, present, and future entities exist. *Presentists* think that only present entities exist. *Growing Block Theorists* think that past and present entities, but not future ones, exist. These exhaust the views taken seriously.[6] Whichever one believes, the particular facts will obey a clear pattern. We can easily imagine a view on which the temporal facts obey no pattern at all. For example, a view according to which only entities existing at some scattered collection of times exist, so that Plato exists but Abelard doesn't, and lunar outposts exist but martian outposts don't (even though both will exist). No one would take such a scattered view seriously. Rather, we all believe that there are patterns in the ontology of time; and indeed, a central question in the metaphysics of time is finding out which pattern is the one that holds.

As a final example, consider the facts about the relationship between token brain states and token mental states: for example, facts about how a certain C-fiber firing is connected to a certain pain. Here's a few theories: *Identity theorists* think that the two are identical; *Functionalists* think that mental states are functionally realized by brain states; *Dualists* think that they are distinct and only contingently connected (if at all). These views all express patterns. One could have, of course, a patternless view, on which some pains are identical to C-fibers firing, others are functional states realized by C-fibers firing, and others are dualist pains cooccurring with C-fibers firing. But no self-respecting philosopher of mind would dare put forward such a view. Again, there is a pattern: and the mind-body problem is the problem of figuring out which is the pattern.

It will be easy to verify how in all areas of metaphysical inquiry, this happens. We might as well say, then:

> **Metaphysical Patterns**: there are patterns in the metaphysical facts.

So much for metaphysics. Let's now look at the *ethical* or *normative* or *moral* facts: facts about what is the right thing to do in some situation, or what are the particular reasons in some situation, or what's good and what's bad. And let's focus

[6] Apologies to the Shrinking Block theorists; at any rate, that's also a pattern.

on the *ethical mosaic*, composed of all the "particular" ethical facts. For example, that it's permissible to turn the lever in a certain trolley problem scenario; If we're walking by a pond and we see a drowning child, then we should jump in and save him; or that a certain instance of pain is bad.

There are clearly patterns in the moral mosaic. I won't go through many examples, but the point is clear. If it's permissible to turn the lever in a specific trolley problem, it's also permissible if we duplicate the scenario a bit farther away. If we should save this drowning child, we should also save the next one. Not only this pain is bad, but all pains like this one are bad. As in the case of metaphysics, it's a controversial issue exactly *what* the pattern is. When it comes to right actions, *Consequentialists* would say the pattern is that in any situation, we should produce the best consequences; *Deontologists* would say that the pattern involve things like rights and duties, or the like; *Virtue ethicists* would say that it involves virtues. A patternless view would be one in which, seemingly at random, sometimes we are constrained by rights and sometimes we should bring about the best outcome, even in identical situations. No one seriously has such a patternless view of the moral mosaic. Rather, there are patterns and the point of ethical theory is to find what the patterns are.

So, we can safely say:

Ethical Patterns: there are patterns in the ethical facts.

All in all, then, the metaphysical and ethical mosaics are full of patterns just like the natural mosaic is. In contrast with the natural mosaic, there is widespread disagreement as to what the patterns are (since, alas, this is philosophy and not science); but there is widespread agreement that there are patterns.

What should we make of all these patterns? Just like patterns in the natural mosaic led us to the postulation of laws of nature, we should postulate laws to express the patterns in the metaphysical and ethical mosaics. Metaphysical patterns should lead us to postulate *metaphysical laws*; and ethical patterns should lead us to postulate *ethical laws*. Looking at the examples considered above, we have the following candidate metaphysical laws:

Law of composition: any plurality of objects composes some object;

Law of time: only present things exist;

Law of mind: mental states are identical to brain states.

And the following candidate ethical laws:

Law of right: the right action is the one that has the best consequences;

Law of pain: all pain is bad.

(these are merely candidate laws because you might think they're false; if so, substitute whichever you think is the right theory of composition, time, et cetera)

What makes these things laws is that they play the same role that natural laws play: unification, explanation, prediction and counterfactual reasoning. M/E laws clearly unify a variety of phenomena, by bringing (say) all compositional facts under one simple law. M/E laws explain their instances: the reason why Socrates doesn't exist is that he's not present, and the reason why such-and-such action is right is that it has the best consequences. M/E laws could similarly be used to predict: we can know that the next pain is going to be bad, or that this object won't exist, given the relevant laws; although, admittedly prediction plays a less prominent role in philosophy than in science. Finally, M/E laws support counterfactuals: we can use them to come to know that a certain action would be right in a certain hypothetical scenario, for example (and indeed, considering counterfactual situation is often how we test the plausibility of M/E laws).

My claim is that when we're doing inquiry in metaphysics and ethics our goal is to find the laws, at least when we look for general claims. And in particular, we should understand the general claims that metaphysicians and ethicists argue about as laws. We're arguing about what is the true law of composition, of rightness, and so forth. This, I think, can illuminate our inquiry. To see why it's helpful to see things this way, consider the alternative proposals of what exactly is the goal in metaphysics and ethics when we make general claims.

One could say that the goal is just to find true general claims. So that in metaphysics and ethics we want to find the true universal generalizations, whichever they are. So for example, we want to know whether all things that exist are present, understood as a purely extensional claim. This, however, is inadequate, because there could be accidental generalizations in metaphysics and ethics. If, say, the universe lasted only one instant, then all things are present. But that would be accidental, and not what philosophers of time would care about; an Eternalist could still be an Eternalist in that world.

So the claims that metaphysicians and ethicist are interested in aren't universal generalizations. Of course, they are aware of this. Typically, when putting forward a view, what happens is a familiar dance. First, they state their claim extensionally. Then they say that what they really mean is that it's *necessary* that *p*, presumably to shield against accidental generalizations. One could then avoid law-talk and say that what we're interested in aren't true universal generalizations, but necessarily true universal generalizations.

This is also inadequate. For one thing, it requires thinking that the general metaphysical truths are necessary, but this isn't clear or uncontroversial. Those who are contingentists about metaphysics or ethics seek a middle ground between true generalizations and necessarily true ones; and this middle ground is given by the notion of law. Secondly, many necessarily true metaphysical generalizations don't seem very interesting. For example, that all objects in this room are present. This necessarily true generalization isn't very interesting *per se*; and the reason it's not interesting is that it doesn't correspond to a metaphysical law (similarly, it's not a law of nature that all massive objects in the United States attract each other).

Rather, the claim that we're interested in is that it's a *law* that all things are present. From this perspective, all the various positions in metaphysics and ethics are really laws. Physicalism? A law. Deontology? A law. Perdurantism? A law.

When we take this approach to metaphysics and ethics, we end up with a nicely unified picture of inquiry. Just as scientists study the scientific mosaic, and want to find the laws therein, metaphysicians study the metaphysical mosaic and want to find the laws therein, and ethicists study the ethical mosaic and want to find the laws therein. Really, we're plunged into reality and find a bunch of facts that concern all sorts of things, from massive particles to composition to rightness. Our goal in inquiry is to find what these facts are, and if we find patterns, to find the laws that unify these facts. This is a fully topic neutral conception of inquiry. As it happens, some of these facts are natural facts, and so those laws are laws of nature; other facts are M/E facts and so those laws are M/E laws. But the picture I'm putting forward is unificationist: the goal of inquiry is to find the laws of reality simpliciter.

In a way, this idea is not especially new. People do talk of "metaphysical principles" and "metaphysical laws", and of "moral principles" and "moral laws", and this is exactly what I'm getting at. I would like to stress, however, the unificationist point mentioned above. In addition, the concept of M/E law that I'm talking about might be more general than the one that has been discussed in the literature. Some of the recent work on metaphysical laws (e.g. Schaffer 2017 and Wilsch 2020) treats metaphysical laws as being laws that back metaphysical explanation, in the analogous way in which scientific laws back causal explanations.

Now, for sure, inasmuch as there are metaphysical explanation facts, we can look at the mosaic of such facts, and we will find patterns, and so we will find laws. For example, everytime that a certain fact *p* is true, it's also true that *p* metaphysically explains "*p* or *q*". So we can say that there is a metaphysical law of disjunction, that *p* metaphysically explains "*p* or *q*". One may then think that all metaphysical laws are like that. They might say that the universalist law of composition, for example, is properly understood as a grounding law: for any collection of objects, they ground the existence of a composite objects.

Is this the proper way to understand the law in question? Maybe, but I want to remain neutral on that. My notion of law is more general, and it's compatible with someone who denies such grounding facts altogether. One could believe that the metaphysical law just is that any collection of objects compose an object, without endorsing the corresponding explanatory claim. Furthermore, such "explanatory metaphysical laws" seem to be plausible only when we deal with "input-output" laws: laws that take an input and produce an output. This seems plausible (though not obligatory) in the case of composition: the law "produces" or "generates" a composite from its parts. But this model of law doesn't seem to apply to all the metaphysical patterns. Take the second example, the law of time: that all and only present things exist. It doesn't seem natural to understand this law as an input-output law; rather, it seems that we should understand this law as a *constraint* on existence: that only things in the present time can exist (similarly for the law that everything is physical). Similar considerations apply to ethical laws: the ethical law that pain is bad could be understood as being explanatory (i.e. it's a law that something is bad *because* it's a pain), but needn't be. The same distinction arises when it comes to laws of nature. Some laws of nature can be naturally interpreted as input-output, for example the dynamical laws of physics, since they specify an object's future states (the output) given the current state (the input). But one could still interpret this laws as just being constraints on all the admissible paths; and in the case of some laws, the constraint interpretation is the most natural one: for example, equilibrium laws, or the law of conservation of energy.

For these reasons, the notion of law that I'm working with is more general than the one mostly discussed in the recent literature. It's a minimal conception, on which law codify the patterns in the mosaic that play the special role laid out above with respect to unification, explanation, prediction and counterfactual reasoning.

2.3 Humeanism and Anti-Humeanism Generalized

As seen above, there are two competing metaphysical conceptions of the laws of nature: the Humean one and the Anti-Humean one. In the past section I've generalized the concept of law to metaphysical and ethical laws. It's only natural, then, to try to generalize the Humean and Anti-Humean conception to M/E laws.

The distinction is the same: the metaphysical Humean thinks that the metaphysical laws are nothing over and above the particular facts, and just describe the underlying metaphysical facts; while the metaphysical Anti-Humean thinks that the metaphysical laws govern the metaphysical facts. Ditto for the ethical Humean and Anti-Humean.

The precise definition will involve, as said earlier, the notion of Humean mosaic and a certain notion of metaphysical priority, where I generically use determination. When it comes to metaphysics, we have the following two views:

> **Metaphysical Humeanism**: metaphysical laws are determined by the metaphysical mosaic;
>
> **Metaphysical Anti-Humeanism**: metaphysical laws aren't determined

by the metaphysical mosaic.

And we have the corresponding two views about ethics:

> **Ethical Humeanism**: ethical laws are determined by the ethical mosaic;
>
> **Ethical Anti-Humeanism**: ethical laws aren't determined by the ethical mosaic.

Something needs to be said about these definitions. While the notion of metaphysical priority can be directly carried over from the debate about laws of nature, the notion of mosaic needs to be tweaked. The Humean mosaic consists, roughly, of facts about the intrinsic non-modal properties of spacetime points. This won't do in the case of metaphysics and ethics: compositional facts or the facts about the existence of Socrates or about which action is right don't belong to such a mosaic.

But the problem can be overcome. What we want from the mosaic is that it's composed of "particular facts": these are facts that are localized. The precise definition will depend on the particular kind of fact in question, but the general idea is these facts are the "smallest" available ones in any particular domain. A reasonable thing to say is that they're the facts about one particular individual, or event, or one particular instance of a property or relation. In the case of composition, one tile of the mosaic is that certain particles compose something: in this case, the particularity comes from the fact that we're talking about one single instance of the composition relation. The conjunctive fact that these particles compose this object and that those particles compose that objects doesn't belong to the compositional mosaic, because it involves more than one instance of this relation. Similarly, that Socrates exists is a tile of the time mosaic, because it's about a particular individual, while that all ancient Greeks exist is not. And that a certain action was right or wrong is again a tile of the rightness mosaic, because it's a property of an individual entity.

This may not be the ultimate characterization. However, it provides a useful and intuitively compelling way of understanding the metaphysical and ethical mosaics: they are the collection of facts that are about one particular individual or instance of relation or property.

So we have our M/E Humeanism and Anti-Humeanism, which parallel the natural ones. In keeping in line with the unificationist picture sketched above, we can define general theses about the laws:

> **Generalized Humeanism**: laws are determined by the mosaic;
>
> **Generalized Anti-Humeanism**: laws aren't determined by the mosaic.

So we could really conceive of Humeanism and Anti-Humeanism as general theses about the universe. They are two radically different conceptions of reality. For the Humea, reality fundamentally consists of a bunch of scattered facts that we then summarize; the laws don't have much force. For the Anti-Humean, all these scattered facts are governed by laws and principles which are at least as fundamental or more fundamental.

