The Empirical Relevance of Metaphysics

by

David Builes

B.S. Mathematics, B.A. Philosophy Duke University, 2016

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

MAY 2020

© 2020 Massachusetts Institute of Technology

All rights reserved

Signature of Author.....

Certified by

Bradford Skow Laurance S. Rockefeller Professor of Philosophy Dissertation Supervisor

Accepted by.....

Bradford Skow Laurance S. Rockefeller Professor of Philosophy Chair of the Committee on Graduate Students

The Empirical Relevance of Metaphysics

by

David Builes

Submitted to the Department of Linguistics and Philosophy on May 22, 2020 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Philosophy

Abstract

Are metaphysical debates relevant to ordinary empirical inquiry? This dissertation collects a series of papers which answers in the affirmative. The first part of the dissertation is concerned with inductive inference. I argue that we shouldn't expect the world to be amenable to induction if orthodox versions of Humeanism or Non-Humeanism are correct. I then develop and defend a hybrid view, a 'Humean Non-Humeanism', which has a better hope of vindicating inductive inference. The second part of the dissertation is concerned with self-locating belief. While puzzles regarding self-locating belief are often motivated by certain fanciful thought experiments, it has recently been argued that the epistemology of self-locating belief is of central concern to many of the deepest questions in fundamental physics: including the interpretation of quantum mechanics, large-scale cosmology, and the (alleged) fine-tuning of the universe. I begin by arguing that the correct epistemology of self-locating belief is also relevant to classic debates in the metaphysics of time. By exploiting the fact that different theories in the metaphysics of time classify different sorts of facts as 'merely indexical' facts, it can be shown that different views in the metaphysics of time make different empirical predictions. I then turn to argue for the correct epistemology of self-locating belief on metaphysical grounds. I first argue for some conditional claims: if one holds certain (controversial) metaphysical views regarding the nature of objects, properties, and identity across time, then one should uphold a particular theory of self-locating belief. I then go on to argue for an overall metaphysical picture that vindicates these views concerning the nature of objects, properties, and identity across time.

Dissertation Committee

Bradford Skow (Chair), Laurance S. Rockefeller Professor of Philosophy Roger White, Professor of Philosophy Miriam Schoenfield, Associate Professor of Philosophy Jack Spencer, Associate Professor of Philosophy

Table of Contents

Front Matter	
Acknowledgements	4
Preface	5
The Empirical Relevance of Metaphysics	
Part One	6
1. A Humean Non-Humeanism	7-38
2. The Ineffability of Induction	39-66
Part Two	67
3. Self-Locating Evidence and the Metaphysics of Time	68-85
4. Time-Slice Rationality and Self-Locating Belief	86-108
5. No-Self Location	109-137
6. The World Just Is The Way It Is	138-168

Acknowledgements

MIT has been a wonderful place to pursue my philosophical interests over the past few years. I would like to thank everyone in the MIT community for making the department so welcoming and supportive. For giving me helpful feedback on the ideas present in this dissertation, thanks go to David Balcarras, Tyler Brooke-Wilson, Alex Byrne, Kevin Dorst, Caspar Hare, Sally Haslanger, Michele Odisseas Impagnatiello, Agustín Rayo, Haley Schilling, Alex Skiles, Robert Stalnaker, and Stephen Yablo. Special thanks go to my committee members – Miriam Schoenfield, Brad Skow, Jack Spencer, and Roger White. Many of the philosophical views that I have pursued at MIT are very speculative and counterintuitive. Having a committee that has encouraged me to explore these ideas, while at the same time exposing me to all sorts of different criticisms (both in style and substance), has been invaluable.

For those who helped me on a more personal level during my time at MIT, special thanks goes to David Balcarras, Allison Balin, Nathaniel Baron-Schmitt, Tyler Brooke-Wilson, Thomas Byrne, Shannon Doberneck, Kevin Dorst, Kelly Gaus, Kat Hintikka, Michele Odisseas Impagnatiello, Daniel Muñoz, Haley Schilling, Miriam Schoenfield, Ginger Schultheis, Jack Spencer, and Kirsi Teppo.

Lastly, thanks to my family for providing me the opportunities and support for pursuing my passions.

Preface

Metaphysics has always been particularly vulnerable to skeptics. The notions that metaphysicians work with (e.g. *possibility*, *fundamentality*, *existence*) are so basic that it's hard to elucidate them in more basic terms. Many philosophers profess to not even *understand* the central questions of metaphysics. Those who do understand the questions give wildly incompatible answers, with no consensus in sight. Moreover, given that rival metaphysical views are often 'empirically equivalent', it's very unclear what the correct methodology for metaphysics is supposed to be. How are we supposed to *tell* if atoms-arranged-table-wise compose a table or not?

As a practicing metaphysician, I have often wrestled with these doubts. One of the primary aims of this dissertation is to try to respond to these doubts. If it can be shown that metaphysical debates regarding modality (chapter 2), casual powers (chapter 2), time (chapter 3), persistence (chapter 4), mereology and ontological categories (chapter 5), *etc.* should be taken to have *observable*, *empirical* consequences, then not even the most hardcore logical positivist should be skeptical of these debates. If this claim can be substantiated, it would also reveal a new methodological approach to metaphysics. If different views in (say) the metaphysics of modality have empirical consequences (chapter 2), then we may be able to get purely empirical evidence into the nature of modal space, just as we get empirical evidence for ordinary scientific theories.

A second aim of this dissertation is purely epistemological. Formal epistemologists have done a lot of work in justifying various purely formal constraints on epistemic rationality, such as probabilism and conditionalization. However, these purely formal constraints aren't much help in scientific theorizing. Merely having probabilistic credences isn't going to make you a good scientist. Perhaps the central epistemological task with respect to science consists in trying to solve the so-called problem of induction. Why are the inductive inferences that scientists make rational? In the first part of my dissertation, I argue for the negative point that orthodox Humean and Non-Humean views cannot justify induction (chapter 2). I also develop and defend an alternative metaphysical picture that may have the resources to justify induction (chapters 1 and 2). The positive proposal I offer is in many ways incomplete however, and I hope to flesh out the details in future work.

Another exciting area in epistemology with relevance to science is the problem of self-locating belief. Some of the deepest questions in contemporary fundamental physics - concerning the correct interpretation of quantum theory, the apparent fine-tuning of the universe, and the large-scale cosmological structure of the universe - may (surprisingly) all turn on the correct theory of self-locating belief! In the second part of my dissertation, I try to derive the correct theory of self-locating belief from metaphysical principles (chapters 4-6).

PART ONE

The Problem of Induction

Chapter 1: A Humean Non-Humeanism

The world is an astonishingly regular place. Apples fall, planes fly, and bread nourishes. The astonishment only increases when one looks at fundamental physics. Fundamental physical entities *always* seem to behave according to certain elegant, ironclad, mathematically precise rules. This is utterly baffling. Why should this be so?

Standard answers to this question are usually divided into two camps: Humean accounts and non-Humean accounts. Both camps are driven by different core commitments, and as a result, Humeans and non-Humeans end up with very different metaphysical pictures. The goal of this chapter is to provide a *synthesis* of Humean and non-Humean views, which seeks to accommodate the core commitments of both views.

My first task will be to say something about what these core commitments amount to. I do this by briefly going over orthodox Humean and non-Humean views, together with the main objections each side makes to the other (sections 1 and 2). I then go on to develop the corresponding synthesis in three stages. First, I argue that any such synthesis is committed to a specific view regarding the source of causal power in the world (sections 3 and 4). Second, I argue that any such synthesis is committed to a particular view in the metaphysics of persistence (section 5). Third, I argue that any such synthesis is concerning the mereological structure of the world (section 6). Lastly, I argue that none of the objections canvassed in sections 1 and 2 have any force against the resulting synthesis (section 7).

Before we begin, some caveats are in order. Since my goal is to synthesize two very general metaphysical pictures, the Humean picture and the non-Humean picture, I will be operating at a fairly high level of abstraction throughout this paper, especially in sections 1 and 2. Humeans and non-Humeans of course come in many different stripes, and they are each motivated by subtly different considerations. Because of this, when I talk about the views and motivations of "Humeans" and "non-Humeans" in the abstract, I will inevitably not be doing justice to every member of those two categories. Furthermore, in trying to synthesize these two very general pictures, inevitably some aspects of both metaphysical pictures must be given up. This runs the

risk of developing a view which is unattractive to both camps, rather than a view which is attractive to both camps. Ultimately, which reaction is warranted will depend on the details of the resulting synthesis.

1. The Non-Humean Picture

A natural first reaction to the astonishing regularity of the world is to think that things don't just *happen* to behave as they do, but, in some sense, they *have* to behave as they do. This "have to" is cashed out in different ways by different non-Humeans. Non-Humeans typically cash this out by picking one of a family of different modal notions and regarding it as fundamental. One might take *laws* to be fundamental, or *dispositions* or *powers* to be fundamental, or *counterfactuals* to be fundamental, or *necessitation relations* among universals to be fundamental.¹ In other words, non-Humeans are typically *anti-reductionists* about the nomic.

Each of these views faces its own specific *prima facie* problems. By way of introducing these views, I will first briefly go over some of these more specific problems before turning to the central Humean objection which applies to all of these views.

Against the fundamental laws view, Sider (2011: 22) has objected that a sharp distinction between laws and non-laws is nowhere to be found in physics. For example, Sider asks us to consider the claims "spacetime is a four-dimensional Lorentzian manifold" or "the universe began with an initial singularity" or "the universe began in a state of very low entropy". These claims are all central to different physical theories, but it is not clear whether these should be enshrined as "laws". Those who take laws to be non-fundamental might plead vagueness in the concept of "law", but those who take laws to be "ontological primitives" seem to be forced to draw a sharp distinction which isn't justified by scientific practice.

Among those who take dispositions or powers to be fundamental, *dispositional monists* or *causal structuralists* hold that all physical properties are fundamentally dispositional or "powerful" (e.g. Mumford (2004), and Bird (2007b)). On this view, properties are "pure powers", which is roughly to say that there is *nothing more* to the properties than their capacity to bring about certain effects.

¹ Carroll (1994, 2008) and Maudlin (2007) defend the view that laws are fundamental. Mumford (2004), Bird (2007b), and Shoemaker (1980) defend the view that there are fundamental dispositions/powers. Lange (2000, 2009) defends the view that counterfactuals are fundamental. Armstrong (1983, 1991, 1993), Dretske (1977), and Tooley (1977, 1987) defend the view that necessitation relations among universals are fundamental.

Many philosophers have thought that this sort of picture is ultimately unintelligible. There needs to be more to reality than just a network of "bare powers"; reality needs to have some intrinsic, categorical nature. Philosophers have pressed this point by arguing that it is hard to see how we can capture the nature of any causal power without referring to properties that are not themselves causal powers.² Goff (2017) puts the point as follows:

[T]he nature of a causal power is given with reference to its manifestation, that is, the property it gives rise to. For example, the manifestation of *fragility* is *breaking*; the manifestation of *flammability* is *catching fire*. Arguably the manifestation of a causal power is a categorical property . . . The obvious response for the causal structuralist is to deny that the manifestation of a causal power is a categorical property. Why can't the manifestation of a given causal power be another causal power? The problem is that if the manifestation of causal power F is itself a causal power—call it "G"—then we can only understand the nature of G by understanding its manifestation—call it "H." If H is also a causal power, then we can understand its nature only by understanding its manifestation that is not itself a causal power, we will never reach an adequate specification of the nature of F. (137-138)

In response, one could certainly posit extra fundamental intrinsic categorical properties distinct from these irreducible powers. However, the resulting view offers us a fragmented picture of reality: all of the causal work is done by the irreducible powers and the intrinsic, categorical properties of physical entities might as well be anything at all. Such categorical properties are mere explanatory danglers that don't do any work.³

On the fundamental counterfactuals view, Lange (2009) argues that one ought to understand lawhood as follows:

² Some philosophers that have defended this point include Russell (1927), Campbell (1976), Robinson (1982), Blackburn (1990), Armstrong (1997), Heil (2003), Lowe (2006), Goff (2017). See Bird (2007a) for a response to this objection.

³ There is also the *powerful qualities* view, which is the view that the causal powers of objects are *identical* to the categorical properties of objects (e.g. Martin and Heil (1999), Heil (2003, 2012), and Strawson (2008)). I will return to this view in section 3.

m is a law if and only if for any conversational context, and for any *p* that is relevant as a counterfactual antecedent in that context and logically consistent with all of the laws (taken together), the propositions expressed in that context by $p \square \rightarrow m$ is true. (15)

In response, many critics have objected that counterfactual conditionals seem to be much too slippery, vague, and generally semantically ill-behaved to ground the fundamental laws of nature (e.g. Hall (2011), Carroll (2011), and Demarest (2012)). Carroll gives voice to this objection when he writes, "On the face of it, [Lange's account] is troubling. What regularities are laws is tied to whether certain sentences are true in all contexts. This is startling; I find myself imagining Kip Thorne phoning Robert Stalnaker for a consult on the laws of quantum gravity!" (14).⁴

On the universals account, laws of the form "All Fs are Gs" are grounded in a fundamental "necessitation relation" N relating the universals of F-ness and G-ness. Besides committing itself to a controversial ontology of properties, Maudlin (2007) has objected to this view on the grounds that actual fundamental physical laws do not naturally come in the form "All Fs are Gs". In investigating the structure of the laws of physics, Maudlin writes:

An appropriate place to begin is with some real scientific theories rather than with cooked examples of the 'All ravens are black' variety. Let us look at some laws without formal prejudices and see what we find. The fundamental law of Newtonian mechanics, the mathematical consequence of Newton's first two laws, is $\mathbf{F} = \mathbf{ma}...$ The fundamental law of non-relativistic quantum mechanics, Schrodinger's equation, is ih $\partial/\partial t |\psi\rangle = H |\psi\rangle$. No doubt these can be tortured into a form similar to (x)(Fx \supset Gx), but it is hard to see what the purpose of the exercise would be. (11)

Bracketing all of these more specific worries, the central Humean objection against all of these accounts is that they traffick in unreduced, modal primitives. These modal primitives then give rise to brute necessities, which the Humean strongly wishes to avoid. Consider a world with only two electrons. According to the necessitation relation theorist, in the absence of any such relations, these electrons may take any trajectories they want to. However, given the appropriate necessitation relation among universals, it is *metaphysically necessary* that the electrons take very

⁴ For a response to this criticism (and others), see Lange's reply in Woodward et al. (2011). Demarest (2012) suggests formulating the view in terms of propositions rather than sentences, but she ultimately argues that such a reformulation predicts the wrong results for certain nested counterfactuals.

specific, mathematically precise trajectories. But how does *N* manage to secure these necessities? Lewis (1983) famously writes:

The mystery is somewhat hidden by Armstrong's terminology. He uses 'necessitates' as a name for the lawmaking universal N... But I say that N deserves the name of 'necessitation' only if, somehow, it really can enter into the requisite necessary connections. It can't enter into them just by bearing a name, any more than one can have mighty biceps just by being called 'Armstrong'. (366)

Without a story that explains *how N* manages to necessitate what it does, the universals theorist is committed to unexplained, brute necessities. Similar things can be said about every other view. The fundamental law theorist also believes that, in the presence of an appropriate law, these electrons *have* to follow very specific trajectories, but no story is given about how these laws can accomplish such an extraordinary feat. Why should the electrons "care" about these metaphysical primitives? On the fundamental counterfactuals account, these modal facts are just posited as fundamental in the form of fundamental, unexplained counterfactuals. Lastly, the fundamental disposition theorist faces a similar problem. In saying that the dispositions or powers of things are *fundamental*, one is saying that these powers are unexplained.⁵

Another important Humean objection goes by way of *Hume's Dictum*. Here is a standard formulation of Hume's Dictum from Wilson (2010):

HD: There are no metaphysically necessary connections between wholly distinct, intrinsically typed, entities.⁶

The restriction to "intrinsically typed" entities is to avoid counterexamples having to do with relationally characterized entities. For example, the existence of a planet might necessitate the existence of a corresponding sun, or the existence of a brother might necessitate the existence of a corresponding sibling. For two entities to be "wholly distinct" is for them to be distinct entities that do not have any mereological or spatio-temporal overlap. Clearly, there are necessary

⁵ For a recent response to these sorts of challenges on behalf of the Non-Humean, see Schaffer (2016).