Let's now look at some specific forms of M/E Humeanism and Anti-Humeanism. The various views about laws of nature carry over to metaphysics and ethics.

Starting with Humeanism, the natural view is an adaptation of the Best System Theory of laws. As before, the M/E laws are the universal generalization that figure in the best systematization of the M/E facts, where "best" means the best balance between informativeness and simplicity. Taking a look at composition, for example, the mosaic is (let's pretend) a universalist one, where every collection of particles composes something. A very informative, but not simple, system is the full list of all these facts. A very simple, but not informative, system is the mere claim that sometimes, composition occurs in the United States. The claim that any objects compose an object is both very informative and very simple; in fact, it strikes the optimal balance between these two virtues. Hence, Universalism is a law of metaphysics. The same story can be told for the others M/E laws. On this view, M/E laws are just efficient summaries of the mosaic.

On to Anti-Humeanism. We can first adapt the Universalist approach, and say that M/E laws are relations between universals. For example, the law that the right action is the one that has the best consequences consists in a relation of necessitation holding between the universals having-the-best-consequences-ness and rightness. This view might have a harder time expressing other laws, like the ones of composition or time.

Secondly, we can easily adapt the Primitivist theory. One can say that M/E laws are just primitive entities that cannot be analyzed further.

Finally, we can adapt the Dispositionalist theory. The idea is that the laws flow from the nature of things. In the case of scientific laws, they come from the dispositions or powers of things, like mass' disposition to attract other masses. We can similarly say that M/E laws flow from the nature of things. For example, it's in the nature of composition that any objects compose an object; it's in the nature of pain that pain is bad, and so on. In the M/E case, however, the nature of things isn't a disposition, but rather it's a more general *essence*, or a metaphysical or ethical power. So a more appropriate name for this kind of Anti-Humean is *Essentialism*.

Anti-Humeans are going to have to choose which particular version they like. In this framework, we can understand some of the preexisting conceptions of metaphysical laws as being either Primitivist or Essentialist.

Let me now consider objections to generalizing the Humean/Anti-Humean distinction to metaphysics and ethics. In particular, one may think that there is a crucial difference between the natural laws and the metaphysical/ethical laws which somehow makes the distinction unapplicable in the M/E case. I can think of two such candidate differences.

The first is an *epistemic* difference: the laws of nature are *a posteriori*, while the laws of metaphysics and ethics are *a priori*. The patterns that we find in the natural world couldn't possibly be discovered from the armchair, but they require empirical investigation. In contrast, the patterns that we find in the metaphysical and ethical mosaic can be found without leaving the armchair, just by thinking hard, giving arguments, and considering thought experiments.

Is this a problem for me? I don't think so. First of all, one could reject the epistemic difference altogether. While the "classical" view is that metaphysics and ethics are *a priori*, there is room to deny this claim. It's been claimed, for example, that perception is a source of knowledge in those domains just as it is in the case of science.[7] If that's right, then the epistemology of metaphysics and ethics isn't that

[7] For example, Byrne (2018) argues that we can know by perception that there are ordinary objects (and so that compositional nihilism is false). And several philosophers have argued that

different from the epistemology of nature. But even if that weren't so, it's unclear why it would matter: the Humean/Anti-Humean distinction is a metaphysical distinction that doesn't turn on epistemic issues; it's unclear why an epistemic difference should be relevant.

The second difference is a metaphysical difference: the laws of nature are *contingent*, while the laws of metaphysics and ethics are *necessary*. This difference in modal status may seem more threatening, since it is a metaphysical difference. As before, there is room to deny that there is such a difference at all.[8] First of all we need to be clear on what is the "standard" view. Plausibly, the majority opinion is that scientific laws are contingent, and ethical laws are necessary. On metaphysical laws, the opinions are mixed.

Of course, one could reject the modal difference. Against orthodoxy, *nomic necessitarians* believe that the laws of nature are metaphysically necessary (e.g. Bird 2007), and metaphysical and ethical *contingenstists* believe that metaphysical and ethical laws are contingent (e.g. Rosen 2006, 2020, 2021). If either (but not both) of these views are right, then modal equivalence is reestablished, as natural, metaphysical and ethical laws would be all contingent or all necessary. Another way of establishing modal equivalence is through the claim that each of these laws hold with a different kind of necessity: scientific laws with natural necessity, metaphysical laws with metaphysical necessity, and moral laws with normative necessity, along with the claim that none of these is reducible to any other (Fine 2002).

But even conceding that there is indeed a difference in modal status, what's the problem? Maybe the problem is that the Humean/Anti-Humean distinction doesn't make sense if the relevant laws are necessary. This is true if the distinction is spelled out in terms of supervenience: necessary laws trivially supervene on the mosaic, and so Humeanism becomes trivially true of them. But I've already noted that the distinction is best expressed not in terms of supervenience, but in terms of metaphysical priority. Furthermore, a paradigmatic Anti-Humean

perception plays an important role in our knowledge of moral truths (e.g. McGrath 2004).

[8] See Bhogal (forthcoming) for an investigation of why many philosophers believe the combination of contingentism about scientific laws and necessitism about moral laws.

view about the laws of nature is the dispositionalist view, on which laws of nature are necessary; surely we don't think that dispositionalists can't make sense of the Humean/Anti-Humean distinction about the laws of nature, when their view is a paradigmatic example of a view that picks a side. If anything, then, the necessary status of these laws seems to be somewhat in favor of their Anti-Humean nature. But this is not a problem; it's just an argument in favor of metaphysical and ethical Anti-Humeanism. It's true, indeed, that if the metaphysical and ethical laws are necessary then the arguments that have been given in favor of natural Anti-Humeanism may not carry over; I will examine it case by case in the next section.

All in all, then, the (alleged) apriority and necessity of metaphysical and ethical laws doesn't prevent the Humean/Anti-Humean distinction to be generalized to them. They may for sure be relevant to whether such laws are Humean and Anti-Humean.

2.4 In Favor of Generalized Anti-Humeanism

O.K. We have two views laid out on the table: generalized Humeanism and generalized Anti-Humeanism. Which view is right? Inasmuch as the distinction is novel when applied to metaphysics and ethics, not much has been written on the subject. Some of the literature can be interpreted as taking a stance that, in the terms I've described, count as Humean or Anti-Humean.[9]

Of course, no uniform view is forced upon us. One could believe any combination of Humeanism and Anti-Humeanism vis-à-vis nature, metaphysics, and ethics. And indeed, maybe the modal differences between these domains do point to a mixed view. I, for one, prefer a uniform treatment; and might as well come out of the closet and say that I'm a generalized Anti-Humean. I believe that the laws of nature are Anti-Humean, and so are the laws of metaphysics and ethics.

A mixed result, especially one that posits a different in Humean status between metaphysics and ethics, may be reached upon investigation of the particularities

[9] For example, Berker (2018) can be somehow read as (and he recognizes it) a Humean about moral laws.

of these two domains. Here, I'm more concerned with the general claim: that is, I wonder whether we should be Humeans or not about metaphysics and ethics. I will consider the standard battery of four arguments that have been given in favor of Anti-Humeanism about laws of nature, and see whether they also support Anti-Humeanism about metaphysics and ethics. I will then consider two new arguments, that only work for the metaphysics and ethics case and not for the laws of nature case.

2.4.1 The Argument from Small Worlds

A classic argument for Anti-Humeanism about natural laws is what I call the argument from small worlds.[10]

The argument is as follows: Humeanism has the consequence that there can be no difference in laws of nature without a difference in the mosaic. But intuitively, there can be such a difference in laws. This comes about, for example, when the mosaic is especially "small" so that there could be uninstantiated laws. For example, consider a world containing just one massive particle, standing still at a certain point in space. Intuitively, this world is compatible with many laws of nature: it could be a world in which massive particles attract according to Newton's law of gravitation, but it could also be a world in which massive particles attract each other, and it could be countless more. The Humean can't account for these multiple possible laws, since for the Humean the laws come from the mosaic. The Anti-Humean, instead, can easily account for the possibility of distinct laws within the same mosaic (with different primitive laws, a necessitation relation holding between different universals, or different dispositions of mass properties). Hence, Humeanism is false.

The Humean has little recourse but to deny that there are such multiple possibilities corresponding to a single mosaic (see, for example, Beebee 2000). To me, this is a huge cost, and the argument a damning one for the Humean; but everyone can pay the price they want for their view. The question is: does this argument also apply to metaphysical and ethical laws?

[10] Statements of the argument can be found in Tooley (1977) and Carroll (1994).

I think so. To do it, we need, for each particular law, to build a "small world" that's compatible with many laws. Take the law of composition, and now consider a possible world which contains, at all times, just one fundamental particle and nothing else. This is compatible with multiple laws of composition: it's at least compatible with mereological nihilism (since there is only one thing, no collection of things composes anything) and with mereological universalism (since there is only one thing, all collections of things compose something). The Humean can't account for these differences, since the laws come from the mosaic; the Anti-Humean can. Or take the law of time. Do past and future times exist? Consider a possible world which exists for only one instant. This is compatible with Presentism (since only the present time exists), with Eternalism (since there is only one time and it's the present, all times exist) and with the Growing Block Theory (since all times up to the present exist). In both these cases, the Humean mosaic seems insufficient to settle what the law is. What about ethical laws? Take the law that pain is bad, and now consider a world wholly devoid of pain. This world is compatible with the law that pain is bad, but also with the law that pain is good, since both would be uninstantiated laws. So, the small worlds argument applies in the M/E case.

Here, the Humean has a natural way of resisting, namely by appealing to the necessity of M/E laws. If M/E laws are necessary, then there aren't multiple possibilities corresponding to the same mosaic. If (say) mereological nihilism is the true law of composition, then that is the only law that could hold of the small world containing only one particle; mereological universalism is impossible. The same for ethics. Pain is necessarily bad: so there aren't two possibilities, one in which the law is that pain is bad and one in which the law is that pain is good.

This response is compelling only inasmuch as the laws seems necessary: more in the case of ethics than in that of metaphysics. But more importantly, while this response does defeat the letter of the argument, it doesn't defeat its spirit. The small world argument is standardly put in modal terms: for the Humean, laws supervene on the mosaic, but intuitively they don't. The necessity of M/E laws saves the Humean, by making the supervenience trivial. But just like the Humean/Anti-Humean distinction isn't properly captured with supervenience talk, so the true force of the small worlds argument isn't properly captured with supervenience talk. For the Humean, laws come from the mosaic. But if the mosaic is too poor, how could we possibly extract laws from it? Go back to the single floating particle. We just cannot extract, from all the facts about this particle, anything like the actual laws of physics. The particle is standing still forever. A Humean best system summary would not mention interactions with other particles, Newtonian gravitation or the like. That's the real Humean problem with small worlds. The core of the argument, then, is that it seems that sometimes there are law fact that cannot be extracted from the mosaic, that are there *over and above it*; and the Humean cannot capture that. Small worlds *underdetermine* the laws.

When the argument is so understood, it also applies to metaphysics and ethics. In the world with just one particle, there aren't enough facts to derive the true law of composition, whichever it is. The simplest summary of the compositional facts in such a world may be compositional nihilism, but surely it would be very surprising if Humeanism alone can rule out universalism or restrictivism about composition. The same is true for the other cases: in the world that lasts only one instant, there aren't enough tensed facts to derive, by a best system analysis, the true theory of time. And the same is true in the ethical cases: in a world devoid of pain or of promises or of acts, there aren't enough ethical facts to derive the law that pain is bad or that we should keep our promises or that we should do the act with the best consequences. And so, the true core of the argument stands even given the necessity of M/E laws.

The Humean can use modality to push back yet more. In the cases under consideration, the actual world is small. But modal space is big: and if the laws are necessary, the Humean can find in other possible worlds what was missing from the actual mosaic. If there is only one particle, the actual world underdetermines the law of composition. But it's certainly possible that there are many particles. If the law of composition is necessary, it will be true in the world with many particles: if (say) nihilism is true, even in a world with many particles there are no composite objects. The possible mosaics *do* determine that nihilism is true. The Humean can then say that laws are grounded not in the actual mosaic, but in all the possible mosaics: so that M/E laws are summaries of patterns in modal space. Similarly, even if the world under consideration lasts only one instant, other possible worlds last longer; and the Humean can say that the law of time is the summary of which times exist across modal space. And the same can be naturally said about ethical cases: even if the world under consideration is devoid of pain and of agents, it's possible that there are. Ethical laws summarize ethical patterns across all these possibilities.

Call this view "Modal Humeanism". It does successfully resist the argument from small worlds. However, there are several things that are unappealing about such a view. Firstly, it relies on assuming the necessity of M/E laws, which is controversial. Secondly, it results in an undesirable asymmetry between laws of nature and M/E laws: why are Humean laws of nature summaries of the actual mosaic, while Humean M/E laws summaries of the modal mosaic? One should prefer a unified concept of laws for all these domains.

Thirdly, this view is quite *ad hoc* and lacks independent motivation. Modal Humeanism says that M/E laws are summaries of patterns in all the (localized) possible facts, that is, in facts with the possibility operator in front of it. But there are many operators that attach to facts. For example, there's the operator "Someone believes that", or "according to some work of fiction", which can form their own facts; relatedly, we have all the doxastic possible worlds, and all the fictional possible worlds. Why should Humeanism be formulated using the possibility operator rather than any other operator? We could easily define "Doxastic Humeanism", where the laws are summaries of patterns in all the doxastic possibilities; and "Fictional Humeanism", where the laws are summaries of patterns in all the fictional possibilities. Whichever laws would come out would be very different from the Modal Humean ones. But then, what's so special about Modal Humeanism? Standard Humeanism says that laws are summaries of the simple, non-operator mosaic; that is a simple, clean cut set of facts, and one could see why that would give rise to distinguished laws. Neither Doxastic nor Fictional Humeanism deserves a place in our theory of laws, so why is Modal Humeanism interesting?[11]

[11] If one is a Modal Realist as Lewis (1986a), modal facts are just plain facts and so Modal Humeanism does avoid this challenge. It would be interesting if Humeans had to be Modal Realists,

Fourthly and finally, Modal Humeanism just gives the game away. The point of Humeanism is to eschew modal facts at the bottom. No self-respecting Humean would respond to the original small world argument about natural laws by saying that there are counterfactuals about what the lone particle would do if there were other particles, and that Humean laws summarize these counterfactuals: the Humean game is to reduce counterfactuals to the mosaic as well. Appealing to modal resources to expand the mosaic is cheating in the case of natural laws; and it is cheating in the case of M/E laws as well.

All in all, then, the Humean should spare us this modalizing attitude; it won't help. The small worlds argument is also in favor of M/E Anti-Humeanism. I will add, indeed, that it's stronger in this case that in the case of natural laws. In that case, we've seen, the Humean can deny that there are multiple possibilities corresponding to the small world; and when the argument is put non-modally, the Humean can just accept that such small world has no laws at all or just the very simple laws that summarize the poor mosaic. But this move is unacceptable in the case of M/E laws, if those laws are necessary. Even the small world, by the necessity of laws, must have the same law as the rich world: so "biting the bullet" cannot be done in this case.

2.4.2 The Argument from Explanatory Circularity

Another standard argument against Humeanism about laws of nature is that it leads to explanatory circularity. Here it is: laws of nature explain individual events (that's part of what makes them laws). But according to Humeanism, laws themselves are explained by the pattern of events. But laws can't both explain and be explained by the mosaic; hence, Humeanism is false.[12]

The argument can easily be transposed to M/E laws. M/E laws explain their instances, just like laws of nature do. But according to M/E Humeanism, they are explained by the patterns. So, the circularity objection can be leveraged against M/E Humeanism as well.

however. Lewis was a Humean and a modal realist, but I doubt many Humeans want to follow him down this path.

[12] See Armstrong (1983, p.40) for a statement of this argument.

The most popular response to the original argument, due to Loewer (1996), is to say that it equivocates between two distinct notions of explanation. Laws of nature, he claims, are indeed *metaphysically* explained by the mosaic, according to Humeanism; but they *scientifically* explain their instances. Given this disambiguation, the explanatory circularity disappears. A lot has been written on whether this response is successful.[13] I won't evaluate the situation; I will just note that I find the notion of scientific explanation unsatisfying on the Humean view; although I cannot convince you of that, I can at least establish the parallel with the metaphysics and ethics case.

The ethical Humean can easily adapt Loewer's response. They can say that while ethical laws are metaphysically explained by the mosaic, they *normatively* explain their instances. The viability of this move depends on whether we can make sense of this special notion of explanation.

The metaphysical Humean, in contrast, may have a harder time adapting the response. This time, since they're metaphysical laws, the disambiguation doesn't seem to work: metaphysical laws would *metaphysically* explain their instances, but that's the same kind of explanation that we have when we say that for the Humean, the mosaic metaphysically explain the laws. The Humean's only move is to say that there is a *sui generis* form of metaphysical explanation, the one mediated by metaphysical laws, that is distinct from the way in which the mosaic explains the metaphysical laws. The plausibility of this response depends on whether we're happy accepting such a fine-grained distinction within the notion of metaphysical explanation.

In conclusion, the argument from explanatory circularity is at least as compelling in the case of M/E laws as it is in the case of natural laws; and in the case of metaphysical laws, it may be more compelling, by undercutting the standard objection to the argument.

[13] Lange (2013) objects that a transitivity principle connecting scientific and metaphysical explanation make the circularity reappear. See Bhogal (2020) for a summary of the debate.

2.4.3 The Argument from the Regularity of the World

Another standard argument for Anti-Humeanism about laws of nature is the argument from the regularity of the world, or the argument from coincidence.

The basic idea is simple. As I said in the opening of this paper, there are patterns in nature: the universe is extremely regular. But if the mosaic is really fundamental, and there are no constraints in it, then how come it is so regular? How come massive particles always attract each other? In the absence of constraints, we would expect a "random" mosaic, not a well-ordered one. On the Humean view, it's an astonishing coincidence that the universe is so regular and that are laws of nature at all. The Anti-Humean, in contrast, can explain the regularity of the world by appealing to their laws: they're governing laws, hence they can govern the mosaic and make it regular.

The argument can be put both in informal and formal epistemology language. In informal epistemology, the idea is that the regularity of the world is the kind of thing that cries out for explanation, and the best explanation is that laws are Anti-Humean. The point isn't that things that cry out for explanation *must* be explained; however, we should really prefer a theory that provides an explanation. As an analogy: if we flip a coin a thousand times and it always lands heads, this cries out for explanation. The best explanation is that the coin is biased towards heads, and indeed that's the rational thing to believe. It's possible, of course, that the coin is fair and that it was just a fluke; but it would be irrational to believe that given that there is a better explanation.

In formal epistemology language, the argument can be put in probabilistic terms. The vast majority of the Humean mosaics are irregular: for any arbitrary distribution of intrinsic properties throughout spacetime, there is a possible mosaic that is like that. Hence, the Humean should assign an extremely small probability to the actual mosaic being one of the few regular ones. So, the likelihood P(the universe is regular | Humeanism) is extremely low. In contrast, the Anti-Humean can say that Anti-Humean laws guarantee that the mosaic be regular, or at least they make it somewhat likely. So, the likelihood P(the universe is regular | Anti-Humeanism) is decently high, and at any rate much higher than the former likelihood. So, applying standard Bayesian machinery, we get the result that the regularity of the universe confirms Anti-Humeanism over Humeanism.

To me, this is the decisive objection to Humeanism. It may also be the most intuitively compelling one. Can it be adapted to M/E Humeanism?

Seemingly, it can. For the M/E Humean, the M/E mosaics are fundamental. Then how come they are so regular, and can be summarized in simple laws such as "all pluralities compose something"? We wouldn't expect such a neat mosaic; rather, we would expect a "random" one, with some arbitrary pluralities composing something and others not, and with an arbitrary selection of times containing the existent objects. Yet, that's not how the M/E mosaic is. Even if we disagree on what the regularities are (since, alas, this is philosophy and not science), we all agree that the mosaic is regular. For the Humean, it's a stunning coincidence that particles arranged table-wise always compose tables. For the Anti-Humean, this is no coincidence: the governing law of composition is responsible for this regularity. And so, we should prefer M/E Anti-Humeanism to M/E Humeanism.

The Humean can respond that the argument is less appealing in this case, given the peculiarities of M/E laws.

First, the Humean can appeal to the alleged necessity of M/E laws, and argue that it undermines both the explanatory (informal) and the probabilistic (formal) version of the argument. For the explanatory version, one might think that the necessity of the laws is a good enough explanation of the regularities. We cry out for an explanation of the regularities, and say that they're a coincidence: but the Humean replies "what do you mean it's a coincidence? It *had to be* so!". But while this response has a certain ring to it, it's ultimately satisfying. Perhaps this shows that the notion of coincidence is not a modal one. Intuitively, the mere necessity of a pattern does not make it less in need for explanation. Think about it this way: we first see a pattern, and demand an explanation. Then we're being told that the pattern is necessary. This doesn't make the demand for explanation go away: in fact, it makes it even more pressing, because now the pattern holds across modal space, and so it *definitely* cannot be a coincidence, but it must be explained somehow (I'll return to this point in section 4.5).

As for the probabilistic version, a crucial premise of that argument is that any arbitrary mosaic is possible. But if the laws are necessary, then only the regular ones are possible (in fact, only the ones with the actual regular laws are possible); and so, the probability of regularity is 1, regardless of Humeanism or Anti-Humeanism, and the argument is undermined. This, however, is a mistake. It's true that if the M/E laws are necessary the only metaphysically possible mosaics are regular. But what the argument requires is that there is a large amount of irregular mosaics that are *epistemically* possible. It's the epistemic possibilities that are used to compute the relevant probabilities.

The Humean can now appeal to the other alleged difference, namely the apriority of M/E laws. If the M/E laws are apriori, then the epistemically possible mosaics are just the metaphysically possible ones, namely the regular ones, and the objection to the argument succeeds. This, however, is yet another mistake. Apriority alone doesn't get rid of the gerrymandered mosaics; what's needed is *certainty* in the laws. If we're not certain that the laws are such-and-such, then many more mosaics will be epistemically possible. Apriority as such doesn't entail certainty: mathematics is apriori, but not all certain, in fact often unknown. Some M/E laws are certain: that pain is bad, for example. In this case, I can concede that the only epistemically possible mosaics are the regular ones where all pain is bad. But in the case of most M/E laws, we don't have such certainty: the amount of controversy in metaphysics and ethics is enough to settle this point. Who is really certain beyond all doubt in mereological universalism as opposed to nihilism, or in deontology as opposed to consequentialism? Even is one is convinced of any such thesis, one cannot be *certain* of it. If they are, they would be irrational.

So the epistemically possible mosaics include all the mosaics with laws that we cannot rule out with certainty. This itself is not too bad; all these mosaics are still regular, so the probabilistic argument doesn't yet threaten us. And this is indeed the situation that the Anti-Humean who is uncertain about the laws is in. But for the Humean, things are worse. If regular mosaics corresponding to different laws are epistemically possible for the Humean, then gerrymandered mosaic must also be epistemically possible. Why is that? Let's take a simple case. Imagine there are two groups of particles arranged chair-wise, next to each other; let's call these groups the As and the Bs. As said above, it's both epistemically possible that the As composes something and that the As don't compose anything. Given this, the Humean thinks that for an individual tile of the mosaic, in this case the As, there are two epistemic possibilities. The Humean also thinks that the mosaic is fundamental, and that there is nothing that's "behind" the As composing or not composing something: that's just a fundamental fact about the universe. But then, what's there to stop the As from composing something while the Bs compose nothing (or viceversa)? There's nothing pushing or pulling, or any constraint that that As exert on the Bs. They are both fundamental, independent tiles of the mosaic.

The only way to eliminate the gerrymandered epistemic possibilities is if the uncertainty is about some feature which guarantee the regularities. To take a mathematical example, I'm uncertain about whether 4814691839 (let's call it a) is a square number. So, there are two epistemic possibilities; a is square, and a is not. Considering now $4a$, I'm also uncertain about whether it's a square or not. But there is no epistemic possibility in which a is a square and $4a$ is not, or viceversa (because if $a = b^2$, then $4a = (2b)^2$). So in this case, we can't gerrymander the epistemic possibilities, and this is because we're uncertain about a fact that constrains the other epistemic possibilities.

The Anti-Humean is in such a predicament: they're uncertain about the Anti-Humean laws, which constrains the other epistemic possibilities. Going back to the As and Bs, The Anti-Humean can say that the only two epistemic possibilities are that either both the As and Bs compose something or they both don't, because ultimately it's the laws that govern the mosaic, and the Anti-Humean is uncertain on whether universalism or nihilism is the law. But the Humean cannot say that. The Humean has no principled reason to exclude the gerrymandered mosaic. If they can coherently conceive of the As composing something and the As not composing anything, and this is just where things bottom out, they should be able to recombine the fundamental facts, at least as far as the epistemic possibilities are concerned. And so, as before the Humean is faced with an enormous amount of gerrymandered mosaics and an extremely tiny amount of regular ones. And so, the probabilistic argument does apply to the M/E Humean as well.

All in all, then, unless the M/E law in question is beyond doubt (as it rarely, if ever, is), the M/E Humean should be shocked to see the regularity of the universe; and the argument carries over. Of course, the Humean can reply in any of the ways that they could in the case of laws of nature. The best response is to deny that the Anti-Humean can properly explain the regularities or rule out gerrymandered mosaics. What if the Anti-Humean laws themselves are gerrymandered? The force of this response, whichever it is, will be the same in the M/E case. It's clear that the Humean is the one on the defense here.