⁶ As formulated here, HD is vulnerable to certain sorts of counterexamples that some Humeans would want to allow for (e.g. the necessary connection between Socrates and his singleton set). These sorts of counterexamples will not be relevant for our purposes. For a wide-ranging critical discussion of various different formulations of Hume's Dictum, see Wilson (2010).

connections between an entity and itself (e.g. Necessarily, if A is F, then A is F), and there are also necessary connections between entities that overlap (e.g. Necessarily, if half of an apple is red, then the whole apple must be red-on-one-half). The Humean even acknowledges that there might be not-so-trivial necessary connections between an entity and itself: for example, an object's having 5 units of mass necessitates that it does not have 7 units of mass.

Again, it is easy to feel the intuitive force behind HD by considering some examples. Suppose there were two lights such that it was *metaphysically necessary* that for any time *t*, the first light is on at *t* iff the second light is on at *t*. This connection cries out for explanation. How can the second light even "know" about the first light? Even if it could "know" about the first light, why is it *impossible* for the second light to be out of sync with the first light? The necessary connections between the trajectories of our two electrons seem to similarly cry out for explanation.

2. The Humean Picture

In contrast to Non-Humeans, Humeans eschew any primitive modal notions. Lewis sums up the Humean view as follows:

[A]ll there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another . . . For short: we have an arrangement of qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that." (1986: ix-x).

Since there are no primitive modal notions constraining these local qualities, the instantiation of any quality at a space-time point puts *absolutely no constraints* on what other qualities are instantiated elsewhere in the Humean mosaic. Humeans typically accompany this metaphysics with a fairly deflationary account of lawhood. For Humeans, laws are mere summaries of the distribution of qualities, albeit very simple and informative such summaries.⁷

Recall the question we started off with: why is it that the world is such a regular place? The central non-Humean objection to the Humean view is that the Humean view provides no explanation for this central datum. For the Humean, it is mere coincidence that the world is regular; we just lucked

⁷ This account is precisified in the "best system" account of lawhood. See Lewis (1973: 73-77, 1999: 8-55, 224-247) for early versions of the view. For different contemporary formulations of the best system account, see Schrenk (2006), Cohen and Callender (2009), and Dorst (2017).

out.⁸ Typically non-Humeans attempt to explain the regularity of the world by appeal to fundamental laws, causal powers, counterfactuals, etc. But Humeans typically *reduce* all such notions to the pattern of instantiation of local qualities. Facts about laws, dispositions, causation, and counterfactuals all ultimately hold in virtue of the distribution of local qualities. Consequently, many non-Humeans have argued that it is objectionably circular for the Humean to try to explain the distribution of qualities by appeal to such nomic notions, since all of these nomic notions are themselves explained by the distribution of qualities. How could a *mere summary* of the distribution of local qualities explain why the distribution of qualities is as it is?⁹

Sometimes this worry is given a more straightforwardly empirical gloss. If a metaphysical theory's predictions go against observational fact, then that is very strong reason to think that the metaphysical theory is false. Many non-Humeans have argued that Humeanism is incompatible with our inductive predictions about the future; if Humeanism were true, we should expect the world to be entirely irregular. Given that it is an observational fact that the world is an exquisitely ordered place, this gives us strong empirical reason to reject Humeanism.

This worry may be precisified in a number of different ways. For example, many non-Humeans have argued that inductive inference is only valid as an instance of inference to the best explanation.¹⁰ The basic thought is that, given various observational facts that fit a very general pattern, one may use inference to the best explanation to conclude that there must be some corresponding law that explains the pattern. Having inferred the existence of such a law, one may then appeal to the law to predict unobserved instances of the law. However, given the fact that Humean laws cannot explain their instances (which the non-Humean typically accepts), Humeans are not able to make any such inferences to the best explanation, and hence they are not justified in making the inductive predictions that the non-Humean is licensed in making. Given that the

⁸ Armstrong (1983) makes this charge when he says that the Humean view must consider laws to be a "cosmic coincidence" (161). Strawson (2014) writes, "According to the [Humean], then, the regularity of the world's behaviour is, in a clear sense, a complete and continuous fluke" (23). Chalmers (2012) also writes that "the hypothesis that our world has only [Humean] laws is coherent, but I think that it is extremely unlikely, as it in effect renders our own world a cosmic-coincidence world on which pervasive regularities are not explained by anything more basic than themselves" (338).

⁹ Non-Humeans who have defended this argument include Maudlin (2007: 170-183), Lange (2013, 2018), Shumener (2017), and Emery (forthcoming). Humeans who have responded to this argument include Loewer (2012), Dorst (forthcoming), and Bhogal (forthcoming).

¹⁰ Dretske (1977: 261-262), Armstrong (1983: 52-59), and Foster (1983, 2004) have all argued for this connection. In response, Beebee (2011) has argued that non-Humean approaches are no better than Humean approaches to induction. See Hildebrand (2018) for a non-Humean response.

non-Humean inductive predictions have been extremely well confirmed, this is extremely strong empirical evidence in favor of non-Humeanism.

Other objections to Humeanism have revolved around the fact that Humeanism does considerable violence to our pre-theoretic views regarding laws, counterfactuals, chance, and other such nomic concepts; views which philosophers and working scientists alike share. Let us close by considering two such examples of this phenomenon.

First, take laws. Tooley (1977: 669) asks us to consider a world with 10 different kinds of particles. In such a world, there would be fifty-five different kinds of two-particle interactions. Suppose that in the history of the universe only fifty-four of these two-particle interactions have been realized. We might naturally wonder: what would have happened had the last pair of particles interacted? We are naturally inclined to think that there must be an answer to this question, but Humeans cannot supply such an answer. Since there is no regularity in the world corresponding to their interaction, there is no corresponding law about their interaction.

Second, take objective chance. Humeans think the chances that figure in the laws are what they are because they accurately match the global frequencies of various events, subject to the constraint of simplicity and informativeness. However, unlike Humeans, we naturally think that it's possible that the chances of various events can drastically come apart from the associated frequencies. This gives rise to the so-called *problem of undermining futures*. Lewis was well aware of the problem that this posed for Humeanism. He writes:

There is one big bad bug: chance. It is here, and here alone, that I fear defeat. But if I'm beaten here, the entire campaign goes kaput. (1986: xiv)

Consider the following case. The half-life of a tritium atom is 12.26 years. This means that a newly created tritium atom has an objective chance of 0.5 of decaying in 12.26 years. We are naturally inclined to think that the objective chances might in fact be like this, even if it turns out that all future tritium atoms decay extremely quickly. Let *H* be a total description of the history of the universe according to which every newly created tritium atom in the future decays in under one minute. Letting *t* be the present time, the objective chance of *H* at *t* will be some extremely low, but non-zero number, call it *c*. So, $Ch_t(H) = c$. Given that rational agents ought to align their credences with the known chances, it follows that, for every rational prior credence function C_0 ,

 $C_0(H | Ch_t(H) = c) = c$. Humeans cannot account for this data point. Given the Humean account of chance, $C_0(H | Ch_t(H) = c) = 0$. According to Humeans, if *H* were to obtain, the half-life of tritium atoms would be much *less* than 12.26 years, since the objective chances cannot come too far apart from global frequencies. But, if the half-life of tritium atoms was much less, then $Ch_t(H)$ would no longer be equal to *c*. Given the fact that *H* and $Ch_t(H) = c$ are *a priori* incompatible on the Humean analysis of chance, Humeans must set $C_0(H | Ch_t(H) = c) = 0$. This violates the truism that rational agents ought to align their credences with the known chances.¹¹

3. Powerful Qualities

I have canvassed some *prima facie* problems for standard Humean and non-Humean views. On their own, certainly none of these problems are fatal to any of these views, and there are ongoing debates surrounding many of these problems. However, it would be nice if we could come up with a view that simply avoided all of the objections we have encountered. Fortunately, I think we can.

What would such a view have to look like? Well, in order to abide by Humean scruples, the view could only work with Humean-friendly resources, namely the qualities themselves. The view would have to avoid commitment to primitive laws, necessitation relations, dispositions, and other "spooky" non-Humean what-nots. But, if we only have the qualities to work with, how can we possibly explain why things do what they do? How can we explain why apples fall, planes fly, and bread nourishes? It would somehow have to be that the *qualities themselves* fully explain why things do what they do.

Such a view, however, seems to be a non-starter. Qualities, unlike dispositions or powers, are supposed to be entirely non-modal properties. Suppose we thought that *being red* was an example of a quality.¹² It's hard to see how an object's being red could explain why it does what it does. In other words, it intuitively seems like qualities are *powerless*: all on their own, they can't explain why their bearers do what they do. This is why the Humean wants to say that the instantiation of

¹¹ Humeans have responded to this sort of challenge in different ways. For a representative survey, see Lewis (1994), Hall (1994, 2004), Roberts (2001), Hoefer (2007), and Ismael (2008). For a comprehensive criticism of these responses, see Briggs (2009).

¹² Of course, there is much controversy about the actual natures of colors. For example, those who identify colors with certain sorts of dispositions will deny that colors are qualities (e.g. see Peacocke (1984), Johnston (1992), Levin (2000)). For the purposes of this example, I have in mind a sort of naive realism about colors, according to which colors are objective, intrinsic, non-reducible qualities of objects (e.g. see Watkins (2005), Campbell (2005), Gert (2006, 2008), and Allen (2017)).

a quality at some time and place puts absolutely no constraints on what qualities are instantiated elsewhere and elsewhen. Nonetheless, there at least seems to be some theoretical space for a view according to which the fundamental qualities themselves are *powerful*: all on their own, they can fully explain why their bearers do what they do. Let us call such a view the Powerful Qualities View (PQ).¹³ PQ can be succinctly described in the following slogan: *things do what they do because of what they're like*.¹⁴

The easiest way to get a grip on PQ is by way of a concrete example. In the philosophy of mind, a number of different philosophers have been attracted to *Russellian Panpsychism* (e.g. Russell (1927), Strawson (2006), Chalmers (2015), Goff (2017)), which is the view that (at least some of) the categorical properties of fundamental physical entities are phenomenal properties (i.e. properties that characterize what it's like to be a certain entity). The *Phenomenal Powers View*, defended in different ways by Mørch (2014, 2018, forthcoming-a, forthcoming-b), Goff (2017, forthcoming), and Langsam (2011), is the view that categorical phenomenal properties may ground certain physical dispositions purely in virtue of their intrinsic phenomenal character (i.e. what they feel like).¹⁵

The Phenomenal Powers view is, of course, a very controversial view. However, it is a view that panpsychists should take very seriously. Since panpsychists already have independent reason to posit fundamental phenomenal qualities at the most basic level of reality, they all face an important choice point. Are these basic phenomenal qualities powerful or powerless? The powerful view promises to integrate categorical, phenomenal qualities much more tightly into the causal network of physics than the powerless view.

While a full defense of the Phenomenal Powers view is far beyond the scope of this paper, some initial motivation for the view can easily be found by reflecting on our own phenomenological experiences. It is natural to think, for example, that *pain* makes subjects who experience it try to

¹³ In some respects, this terminology is unfortunate, because the view is somewhat different than an already existing view with the same name. I'll address this terminological matter at the end of this section.

¹⁴ Here one might object: aren't the bearers of the qualities space-time points on the Humean view? Space-time points can't be disposed to "behave" in any way since they don't even persist through time! This objection will be addressed in sections 5 and 6, when the other two components of the synthesis will be developed. For the purposes of this section, I only want to get PQ on the table. The synthesis as a whole will be assessed in section 7.

¹⁵ Importantly, Goff (2017, forthcoming) argues that only "consciousness +" properties - properties which have both phenomenal and non-phenomenal aspects - necessitate their corresponding effects.

avoid it simply in virtue of how bad it feels. *Pleasure* might make subjects who experience it try to pursue it simply in virtue of how good it feels. Someone who feels *tired* or *exhausted* might be disposed to stay in bed simply in virtue of how tired they feel. These connections between these experiences and their effects are certainly *defeasible*. Of course, in our own case, human subjects endure pain for all sorts of reasons. Someone might endure short-term pain in order to avoid more pain in the future (e.g. going to the dentist), or in order to experience a greater pleasure (e.g. masochism), etc.¹⁶ However, in the absence of *any* interfering causes like these, it seems like pain disposes subjects to at least *try* to avoid it. Similarly, throwing a ball at a window might cause the window to break, but only in the *absence* of any interfering causes. Of course, the wind might blow the ball away before it hits the window, or the window might be shattered by a rock before the ball gets to it, etc.

One needn't be a panpsychist to find the Phenomenal Powers View intuitive. Even extremely deflationary views about consciousness - such as *behaviorists* and *analytic functionalists* - have recognized a very close (even analytic!) connection between these mental properties and their effects. It is no coincidence that *epiphenomenalism* - the view that mental properties have no effects at all - is widely regarded as implausible.

Having gone through this example, it should be stressed that *none of the arguments in this paper presuppose the truth of Panpsychism or the Phenomenal Powers View*. In fact, one can perfectly well think that both views are extremely implausible while still endorsing PQ. Why can't non-phenomenal qualities be powerful too? As long as one finds the Phenomenal Powers View intelligible, then one should find PQ to be intelligible in general. I only chose to discuss the panpsychist version because phenomenal qualities like pains and pleasures are especially familiar and intuitive examples of (allegedly) powerful qualities. Because of its familiarity, I will continue to use pain as my toy example of a powerful quality throughout this paper. I will let "PQ" stand

¹⁶ Another common objection to the claim that pain essentially disposes one to avoidance behavior is the phenomenon of pain asymbolia. Those who have pain asymbolia may experience pain without the corresponding feeling of unpleasantness, so they are not motivated to avoid the pain. For this objection to go through, it would have to be maintained that pain experiences for those who have pain asymbolia have the *very same* phenomenal character as ordinary experiences of pain. However, this is doubtful. It is natural to think that the *unpleasantness* of pain is part of the phenomenal character of ordinary experiences of pain. On this view, what the phenomenon of pain asymbolia shows is that ordinary phenomenology of pain has two components, a sensory component and an affective component. For those who have pain asymbolia, these two components come apart, and they only experience the sensory component without the affective component. For much more on this, see Grahek (2007).

for the general Powerful Qualities View and "PPQ" stand for the panpsychist version of the Powerful Qualities View.

Before going further, an exegetical point should be made. Several philosophers have already argued for a view associated with the name "Powerful Qualities", including Heil (2003, 2010, 2012), Martin (2008), Martin and Heil (1999), and Strawson (2008). The view I am defending here is very much in the spirit of the views of these other philosophers. However, the main claim associated with these philosophers seems to me to be false. The main claim of these philosophers is that the dispositions of an object should be *identified* with the qualities of that object.¹⁷ However, according to PQ, this claim must be false. According to PQ, fundamental physical entities have their dispositional properties *in virtue of* their qualities. The qualities of objects *fully explain* their corresponding powers. This relation of asymmetric dependence between an object's dispositions and its qualities is incompatible with the claim that they are literally identical.¹⁸

4. Three Roadblocks to Powerful Qualities

Having put PQ on the table, we should ask: why isn't PQ a far more popular view? PQ can almost seem like a piece of common sense: things do what they do because of what they're like. Of course, how could it be otherwise?

I think the first explanation for why PQ hasn't proven to be popular has to do with very general features about how explanation works in science. The standard way we explain why things do what they do in higher-level sciences is by taking for granted various facts about the causal powers of things in a more fundamental science. For example, biologists might explain why biological cells do what they do by taking for granted all sorts of facts in chemistry, and chemists might explain why certain chemical reactions occur by taking for granted all sorts of facts in physics. This pattern is pervasive in non-scientific cases as well. We might explain why a car manages to run, for example, by taking for granted various facts about how the car's engine interacts with other parts of the car. But, when we are pushed all the way to fundamental physics, we can no

¹⁷ For example, Martin and Heil write, "Our suggestion is that dispositionality and qualitatively are to be identified" (1999: 47). Later, Martin writes, "the qualitative and dispositional are identical with one another and with the unitary intrinsic property itself" (2008: 65). For Strawson (2008), this claim is right in the title of his paper "The Identity of the Categorical and the Dispositional".

¹⁸ Although most philosophers would agree that that in-virtue-of explanations are irreflexive, see Jenkins (2011) for a defense of the view that such explanations can be reflexive.

longer use this strategy. So, why do fundamental physical entities do what they do? Why does, say, the quantum wave function of the universe evolve through time as it does? Given that standard explanatory strategies fail here, many non-Humeans have thought that there must be some primitive metaphysical posit that somehow *makes* the fundamental physical entities do what they do (e.g. laws or necessitation relations among universals), or otherwise it must be that it's just an unexplained, brute fact that the fundamental entities have certain dispositions to behave in certain ways (e.g. fundamental dispositions/powers). On the other hand, Humeans are content in saying that there really are no constraints at all on how the fundamental stuff in the universe is arranged, and so there is no need to answer this question.

According to PQ, we need not swallow any primitive modal notions in order to explain why the fundamental entities do what they do. The fundamental entities simply do what they do because of what those entities are actually like! To get a better grip on this sort of reply, it is helpful to consider an analogy with the problem of other minds. The problem of other minds is the problem of how we can come to know about the existence and nature of the mental lives of others, given that we can only observe their corresponding behaviors. Here, many of us feel comfortable attributing various mental qualities to others as a sort of inference to the best explanation. Attributing such mental qualities as *pain*, *hunger*, *happiness*, and *tiredness* to agents lets us explain why other agents do what they do in a satisfying way. The situation we find ourselves in with respect to explaining the behavior of fundamental physical entities is, according to PQ, highly analogous.¹⁹ In the problem of other minds, we are unable to directly observe the mental qualities of others, but we can reasonably infer their existence by positing them to explain various bodily movements that we can observe. Similarly, in the case of physical entities, we can't "directly observe" the fundamental qualities in nature. For example, a region of space with a stronger electromagnetic field "looks" the same as a region of space with a weaker electromagnetic field. Nonetheless, we can infer that the fundamental qualities instantiated in those regions of space must be different by, for example, observing the effects of those fields on various magnetic objects. Just as we may naturally think that the quality of *pain* can explain certain dispositions concerning

¹⁹ Of course, the situation is far from completely analogous. In the problem of other minds, we can also use an argument from analogy (since we are conscious and others are relevantly like us, they must be conscious to). This argument from analogy starts becoming more problematic when attributing mentality to things very unlike us (e.g. various non-human animals, plants, artificial intelligence, etc.).

avoidance behavior, we should also think that the fundamental qualities instantiated in our regions of space (which may be mental according to the panpsychist) fully explain the causal powers of that region of space.

The big difference between the two cases is, of course, that we have a much better positive grasp on the nature of mental qualities instantiated in other human beings than we have a positive grasp on the fundamental qualities instantiated in our regions of space. Perhaps we can never really know what these fundamental qualities are like. After all, many philosophers have been tempted by the view that science can only ever tell us about the causal *structure* of the world, and not how the world is "in itself".²⁰ Nevertheless, according to PQ, we should think that the fundamental qualities *beneath* this causal structure have a powerful nature. Just as anyone who grasps what is to be in pain can see why that quality might dispose one towards certain avoidance behaviors, we should also believe that the intrinsic natures of the fundamental physical qualities, whatever they are, fully explain their corresponding causal powers.

A second explanation for why PQ has not proven popular among contemporary metaphysicians has to do with how the distinction between the categorical and the dispositional is usually drawn. Unfortunately, many philosophers have characterized the distinction in merely modal terms. Traditionally, philosophers tended to think that some property is a dispositional property just in case the instantiation of that property entailed certain counterfactual conditionals (e.g. *x* being *F* should entail that *x* would produce a certain manifestation given a certain stimulus condition).²¹ Some more modern ways of drawing the distinction don't appeal to counterfactuals but they are still merely modal. Bird (2007b) draws the distinction as follows:

What we mean by 'categorical' must be understood in negative terms. That is, a categorical property does not confer of necessity any power or disposition . . . The categorical versus

²⁰ Langton (2001, 2004) defends an interpretation of Kant in which we can never have knowledge of things "in themselves". Contemporary Russellian Monists follow Russell (1927) in endorsing a similar conclusion. Lewis (2008) also defends a similar claim. Many who identify as "structural realists" in the philosophy of science are also sympathetic to this claim. For an overview of Structural Realism, see Ladyman (2016).

²¹ Ryle (1949), Goodman (1954), Quine (1960), Mackie (1973), Prior (1985), and Mumford (1998) all endorse accounts of this kind. For a survey of some different challenges to counterfactual analyses of dispositions, see Choi and Fara (2018).

essentially dispositional distinction is a modal one. To say that a property is categorical is to deny that it is necessarily dispositional. (66-67)

Working with these definitions closes off any theoretical space for PQ. According to these definitions, PQ simply collapses into a version of the fundamental dispositions view, since fundamental powerful qualities would count as fundamental "dispositions". But according to PQ, dispositions are emphatically not fundamental; they hold in virtue of the qualities of things! Recall that on the dispositional monist or causal structuralist view, the properties of fundamental physical entities are "bare powers". In other words, there is nothing more to these properties than their capacity to bring about certain effects. According to PPQ, however, phenomenal properties have their capacity to bring about their effects in virtue of their phenomenal character. Contra behaviorists, there is much more to pain than simply being disposed towards avoidance behavior! It is simply a mistake to classify how pain actually feels like as a "disposition". Sure, it might ground, and hence necessitate, a certain disposition if PPQ is correct (absent defeaters), but that doesn't mean that it should count as a "disposition" itself. These merely modal ways to draw the distinction between the categorical and the dispositional miss the mark. As it has done in so many other places, this is another place where the so-called "post-modal" revolution in contemporary metaphysics - the shift towards more fine-grained theoretical tools like *fundamentality*, grounding, and *essence* - should change how the basic terms of the debate should be set up. 22

Partly because of the prevalence of these merely modal ways of characterizing "categorical", I have abstained from calling PQ the "Powerful Categorical View", because for many philosophers that would be a contradiction in terms. However, using "quality" instead of "categorical" invites the question of what is exactly meant by "quality". Although I am not sure how to analyze the notion in even more basic terms, much can be done to clarify the notion. Qualities are both intrinsic and qualitative (as opposed to haeccaetistic) properties. The *essential nature* of a quality is entirely non-modal and non-potential; qualities only serve to specify what an object is *actually like*, rather than *what it does*. In this sense, it is a notion that Humeans should feel perfectly comfortable with. According to PQ, the essential nature of a quality might necessitate certain dispositions, but following Fine (1994), we should think essence is more fine-grained than mere necessity. Lastly,

²² For much more on how the post-modal revolution changes various metaphysical debates, see Nolan (2014) and Sider (forthcoming).

we have a concrete example of PQ, namely PPQ, because phenomenal properties are paradigm examples of qualities, namely mental qualities. Ultimately, I take it that the notion of an object's qualities is a deeply intuitive one; for example, it is a notion that Lewis (1986: ix) easily felt that he could use in informally characterizing his own metaphysical views.

Let us lastly turn to a third possible explanation for why PQ has proven to be unpopular amongst contemporary metaphysicians. This third explanation boils down to the intuition that powerful qualities are just impossible. This is certainly a worry that needs to be taken seriously. How could facts about what something is *actually like* fully ground *modal* facts about how it is disposed to behave in various circumstances?

I have three quick responses to this worry. First, I think at the end of the day we should believe in the possibility of powerful qualities if they figure into the best, most explanatory theory regarding the lawful regularity of the world. The overarching goal of this paper is to construct and defend such a theory. If a theory that uses powerful qualities can avoid all of the objections that we canvassed in sections 1 and 2, then that is a strong reason to believe in the existence of powerful qualities. Second, it should again be noted that many philosophers have argued that we are *entirely ignorant* about the fundamental, intrinsic nature of reality (e.g. see footnote 23). If we really have no idea what the fundamental physical qualities are like, on what grounds can we be so confident that they are powerless as opposed to powerful? It seems like we simply have no direct evidence either way. It is somewhat ironic that Lewis (2008) himself argued that we are entirely ignorant as to the nature of these fundamental qualities, yet it was crucial for his Humeanism that these qualities were entirely powerless as opposed to powerful. Third, although I have not defended this at any length here, it seems to me that there is a strong case to be made that the qualities that we are most intimately acquainted with, phenomenal qualities like exhaustion, hunger, sexual and romantic attraction, pain, and pleasure, are all powerful qualities. Again, the reader is referred to the work of Mørch (2014, 2018, forthcoming-a, forthcoming-b) and Langsam (2011) for a defense of this view.

5. From PQ to Endurantism

I have said that powerful qualities can be used to explain the behavior of their bearers. Crucially, they do not explain the behavior of anything *other* than their bearers. If we are to satisfy HD, there

can be no necessary connections between wholly distinct existences. It is perfectly consistent with HD for powerful qualities to ground certain dispositions in their bearers. This is not a necessary connection between distinct existences. Everyone, even Humeans, agree that there can be necessary "connections" between an entity and itself (e.g. necessarily, if x has exactly 5 units of mass then x does not have exactly 7 units of mass). However, if powerful qualities were to ground certain dispositions in some other wholly distinct object, this would be a flagrant violation of HD.

This restriction should be independently plausible anyway. My being in pain might explain *my* avoidance behavior. But how could the painfulness of my pain, all by itself, make someone *else* behave in any particular way? How could this wholly distinct person even "know" that I'm in pain in the first place? It doesn't matter how close or far away this other person may be, my pain can only make me behave in certain ways. Suppose we considered the mereological fusion of that other person, call her Alice, with me. Does this help? No. The mere fact that we are both spatial parts of some mereological fusion does nothing to avoid violations of HD. As long as it is *me* that is in pain and not Alice, then there would be violations of HD.

All of this also goes for the temporal dimension. Because of this, PQ (together with HD) entails Endurantism. Below, I will go through the three main views in the metaphysics of persistence -Endurantism, Perdurantism, and Exdurantism - and argue that this is so.²³

Enduring objects persist by being wholly present at every time in which they exist. They are threedimensional, and they lack temporal parts.²⁴ For the Endurantist, the physical properties that are responsible for the fundamental causal powers of things are instantiated *at a time*. For example, a particle might have a certain mass or charge at a time, or a field might have some particular value at some space-time point. By HD, the qualities had by (say) a point particle at a time can only constrain the behavior of *that very point particle*. It cannot explain the behavior of any distinct point particle. So, if the qualities had by a point particle at a time are to constrain the future in any way, it must be that *that very point particle* exists in the future. After all, qualities can only explain the behavior of their bearers. This is all fine for the Endurantist. The Endurantist can just believe

²³ For an overview of the persistence debate, see Haslanger and Kurz (2006) for a collection of influential papers on the subject.