2.4.4 The Argument from the Rationality of Induction

The last traditional complaint against Humeanism is the complaint that Humeanism can't account for the rationality of inductive inferences.[14] This argument is in many ways analogous to the previous one, so many of the moves will be the same.

Inductive inferences are inferences from the observed to the unobserved. For example, we've observed in the past that all massive objects attract each other, and on this grounds we infer that in the future, massive objects will attract each other. Intuitively, this is a rational inference, and it plays a crucial role in scientific investigation. However, philosophers have had a hard time explaining why such inferences are rational, ever since Hume brought this problem to the fore. After all, this inference is not logically valid. And furthermore, there are many possible futures consistent with the past: those in which massive objects stop attracting each other altogether, or those in which they attract each other only in some specific conditions, and so on. Given this, why is induction rational? Given the centrality of induction in scientific practice, this question is among the most urgent in epistemology.

The argument is then that Humeanism is especially unsuitable to justify inductive reasoning. After all, as said discussing the argument from regularity, the Humean admits of an enormous amounts of possible mosaics, the vast majority of which are not regular at all. Then, why should we expect the world to be regular in the future? If anything we should expect things *not* to keep going as they have in the past. The Anti-Humean, instead, has things to say. The Anti-Humean can

[14] See Armstrong (1983), Builes (2020), and Segal (2020) for versions of this argument.

say that the laws constrain the possible mosaics so that they're all regular; hence, induction is secured. This is a rough story, of course, and much more needs to be said: what about gerrymandered laws or the like? But at least I can see – dimly but well enough – how the Anti-Humean can try to secure induction. But I can't see how the Humean can do it.

Does the argument generalize to M/E laws? First, we need to pose the problem. In the past, particles arranged chair-wise composed a chair. Will the next plurality of particles arranged chair-wise compose a chair? That's the problem of *metaphysical induction*. In the past, promises had to be kept. Will the next promise have to be kept? That's the problem of *ethical induction*.

Given this problem, one can adapt the argument that Humeanism leads to inductive skepticism. For the Humean, the vast majority of mosaics are gerrymandered; hence, we shouldn't expect the next particles arranged chair-wise to compose a chair.

Some of the things the Humean can say in response mirror the previous argument. The Humean can say that the laws are necessary, hence there aren't all these gerrymandered mosaics. But the problem of induction arises given the gerrymandered epistemic possibilities; the Humean is still forced to countenance those, while the Anti-Humean has things to say. How good these various moves are is unchanged with respect to the previous argument.

Is there, however, something else the Humean can say that blocks the argument from induction specifically? Maybe. There is something strange about M/E induction. The problem of induction is a specifically *empirical* thing; and the domains of metaphysics and ethics are *a priori*. We just don't do induction in those domains. It's not like we establish, in some way or other, that these particles compose a chair, and then every time we find some particles arranged chair-wise we need to somehow re-establish it from scratch, so that it's an open question whether the next particles arranged chair-wise will compose a chair. Rather, it seems that *if* we've established that these particles arranged chair-wise compose a chair, then we've automatically established that so will future particles similarly arranged.

Maybe that's right, but it doesn't affect the argument. Given that the laws aren't certain, for the Humean various regular mosaics are possible. But as argued when discussing the previous argument, this leads to the epistemic possibility of gerrymandered mosaics, which is all that's needed for induction to be a problem. Perhaps it's not a problem in the sense that induction doesn't play a central role in metaphysical and ethical inquiry in the way that it does in science; however, it's still a problem in the sense that the Humean has no reason to expect that the future will conform to the past, even in the domain of metaphysics and ethics.

The M/E Humean, then, cannot rationalize M/E induction. The Anti-Humean is in better shape. They have work to do: there needs to be, for example, a restriction on gerrymandered laws, or at least a reason to prefer simple laws. The chances of success depend on the particular Anti-Humean version at stake. For example, Essence-based Anti-humeanism may satisfy this requirement better than Primitivist Anti-Humeanism. I won't get into the details. For sure, the Anti-Humean has a long road ahead to secure induction. But at least, there's a road. The Humean, instead, is at a dead end.

2.4.5 The Argument from Necessity

So far, I've considered preexisting argument for Anti-Humeanism about natural laws, and I've argued that they mostly carry over to metaphysics and ethics. In this section and the next one, instead, I put forward two new arguments for Anti-Humeanism that only apply to M/E laws. These two arguments are M/E specific because they turn on the two features that (allegedly) distinguish M/E laws from natural laws, namely their necessity and apriority. In the previous sections, these two features were invoked as a way of resisting the classical Anti-Humean arguments. I've argued that such invocations were unsuccessful; and now, I'll argue that the necessity and apriority of M/E laws actually give us positive reason in favor of M/E Anti-Humeanism.

Let's start with the argument from necessity. It's alleged that M/E laws are necessary, and many believe that. My claim is that if they're necessary, then M/E Anti-Humeanism is true.

Why is that? To get the intuition, consider the case of the laws of nature. What should we think if we learned that the laws of nature are necessary? We could, in principle, be Humean. We could believe that laws are mere summaries of the mosaic and that there are no necessary connections between parts of the mosaic; and yet believe that, somehow, the only possible mosaics are those that can be neatly summarized in terms of actual physics. But this would be weird: why would modal space be like that? Rather, the natural response would be to adopt some variety of Anti-Humeanism, one that has the resources to explain why modal space has such structure.

Similarly in the case of M/E laws. The idea is that there is something wrong with *brute necessities*. We should be able to explain where necessities come from; and Anti-Humean laws are such an explanation. Indeed, this allows us to reinterpret in a new light a standard approach to modality. A popular position is that modal facts come, at least in part, from the essences of things. But the Essentialist view is a version of Anti-Humeanism about laws (specifically, the generalization of the Dispositionalist view). Hence, we can understand the popular essentialism about modality as being a form of Anti-Humeanism about M/E laws.

2.4.6 The Argument from Apriority

The second new argument has to do with the apriority of laws. The argument, as you might see it coming, is as follows: that if the M/E laws are apriori, then M/E Anti-Humeanism is true.

Let me first be clear of what I mean by apriority in this context. There are many ways in which we come to know the M/E laws. As mentioned earlier, some have argued that it needn't be a priori after all. And in cases where it's a priori, there are several ways. Sometimes, one gives an argument involving several steps; for example, the gunk argument against compositional nihilism. Other times, we can directly come to know the relevant law; for example, we can directly come to know that pain is bad. What's really relevant for my argument is that in these cases we can come to know the general fact as such, without going through the individual cases.

I claim that M/E Humeanism is incompatible with the apriority of M/E laws.

Why is that true? To get the intuition going, we can again make a parallel with the laws of nature. Suppose that the laws of nature were a priori, and that just by sitting down and thinking hard, we could derive quantum mechanics, without ever entering in a lab. I think that intuitively, we would take that as evidence for Anti-Humeanism. Let's now try to see why.

According to M/E Humeanism, M/E laws are summaries of the mosaic. There are, then, two ways of coming to know them. Either we come to know the mosaic and then apply some sort of summarizing algorithm to come to know the summary, or we somehow directly come to know the summary.[15] This can be illustrated using as example the summary of a book. If someone knows the summary of the book, then either they read the whole book and then summarized it, or they somehow directly got ahold of the summary: perhaps they found the summary online, or someone who read the book summarized it for them.

The problem is that neither of those are a plausible way in which our apriori knowledge of M/E laws could be manifested. Consider, for example, the ethical law that pain is bad which we plausibly know it *a priori*. We don't know it by examining the mosaic of pain facts and of badness facts, and by finding out that pain and badness go together, and then summarizing this pattern. We come to know it directly; indeed that's why it's a priori law. But it also doesn't seem right to say that we directly come to know how to *summarize* the distribution of pain facts and badness facts. Our apriori knowledge couldn't have that object. To think that would be to think that we have some kind of apriori connection to summaries of these facts about the distribution of properties throughout the universe. But how could we have such a connection? Some of these facts are extraordinarily remote – for example, the pain facts in some remote corner of the universe, billions of years ago. It is magical to think that somehow, we have an apriori faculty to directly grasp summaries that encompass such facts.

[15] One might reply that another way we come to know such laws is by induction, namely by generalizing after looking at some instances in which the laws hold. Even granting that M/E Humeanism can secure inductive reasoning (in spite of the argument mentioned a few sections back), some M/E laws are definitely not known in this way. We don't think that all collections of particles arranged chair-wise compose a chair because all collections until now did; and we don't think that all pains are bad because all pains until now were bad.

Rather, the most natural explanation of our a priori knowledge is that we know some simple, unified principle that *produces* the mosaic, which an Anti-Humean law. It's plausible that a priori knowledge be directed at generative principles. The details, of course, will depend on the particular version of Anti-Humeanism. Perhaps the theory with the easiest way to account for it is Essentialism, where the laws flow from the natures or essences of things. On this view, our a priori knowledge of the laws comes from a privileged access that we bear to such essences. This is quite plausible in the case of pain mentioned above: we know the law that pain is bad just by introspecting on the nature of pain, which is directly revealed to us. Other versions of Anti-Humeanism will need to tell a different story; but in general, the idea that we can have direct contact with general principles is more appealing than the idea that we can have direct contact with summaries of a mosaic that spans millions of years.

2.5 Conclusion

We think of the natural world as being subsumable under general laws. Some of us think that the natural world is governed by such laws. There's more to reality than the natural world; there's metaphysics and ethics. I say that the metaphysical and ethical world are also governed by laws. All of reality, everything, comes from the laws.

References

Armstrong, D. (1983), *What is a law of nature?*, Cambridge: Cambridge University Press.

Beebee, H. (2000), "The non-governing conception of laws of nature", *Philosophy and Phenomenological Research*, 61(3), 571–594.

Berker, S. (2018), "The Explanatory Ambitions of Moral Principles", *Noûs* 53 (4):904-936.

Bhogal, H. (2017), "Minimal Anti-Humeanism", *Australian Journal of Philosophy*, 95 (3), 447–460.

——— (2020), “Humeanism about laws of nature”, *Philosophy Compass*, 15(8), 1–10.

——— (forthcoming), “Moral Necessitism and Scientific Contingentism”, *Oxford Studies in Metaethics*.

Bird, A. (2007), *Nature’s Metaphysics: Laws and Properties*. New York: Oxford University Press.

Builes, D. (2020), “The Ineffability of Induction”, *Philosophy and Phenomenological Research* 104 (1):129-149.

Byrne, A. (2018), “Perception and ordinary objects”, in Javier Cumpa & Bill Brewer (eds.), *The Nature of Ordinary Objects*. New York: Cambridge University Press.

Carroll, J. (1994), *Laws of nature*, Cambridge: Cambridge University Press.

Chen, E. K., & Goldstein, S. (2022), “Governing without a fundamental direction of time: Minimal primitivism about laws of nature”, In Y. Ben-Menahem (Ed.), *Rethinking laws of nature* (pp. 21–64). Cham: Springer.

Dretske, F. (1977), “Laws of Nature”, *Philosophy of Science*, 44: 248–268.

Emery, N. (2019), “Laws and their instances”, *Philosophical Studies*, 176(6), 1535–1561.

——— (2020), “Laws of Nature”, in *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. pp. 437-338.

Fine, K. (2002), “Varieties of Necessity”, in Tamar Gendler & John Hawthorne (eds.), *Conceivability and Possibility*. New York: Oxford University Press. pp. 253-281.

Hildebrand, T. (2020), “Non-Humean theories of natural necessity”, *Philosophy Compass* 15 (5):e12662.

Lange, M. (2013), “Grounding, scientific explanation, and Humean laws”, *Philosophical Studies*, 164, 255–261.

Lewis, D. (1973), *Counterfactuals*. Cambridge, MA: Harvard University Press.

——— (1986a), *On the Plurality of Worlds*. Oxford: Blackwell.

——— (1986b), Introduction to *Philosophical Papers, Vol. 2*. Oxford: Oxford University Press.

Loewer, B. (1996), “Humean supervenience”, *Philosophical Topics*, 24, 101–127.

McGrath, S. (2004), "Moral knowledge by perception", *Philosophical Perspectives*, 18(1), 209–228.

Rosen, G. (2006), "The limits of contingency", in Fraser MacBride (ed.), *Identity and Modality*. Oxford University Press. pp. 13–39.

——— (2017), "What is a Moral Law?", *Oxford Studies in Metaethics* 12.

——— (2020), "What is normative necessity?", in *Metaphysics, Meaning, and Modality*, pp. 205–233. Oxford University Press.

——— (2021), "The modal status of moral principles", *Oxford Studies in Metaethics*, 16:257–279.