²⁴ For our purposes, this informal characterization of Endurance will suffice. Many philosophers have developed more formal characterizations, however. See, for example, Sider (2001: ch 3) for a widely used characterization of Endurantism and Perdurantism.

that the bearer of the powerful qualities endures through time. The *very same* entity that had the quality at this time will go on to exist at future times. No violation of HD.

Among so-called "four-dimensionalists", there are Perdurantists and Exdurantists. Although the fundamental metaphysics underlying these two theories are similar, they have different accounts of what it takes for objects to persist through time. Perduring objects are supposed to be fourdimensional objects that are spread out through time, and they persist by having different temporal parts at different times. Just as I exist at different spatial locations by being spread out in space, four-dimensional "worms" achieve the same feat by being spread out in time. However, Perdurantists typically follow Lewis in saying that the fundamental qualities are had by objects that are not spread out through time, and for good reason. Again, the fundamental properties that physics describes aren't had by "spread out" objects; they are had by instantaneous objects (like space-time points). A four-dimensional worm might have properties like *having mass x at time t*, but these properties are always had in virtue of the fact that the temporal part of that worm at t has mass x simpliciter. These "simpliciter" properties are responsible for the fundamental causal powers of things. However, the instantaneous objects that have these fundamental properties emphatically do not persist into the future according to the Perdurantist. This is the problem. Given that fundamental qualities can only explain the behavior of their bearers, and the bearers of these fundamental qualities are instantaneous objects that do not persist into the future, no fundamental qualities can constrain the future in any way according to the Perdurantist.

Next, consider Exdurantism. Exduring objects are three-dimensional objects that persist into the future by having "counterparts" that exist at later times. These counterpart objects are numerically distinct objects that are in some way similar to, or continuous with, the original object in question. It is not hard to see how this view faces the same objection as the Perdurantist. Again, let o be an object that bears a fundamental causal power by bearing some fundamental quality. If the causal power had by o is to constrain how things will be in o's future, then it must be that o exists in o's future, since o can only constrain o by HD. However, the Exdurnatist only believes that there are wholly distinct (i.e. numerically distinct and non-overlapping) counterparts of o in the future. One cannot violate HD just by saying that the relevant necessary connections between wholly distinct existences involve "counterparts".

The intuition behind this argument against perdurantism and exdurantism can be brought out by continuing with the analogy of our pain case. Suppose I'm in pain right now. The thing that's in pain isn't some four-dimensional worm, it is a particular temporal part of that worm: it is me, here and now. Similarly, if we were to go back and consider our spatial mereological fusion of me with Alice, call it Alice*, then Alice* isn't the thing experiencing the pain. I am the one experiencing pain. Just as it would involve an HD-violating necessary connection if *my* pain were to make Alice behave in various ways (even if we are both parts of Alice*), it would involve an HD-violating necessary connection if the pain of this particular temporal part made some future *wholly distinct* temporal part(s) behave in various ways (even if all of these temporal parts are part of some larger space-time worm).

Thomson (1983) famously argued that four-dimensionalism is a "crazy-metaphysic" because it implies that "new stuff... keeps constantly coming into existence *ex nihilo*" (213). This is exactly what's happening here. Since numerically distinct subjects of experience are constantly coming into and going out of existence on the four-dimensionalist picture (a four-dimensional worm isn't a single subject of experience!), the powerful qualities of each of these infinitely many subjects of experience have no time to explain anything. Their bearers only exist for an instant - different ones keep on popping into and out of existence! However, on the endurantist picture, since one and the same subject endures through time, powerful qualities perfectly well have the chance to explain the behavior of their bearer.

In wrapping up this section, it should be pointed out that Haslanger (1989) has made a similar sort of argument, although she doesn't explicitly appeal to HD or PQ. Haslanger argues that if the past is to genuinely constrain how the present is, Endurantism must be true. The main idea behind her argument is similar: "the past can get a hold on the present only through things presently existing. If an object does not exist at t, then it itself cannot 'make demands' on things at t" (17). She summarizes her conclusions as follows (by "persists" and "persistence" she means "endures" and "endurance"):

...if nothing from the past persists to the present, the past can set no constraints on the present; the "causal message" cannot be communicated across the gap. Thus, on the [four-dimensionalists] model of change, because the past is causally ineffective, it is "from the point of view of the present" nothing. From this we can see that persistence does provide

us intelligibility in explanations of change. Natural explanations work by showing the systematic causal interconnections between things. Without persistence, the causal story becomes unconnected; neither the past nor the future can get a hold on the present in a way that is causally efficacious. (21)

I entirely agree with Haslanger. Only Endurantism can render the causal connections in the universe intelligible.

6. From PQ to Monism

There is a dispute in contemporary metaphysics about whether the world is fundamentally built out of an uncountable infinity of "tiny" things (e.g. extensionless space-time points or point particles) or whether the world is fundamentally a single, unified whole. While orthodoxy endorses the former position, Monists endorse the latter position. Schaffer (2018) distinguishes between two different types of Monists, *Existence Monists* and *Priority Monists*. Existence Monists believe that there is only a single entity, the Cosmos. Priority Monists believe that there is only one fundamental entity, the Cosmos. However, Priority Monists also typically believe that there are various non-fundamental entities, which are all parts of the Cosmos. Let *Monism* (simpliciter) be neutral between these two views; that is, let Monism be the view that (at least fundamentally) the only entity is the Cosmos. The third aspect of my synthesis will be a commitment to Monism. I will remain neutral between Existence Monism and Priority Monism.

Recall the explanatory ambitions of standard non-Humeans. Non-Humeans don't just want to say that the world *happens* to behave in this extraordinarily regular way. Rather, they want to say that, in some sense, things *had* to behave as they do. Our goal has been to satisfy these explanatory ambitions without violating HD or appealing to brute necessities. It turns out that this can only be done given Monism.

According to the orthodoxy, the fundamental ingredients of reality are things like space-time points or point particles. However, it is just a scientific fact that, say, the behavior of a particular electron cannot be fully accounted for *purely on the basis of the intrinsic nature of that electron*. If I were to put the very same electron in different external environments, it would behave in different ways depending on which environment it was placed in. So, its own powerful qualities cannot necessitate its behavior in any given circumstance. Rather, there must be a necessary

connection between the electron and the external environment it's placed in that makes it take the trajectory it eventually travels. This is a flagrant violation of HD.

However, there is no corresponding problem for the Monist. The Cosmos can perfectly well evolve the way it does *purely because* of its own intrinsic powerful qualities. There are no necessary connections needed between the Cosmos and its "external environment", since there is nothing external to the Cosmos! This echoes a point that has been made by Schaffer (2013): it is the Cosmos as a whole that evolves by the fundamental laws of nature.

Again, this result can be made intuitively plausible independent of any appeal to HD. This is easiest to see on the panpsychist version of the view. From the perspective of PPQ, the orthodox non-Monist picture is one where there are an infinity of wholly distinct subjects of experience being the bearers of fundamental phenomenal qualities (e.g. space-time points or point particles). On this view, it really is mysterious how there could be necessary connections between these wholly distinct subjects of experience - how could one even "know" about any other? However, if the Cosmos as a whole was the bearer of the fundamental, powerful phenomenal qualities, there would be much less of a mystery. Of course a single subject of experience can constrain its own behavior - our own minds influence themselves all the time!

Even bracketing panpsychism, there is some intuitive force to this thought. On the non-Monist picture, the Cosmos is a "mere aggregate" of an uncountably infinite number of tiny (even extensionless!) entities. On this picture, it is natural for the Humean to think that all of these tiny building blocks can simply be mixed and matched in various ways, and any necessary connection between wholly distinct building blocks would be entirely mysterious. However, on the Monist conception, the Cosmos is a single integrated and unified whole. If it does have any parts, they would be "mere abstractions" from the Cosmos. The Cosmos isn't "built up" out of anything that can be "mixed and matched". Rather, the Cosmos simply evolves of its own accord, following its own powerful nature.

So, all is well and good for the Existence Monist. But, there is a wrinkle for the Priority Monist. If the Priority Monist grants that the Cosmos has non-fundamental parts, then there may well be HDviolating necessary connections between those parts. For example, if our electron is one part of the Cosmos and the external environment to the electron is another part, then there will be necessary connections between those two wholly distinct parts of the Cosmos. This is a violation of HD.

In response, the Priority Monist should restrict HD as follows:

HD (fundamental): There are no metaphysically necessary connections between wholly distinct, intrinsically typed, fundamental entities.

This restriction is perfectly compatible with Priority Monism. Moreover, this restriction need not be seen as an *ad hoc* restriction. The motivating spirit behind HD is that necessary connections between wholly distinct existences would have to be brute and inexplicable. This does seem to be true if the existences in question are fundamental. But, if the existences in question are nonfundamental, then perhaps their connections can be explained in terms of the fundamental base that they arise from, without ever appealing to such necessary connections at the most basic level. This is exactly the case in Priority Monism.

7. Taking Stock

At this point, I have completed the exposition of my synthesis of Humeanism and non-Humeanism. The view I have argued for is certainly counterintuitive: all that there is, fundamentally speaking, is an enduring Cosmos that evolves according to its own powerful, qualitative nature. Let us turn to see how this synthesis deals with the objections that standard versions of Humeanism and non-Humeanism face.

Unlike the fundamental laws view, this synthesis does not invoke the notion of a "law" anywhere in the theory, and it is not committed to there being a sharp dividing line between facts that count as "laws" and those that count as "non-laws". The view will treat "laws" similarly to other non-Humean theories that don't take such a notion to be fundamental. Demarest (2015), for example, develops a conception of laws for the "pure powers" theorist that works equally well here.

Unlike dispositional monists or causal structuralists, the picture of the world I have defended is more than just a network of "bare powers". The view therefore avoids the worries voiced by many philosophers that such a view is ultimately unintelligible. Unlike those who posit fundamental categorical properties in addition to irreducible powers or dispositions, the view provides an integrated picture of the world where the fundamental qualities fully explain the powers of things. The view therefore avoids the objection that the fundamental categorical properties of things are mere explanatory danglers that do no work.

Unlike the fundamental counterfactuals view, the view does not invoke counterfactuals anywhere in the theory, and so counterfactuals can be as slippery, vague, and semantically ill-behaved as they are.

Unlike the necessitation relation theorist, the view is not committed to an ontology of universals. Moreover, the view need not think that laws must all come in the form "All *F*s are *G*s".

Unlike all non-Humean views, the view is a *reductionist* view about laws, dispositions, causal powers, and counterfactuals. In this respect, the view is similar to the standard Humean view. Fundamentally speaking, there is only the *qualitative* nature of the Cosmos. Furthermore, the view avoids brute necessities and conforms to Hume's Dictum.

Unlike Humean views, there is no explanatory circularity problem. The reason why the world had to evolve in this particular way is because of the intrinsic causal powers possessed by the world, which hold in virtue of the world's qualitative nature.

Unlike Humean views, approaches to the problem of induction that appeal to inference to the best explanation are unproblematic, since powerful qualities do not circularly explain.

Unlike Humean views, the view allows for physical histories of the universe that underdetermine the laws of nature (so long as the underlying powerful qualities of things are not specified in the corresponding physical histories). For example, consider Tooley's world with 10 kinds of particles and fifty-five possible two-particle interactions. Even if two particular kinds of particles had not interacted, then (just as with the other particles) the particular way they would have interacted would be determined by the powerful qualities inherent in those two kinds of particles according to PQ. Much more generally, the global causal structure that physics investigates might well underdetermine the powerful qualities that are beneath this causal structure that give rise to it.

Unlike Humean views, there is no problem of undermining futures for objective chances. On this view, the objective chances are not even partly determined by the global frequencies of events. Rather, the chancy evolution of the world would be grounded in the powerful qualities of the Cosmos. Instead of the powerful qualities fixing how the Cosmos will evolve in a deterministic

way, the powerful qualities would determine a corresponding chance-distribution for how the Cosmos might evolve.

At least on the way I have been presenting the dialectic between the Humean and the non-Humean, both are ultimately motivated by explanatory considerations. The Humean wants to avoid unexplained primitive modal notions and corresponding brute necessities, a desiderata that is enshrined in HD. Sider (2011), for example, puts the Humean vision as follows:

At bottom, the world is an amodal place. Necessity and possibility do not carve at the joints; reality is not "full of threats and promises". The book of the world says how things are, not how they must or might be. (266)

On the other hand, the Non-Humean wants there to be some robust explanation for the regularity of the world. The Non-Humean is repulsed by the Humean thought that the world is simply "one damn thing after another". At first glance, it might seem like these two visions are incompatible. We just have to pick our poison. We can either posit brute necessities or give up on explaining the regularity of the world. The view I have been arguing for tries to reconcile these two visions. Maybe we needn't give up on any of our explanatory ambitions. Maybe we can have our cake and eat it too.

References

Allen, Keith. 2017. A Naive Realist Theory of Colour. Oxford University Press UK.

Armstrong, David. 1983. What Is a Law of Nature? Cambridge University Press.

Armstrong, David. 1991. What Makes Induction Rational? *Dialogue* 30, no. 4: 503–11.

Armstrong, David. 1993. The Identification Problem and the Inference Problem. Philosophy and

Phenomenological Research 53, no. 2: 421–422.

Armstrong, David. 1997. A World of States of Affairs. Cambridge University Press.

Beebee, Helen. 2011. Necessary Connections and the Problem of Induction. Noûs 45, no. 3:

504–527.

- Bhogal, Harjit. Forthcoming. Nomothetic Explanation and Humeanism About Laws of Nature. Oxford Studies in Metaphysics.
- Bird, Alexander. 2007a. The Regress of Pure Powers? *Philosophical Quarterly* 57, no. 229: 513–534.
- Bird, Alexander. 2007b. *Nature's Metaphysics: Laws and Properties*. Oxford University Press.Blackburn, Simon. 1990. Filling in Space. *Analysis* 50, no. 2: 62–5.