Schaffer, J. (2008), "Causation and Laws of Nature: Reductionism", in J. Hawthorne, T. Sider, and D. Zimmerman (eds.), *Contemporary Debates in Metaphysics*. Oxford: Basil Blackwell.

——— (2017), "Laws for Metaphysical Explanation", *Philosophical Issues* 27, 302-21.

Segal, A. (2020), "Humeanisms: Metaphysical and epistemological", *Synthese*, 199(1–2), 905–925.

Tooley, M. (1977), "The nature of laws", *Canadian Journal of Philosophy*, 7(4), 667–698.

van Inwagen, P. (1990), *Material Beings*, Ithaca: Cornell University Press.

Wilsch, T. (2020), "Laws of Metaphysics", in *The Routledge Handbook of Metaphysical Grounding*. New York: Routledge. pp. 425-436.

Chapter 3

Ethics vs. Metaphysics

3.1 Introduction: Ethics Meets Metaphysics

Metaphysics is about the way things are: what there is, what it's like, the nature of things, et cetera. Ethics is normative: it's about what we should and shouldn't do, what's good and bad, et cetera. There are connections between the two: what we should do partly depends on how things are, after all. I shouldn't punch Jimmy because that would cause him pain. But some of these connections appear to be far reaching, and some metaphysical theories appear to have highly revisionary ethical consequences. For example, some have thought that if modal realism is true, then we don't have any moral obligations;[1] others have claimed that for there to be moral truths, God must exist.[2]

In these cases, one might be tempted to draw an inference from ethical premises to metaphysical conclusions. Surely, we have some moral obligations. If modal realism is incompatible with that, then so much the worse for modal realism. In general, such an argument has the following form:

P1 [ethical claim];
P2 if [metaphysical theory], then not-[ethical claim]; therefore,
C not-[metaphysical theory].

[1] Adams (1974); Heller (2003).

[2] See Baggett & Walls (2019) for a history of this kind of argument for God's existence.

Is such an argument any good? If you think so, then you're a *Moralist*. You think that ethics can dictate metaphysics, or at least it can give us some reason for or against a metaphysical theory. If you think such an inference is not good, then you're an *Anti-Moralist*. You think that we cannot use ethical premises in arguing for or against a metaphysical conclusion.

Which view is right? You may think there is no general answer, and that it really depends on the subtleties of each particular case. Or you may think that the answer depends on one's prior methodological standpoint: some people might give more weight to ethics, and be happy to let ethics dictate metaphysics, while others might prefer to leave metaphysics alone.

In this paper, I'll argue that the situation is less flexible than this, and that there is a principled answer as to whether Moralism or Anti-Moralism is true. The answer depends on the structure of the ethics to metaphysics inference in question. For the main class of such cases, I'll argue that Anti-Moralism is surprisingly right. In some limited cases, Moralism may be right.

In section 2 I put forward the main kind of ethics to metaphysics inference. In sections 3, 4 and 5 I argue that, surprisingly, Anti-Moralism is the right account of those inferences. In sections 6 and 7 I consider two different kinds of ethics to metaphysics inferences, where Moralism might be the right response.

3.2 Commonsense Ethics against Revisionary Metaphysics

Metaphysicians say wild things at times. Some metaphysicians believe that there are no people or tables or chairs, or that there is an object composed of a person plus the Eiffel tower, or that all possible worlds are equally real, or that nothing is conscious, or that everything is conscious, or that near me there's a myriad of people-like objects, slightly different from me. Normally, that doesn't bother us commonsensical chaps very much. However, in some cases, these wild metaphysical theories have consequences for things we care about.

For example, consider modal realism. That's the view, famously defended by

Lewis (1986), according to which all possible worlds exist and are just as real and concrete as our world. So, just like me and you exist, there are flying pigs and purple cows; they're just spatiotemporally disconnected from us. Never mind whether this view is true. Suppose it is true, as Lewis thought. What of it?

Adams (1974) pointed out that modal realism appears to have consequences for things we care about. For example, we typically only care about actual sufferings, and not about merely possible sufferings. Joe is happy; the fact that Joe *could* be in great pain doesn't worry us, given that he's happy. And the reason we don't care about possible sufferings is that they're not *real*. But if modal realism is true, then there is a concrete possible world, just as real as this one, in which a counterpart of Joe *is* in great pain. This suffering Joe is outside our spacetime, but intuitively, that isn't morally relevant. It's *as if* Joe is happy in this room, but in the next room over a clone of Joe is in pain. Since we care about people suffering, modal realism does matter for things we care about.

This clearly generalizes. For all the ways things could be that we care about, there is a way like that, both in the good and the bad. There are communities of perfectly happy people, and communities of people tortured forever. All of it is real; it's just the "next room over". This already is a bit unsettling; a sense of moral vertigo enters into me when I consider that.

But the problem trickles down to more practical matters – to actual decisions, as Heller (2003) argued. To take the standard ethics cliché, suppose I'm strolling by a lake, and I see a drowning child crying for help. Now, we might disagree about morality – consequentialism, deontology and all that – but I clearly have to jump in and save the child. It might ruin my clothes, but I have to do it. For simplicity, we can assume that there are only two possibilities: I jump in and the child lives, or I do nothing and the child dies. Normally, I would try to actualize the first one, because it's the better one. But if modal realism is true, both these possibilities are real. In one world, I jump in and save the child; in another, I don't and the child dies. No matter what, one child dies and another one doesn't. What's the point of saving this child? I'm not making reality a better place. I'm just choosing whether this spatiotemporal mass is the one with the child who drowns or the child who lives. In a way, it's as if there are two towns I could live in. In one of them, a child lives; in the other, a child drowns. Do I really have an obligation to live in the one where the child lives? It doesn't seem so.

If this is right, then modal realism has the consequence that, contrary to what we thought, I actually don't have to save the drowning child. And so, one could give the following argument:

P1 Modal realism is true;
P2 If modal realism is true, I don't have to save the drowning child; therefore,
C I don't have to save the drowning child.

The conclusion is very surprising, for it seems pretty obvious that we have to save the drowning child. Not only surprising, but unsettling and disturbing. It seems like metaphysics can teach us a whole lot about our ethical obligations, and radically alter our view of what we should do: after all, the argument clearly generalizes. Not only do I not have to save the drowning child, but I don't have to help anyone in need, and it's not wrong to murder or torture, et cetera.

Of course, this argument isn't airtight. The reasoning sketched above isn't flawless, and had a consequentialist flavor. Maybe it really matters whether *we* cause good outcomes, even if the overall consequences are the same; maybe people who are spatiotemporally related to us matter more than people who aren't. In other words, maybe P2 isn't true. Maybe. Still, the reasoning is plausible, and has at least some credibility. For the argument's sake, let's pretend that P2 is true.

This argument here is an argument from a metaphysical premise to a surprising ethical conclusion. But rather than surprising or unpleasant, one might think the conclusion is just false. As is well-known in philosophy, someone's modus ponens is someone else's modus tollens. If you don't want to accept the conclusion, all you need to do is "flip" the argument, as follows:

P1 I have to save the drowning child;
P2 If modal realism is true, I don't have to save the drowning child; therefore,
C Modal realism is false.

There. Now, we have an argument against modal realism, with ethical premises. This is an example of inference from ethics to metaphysics, as exposed in the previous section. Let's call this kind of argument the "Moralist Argument". There are other cases like this one, but for simplicity I will just focus on this example.[3]

Is it a good argument? Let me first be clear on what is the question. Sometimes, in philosophy a good argument is understood as one that is logically valid and sound. The moralist argument above is clearly valid, and if modal realism is false it's sound. But there is another sense of good argument. That's the sense in which an argument is *successful*, in that it can rationally convince someone of its conclusion. Another way of saying it is that a good argument is one that gives us reason to believe the conclusion on the basis of the premises. An argument being valid and sound isn't enough for it being good in this sense. For example, the argument "*p*; therefore *p*" is valid and sound if *p* is true, but doesn't give us reason to believe *p*. As we'll see, arguments aren't bad or good simpliciter; they're only good or bad relative to people in certain epistemic positions. In the rest of the paper, by "good argument" I mean one that is good in this sense. This is the sense that is relevant, because we're wondering whether we gain reason to reject a metaphysical theory on the basis of ethical premises.

Moralists think that this argument is a good argument. It gives us at least *some* reason to disbelieve modal realism. Some Moralists might be punchier, and would say that this argument basically defeats modal realism in and of itself. Anti-Moralists, in contrast, think that this argument gives us *no reason at all* against modal realism. According to them, in whatever high or low regard we held modal realism, we shouldn't change our mind at all given the above argument.

Which is the right view? At first glance, it's hard to resist the Moralist temptation. The argument above is valid after all, and presumably sound as well. Surely, the fact that modal realism has this strange consequence should tell against modal realism at least to some degree! As long as it's *a* reason, no matter how weak, against modal realism, then Moralism is vindicated. And surely, we do gain some reason. Isn't it obvious that we should save the child? An uncompromising Moorean will

[3] Another widely discussed case is the *personite problem*. Johnston (2016) has argued that if four-dimensionalism is true, then it's not morally permissible to, say, study Hungarian. The Moralist argument is then to infer that four-dimensionalism is false, from the premise that it's permissible to study Hungarian. I won't discuss the personite problem in this paper, but the conclusions I reach apply to this Moralist argument as well.

say that it's so obvious, that modal realism just goes out of the window. But even without going that far, it's plausible that in our theorizing we should balance all the relevant considerations. It seems that we should save the child, and this is going to exert some pull away from modal realism. Lewis himself though that if modal realism has strange ethical consequences, then this is some reason to disbelieve it (his own solution was to deny that modal realism has such consequences: he denied P2 in the Moralist argument).

I will argue, however, that we have to resist the Moralist temptation. Anti-Moralism is right: the argument above gives us no reason at all against modal realism. This might seem surprising. As a warm-up, let me put forward a similar inference, from ethical premises to a different descriptive conclusion, in which Anti-Moralism is clearly the right answer.

Take the drowning child again. We have to save the child; that's true if anything is. But even putting modal realism aside, there's a variety of non-metaphysical hypotheses about the way the world is that, if true, would contradict that claim. Suppose that at the other end of the lake a hundred children were drowning (and could be saved by the pull of a lever, say). Then we don't have to save this child; rather, we have to run to the other end of the lake and save the one hundred children before it's too late. Or suppose that this isn't actually a child, but a childlike doll. Then we wouldn't have to save it; dolls don't matter that much.

But clearly, we have to save the child! So, here's the Lake argument:

P1 I have to save the child;
P2 If there are one hundred drowning children over there, I don't have to save the child; therefore,
C There aren't one hundred children drowning at the other end of the lake.

And here's the Doll argument:

P1 It's wrong to sit down and do nothing;
P2 If this is a childlike doll, it's not wrong to sit down and do nothing; therefore,
C This is not a childlike doll.

I trust that these two arguments aren't good. They are valid, and they may be sound. But something seems off. We can't really, on ethical grounds, conclude that the lake doesn't contain drowning children over there. That it has a strange ethical consequence is *no reason at all* against the hypothesis. Anti-Moralism is clearly true of these arguments. For these reasons, let's call these "easy cases" of Anti-Moralism.

Of course, these arguments are different from the modal realism one. The conclusion here isn't a metaphysical claim, but a perfectly ordinary, empirical claim about the distribution of drowning children. This might make a big difference. As it happens, it doesn't. Indeed, we certainly need an explanation of why Anti-Moralism is intuitively right in the Lake and Doll cases. I'll provide such an explanation. Once the explanation is given, we will see how it also applies to the original argument targeting modal realism. The Lake and Doll arguments are bad arguments in that they don't give us reasons to believe the conclusion. I'll argue that this is because those two arguments *fail to transmit justification* from the premises to the conclusion. To see why, we need to take a detour into epistemology.

3.3 Transmission Failures and Inferential Justification

Let's leave metaphysics and ethics behind, just for a section. And let's look at another argument, which, I hope, will prove instructive with respect to the Moralist arguments we've been considering.

I live in a pretty modest lodging; I'd prefer to live in a giant mansion. Will I ever? Maybe some day. But in the short term, I'm pretty sure I won't. In other words, I'm pretty sure that next month I won't live in a giant mansion. That's something I justifiably believe (arguably, it's something I know). Now suppose that we're talking about my finances. And somehow, the question comes up: will I become rich in the next month? I say no. You ask me for a reason. I oblige, and give this argument:

P1 I won't live in a mansion anytime soon;
P2 If I become rich next week, I will live in a mansion soon enough; therefore,
C I won't become rich next week.

Let's call this the Rich argument.[4] There is a sense in which the argument is good. P1 is very reasonable. P2 is reasonable as well: I'd love to live in a giant mansion, so if I do become rich I'll certainly buy one. And C follows from P1 and P2, from a simple modus ponens. So we have a valid argument (most likely a sound one), with premises which are justified and even known!

Still, it's pretty clear that something is wrong with the Rich argument. Suppose someone raised the possibility that I *will* become rich, if, for example, some random stranger decides to buy my hat for a few million dollars. I can't rule that possibility out, or even reduce my confidence in it by any amount, using this argument. I can't really give, as a reason to believe that I won't become rich, that I won't live in a giant mansion.

Why is that? Well, one of premises of the argument is that I won't live in a giant mansion. That's a justified belief, to be sure. But what justifies it? What justifies it is, I think, something like the following two beliefs: first, that I won't become rich anytime soon. And second, that if I won't become rich anytime soon, I won't have a mansion anytime soon. Both these beliefs are justified: why would I become rich soon? I have a low paying job, and no great prospect in sight. And I know that the most likely way to get a giant mansion is by being rich; it's not like many non-rich people somehow have giant mansions.

But then it's pretty easy to identify what's wrong with the Rich argument. What's wrong is that I believe I won't live in a mansion *because* I believe I won't become rich soon. To be more precise, I'm justified in thinking that I won't live in a mansion on the basis of my antecedent justification in thinking that I won't get rich. I can't then justify the latter on the grounds of the former. To be so would be epistemically circular.

Here, I'm appealing to the notion of *epistemic priority* or *epistemic dependence*. While no account of epistemic priority is widely agreed upon, the notion is an intuitive one. Some of our justification is dependent on other justification. In the case at hand, my justification for the proposition that I won't live in a mansion is *dependent*, or *inferential*, on my justification for the proposition that I won't be

[4] Arguments similar to this one were examined by Vogel (1990) as possible counterexamples to the principle of knowledge closure.

rich. The notion of justification at play here is *propositional justification*. This is the kind of justification that a proposition has for a subject when the subject is justified in believing it (if, for example, the subject has sufficient evidence to believe it), whether or not they believe the proposition. This is in contrast with *doxastic justification*, which is the justification that the subject's belief has (if, for example, it's formed on the basis of the right evidence).[5] Talk of "inferential justification" may suggest a doxastic reading, where a belief is formed inferentially. Here, and in what follows, I use "inferential" to apply to propositional justification, when the justification rests on other justification.

We can see the problem if we ask for the larger argument that makes clear the source of out justification for the first premise:

P0 I won't become rich next week;
P0.5 If I don't become rich next week, then I won't live in a mansion; therefore,
P1 I won't live in a mansion;
P2 If I become rich next week, then I will live in a mansion; therefore,
C I won't become rich next week.

Where we can directly see that the conclusion of the argument is one of the premises.

This is an example of a *transmission failure*. The justification that I have for P1 and P2 doesn't transmit to C, even if P1 and P2 logically entail C. Even if the Rich Argument is valid and sound, it fails to transmit justification from the premises to the conclusion. The issue of transmission failure has received a lot of attention in epistemology.[6] Sometimes, it's controversial whether an argument fails to transmit justification: for example, it's an open question whether Moore's proof of an external world transmits justification (see e.g. Pryor (2004) and White (2006)). It should be uncontroversial, however, that there is transmission failure in the Rich argument.

It might seem strange that the Rich argument doesn't transmit justification. After all, am I not justified in thinking that I won't become rich? And isn't that

[5] On propositional vs doxastic justification, see Silva & Oliveira (forthcoming).

[6] See, for instance, Wright (1985) and Davies (1998). For overviews on transmission and transmission failure, see Tucker (2010) and Moretti & Piazza (2023).

the conclusion of the argument? Yes. I *am* justified in thinking that I won't live in a mansion, and I *am* justified in thinking that I won't be rich. But I am *not* justified in thinking that I won't be rich based on the fact that I won't live in a mansion. It's the other way around: I'm justified in thinking that I won't live in a mansion in virtue of my prior justification that I won't get rich. I was *already* justified in thinking that I won't get rich.

Crucially, the Rich argument is bad because I don't have any *independent* justification for P1. If I did, then the argument would be fine. For example, suppose I have a crystal ball that allows me to directly see the future. I look into it, and I see myself living in a non-mansion. Now intuitively this is a good reasoning: I won't live in a mansion, therefore I won't get rich. Now I *do* get reason to believe I won't get rich. Now the Rich Argument works. And it does because the crystal ball gives us some independent justification for P1. This by itself shows that whether the Rich Argument is good or bad doesn't just depend on its formal features, i.e. on its structure, nor on what propositions figure in the argument. Structurally, it's a modus tollens – it doesn't get any better than that. But the very same modus tollens is bad in the original scenario, and great if we have a crystal ball.

In the Rich argument, the epistemic circularity is transparent: I believe P1 on the explicit basis of C. But an argument can be similarly problematic without this explicit circularity. Consider the Lottery Argument:[7]

P1 I won't be rich next week;
P2 If I win the lottery tomorrow, I will be rich next week; therefore,
C I won't win the lottery tomorrow.

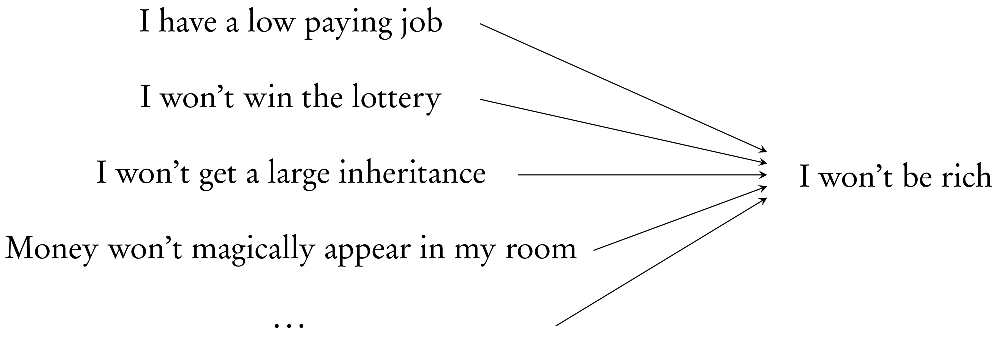
I trust that this argument feels intuitively just as bad as the Rich argument. However, the same diagnosis can't exactly apply, since C isn't what directly justifies P1. If asked for justification that I won't be rich next week, I might not say that it's because I won't win the lottery. Luckily, the diagnosis can be extended to account for the badness of this argument.

[7] Hawthorne (2004) consider such an argument, in the context of knowledge closure.

What is it that justifies my belief that I won't be rich? If someone asked me why I thought that, I would probably say that it's because I have a low paying job. In other words, I cite a cause of my future economic situation, and I am justified in believing that that cause holds. But there are many conditions that are required for me not being rich in the future, in addition to me having a low paying job. One of the conditions is that I won't unexpectedly get a large inheritance from a distant uncle. Yet another is that a big pile of money won't magically materialize into my room. And another one is that I won't win the lottery.

In order to be justified in thinking that I won't be rich, I must be justified in all of these conditions. If I wasn't justified in thinking that money won't magically appear in my room, then I wouldn't be justified in thinking that I won't be rich. Of course, such faraway conditions, such as money not appearing out of nowhere, or a distant relative giving me a large inheritance, needn't be explicitly present as such in my thinking (maybe they're all filed under a big "no strange thing will happen" proposition. And I am justified in that catchall proposition, for strange things are unlikely). This isn't a problem, since propositional justification is what is at stake. In order to be justified in thinking that I won't be rich, I must have justification, implicitly or explicitly, for believing in all the conditions.

This crucially depends on the fact that our knowledge of the future is fully inferential on our knowledge of the present (and of the past) and of how it will bring about the future. The only way we could be justified in what could happen tomorrow is by being justified in what happens now (or in the past), and in how it brings about what happens tomorrow. The situation can be represented as follows:



I have a low paying job

I won't win the lottery

I won't get a large inheritance

Money won't magically appear in my room

. . .

I won't be rich

These are (some of) the causally relevant factors of my not being rich. And, when it comes to my not being rich in the future, I am only justified in believing it by means of having justification in believing its causes.

This, then, is what's wrong with the Lottery argument. We are justified in believing "I won't be rich" wholly inferentially, through its causes (explicitly or implicitly); we cannot, then, infer back from it to any one of its causes. To do so would be implicitly circular. In general, we can give the following criterion of transmission failure:

> **Transmission Failure**. If our justification to believe a proposition P is fully inferential on its conditions C1, . . . , Cn, then we cannot infer any of its conditions from P. An argument to that effect would fail to transmit justification.

A crucial requirement of this account of the failure of the Rich and Lottery argument is that the justification for P is fully inferential on P's conditions.[8] There are two ways in which this criterion could not be met, and in those cases the argument would not fail to transmit justification.

The first way is that we could have *direct* justification in P. In the Lottery case, it's hard to imagine having direct justification concerning facts about the future. But let's imagine that time has passed, so that the question is whether I have won the lottery last month. And suppose that I have amnesia and suddenly lost all my memories of the past. I can still be directly justified in believing that I'm not rich, by looking at my bank account and my living situation. Then, it would be rational to infer that I haven't won the lottery in the past (and that money didn't magically materialize, etc.).

Secondly, we could have inferential justification that doesn't wholly come from P's conditions. Suppose a time traveler from the future pops in and tells me that,

[8] What's a "condition"? In the cases seen so far, conditions are causes or causally relevant factors. But transmission failures occur more generally, when conditions are the reasons or grounds that make a fact hold. For example, if my justification for believing a conjunction "A and B and . . . and Z" is fully inferential on the individual conjuncts, an inference from the conjunction to an individual conjunct would intuitively fail to transmit justification, and the criterion I gave accounts for it.

indeed, I won't be rich. Then it seems that now I can infer that I won't win the lottery next week (and the other conditions). This is because the time traveler's testimony provides me with independent justification that I won't be rich, one that is not inferential on its causes.

As a matter of fact, our knowledge of the future is always mediated by its causes, and that's why arguments like the Lottery naturally strike us as bad. But, as those two variations show, they needn't be: it all depends on how we are justified in the relevant premises. And my account of transmission failure explains why in the time-traveler and amnesia case the argument is good, while in the normal case, it's bad.

There are a lot of interesting issues that arise here. A full theory of the transmission of justification is certainly needed to understand all that's going on. But for the present purposes, a fragment of that theory will be enough. We have enough to understand why the Rich Argument and the Lottery Argument are bad arguments. I will argue in the next section that we can apply the same analysis to the various Moralist arguments I've put forward.

3.4 Transmission Failures and Anti-Moralism

Let's begin with one of the easy cases of anti-Moralism. Here's the Lake argument again:

P1 I have to save the child;

P2 If there are one hundred drowning children over there, I don't have to save the child; therefore,

C There aren't one hundred children drowning at the other end of the lake.

It's pretty obvious that I should save the child. But what justifies me in thinking that?

The conditions for this claim are the descriptive facts of the case, and at least partly, my justification is posterior to the justification that I have for believing the descriptive facts. Some of these descriptive facts readily come to mind. That a child is drowning, for example. I'm justified in thinking that because I see him.

And if I didn't think that, I wouldn't think I'd have to save him: if the child was happily swimming, I wouldn't think there's anything I had to do. Furthermore, I think that that I'm able to jump in and save the child. If I didn't think that, I wouldn't think that I have to save him.

How exactly do these two claims provide justification for the claim that I have to save the child? There are two justificatory models of the situation. On the first one, there are overarching moral principles, and these descriptive claims first justify the intermediary proposition that the conditions for the moral principles are met. For example, a consequentialist will say that these propositions justify the intermediary proposition that saving this child has the best consequences, and in turn this proposition justifies the proposition we have to save the child.[9] We may call this the "generalist" model. In contrast, someone might deny that moral principles play a role in justifying the claim that we have to save the child. This person, which we can call the "particularist", will say that these descriptive propositions directly justify the belief that we have to save the child.[10] Either way, the justification that we have for believing these two descriptive propositions is prior to that that we have for believing that we have to save the child.

What about the hypothesis that appears in the Lake argument, namely that there aren't one hundred children drowning at the other end of the lake? This is another descriptive claim that is relevant to the question of whether we have to save the child. And indeed, it's another condition for it to be true. On the generalist model, this claim is another condition for the intermediary proposition. Using consequentialism as an example, that the lake doesn't have a hundred children drowning over there is a condition for it to be the case that saving this child has the best consequences (if there were those children, then saving those children instead would have the best consequences). On the particularist model, this claim is another direct condition for the claim that we have to save the child (otherwise,

[9] Different moral theories will plug in different intermediary propositions.

[10] In the literature, generalism and particularism are theses about the role that moral principles play in moral theory more generally. On that, see Ridge & McKeever (2016). Here, I'm using those labels only to refer to the role that moral principles play in the structure of our justification for believing moral facts. Someone could be a particularist in this sense even if they believe that there are general moral principles, as long as our justification doesn't go through them.

it would be a mystery why it's relevant to it). On either model, this claim is epistemically just like the claim that a child is drowning: it's prior to the claim the we have to save the child. We're justified in thinking that we have to save the child on the basis of this claim (among others).