Briggs, Rachael. 2009. The Anatomy of the Big Bad Bug. Noûs 43, no. 3: 428-449.

Campbell, John. 2005. Transparency Versus Revelation in Color Perception. *Philosophical Topics* 33, no. 1: 105–115.

Campbell, Keith. 1976. Metaphysics: An Introduction. Encino, CA: Dickenson.

Carroll, John. 1994. Laws of Nature. Cambridge University Press.

- Carroll, John. 2008. Nailed to Hume's Cross? In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean W. Zimmerman, 67–81. Blackwell.
- Carroll, John. 2011. Counterfactuals All the Way Down? Book symposium on Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature, by Marc Lange. Metascience 20, no. 1: 39–45

Chalmers, David. 2012. Constructing the World. Oxford University Press.

Chalmers, David. 2015. Panpsychism and Panprotopsychism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, edited by T. Alter and Y. Nagasawa, 246-76.
Oxford University Press.

- Choi, Sungho and Fara, Michael. 2018. Dispositions. In *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), edited by Edward N. Zalta. URL = <https://plato.stanford.edu/archives/fall2018/entries/dispositions/>.
- Cohen, Jonathan, and Craig Callender. 2009. A Better Best System Account of Lawhood. *Philosophical Studies* 145, no. 1: 1–34.
- Demarest, Heather. 2012. Do Counterfactuals Ground the Laws of Nature? A Critique of Lange. *Philosophy of Science* 79, no. 3: 333–344.
- Demarest, Heather. 2015. Powerful Properties, Powerless Laws. In *Putting Powers to Work: Causal Powers in Contemporary Metaphysics*, edited by Jonathan Jacobs. Oxford University Press.
- Dorst, Chris. Forthcoming. Humean Laws, Explanatory Circularity, and the Aim of Scientific Explanation. *Philosophical Studies*: 1–23.
- Dorst, Chris. 2017. Toward a Best Predictive System Account of Laws of Nature." British Journal for the Philosophy of Science: axy016.
- Dretske, Fred. 1977. Laws of Nature. Philosophy of Science 44, no. 2: 248–268.
- Emery, Nina. Forthcoming. Laws and Their Instances. Philosophical Studies: 1–27.
- Fine, Kit. 1994. Essence and Modality. *Philosophical Perspectives* 8: 1–16.
- Foster, John. 1983. Induction, Explanation, and Natural Necessity. *Proceedings of the Aristotelian Society* 83: 87–101.
- Foster, John. 2004. The Divine Lawmaker: Lectures on Induction, Laws of Nature, and the Existence of God. Oxford University Press.

- Gert, Joshua. 2006. A Realistic Colour Realism. *Australasian Journal of Philosophy* 84, no. 4: 565–589.
- Gert, Joshua. 2008. What Colors Could Not Be. Journal of Philosophy 105, no. 3: 128–155.
- Goff, Philip. 2017. Consciousness and Fundamental Reality. New York, USA: Oxford University Press.
- Goff, Philip. Forthcoming. Revelation, Consciousness+ and the Phenomenal Powers View. *Topoi*: 1–4.
- Grahek, Nikola. 2007. Feeling Pain and Being in Pain. Cambridge, MA: MIT Press.
- Hall, Ned. 1994. Correcting the Guide to Objective Chance. Mind 103, no. 412: 505-518.
- Hall, Ned. 2004. Two Mistakes About Credence and Chance. *Australasian Journal of Philosophy* 82, no. 1: 93–111.
- Hall, Ned. 2011. Review of Laws & Lawmakers: Science, Metaphysics, and the Laws of Nature by Marc Lange. Notre Dame Philosophical Reviews.

https://ndpr.nd.edu/news/laws-lawmakers-science-metaphysics-and-the-laws-of-nature/.

Haslanger, Sally. 1989. Persistence, Change, and Explanation. Philosophical Studies 56: 1-28.

Haslanger, Sally, and Roxanne Marie Kurtz. 2006. *Persistence: Contemporary Readings*. Bradford.

Heil, John. 2003. From an Ontological Point of View. Oxford University Press, 2003.

Heil, John. 2010. Powerful Qualities. In *The Metaphysics of Powers: Their Grounding and Their Manifestations*, edited by Anna Marmodoro. Routledge.

Heil, John. 2012. The Universe as We Find It. Oxford University Press.

- Hildebrand, Tyler. 2018. Natural Properties, Necessary Connections, and the Problem of Induction. *Philosophy and Phenomenological Research* 96, no. 3: 668-689.
- Hoefer, Carl. 2007. The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance. *Mind* 116, no. 463: 549–596.

Ismael, Jenann. 2008. Raid! Dissolving the Big, Bad Bug. Noûs 42, no. 2: 292–307.

Jenkins, C. S. 2011. Is Metaphysical Dependence Irreflexive? The Monist 94, no. 2: 267–276.

Johnston, Mark. 1992. How to Speak of the Colors. Philosophical Studies 68, no. 3: 221-263.

Ladyman, James. 2016. Structural Realism. In The Stanford Encyclopedia of Philosophy (Winter

2016 Edition), edited by Edward N. Zalta. URL =

<https://plato.stanford.edu/archives/win2016/entries/structural-realism/>.

Lange, Marc. 2000. Natural Laws in Scientific Practice. Oxford University Press.

- Lange, Marc. 2009. Laws and Lawmakers Science, Metaphysics, and the Laws of Nature. Oxford University Press.
- Lange, Marc. 2013. Grounding, Scientific Explanation, and Humean Laws. *Philosophical Studies* 164, no. 1: 255–261.
- Lange, Marc. 2018. Transitivity, Self-Explanation, and the Explanatory Circularity Argument Against Humean Accounts of Natural Law. *Synthese* 195, no. 3: 1337–1353.
- Langsam, Harold. 2011. The Wonder of Consciousness: Understanding the Mind Through Philosophical Reflection. MIT Press, 2011.
- Langton, Rae. 2001. Kantian Humility: Our Ignorance of Things in Themselves. Oxford University Press.

Langton, Rae. 2004. Elusive Knowledge of Things in Themselves. Australasian Journal of Philosophy 82, no. 1: 129–136.

Levin, Janet. 2000. Dispositional Theories of Color and the Claims of Common Sense. *Philosophical Studies* 100, no. 2: 151–174.

Lewis, David. 1973. Counterfactuals. Blackwell.

Lewis, David. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61: 343–377.

Lewis, David. 1986. Philosophical Papers Vol. II. Oxford University Press.

- Lewis, David. 1994. Humean Supervenience Debugged. Mind 103, no. 412: 473-490.
- Lewis, David. 1999. Papers in Metaphysics and Epistemology. Cambridge: Cambridge University Press.
- Lewis, David. 2008. Ramseyan Humility. In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, 203–222. MIT Press.
- Loewer, Barry. 2012. Two Accounts of Laws and Time. *Philosophical Studies* 160, no. 1: 115–137.
- Lowe, E. J. 2006. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Clarendon Press.

Mackie, J.L. 1973. Truth, Probability and Paradox. Oxford: Oxford University Press.

Martin, C. B., and John Heil. 1999. The Ontological Turn. *Midwest Studies in Philosophy* 23, no. 1: 34–60.

Maudlin, Tim. 2007. The Metaphysics Within Physics. Oxford University Press.

Mørch, Hedda Hassel. 2014. Panpsychism and Causation: A New Argument and a Solution to the Combination Problem. PhD Thesis, Oslo.

Mørch, Hedda Hassel. 2018. The Evolutionary Argument for Phenomenal Powers. *Philosophical Perspectives*, 293–316.

Mørch, Hedda Hassel. Forthcoming-a. Does Dispositionalism Entail Panpsychism? Topoi: 1-16.

- Mørch, Hedda Hassel. Forthcoming-b. The Argument for Panpsychism From Experience of Causation. In *The Routledge Handbook of Panpsychism*, edited by William Seager. Routledge.
- Mumford, Stephen. 1998. Dispositions. Oxford University Press.
- Mumford, Stephen. 2004. Laws in Nature. Routledge.
- Nolan, Daniel. 2014. Hyperintensional Metaphysics. *Philosophical Studies* 171, no. 1: 149–160.
- Peacocke, Christopher. 1984. Colour Concepts and Colour Experience. *Synthese* 58, no. 3: 365–82.
- Prior, E. 1985. Dispositions. Aberdeen: Aberdeen University Press.
- Quine, Willard Van Orman. 1960. Word & Object. MIT Press.
- Roberts, John T. 2001. Undermining Undermined: Why Humean Supervenience Never Needed to Be Debugged (Even If It's a Necessary Truth). *Proceedings of the Philosophy of Science Association*, no. 3: S98-108.
- Robinson, Howard. 1982. Matter and Sense: A Critique of Contemporary Materialism. Cambridge University Press.

Russell, Bertrand. 1927. The Analysis of Matter. London: Kegan Paul.
Ryle, Gilbert. 1949. The Concept of Mind. London: Penguin.

Schaffer, Jonathan. 2016. It Is the Business of Laws to Govern. Dialectica 70, no. 4: 577–588.

Schaffer, Jonathan, 2018. Monism. In *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), edited by Edward N. Zalta, URL =

<https://plato.stanford.edu/archives/win2018/entries/monism/>.

- Schrenk, Markus. 2006. A Theory for Special Science Laws. In *Selected Papers Contributed to the Sections of Gap.6*, edited by H. Bohse and S. Walter. Mentis.
- Shoemaker, Sydney. 1980. Causality and Properties. In *Time and Cause*, edited by Peter van Inwagen, 109–35.
- Shumener, Erica. 2017. Laws of Nature, Explanation, and Semantic Circularity. *British Journal* for the Philosophy of Science: axx020
- Sider, Theodore. 2001. Four Dimensionalism: An Ontology of Persistence and Time. Oxford University Press.
- Sider, Theodore. 2011. Writing the Book of the World. Oxford University Press.
- Sider, Theodore. Forthcoming. *The Tools of Metaphysics and the Metaphysics of Science*. Oxford University Press.
- Strawson, Galen. 2006. Realistic Monism Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies* 13, no. 10–11: 3–31.
- Strawson, Galen. 2008. The Identity of the Categorical and the Dispositional. *Analysis* 68, no. 4: 271–82.

Strawson, Galen. 2014. The Secret Connexion: Causation, Realism, and David Hume: Revised

Edition. Oxford University Press UK.

- Thomson, Judith Jarvis. 1983. Parthood and Identity Across Time. *Journal of Philosophy* 80, no. 4: 201–220.
- Tooley, Michael. 1977. The Nature of Laws. Canadian Journal of Philosophy 7, no. 4: 667–98.
- Tooley, Michael. 1987. Causation: A Realist Approach. Oxford University Press.
- Watkins, Michael. 2005. Seeing Red: The Metaphysics of Colours Without the Physics. Australasian Journal of Philosophy 83, no. 1: 33–52.
- Wilson, Jessica. 2010. What Is Hume's Dictum, and Why Believe It? *Philosophy and Phenomenological Research* 80, no. 3: 595–637.
- Woodward, Jim, Barry Loewer, John Carroll, and Marc Lange. 2011. Counterfactuals All the Way Down? *Metascience* 20, no. 1: 27–52.

Chapter Two: The Ineffability of Induction

Every time we eat bread, it nourishes. Every time we drop an apple, it falls at 9.8 m/s^2 . Those who haven't felt the force of skeptical arguments might find these facts to be fairly mundane. However, for those who have experienced the inductive vertigo that Hume famously did, these facts are shocking. After all, we can certainly imagine that the next time we eat bread, it might poison us. The next time we drop an apple, it could very well fall up at 9.8 m/s^2 . There are an infinity of ways in which the world could go completely haywire in the next second: gravity might reverse, tigers might start popping into existence, or the whole planet might disappear in a puff of smoke. What reason do we have for thinking that the world won't go haywire in the next second? The problem of induction, at least on one formulation, is the problem of giving some sort of story as to why we're justified in thinking that the world will continue to go on as we expect it to.

I have two main goals in this paper. My first goal is to formulate a mathematically precise, probabilistic version of the problem of induction (sections I-III). My formulation will take the form of an inconsistent triad: a set of three premises, each of which seems very plausible, that together imply a contradiction. This more precise version of the problem of induction has two main virtues. First, it gives a certain sense of *urgency* to the problem of induction. As we will see, the only way to recover anything like our inductive practices is either to swallow some extremely strong and seemingly unmotivated probabilistic assumptions or to pursue an ambitious rationalistic metaphysical project, akin to the kind of project that Descartes carried out in his *Meditations* to ward off his own skeptical doubts. Second, the inconsistent triad provides a taxonomy of ways one might respond to the problem of induction. The three main ways philosophers have tried to respond to the problem of induction nicely correspond to the three different premises one might deny in the inconsistent triad.

My second goal is to make some initial progress towards providing a distinctively metaphysical solution to the problem of induction (section IV-VII). Although my remarks won't amount to anything like a complete solution to the problem, I conclude by illustrating how it might work in a very simple, toy case.

I. The Informal Problem of Induction

We can break up the inductive skeptic's argument into two parts. First, there is the claim that there are an infinity of possible ways that the future might deviate in strange ways from the past. Apples might suddenly start rising, and bread might suddenly start poisoning. Second, there is the claim that we have no good reason for thinking that the actual world won't deviate from the past in one of these strange ways. The skeptic might try to justify this point in different ways. For example, the skeptic might claim that there is nothing we could say to convince the *counter-inductivist*, someone who makes the 'opposite' inferences about the future than the inductivist. After all, the standard circular justification in favor of induction ('induction has always worked before!') can also be given by the counter-inductivist ('since counter-inductive inferences are simply based on the assumption, the faith, that the world is simple or law-like or uniform. Given that we don't have any good reason to make these arbitrary assumptions, the skeptic says that we should revise our beliefs about the future. For example, we should abandon our belief that the sun will rise tomorrow.

The first step of the skeptic's argument is crucial. As a parody argument, suppose a mathematical skeptic argued as follows:

You should abandon your belief that 2+2 = 4. After all, there are an *uncountable infinity* of possibilities in which 2+2 isn't 4. For example, for every real number *r*, there is a possibility in which 2+2 = r. Given all of these possibilities, what reason could we have for thinking that we aren't in one of the worlds where 2+2 isn't 4? The assumption that we are in such a world is simply an arbitrary bias in favor of the number 4.

This skeptical argument isn't going to be keeping anybody up at night. We should of course simply deny that there are any possibilities where 2+2 isn't 4. The reason why inductive skepticism is so much more gripping is because it does seem like a genuine metaphysical possibility that the world could go completely haywire tomorrow. If these sorts of worlds are genuinely possible, then it seems like we need some reason to think that the actual world isn't one of these possible worlds.

Still, even though the problem of inductive skepticism is more compelling, the vast majority of philosophers have not become converted skeptics.²⁵ In part, this is because it is unclear what positive *argument* there is to think that induction is irrational. The skeptic simply challenges us to provide a justification of induction that would convince the skeptic or other fanciful characters like the counter-inductivist. But why think we must be able to convince the skeptic in order for induction to be rational? We might not be able to convince the flat-earther that the earth is round. Should we therefore abandon our belief that the earth is round? Roger White (2015) has coined the phrase 'the problem of the problem of induction' to refer to the problem of figuring out what exactly is the skeptic's positive argument, it is easy to be dismissive of the skeptic. In the next section, I try to formulate a more compelling, probabilistic challenge to our inductive practices.

II. An Inconsistent Triad

I will start with a certain fanciful thought experiment. Suppose God presents you with two books. Both books contain maximally specific descriptions for a way for the world to be. They include, for example, two alternative complete descriptions of space-time and its material occupants. God assures you that both books describe genuine metaphysically possible worlds, and he reveals to you that *one of the books describes the actual world*. After spending some time reading the two books, you realize that they agree on all matters about the present and the past. However, one of the books describes a world that goes on to behave regularly into the future, and the other book describes a world that goes 'haywire' at a certain point in the future. Question: what credence should you have that the 'regular' book describes the actual world?

There are two competing pressures here. On the one hand, the totality of your evidence doesn't discriminate between the two books. After all, both books entail your evidence (which is only about the past and present) with certainty.²⁶ Suppose we let Cr be your prior probability function, E be your evidence concerning the past, and Cr_E be your prior probability function conditioned on

²⁵ According to the philpapers survey conducted by Bourget and Chalmers (2014), 81.6% of philosophers endorse 'non-skeptical realism'.

²⁶ Some philosophers would disagree that our evidence is only about the past and present. For example, one might think that our evidence is just our knowledge and that we know things about the future (e.g. Williamson (2000)). In response, one can modify the case so that the two books agree on the totality of our evidence, whatever it is. So long as our evidence leaves open an uncountable infinity of future possibilities, the three premises of the inconsistent triad can be easily reformulated.

your evidence. Furthermore, let's call 'Regular' the proposition that the regular book describes the actual world and 'Irregular' the proposition that the non-regular book describes the actual world. In a Bayesian framework, since both Regular and Irregular entail your evidence, we have that Cr (Regular | Regular or Irregular) = Cr (Regular | E and (Regular or Irregular)) = Cr_E (Regular | Regular or Irregular). In other words, your evidence about the past is entirely *neutral* with respect to whether Regular or Irregular is true. So, if you are to assign a higher probability to Regular, it must be for some wholly a priori, non-evidential reason. Since your evidence doesn't count in favor of either possibility, a natural thought is that we should apply the Principle of Indifference and assign a 0.5 credence to both Regular and Irregular.²⁷ According to one formulation of the Principle of Indifference, you shouldn't assign *different* probabilities to possibilities which are 'evidentially symmetric', where two possibilities are evidentially symmetric just in case you have *no more reason* to suppose that one is true rather than the other (White 2010: 161-162). In this case, since we have no evidence that has any bearing on the question, it is natural to think that we have no reason to privilege either one of these possibilities.

The most powerful objection to the Principle of Indifference is the 'multiple partitions' problem, which claims that one can get contradictory verdicts from the Principle of Indifference by assigning equal probabilities to different partitions of the space of possibilities.²⁸ However, there is no multiple partitions problem here. There are only two ways for the world to be! So, since your evidence is entirely neutral with respect to Regular and Irregular and there is one unique way to 'partition' these two possibilities, there is at least some pressure to assign a 0.5 credence to both Regular and Irregular.

Alternatively, one could violate the Principle of Indifference and 'skew' one's probabilities in favor of Regular, perhaps because Regular is more 'simple' or 'natural'.²⁹ Of course, the inductive skeptic will be suspicious of these wholly a priori skewings of one's probabilities in the absence

²⁷ See van Inwagen (1996) for an explicit defense of the claim that maximally specific possible worlds should be assigned the same probability.

²⁸ For different responses to the multiple partitions problem, see White (2010) and Huemer (2009).

²⁹ See Sober (2015) for an overview of the uses of simplicity in epistemology, and see Bradley (2019) for a recent approach to the problem of induction that favors natural hypotheses. Hedden (2015) argues that we should be biased towards possibilities that are more explanatory.

of any evidence. Moreover, such skewings seem entirely arbitrary. Exactly how much should we skew our probabilities? 60-40? 90-10?

The first premise of our inconsistent triad makes an extremely weak claim about this sort of thought experiment:

For any evidence E about the past, given that w₁ and w₂ are maximally specific metaphysical possibilities that entail E, it is *permissible* to set Cr_E (w₁ | w₁ or w₂) ≠ 1 ≠ Cr_E (w₂ | w₁ or w₂).

All this premise says is that it's *permissible* not to be maximally dogmatic. Or, alternatively, failing to do the *exact opposite* of what the Principle of Indifference recommends (by assigning all of your credence to one of the two possibilities in the absence of any evidence) is permissible.

Let us now turn to a second thought experiment. Imagine the following evidential situation that we might find ourselves in. Suppose that the community of theoretical physicists have unanimously decided on the true deterministic laws of physics, L, describing our universe. Suppose, for example, that billions of years pass without a single counterexample, any measurement can be predicted to arbitrary degrees of precision, the laws are breathtakingly elegant and simple, and physics departments are forced to close down because there is no work left for theoretical physicists to do. Fundamental physics is finally complete! Given this evidential situation, E*, it seems that the scientists are rational in having very high credence that the true deterministic laws of physics are L. Moreover, suppose that IC is some maximally specific characterization of the initial conditions of the universe, and let w_{ICL} be the unique world in which IC describes the exact initial conditions and the deterministic laws of physics L holds. Given that $Cr_{E*}(L)$ should be very high, scientists should also assign a similarly high credence to $Cr_{E*}(w_{ICL} | IC)$. After all, if they were convinced in the truth of L on the basis of E*, then given that the initial conditions of E*.

The second premise of our inconsistent triad also makes an extremely weak claim about this sort of thought experiment:

2. Given E*, one *shouldn't* set $Cr_{E*}(L) = 0$. So, neither should one set $Cr_{E*}(w_{ICL} | IC) = 0.^{30}$

All this premise says is that one shouldn't believe the *exact opposite* of what all the scientific experts believe. If all the scientific experts assign high credence to L given E*, you shouldn't be maximally confident that L is false given E*. Similarly, if all the scientific experts believe in global warming, we shouldn't assign a credence of 0 to global warming.

The final premise of our inconsistent triad appeals to the Humean thought that there are an uncountable infinity of ways in which the world may go haywire tomorrow. For example, suppose I am about to drop an apple. Perhaps we may expect it to fall at 9.8 m/s², but we cannot rule out a priori that it falls at some other rate. It seems like, for every real number $r \in [0, 10]$, there is a world just like this one up to now, except the apple drops at exactly r m/s² when I let it go. In other words:

3. For any initial segment of some world, there are an uncountable infinity of metaphysically possible worlds that exactly match that initial segment and deviate afterwards.

Surprisingly, these three premises imply a contradiction. I will bring out this contradiction by arguing that 1 and 3 entail the falsity of 2. In particular, I will argue that, given 1 and 3, it is permissible to set $Cr_{E*}(w_{ICL} | IC) = 0$.

Before presenting the formal argument, it will be useful to provide an informal gloss. The intuitive idea behind the argument is just that the scientist's evidence E^* leaves open an uncountable infinity of worlds that start out in conditions IC (by premise 3). Only *one* of these worlds will be w_{ICL}. By premise 1, it is permissible not to be 'infinitely' biased towards w_{ICL}: when making a pairwise comparison of the plausibility of w_{ICL} against one of the other worlds left open by the scientist's evidence, one doesn't have to be certain that w_{ICL} is the actual world and the other one isn't. However, if one is only 'finitely' biased towards w_{ICL}, because there are *so* many alternative

³⁰ As long as one doesn't take IC to be evidence *against* the truth of L, the second claim follows from the first. Here is a proof. Assume that $Cr_{E*}(L \mid IC)$ is at least as great as $Cr_{E*}(L)$. We know that $Cr_{E*}(L \mid IC) = Cr_{E*}(w_{ICL} \mid IC)$, since the conjunction of L and IC are equivalent to w_{ICL} . So, supposing it were permissible to assign $Cr_{E*}(w_{ICL} \mid IC) = 0$, then it would be permissible to assign $Cr_{E*}(L \mid IC) = 0$. By our assumption, this would force one to assign $Cr_{E*}(L) = 0$. This, however, is impermissible.

possibilities that are left open (an uncountable infinity of them), the probability calculus forces one to assign a credence of 0 to w_{ICL} .

Here is the more formal argument (readers may skip to the next section to avoid these formal details). Using 3, let $S = \{w_i \mid i \in I\}$ be the uncountably infinite set of worlds that satisfy both E* and IC, where I is some index set.³¹ By 1, let Cr be some permissible prior credence function such that, for any world $w_i \in S$ distinct from w_{ICL} , Cr_{E^*} ($w_{ICL} \mid w_{ICL}$ or w_i) $\neq 1 \neq Cr_{E^*}$ ($w_i \mid w_{ICL}$ or w_i). For every $i \in I$, let r_i be such that the *ratio* of Cr_{E^*} ($w_{ICL} \mid w_{ICL}$ or w_i) to Cr_{E^*} ($w_i \mid w_{ICL}$ or w_i) is equal to the ratio of 1 to r_i . In other words, let r_i satisfy the following equation:

$$\frac{Cr_{E*}(w_{ICL} \mid w_{ICL} \text{ or } w_i)}{Cr_{E*}(w_i \mid w_{ICL} \text{ or } w_i)} = \frac{1}{r_i}$$

For any positive natural number *n*, let $S_n = \{w_i | r_i > 1/n\}$. Since S is uncountable and equal to the union of each of the S_n , there must be some *m* such that S_m is infinite.³²

Abusing notation, let us also treat S (and S_m) as the proposition that the actual world is one of the worlds in S (or S_m). Then, $Cr_{E*}(w_{ICL} | IC) = Cr_{E*}(w_{ICL} | S)$, since S is the same set of worlds as the ones that satisfy E* and IC. We also have that $Cr_{E*}(w_{ICL} | S_m) \ge Cr_{E*}(w_{ICL} | S)$, since S_m eliminates more worlds distinct from w_{ICL} than S does. To simplify things further, let us also let S_m* be some arbitrary *countable* subset of S_m that includes w_{ICL}. Again, we have that $Cr_{E*}(w_{ICL} | S_m^*) \ge Cr_{E*}(w_{ICL} | S)$, since S_m* eliminates more worlds distinct from w_{ICL} that S does. To simplify things further, let us also let S_m* be some arbitrary *countable* subset of S_m that includes w_{ICL}. Again, we have that $Cr_{E*}(w_{ICL} | S_m^*) \ge Cr_{E*}(w_{ICL} | S)$, since S_m* eliminates more worlds distinct from w_{ICL} than S_m does. So, in order to show that $Cr_{E*}(w_{ICL} | IC) = 0$, it suffices to show that $Cr_{E*}(w_{ICL} | S_m^*) = 0$.

Let us label the worlds in S_m^* to be w_{ICL} , w_1 , w_2 , w_3 , w_4 , ... It is a theorem of the probability calculus that $Cr_{E^*}(w_{ICL} | S_m^*) = Cr_{E^*}(w_{ICL} | w_{ICL} \text{ or } w_1 \text{ or } w_2 \text{ or } ...) = 1 / (1 + r_1 + r_2 + ...)$. Since each of these r_i are at least 1/n, it follows that $Cr_{E^*}(w_{ICL} | S_m^*) = 0$, as desired.

³¹ What if there are too many such worlds to form a set? It will suffice for the proof for S to be merely uncountable. So, if there are too many such worlds to form a set, let S be some arbitrary uncountable set-sized collection of worlds that includes w_{ICL} and satisfies E* and IC.

 $^{^{32}}$ This is the only step where we use the assumption that S is uncountable. The proof could work equally well for countable S so long as we assume that there is some infinite S_n. For example, if one thinks there are countably many 'symmetric' ways the world could deviate from w_{ICL}, and one thinks that one should assign an equal ratio between w_{ICL} and each of these symmetric deviant worlds, then the proof would work just the same.

III. Three Ways Forward

How should we respond to this contradiction? There are three ways out, each of which is revisionary of a different aspect of our theorizing about the world.

First, consider premise 1. The primary motivation for premise 1 appealed to the way we should reason about some space of possibilities. Given that we have no evidence favoring either maximally specific possibility, it should at the very least be *permissible* not to be maximally confident in either one! In some ways, denying premise 1 corresponds to the standard way philosophers have tried to respond to the problem of induction. Many philosophers have thought that we should be biased in favor of hypothesis that are natural, simple, or explanatory over ones that are gerrymandered or complex. However, it is easy to feel that it is far too dogmatic to assign *probability 1* to the regular possibility in the absence of any evidence, especially in the absence of some sort of *explanation* as to why we should expect the world to be regular *a priori*. Let's call those who respond to the inconsistent triad by denying premise 1 *Dogmatists*.

Next, consider premise 2. The primary motivation for premise 2 appealed to a certain first-order judgement. If the experimental evidence E* for L was so good that every physicist unanimously endorsed L, then at the very least we shouldn't be *certain* that L is false! Giving up on premise 2 corresponds to a kind of skepticism about inductive reasoning, since inductive reasoning is supposed to align with how scientists actually reason. One way to think of this inconsistent triad is as a positive argument that the skeptic might put forward: since premises 1 and 3 are truths that entail the falsity of 2, our inductive practices are unjustified. Let's call those who respond to the inconsistent triad by denying premise 2 *Skeptics*.