Of course, in this case our justification is implicit. We don't explicitly give, as justification for the claim that we have to save the child, that the lake doesn't contain one hundred children drowning at the other end. But that doesn't matter; we still must be justified in believing it. If we weren't (if, say, we heard some rumors of such hundred drowning children) we wouldn't be justified in thinking that we have to save this child.

We can then see why the Lake argument is a bad argument: it fails to transmit justification just like the Rich argument does.[11] We can't infer, from the claim that I have to save the child, that there aren't one hundred children drowning at the other end of the lake, because the latter is a condition for the former,[12] and our justification for the former is wholly posterior to (or inferential on) the justification we have for its conditions.

This diagnosis accounts for the badness of the Lake argument, and it can be similarly applied to the Doll argument, since "this is not a childlike doll" is similarly a condition for the claim that I have to save the child. Once we're satisfied with this explanation, it's an easy step to conclude that the original Moralist Argument is bad as well. Here's the argument:

P1 I have to save this child;
P2 If modal realism is true, I don't have to save this child; therefore,
C Modal realism is false.

The falsity of modal realism is just another condition for the claim that I have to save this child. For example: for the consequentialist, that there aren't two equally real possible worlds, one in which I save the child and one in which I don't save him, is a condition for it to be true that saving this child has the best consequences.

[11] The same diagnosis applies to the Doll argument. That this is indeed a child and not a doll is a condition for the claim that we have to save the child.

[12] In this case, the descriptive facts are conditions in virtue of being grounds, rather than causes (see footnote 8).

In general, inasmuch as we take Modal realism to be relevant to the question of whether or not we should save the child, either it's a direct condition of it, or it's relevant because by being relevant to the purely descriptive facts about the lives and death of children, in the same way that the hypothesis about the lake containing one hundred drowning children is.

In the justificatory structure of the situation, the falsity of modal realism is prior to the mixed ethical claim. As before, this justification is not present explicitly in our reasoning, but it doesn't matter. We must have justification for thinking that Modal realism is false, for us to be justified in thinking that we have to save the child. If we weren't (suppose that the arguments for modal realism were convincing), then we wouldn't be justified in thinking that I have to save this child. And we *are* justified in thinking that modal realism is false. It's an extremely revisionary view, after all. This is why the Moralist Argument is bad: we can't infer a condition from something that we're only justified in believing inferentially, from its conditions. The Moralist argument is a case of transmission failure.

This might be surprising. Isn't it obvious that I should save this child? Yes, it is. But it's obvious only because it's antecedently obvious, among other things, that there aren't one hundred children drowning over there (why on earth would that be?), and that this is indeed a child and not a doll (dolls so realistic are very rare), and that modal realism is false (that's a wild metaphysic). The obviousness of this mixed ethical claim is parasitic on the obviousness of descriptive facts, empirical and metaphysical alike, that are the conditions for this fact to hold.

The diagnosis can be suitably adapted to all Moralist arguments of this form, namely those in which a mixed ethical fact is the premise.[13] In general, such mixed ethical facts are justified inferentially, from the various descriptive facts of the case, including which metaphysical theory is true. For this reason, one can't reject the metaphysical theory on the basis of this ethical fact, and such an argument doesn't transmit justification. When it comes to this kind of argument, then, Anti-Moralism is true in general.

[13] To illustrate: in the case of personites and four-dimensionalism, our justification to believe that it's permissible to study Hungarian is posterior to our justification to believe that (for example) studying Hungarian benefits me and harms nobody; and in turn our justification to believe this claim is posterior to our justification to believe that four-dimensionalism is false.

In the next section, I consider an objection to my argument, by way of rejecting the structure of the ethical justification that I've put forward.

3.5 Ethical Justification and Moral Perception

My argument for Anti-Moralism crucially rests on the claim that our justification for the mixed ethical claim is inferential from the descriptive facts of the case. It's tempting to resist this picture of ethical justification. One could claim that we have *direct* justification in ethical claims. If that were true, then the arguments wouldn't fail to transmit justification. Rather, it would be like the Lottery argument in the case where I directly observe that I don't live in a mansion.

And indeed, that is a tempting thought. It does seem that when we see the drowning child, we directly get justification that we ought to save it. There's no reasoning involved, no sophisticated deduction. It's just direct intuition. We *feel* that we have to save the child. Some would say that we *see* the moral fact that we have to save the child. They say we have moral perception, that gives us direct justification in normative claims, just like ordinary perception gives us justification in descriptive claims.[14]

Whether moral perception is inconsistent with my diagnosis depends on what moral perception is taken to be. Moral perception could just be the claim that our belief that we have to save the child isn't formed inferentially from prior descriptive beliefs, but rather is formed directly upon seeing the drowning child. But this isn't inconsistent with my account. What matters is whether our *justification* for the belief is epistemically posterior to our justification in the descriptive claim that a child is drowning, etcetera. How the belief is actually formed is independent of that question.

A second way moral perception may seem relevant is that it also doesn't seem that our *justification* is inferential. In particular, the philosopher sympathetic to moral perception may say that we cannot really articulate the descriptive and normative conditions that justify our belief that we have to save the child. But

[14] For a defense of moral perception, see McGrath (2004). For an overview of the debate, see Werner (2020).

again, I can concede that. Whether our justification is epistemically posterior doesn't depend on whether it's explicitly present in the head, because what matters is propositional justification and not doxastic justification. This is true as it can be seen from the case of the Rich argument. I cannot really articulate all the epistemically prior propositions that justify my belief that I won't be rich. But it's still true that it's inferential on all claims such as the claim that I won't win the lottery, as well as claims I never explicitly considered such as the claim that money won't magically appear in my room, or that a stranger won't buy my hat for a billion dollars, and so on. Since that move is right in the case of a propositions like the one that I won't be rich, it can also be right in this case.

Finally, moral perception can be taken as the claim that perception directly justifies us in the mixed ethical claim that we have to save the child. Let's call this view "naïve moral perception". This view *is* a challenge to my view. However, I think that this view is implausible. Firstly, the intuition that our justification is not inferential can be put under pressure. Say we're walking by the lake, we see the drowning child, and you say "we have to save him". I ask you why you believe that. At the beginning, you might say "isn't it obvious? Can't you see?". But if I nudge you, you would eventually say something like "well, the child is drowning!". Clearly, we think we should save the child because we think the child is drowning. We surely don't think the child is drowning because we think we should save him: that would be to take things backwards. This is reason to believe that the justification is actually inferential and not direct, even if at first glance it may appear direct.

Secondly, and more importantly, naïve moral perception has weird consequences. If we gain direct justification that we have to save the child, then Moralism is right about the easy cases too: we can gain justification for the claim that there aren't one hundred children drowning at the other end of the lake, on the basis of the direct moral justification that we have to save this child. But this is absurd. The moral sense may be strong, but it cannot tell us what's at the other end of the lake; it's not that strong. That's something we can only find out by going there and checking, or through background information about the distribution of drowning children in lakes. A way to see why this is implausible is to imagine that we're just wondering for the sake of wondering, before anything dramatic happens, whether there are one hundred children drowning over there. Then we see this drowning child, and we see that we ought to save him. It seems very strange that we suddenly gain extra justification that there aren't one hundred children drowning at the other end of the lake.

So naïve moral perception must be rejected. This isn't to say that the whole theory of moral perception has to be rejected. As said earlier, I can concede that we form the moral belief without doing any inference; my claim is about the structure of justification. Furthermore, we can find an alternative sense in which there *is* direct moral perception. Rather than say that moral perception gives us direct justification in the mixed moral claim, we can say that moral perception gives us direct justification in the particular moral conditional: in the claim that *if* things are as they appear (a child is drowning, no one else is, etc.), then I should save the child. After all, I would have the same moral perception if the whole set-up was an illusion, and there was no child and no lake. It would still be true that *if* things are as they appear, I have to save the child. In this sense, moral perception is direct in that we have direct knowledge of the particular moral conditional fact of the case. This is to be contrasted with an intellectualist view on which in each case we get to the mixed moral fact by applying some overarching moral principles (e.g. "maximize total happiness") to the descriptive facts of the case at hand.

All in all, our justification for the first-order ethical claim is inferential, and so the Moralist argument is a case of transmission failure. However, it's interesting to consider whether it could ever be a good argument. After all, the Lottery argument is a bad argument only in the standard situation in which we know the future through the present. If a time-traveler told us we won't live in a mansion, then we can infer we won't live the lottery. What would be the analogous case for the Moralist argument? It would be a case in which our justification for the first-order ethical claim is either direct, or gained inferentially without passing by the descriptive facts and the relevant moral law. This never happens in the actual world. But could it possibly happen?

We can imagine a case of testimony: someone walks up and tells us "You need to save this child. Trust me". If that person has a child-detector that tells her whether there are other children drowning in the lake, then I could reasonably infer that there aren't one hundred children drowning in the lake. Similarly, if she is better than me at metaphysics, I would think that if she's so sure I should save this child, then she must have concluded that modal realism is false, and so I do gain justification for the claim that modal realism is false. But this isn't genuine Moralism. Ultimately, her knowledge that I should save this child is inferential on her knowledge of the fact that there aren't those other drowning children, and that modal realism is false. It's just that she is empirically and metaphysically better informed than I am, and I indirectly inherit that information through her act of testimony. She could have just directly said that there aren't those drowning children, and that modal realism is false. And even if this was genuine Moralism, it's irrelevant in the context at hand. The question is whether the fact that a metaphysical theory has a counterintuitive ethical consequence is a reason against it. Inasmuch as the consequence is counterintuitive, it's because it seems to us to be counterintuitive, not because someone told us so.

Maybe, stretching our imagination a little bit, we can imagine genuine cases of Moralism. Maybe God directly knows all the facts, including the first-order ethical ones. So, God could "infer" a metaphysical theory from this ethical claim. But since he already knows the true metaphysics, it's unclear what the point of that would be. Or we can imagine that we had a "moral crystal ball": a device that spits out true ethical claims. We question the ball, and it says that we ought to save the child. Then plausibly, we get direct justification for this claim, and we can go on to infer that there aren't other drowning children, and that modal realism is false. But I can't really imagine how this moral crystal ball works. The only way I can make sense of it is if it somehow knows the descriptive facts and infers the first-order ethical facts from them. But then, this is just like the case of testimony above. And at any rate, no crystal ball is telling us ethical facts that could allow us to rule out modal realism.

In conclusion, there may or may not be possible cases of successful Moralist arguments. But all the Moralist arguments that we actually make will fail. Ethical justification is direct only insofar as it pertains to moral laws or particular moral conditionals; ethical justification in first-order ethical claims is always inferential.

And so, modal realism is saved from the ethical objection. And in general, we cannot reject a metaphysical theory because it has a strange first-order ethical conclusion. But there are other kinds of inferences from ethics to metaphysics. Let's now see if Anti-Moralism is true of them as well or not.

3.6 Metaphysical Notions that Are Ethical in Nature

Some metaphysical notions are essentially intertwined with ethical notions. For example, the notion of personal identity is essentially tied to facts about blame, responsibility, punishment, prudential rationality, and so on. One might naturally think that in *those* cases, at least, the inference from ethical premises to metaphysical conclusions will be legitimate.

It is useful to see how that would go with an uncontroversial case. Consider the notion of "personal space". It's a metaphysical notion – a region of spacetime surrounding a person – which is clearly ethical in nature. For simplicity's sake, let's suppose this ethical nature is fully expressed by the following condition: X is someone's personal space iff it's *prima* facie wrong to enter X without the person's consent.

In this context, it seems that we *can* legitimately infer metaphysical facts from ethical premises. For example, we have the intuition that we can't get within one inch of someone without their consent. From this, we can infer that their personal space spans at least one inch from their body. And if a certain metaphysical theory had the consequence that someone's personal space was only half an inch long, then our intuition is definitely some reason against this theory, maybe even conclusive reason. So in this case, Moralism seems right. Why is it, and how does it differ from the diagrams shown before?

The difference is a difference in metaphysical structure, which then translates into a difference in justificatory structure. X is someone's personal space *because* it's wrong to enter X without the person's consent. This is what it is for a metaphysical notion to be ethical in nature: it is for the metaphysical facts to hold in virtue of ethical facts. This is the reverse of what happened before, where the mixed ethical facts holds in virtue of the descriptive facts. This reversal results in a reversal in justificatory structure. We are justified in believing claims about someone's personal space only inferentially, from ethical claims about whether we can or cannot enter some region of space.[15] This is why in this case, we can infer a metaphysical fact (and reject a revisionary metaphysical theory) on the basis of ethical facts.