Lastly, consider premise 3. Premise 3 is a metaphysical claim which is supported by Hume's conceivability argument. Descartes is an example of a philosopher who might deny premise 3. According to Descartes, there is a priori reason to believe in the existence of a perfect God, in the form of the ontological argument. Since a perfect God is not a deceiver, there is a priori reason to doubt the metaphysical possibility of worlds where we are radically deceived, such as irregular worlds where apples start falling up for no apparent reason. Unfortunately, while theistic approaches to the problem of induction still have some contemporary defenders (e.g. Foster

(2004)), not very many philosophers today believe that there is a strong a priori case for the existence of God.

One might think that contemporary non-Humeans, who believe in fundamental laws, necessitation relations, or causal powers, would deny premise 3.³³ However, contemporary non-Humeans typically don't argue for the *impossibility* of Humean worlds, which lack such primitive non-Humean posits. Similarly, most non-Humeans don't argue for the impossibility of worlds with "schmass" and "scharge", which are properties that behave like mass and charge until now but deviate in strange ways afterwards. Typical non-Humean views only argue that we have good reason to think that the *actual* world is governed by (say) primitive laws. Since these non-Humean views don't identify any flaws in Hume's conceivability argument, they are not able to rebut the main argument in favor of premise 3.³⁴

Let's call those who respond to the inconsistent triad by denying premise 3 Rationalists.

Should we be Dogmatists, Skeptics, or Rationalists? It seems to me that we should at least *try* to be Rationalists. Skepticism more or less amounts to giving up on the scientific enterprise. Because Dogmatism posits extremely strong probabilistic biases that don't seem to have an adequate explanation, Dogmatism more or less amounts to giving up on a satisfactory epistemology of induction.

However, taking the Rationalist road isn't easy. The Rationalist needs to provide some metaphysical account that responds to Hume's conceivability argument and casts doubt on premise 3. Ideally, the Rationalist should also try to avoid the main drawback of the Dogmatist account by showing how induction could be vindicated *without* biasing or skewing one's probabilities in arbitrary ways. In what follows, I will try to provide the very beginnings of such an account. Much

³³ For a sampling of contemporary non-Humean views, see Maudlin (2007), Carroll (1994), Armstrong (1983), and Mumford (2004).

³⁴ A small minority of non-Humeans, such as Wilson (2013) and Bird (2007b), have defended 'strong' necessitarian views, according to which the actual laws of physics govern every possible world, which would make Humean worlds and "schmass" worlds impossible. However, rather than responding to Hume's conceivability argument, these views typically reject *a priori* modal epistemology altogether. There is therefore a worry that these views cannot help with the problem of induction, if the only grounds for believing that every possible world is governed by regular laws presuppose that induction works for discovering the laws of the actual world.

like Descartes' own Rationalist account, the view I will end up defending involves some highly speculative metaphysics.

My Rationalist approach to induction involves two key steps. First, in sections IV-VI, I will present and motivate a metaphysical picture on which premise 3 is false. My goal isn't to argue that such a metaphysics is *true*, but merely to argue that it cannot be ruled out *a priori*: such a metaphysics deserves to be a serious and live epistemic possibility. Next, in section VII, I argue that the mere tenability of such a metaphysics may open up a non-Dogmatist solution to induction. I do this by considering a very simple toy case, which illustrates the advantages of my approach over standard Humean and non-Humean approaches to induction.

IV. The Inscrutability of Matter

According to many philosophers, science only gives us a very limited picture of the physical world. This view has been held in different guises by Kant, Russell, and Lewis, as well as contemporary Structuralists and Russellian Monists.³⁵ On this view, science only reveals the abstract causal *structure* of reality, rather than its *intrinsic nature*. For example, science only tells us that electric charge is whatever it is that plays a certain causal role (e.g. attracts opposite charges and repels like charges). Likewise, mass is whatever it is that attracts other massive objects and resists acceleration. However, science leaves us completely ignorant about the intrinsic nature of (say) mass and charge.

As Stephen Hawking (1988) once asked, 'What is it that breathes fire into the equations and makes a universe for them to describe?' (1988: 174). It has seemed to many that science is curiously abstract. It only deals with mathematical models and equations to describe the nomic/causal and spatiotemporal relationships among various properties. But, it seems like there must be *more* to the world than this abstract causal structure. There must be some underlying intrinsic nature -a 'fire' – beneath these equations.

Consider, for example, the following, allegedly complete, description of a possible world:

³⁵ Langton (2001, 2004) defends an interpretation of Kant in which we can never have knowledge of things 'in themselves', and Lewis (2008) endorses a similar conclusion. Contemporary Russellian Monists, such as Strawson (2006), follow Russell (1927) in endorsing a similar conclusion. For an overview of Structural Realism, see Ladyman (2016).

Structuralist World: There is a single entity x which has some property or other, call it F. All there is to say about F is that it makes things that have it have some other property, call it G, at the next time. All there is to say about G is that it makes things that have it have the property of being F at the next time. This cycling goes on forever.

To many philosophers, it seems like this cannot be a complete description of a possible world. The description leaves out the intrinsic nature of F and G, because it only describes the causal *relations* between F and G. Some have thought that in a structuralist world such as this one, the nature of any property would circularly depend on the nature of other properties. Since the nature of F is wholly exhausted by how it effects the nature of G, it seems like to understand the nature of F we would first have to understand the nature of G. But, similarly, since the nature of G is wholly exhausted by how it effects the nature of F, it seems like to understand the nature of G we would first have to understand the nature of F.

To get a better grip on these 'intrinsic natures', consider the following analogy. There are two ways of describing phenomenal properties, which specify *what it's like to be* something. Functionally, pain is whatever it is that occupies a certain causal role. For example, it is typically caused by tissue damage and it typically causes avoidance behavior. However, there is also the intrinsic nature of pain – how it feels, or its 'phenomenal character'. There is what pain *does* on the one hand, and what it is *intrinsically like*, on the other hand. According to many philosophers, the same goes for matter. Physics only describes what physical properties *do*, but not what they are *intrinsically like*.

What are these intrinsic natures of fundamental physical properties like? Philosophers who have advocated for this view typically are skeptical that we can know what these intrinsic natures are like. Continuing with the analogy of phenomenal properties, many have thought that phenomenal properties that are radically different than the ones we have experienced are *ineffable* to us. For a blind person, it is hard to see how the nature of phenomenal colors could be adequately described. Similarly, if we were to meet a radically alien intelligent species, we couldn't begin to describe what their phenomenology is like. Our situation isn't that there is a range of perfectly well-

³⁶ This circularity worry has been advanced by Russell (1927), Robinson (1982), Blackburn (1990), Armstrong (1997), Heil (2003), Lowe (2006), and Goff (2017). See Bird (2007a) for a response to this objection.

understood options for what the alien's phenomenology could be like, and we are simply ignorant of which is the correct option. Rather, the complete truth about the alien's phenomenology is just not describable using the concepts that we have. Similarly, it is natural to think that the intrinsic nature of matter is ineffable to us. It is not that we have a range of perfectly well-understood options for what the intrinsic nature of matter could be like, and we are simply ignorant of which is the correct option. Rather, it seems like the complete truth about the intrinsic nature of matter is just not describable using the concepts we have.

These considerations motivate the following metaphysical thesis, which many philosophers have argued is ultimately an a priori truth:

NATURE: Necessarily, all fundamental physical properties have some intrinsic nature.³⁷

By appealing to NATURE, we can already see that there is a problem with Hume's conceivability argument. When we are conceiving of irregular worlds, where apples start rising for no apparent reason, we are not conceiving of a *complete* possible world, since we are not conceiving of the underlying intrinsic nature of the apple and its surroundings. To continue with the analogy of phenomenal properties, when we are watching a movie, the movie does not reveal the *phenomenology* of the people inside the movie. At best, we can only infer their phenomenology from what the movie shows. Similarly, when we conceive of an apple rising for no apparent reason, it is as if we are watching a mere 'movie' of the apple rising inside of our minds, which leaves out facts about the intrinsic nature of the apple and its surroundings.

Say that an ideally conceivable *structural* possibility is some ideally conceivable maximal specification of the spatiotemporal and causal structure of a world.³⁸ Say that an *intrinsic*

³⁷ Different philosophers use different terms of art that correspond to the phrase 'intrinsic nature', but each of these terms has various connotations that I would like to avoid. For example, some philosophers may describe these intrinsic natures as *quiddities*. Hildebrand (2016) distinguishes between 'bare quiddities' and 'qualitative quiddities'. Bare quiddities only differ from other bare quiddities by being primitively distinct. Qualitative quiddities differ from other qualitative quiddities by being qualitatively distinct. I am using the phrase 'intrinsic nature' synonymously with Hildebrand's notion of a qualitative quiddity. Other philosophers may use the term 'categorical property' to describe these intrinsic natures. However, some philosophers (e.g. Bird (2007b)) stipulate that 'categorical properties' cannot necessitate a corresponding disposition, but I am using the phrase 'intrinsic nature' in such a way so that it is an open question whether some intrinsic nature necessitates a corresponding disposition (see section V). ³⁸ For much more on ideal conceivability, see Chalmers (2002).

possibility is some maximal specification of a metaphysically possible world, including its intrinsic nature.

Question: Are all ideally conceivable *structural* possibilities realized by some *intrinsic* possibility?

If the answer is 'no', then it might well be that the conceivable *structural* possibilities where things go 'haywire' may not be genuinely metaphysical possible after all.

V. Powerful Qualities

We can distinguish two views about the intrinsic nature of physical properties. On one view, these intrinsic natures are *powerful*, and on another view, they are *powerless*. To say that an intrinsic nature is powerful is to say that it grounds a corresponding disposition or causal power. Otherwise, it is powerless.

Again, we can consider an analogy with phenomenal properties. According to some, phenomenal properties are powerful. This view has been most prominently defended by Mørch (2014, 2018a, 2018b, 2019, forthcoming), but different variations of it have also been defended by Goff (2017, 2018) and Langsam (2011)).³⁹ While a full defense of the Phenomenal Powers view is far beyond the scope of this paper, some initial motivation for the view can easily be found by reflecting on our own conscious experiences. It is natural to think, for example, that *pain* makes subjects who experience it try to avoid it simply in virtue of how bad it feels. *Pleasure* might make subjects who experience it try to pursue it simply in virtue of how good it feels. Someone who feels *tired* or *exhausted* might be disposed to stay in bed simply in virtue of how tired they feel. These connections between these experiences and their effects are certainly defeasible. Of course, in our own case, human subjects endure pain for all sorts of reasons. Someone might endure short-term pain in order to avoid more pain in the future (e.g. going to the dentist), or in order to experience a greater pleasure (e.g. masochism), etc.⁴⁰ However, in the *absence* of any interfering causes like

³⁹ Goff (2017, forthcoming) argues that only 'consciousness+' properties – properties which have both phenomenal and non-phenomenal aspects – ground a corresponding disposition.

⁴⁰ Another common objection to the claim that pain essentially disposes one to avoidance behavior is the phenomenon of pain asymbolia. Those who have pain asymbolia may experience pain without the corresponding feeling of unpleasantness, so they are not motivated to avoid the pain. For this objection to go through, it would have to be maintained that pain experiences for those who have pain asymbolia have the *very same* phenomenal character

these, it seems like pain has a certain kind of default power to dispose subjects to at least *try* to avoid it. Similarly, throwing a ball at a window might cause the window to break, but only in the absence of any interfering causes. Of course, the wind might blow the ball away before it hits the window, or the window might be shattered by a rock before the ball gets to it, etc.

It is important to note that the claim that the intrinsic natures of physical properties are powerful doesn't need to be associated with consciousness or phenomenology in any way. Many metaphysicians, including Heil (2003, 2010, 2012), Martin (2008), Jacobs (2011), and Taylor (2013), have argued in favor of a general 'Powerful Qualities' view, according to which the intrinsic nature of physical properties are both qualitative *and* dispositional.⁴¹

Other philosophers have argued in favor the so-called Eleatic Principle, which roughly states that in order for any entity to exist, it must have causal powers.⁴² This principle would preclude the existence of entities which were entirely causally inert (such as abstract objects), perhaps because their intrinsic nature is entirely powerless.

Building on these views, here is a speculative metaphysical hypothesis:

POWER: The possible intrinsic natures of fundamental physical properties are all powerful.

Ultimately, what we should think about POWER will turn on various difficult and controversial metaphysical debates. For our purposes, however, we won't need the claim that POWER is particularly plausible. In fact, as we will see below, it is fine if POWER is fairly *im*plausible. All we will need at this point is that POWER cannot be ruled out a priori. Given that POWER concerns a realm of properties that are entirely ineffable to us, it's not clear how we could possibly rule this

as ordinary experiences of pain. However, this is doubtful. It is natural to think that the *unpleasantness* of pain is part of the phenomenal character of ordinary experiences of pain. On this view, what the phenomenon of pain asymbolia shows is that ordinary phenomenology of pain has two components, a sensory component and an affective component. For those who have pain asymbolia, these two components come apart, and they only experience the sensory component without the affective component. For much more on this, see Grahek (2007) and Mørch (2019).

⁴¹ Strictly speaking, the Powerful Qualities view is associated with the claim that qualities are *identical* to dispositions, whereas I have formulated the claim that intrinsic natures are powerful as the claim that intrinsic natures *ground* a corresponding disposition. However, these metaphysical details will not be relevant. So long as the intrinsic nature of a physical property necessitates a corresponding disposition (whether this is by grounding or identity), the epistemological pay-offs will be similar.

⁴² Armstrong (1978) and Ellis (1990: 22) endorse the Eleatic principle, and Field (1980) is also plausibly interpreted as an Eleatic. See Colyvan (1998) and Cowling (2015) for discussions of the Eleatic Principle.

out a priori, since we don't have a positive conception of *any* of the possible intrinsic natures of fundamental physical properties. Because of this, it seems safe to say that POWER cannot be ruled out a priori.

VI. Regular Powerful Qualities

We've now reached our final section on speculative metaphysics. If POWER is true, what kinds of powers can these possible intrinsic natures give rise to? Can they give rise to entirely 'gerrymandered' powers? Consider the power: *disposed to play the charge-role before 2020 and the mass-role afterwards*. Could there be any intrinsic nature that gives rise to this power?