OK, so Moralism is true of personal space. But that's not a very interesting notion. What about personal identity? The question of personal identity is the question of what are the conditions under which a person persists through time. On some views, personal identity requires a certain kind of psychological connection between the person at different times, involving for example memory. On other views, it requires a physical connection, involving the continuous existence of the body or of the brain.[16]

The notion of personal identity is of special normative importance, in that it appears in a network of normative claims. For example, people ought to be blamed for things *they* did at earlier times; or we may have special reasons to care about what will happen to *us*, as opposed to other people. As John Locke said, "Personal identity is a forensic notion"; and many have followed him. One might then think that personal identity is a normative notion in the way that personal space is; and so that we can draw conclusions about personal identity from ethical grounds, and, if need be, refute a theory of personal identity on ethical grounds.

Can we? Let's try it out. Let's make up a theory of personal identity with wild ethical consequences. For example, consider the *Waking theory of personal identity*. It says that a person only survives while they're awake. When they go to sleep, they die, and each morning it is, strictly speaking, a new person (inhabiting the same body, and with similar memories, et cetera).

Regardless of its merits, this theory exemplifies the main topic of this paper. This is because the Waking theory entails many revisionary ethical claims. As we

[15] Of course, our belief in this ethical fact is itself justified from other descriptive facts. The fact that we cannot get within one inch of someone might be inferred from, for example, facts about the consequences of such an action.

[16] Defenders of the psychological theory include Lewis (1976), Parfit (1984). Its contemporary form descends from Locke. Proponents of physical views (in one form or another) include Thomson (1997), Olson (1997).

saw before, we typically think that people ought to be blamed (and punished) for what they do. Take Joe, who is in jail for brutally murdering several people back in 2018. It is pretty obvious that he ought to be in jail. At least, that's what we think. But if the waking theory is right, Joe didn't commit any crime: Joe came into existence this morning when he woke up. Someone else, years ago, committed those crimes. Joe would be actually innocent, and blameless, and arguably should be let out of jail. So, here comes the Moralist argument:

P1 Joe ought to be blamed and punished;
P2 If the waking theory is right, Joe ought not to be blamed and punished; therefore,
C The waking theory is false.

Is this Moralist argument any good? The thought behind this section is that this argument, unlike the Lake argument, is a good one, because personal identity has an ethical nature.

Is it, though? I don't think it is. The first warning sign is that we can build easy cases in which Anti-Moralism is clearly right. Certainly, there are some factual circumstances which would lead us to believe that Joe is, in fact, innocent. Suppose all the evidence was fabricated, and it was someone else who committed the crime. Or suppose the person who committed the crime (the original Joe), and was put in jail since, dissolved into thin air overnight, and, by chance, a perfect duplicate (the person we now call Joe) materialized out of nowhere into his cell. In either case, we wouldn't have to punish Joe. So, here's the argument:

P1 Joe ought to be blamed and punished;
P2 If Joe materialized tonight out of nowhere, Joe ought not to be punished; therefore,
C Joe didn't materialize tonight out of nowhere.

Clearly, we can't be justified in the conclusion on the basis of this argument. And the diagnosis for this is the now familiar one. We are justified in thinking that Joe ought to be punished because we are justified in thinking that he is the one who committed the crime: and one of the conditions for that is that he didn't appear out of nowhere this morning, while the original criminal disappeared. But one of the conditions for Joe being the one who committed the crime is also that the Waking theory is false. So the Moralist argument involving personal identity is just as bad as the other Moralist arguments. The fact that personal identity is a normatively loaded metaphysical notion didn't change anything.

But we saw that the fact that the notion of personal space is ethical in nature *did* change things. So why doesn't it change things in the case of personal identity? It's because facts about who should be punished are still inferential on facts about personal identity. We figure out who we should punish by first figuring out the facts about personal identity (i.e., who committed the crime). We do things in this order because the facts about who should be punished are derivative from the facts about personal identity.

For the case to be analogous to the case of personal space, it would have to be the opposite: it would have to be that the personal identity facts hold *because* of the normative facts. But this just doesn't seem right. It doesn't seem that the reason that Joe is the person who committed the crime is that he should be punished; it's the other way around. He should be punished because he's the one who committed the crime.

All is well. But there's a puzzling sensation that remains. In what sense is personal identity a normative notion, then, if it doesn't have the structure exemplified by personal space, and so doesn't license inferences from ethics to metaphysics? The straightforward sense in which personal identity is a normative notion is that personal identity facts have normative consequences, because they ground facts about who we should punish. In addition, it could be that we *fix the reference* of personal identity using normative facts. Maybe what goes on is something like the following. There are all sorts of relations between persons at times: psychological continuity, biological continuity, waking continuity. We then stipulate that personal identity is whichever relation R among these is the one that matters, and is such that we should punish people for what their R-related person did. This stipulation ensures that personal identity has ethical consequences; but it doesn't change the fact that these ethical consequences are downstream from the personal identity facts. And so, the justificatory structure is the same as that of the Lake argument. If this is indeed how the reference of personal identity is fixed (a claim I need not endorse), then we can see how one might be tempted to think that it's a strongly normative notion.

Of course, you might disagree. You might think that personal identity is a normative notion in the stronger sense in which personal space is. This seems wrong to me. But at any rate, I've pointed out the criterion by which we can ascertain whether we can infer a metaphysical conclusion from an ethical premise: the facts about the metaphysical notion must hold in virtue of the ethical facts. This is true of some notions, like personal space; I've argued that it's not true in the case of personal identity, which was probably the most interesting candidate for Moralism. On the whole, then, Anti-Moralism is still true.

3.7 The Metaphysics of Pure Ethics

The final category of ethics to metaphysics inference I want to consider is when the metaphysics is relevant to the purely ethical facts in question. This happens, for example, when it's the metaphysics of moral facts themselves. Consider the notion of "moral property", such as the goodness or badness of an action or of a state of affairs. It's a metaphysical notion, but it directly about ethics.

When a notion such as this one is involved, we can get a Moralist argument that might be good. Suppose we thought that the existence of God is required for there to be moral properties and moral facts. Going back to our lake, an atheist could give the following argument:

P1 God doesn't exist;
P2 If God doesn't exist, I don't have to save the child; therefore,
C I don't have to save the child.

And here is the flipped, Moralist argument:

P1 I have to save the child;
P2 If God doesn't exist, I don't have to save the child; therefore,
C God exists.

Many have been tempted by an argument for God's existence along these lines.[17] Is it any good?

Our justification for "I have to save the child" is inferential on the descriptive facts of the case, as I've argued in the previous sections. There is, however, a fact in the vicinity such that the justification we have for it is arguably not inferential: the relevant pure normative fact. For the generalist, it's the fact that we ought to follow this or that moral principle; for the particularist, it's the fact that *if* things are such-and-such (a child is drowning, etc.), then I have to save him. If P2 in the argument is true, then God's existence is a condition for this purely normative fact to be true. If our justification to believe this normative fact is wholly inferential on its conditions, then, we couldn't possibly infer from it that God exists. If, instead, we have direct justification to believe it, then we *can* infer from this fact that God exists. The Moralist argument, then, is really the following:

P1 If a child is drowning (etc.), I have to save the child;
P2 If God doesn't exist, P1 is false; therefore,
C God exists.

The question is then whether P1 is justified inferentially, or directly. The followers of moral perception, or those who say that we intuit general moral principles directly, will say that it is justified directly, and so that this argument is successful. Others might say that we're only justified in thinking that it *appears* that we have moral obligations, unless we're justified in the conditions for that claim, such as the existence of God and of moral properties.

When a metaphysical claim is directly relevant to pure ethics (as opposed to the mixed facts), then Moralism may be the right answer, conditional on the directedness of our justification in pure ethical claims. Even if that were true, that wouldn't do much by itself to overturn a metaphysical theory. The real meat of the argument lies in P2: the claim that God is required for morality. That is an extremely controversial premise, and so the possibility that this ethical premise leads us to this metaphysical conclusion is remote.[18] This is in contrast with

[17] See Evans & Baggett (2014).

[18] For another example of this kind of argument, suppose someone thought that physicalism

the case of modal realism (and other arguments similar to that), where there the premise that links the metaphysical theory to a revisionary ethical consequence is more plausible.

3.8 Conclusion

All right. Let's sum up. We've seen three different kinds of arguments from ethical premises to metaphysical conclusions. In the first kind, we infer a metaphysical fact from a mixed ethical fact. I've argued that these arguments are epistemically circular and fail to transmit justification, and so Anti-Moralism is true of them. In the second kind, we use an ethical fact to infer metaphysical facts involving metaphysical notions that are ethical in nature. I've argued that these arguments are good, but they require a strong conception of "metaphysical notions ethical in nature", and the most interesting and plausible candidate (personal identity) doesn't satisfy such a conception. In the third and last kind, we infer a metaphysical fact from a purely normative fact. In this case, Moralism might be true, conditional on a certain view about our justification we have for believing the purely ethical facts; but in the case I've considered (the moral argument for God) it relies on a very controversial second premise. All in all, then, when it's ethics vs metaphysics, metaphysics stands.

References

Adams, R. (1974), "Theories of Actuality", *Noûs*, 8(3): 211–231.

Baggett, D., and Walls, J., (2019), *The Moral Argument: A History*, Oxford: Oxford University Press.

Davies, M. (1998), "Externalism, Architecturalism, and Epistemic Warrant", in *Knowing Our Own Minds*, Crispin Wright, Barry C. Smith, and Cynthia Macdonald (eds.), Oxford: Clarendon Press, 321–362.

(the claim that everything is, or supervenes on, or is grounded on, the physical) is incompatible with the existence of moral facts. Then one could gain reason to believe that physicalism is false on the basis of P1. And again, the premise that is hard to defend is precisely the linking premise that physicalism is incompatible with the existence of moral facts.

Evans, C. S. and Baggett, D. (2014) "Moral Arguments for the Existence of God", *The Stanford Encyclopedia of Philosophy (Winter 2022 Edition),* Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives /win2022/entries/moral-arguments-god/>.

Hawthorne, J. (2004), *Knowledge and Lotteries*. Oxford: Clarendon Press.

Heller, M. (2003), "The immorality of modal realism, or: How I learned to stop worrying and let the children drown", *Philosophical Studies* 114 (1-2):1 - 22.

Johnston, M. (2016), "The Personite Problem: Should Practical Reason Be Tabled?", *Noûs* 50 (4):617-644.

Lewis, D. (1976), "Survival and Identity", in A. Rorty (ed.), *Identities of Persons*, Berkeley, CA: University of California Press; reprinted in his *Philosophical Papers vol. I*, New York: Oxford University Press, 1983.

——— (1986), *On the Plurality of Worlds*, Oxford: Blackwell Publishers.

Parfit, D. (1984), *Reasons and Persons*. Oxford: Clarendon Press.

Pryor, J. (2004), "What's Wrong with Moore's Argument?", *Philosophical Issues*, 14(1): 349–378.

McGrath, S. (2004), "Moral knowledge by perception", *Philosophical Perspectives*, 18(1), 209–228.

Moretti, L. and Piazza T. (2023), "Transmission of Justification and Warrant", *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/ sum2023/entries/transmission-justification-warrant/>.

Olson, E.T. (1997), *The Human Animal: Personal Identity Without Psychology*, New York: Oxford University Press.

Ridge, M. and McKeever S. (2016), "Moral Particularism and Moral Generalism", *The Stanford Encyclopedia of Philosophy (Summer 2023 Edition)*, Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/ archives/sum2023/entries/moral-particularism-generalism/>.

Silva, P. & Oliveira, L. R. G. (forthcoming), "Propositional Justification and Doxastic Justification", in Maria Lasonen-Aarnio & Clayton M. Littlejohn (eds.), *Routledge Handbook of the Philosophy Evidence*. Routledge.

Stoljar, D. (2001), "Physicalism", *The Stanford Encyclopedia of Philosophy (Summer 2023 Edition)*, Edward N. Zalta & Uri Nodelman (eds.), URL =https://plato.stanford.edu/archives/sum2023/entries/physicalism/

Thomson, J. J. (1997), "People and their Bodies", in *Reading Parfit*, J. Dancy (ed.), Oxford: Blackwell.

Tucker, C. (2010), "Transmission and Transmission Failure in Epistemology", *Internet Encyclopedia of Philosophy* 1.

Vogel, J. (1990), "Are There Counterexamples to the Closure Principle?" in *Doubting: Contemporary Perspectives on Skepticism*, M. Roth and G. Ross (eds.), Dordrecht: Kluwer Academic Publishers.

Werner, PJ. (2020), "Moral perception", *Philosophy Compass.*

White, R. (2006), "Problems for Dogmatism", *Philosophical Studies*, 131(3): 525–557.

Wright, Crispin, 1985, "Facts and Certainty", *Proceedings of the British Academy*, 71: 429–472.