Again, if we consider the analogy with phenomenal properties, none of the (allegedly) powerful phenomenal properties we are acquainted with are anything like this. Pain, for example, disposes one towards avoidance behaviors regardless of where or when it happens. On reflection, it's hard to see how an intrinsic nature *could* give rise to such a gerrymandered power. Reflecting on our experience motivates the following speculative metaphysical hypothesis, which builds on POWER:

REGULAR: The only ideally conceivable structural possibilities realized by an intrinsic possibility are regular.

If all intrinsic possibilities contain intrinsic natures that are powerful in regular ways, then such intrinsic possibilities will only ever realize regular structural possibilities, instead of structural possibilities that go 'haywire' in various ways (e.g. possibilities where things start popping into existence for no reason). Again, we will not be needing the claim that REGULAR is particularly plausible. As we will see below, we will only need the claim that REGULAR cannot be ruled out a priori.

An obvious clarificatory question to ask of REGULAR is what 'regular' means. Although there is some intuitive, vague content to which structural possibilities are regular, the phrase is meant to be a placeholder that must be filled in by different precise physical principles. An important question is exactly how to precisify it. Given that we are speculating about the possible ineffable intrinsic natures of things, it's hard to know exactly how it should be filled. One way it could be filled in is by saying that the only structural possibilities realized by an intrinsic possibility are ones that are governed by our actual physical laws, such as the Schrödinger equation in quantum mechanics. However, this can seem fairly ad hoc. What could explain why every possible intrinsic nature must conform to Schrödinger's equation?

A more promising route is to look to the long history of both physicists and philosophers advancing purportedly *necessary* constraints on the laws of physics of a very general form. Recent books by Mittelstaedt (2013) and Darrigol (2014) survey and assess many of these necessary principles, ranging from the attempts of pre-Newtonian philosophers such as Descartes and Leibniz, to present day rational reconstructions of Quantum Mechanics. Ultimately, the most plausible ways to precisify 'regular' will come from the kind of rationalistic investigation carried out by these physicists and philosophers. Briefly, however, let us go through two simple and non-technical examples of how such purportedly necessary principles may be motivated.

Many philosophers and physicists have thought that certain kinds of *conservation laws* are necessary constraints on the laws of physics. For example, in an article devoted to the history of the conservation of energy, Max Planck (1887) wrote:

Recently it has been proclaimed that a proof of the [conservation of energy] principle is neither possible nor necessary because this principle is valid a priori, namely, because it is a necessary, nature-given form of our intuition and understanding. In this case and for many other truths that have been conquered through century-long efforts, the truths are said to be evident and inborn after the fact, when the force of habitude has acted on us.

In contemporary times, Lange (2007, 2012) has argued that conservation laws should in fact be treated as 'meta-laws': laws that constrain the 'first-order' laws of physics. We might therefore conjecture that one principled way of filling in 'regular' is as follows:

CONSERVATION: The only ideally conceivable structural possibilities realized by an intrinsic possibility conform to various conservation laws (e.g. conservation of energy, conservation of linear and angular momentum, etc.)⁴³

⁴³ The details of exactly how to formulate these different conservation laws will sensitively depend on the physical framework that they apply to (for example, these conservation laws look very different in General Relativity than in Newtonian Mechanics). Although these physical details are metaphysically important, they won't matter for the

Rationalist philosophers such as Descartes and Leibniz typically argued for the necessity of conservation laws on theistic grounds.⁴⁴ However, in modern times a plausible a priori case can be made for CONSERVATION by appealing to certain mathematical results proven by Emmy Noether connecting conservation laws to symmetry principles. Using these mathematical results, it can be shown that conservation of energy follows from the fact that the laws are invariant across time, and conservation of linear and angular momentum follow from the fact that the laws are invariant across spatial translations and rotations. We have already seen that the causal powers of (say) pain seem to be invariant across time and space. Whether or not pain causes avoidance behavior is insensitive to where or when that pain is instantiated. But, we can go further. If the causal powers of physical properties are to be *fully grounded* in their intrinsic natures, then these causal powers of physical properties would only be *partly* grounded in their intrinsic natures and *partly* grounded in where or when they happen to be instantiated. At least given POWER, then, it seems like a plausible a priori case can be made in favor of CONSERVATION.

Another kind of principle that has been pervasive in our physical theories is a principle of *continuity*. Can a physical system transition from one state to another, without traversing states that are intermediate between the two? One way to make this principle precise is as follows:

CONTINUITY: The only ideally conceivable structural possibilities realized by an intrinsic possibility are ones whose trajectory across phase space is continuous.⁴⁵

One might also consider different strengthening of this principle:

general epistemological upshots we will discuss later. The main purpose of this section is merely to illustrate some possibilities for how to precisify REGULAR, and so I will consider the simplified case of Newtonian mechanics. ⁴⁴ See Garber (1992: ch. 7) for a discussion of Descartes' arguments for conservation principles, and Garber (1995) for a discussion of Leibniz's arguments. The modal status of Leibniz' arguments for conservation laws is controversial. In the *Theodicy*, Leibniz wrote, '[The laws of motion] do not derive entirely from the principle of necessity, but from the principle of perfection and order; they are an effect of the choice and the wisdom of God'. Leibniz scholars disagree about whether Leibniz is best interpreted as thinking that God could have chosen to make a different world than the actual world (e.g. see Griffin (2012)).

⁴⁵ Different physical theories have different phase spaces. The *phase space* of a physical theory represents all of the physically possible states of a system, where each point represents a possible state. Each axis of a multi-dimensional phase space corresponds to one of the degrees of freedom of a system (e.g. a particle's position or momentum).

56

DIFFERNTIABILITY: The only ideally conceivable structural possibilities realized by an intrinsic possibility are ones whose trajectory across phase space is differentiable.

SMOOTHNESS: The only ideally conceivable structural possibilities realized by an intrinsic possibility are ones whose trajectory across phase space is smooth (i.e. infinitely differentiable).

These kinds of continuity principles were perhaps most famously defended by Leibniz (*NE*), who wrote that 'Nothing takes place suddenly, and it is one of my great and best confirmed maxims that *nature never makes leaps*. I call this the Law of Continuity' (56). One can try to motivate CONTINUITY on the grounds that, given POWER, it seems like small changes in the intrinsic nature of a physical system should only give rise to small changes in the kind of effects that those intrinsic natures give rise to. Otherwise, it would seem that the power of that physical system cannot be intelligibly traced back to its intrinsic nature. Suppose, for example, that some physical system evolved continuously within the interval of time [t_1 , t_2], but then sharply and discontinuously 'leapt' to some other state right after t_2 . Then, the intrinsic nature of the system at t_2 would have caused the system to 'leap', but the intrinsic nature of the system at any time arbitrarily close to t_2 did *not* cause the system (the state at t_2 compared to the state at any time arbitrarily close to t_2) give rise to radically different kinds of causal effects (leaps vs non-leaps). A very natural hypothesis is that none of the possible intrinsic natures of matter can be like this.⁴⁶

Of course, much more should be said about what kinds of physical principles should constrain modal space according to REGULAR (e.g. see Mittelstaedt (2013) and Darrigol (2014)). For our epistemological purposes, we will only need the claim that it is possible to understand REGULAR in a precise and principled way, and it is not an a priori truth that REGULAR is false.

⁴⁶ In contemporary physics, there are some interpretations of quantum mechanics that violate these continuity principles (such as objective collapse interpretations like GRW), but there are also popular interpretations of quantum mechanics that are perfectly consistent with these kinds of principles (such as Bohmian Mechanics and the Many Worlds Interpretation). See Maudlin (2019) for an introduction to these three different interpretations of quantum mechanics.

VII. Our Predicament: A Toy Case

Having finished describing a metaphysics on which premise 3 is false, let us finally turn back to our inconsistent triad. Orthodox modal metaphysics, inspired by Hume's a priori conceivability argument, endorses the following:

PLENITUDE: Every ideally conceivable structural possibility is realized by a genuine metaphysical possibility.

Because of PLENITUDE, orthodox metaphysics implies that premise 3 in our inconsistent triad is an a priori truth:

3. For any initial segment of some world, there are an uncountable infinity of metaphysically possible worlds that exactly match that initial segment and deviate afterwards.

I have argued that PLENITUDE is not an a priori truth. Hume's conceivability argument fails because NATURE is an a priori truth, and we fail to conceive of the intrinsic natures of things when we are conceiving of possibilities in which things go 'haywire'. In contrast to PLENITUDE, I have put forward a metaphysical picture in which REGULAR is true, which builds on NATURE and POWER. While I have argued that REGULAR cannot be ruled out a priori, I don't think that we are in a position to know that REGULAR is true a priori either. While many philosophers have given a priori arguments that support principles like NATURE, POWER, and REGULAR, the truth of REGULAR ultimately turns on the possible ineffable intrinsic natures of physical properties, which we are largely ignorant of. Since there are principled metaphysical pictures which give rise to both PLENITUDE and REGULAR, I believe we should have non-zero credence in both pictures.

However, if we don't know REGULAR a priori, how does this help with the problem of induction? Do we still need to 'skew' our probabilities towards regular possibilities to secure the rationality of induction? One of the main motivations for pursuing a Rationalist response to inductive skepticism was precisely to avoid the Dogmatist reply to our inconsistent triad, which needed to arbitrarily skew our prior probabilities in extremely strong ways. I believe that once REGULAR is open to us, there is hope for recovering induction without having to skew our probabilities at all. Perhaps we could recover induction by simply being *indifferent* among the possibilities within any given Modal Space! Here is a very toy example that illustrates this hope:

<u>Grid Modal Space</u>: Suppose that every possible world was a 10x10 grid. Each square may be black and white, and "times" are represented by the different columns. Suppose the actual world is such that the first five columns are white except for the top square, which is black. What will happen in the future?



In this situation, the analog of PLENITUDE is the hypothesis that every combination of white and black, all 2^{100} of them, are genuine metaphysical possibilities. Just to run the example, let us suppose that REGULAR corresponds to a regular modal space where *Inertia* is true: the color distribution of every column is identical to the color distribution of the first column.

There are three different background metaphysical approaches one might have to <u>Grid Modal</u> <u>Space</u>. I will discuss the epistemological consequences of each in turn.

Suppose, first, that we were convinced of a 'Humean' metaphysics, according to which all there is to any possible world is the distribution of colors. Then, there would be exactly 2^{100} worlds in our modal space. Natural indifference reasoning gives us that there is a $1/2^{50}$ chance that the regularity will continue to hold, since only 1 of the 2^{50} possible futures correspond to a square

where only the top row is black. This is a chance of ~0.0000000000000000. Of course, insofar as one follows the Dogmatist in skewing one's prior probabilities in favor of inductive friendly worlds, one will reach more inductively friendly conclusions.

Next, suppose we were convinced of an underlying modal space analogous to the typical Non-Humean position. On this view, there are all the worlds that the Humean countenances, but there are also *additional* regular worlds that are governed by fundamental laws. More precisely, for every grid that obeys Inertia, there is a corresponding Humean world *and* a corresponding non-Humean world that is governed by fundamental laws. On this view, there are exactly $2^{100} + 2^{10}$ worlds (since there are 2^{10} grids that obey *Inertia*). Applying natural indifference reasoning to this modal space gives us that there is a $2/(2^{50} + 1)$ chance that the regularity will continue, since there are two squares (a Humean one and a Non-Humean one) in which the regularity continues as before. This is a chance of ~0.0000000000002. Unfortunately, the Non-Humean position barely helps at all with the problem of induction! Again, one could always skew one's prior probabilities, the more inductively friendly conclusions one will reach.

Finally, here is the suggestion that I think we ought to adopt. Suppose one thinks that there are two epistemic possibilities for what the true metaphysical modal space consists in. One epistemically possible modal space obeys PLENITUDE and the other obeys REGULAR (i.e. *Inertia* is true). On this approach, we first need to assign prior probabilities to PLENITUDE and REGULAR. The appropriate probabilities to assign to these metaphysical hypotheses will crucially turn on the metaphysical issues we were surveying before. Insofar as the a priori metaphysical arguments in favor of principles like NATURE, POWER, and REGULAR are persuasive, we should assign higher probability to REGULAR. Insofar as they are unpersuasive, we should assign lower probability to REGULAR. Just to make things difficult, let us suppose that these metaphysical arguments aren't all that strong, and so one finds REGULAR to be very implausible. To make things precise, suppose one thought that there was a *one in a million* chance that REGULAR was true. After one assigns probabilities to REGULAR and PLENITUDE, one should proportion one's credence of 0.000001 indifferently among the REGULAR worlds, as well as one's credence of 0.999999 indifferently among the VAR worlds. Conditional on the truth of REGULAR, one should be indifferent among the various possibilities in REGULAR, and conditional on the

truth of PLENITUDE, one should be indifferent among the various possibilities in PLENITUDE. By doing this, because there are *so* many more worlds in PLENITUDE than in REGULAR, any particular world in REGULAR will be assigned *much* more credence than any particular world in PLENITUDE. After crunching the numbers, it turns out that observing the first five columns of the grid makes the chance that the regularity will continue to be ~0.999999999!

This strategy recovers everything we wanted in inductive inference. First, we don't want to be *certain* that the world won't go haywire tomorrow. We should always leave that possibility open. Second, we want to be very confident that past regularities will continue into the future. Third, we don't want to have to skew our probabilities arbitrarily in favor of regular worlds. All three of these desiderata are perfectly captured by this sort of approach!

It is important to note that this approach uses a version of the principle of indifference over *metaphysical* possibilities rather than *epistemic* possibilities. Here are two conflicting constraints one might place on rational prior probabilities:

Epistemic Indifference: For any two maximally specific epistemically possible worlds w_1 and w_2 , one should assign $Cr(w_1 | w_1 \text{ or } w_2) = 0.5 = Cr(w_2 | w_1 \text{ or } w_2)$.

Metaphysical Indifference: Given that it is epistemically possible that M describes the correct space of metaphysically possible worlds, for any two worlds w_1 and w_2 within M, one should assign $Cr(w_1 | M \text{ and } (w_1 \text{ or } w_2)) = 0.5 = Cr(w_2 | M \text{ and } (w_1 \text{ or } w_2))$.

Epistemic Indifference fails because we may have reason to favor w_1 over w_2 insofar as we think that it is more likely that w_1 is metaphysically possible. For example, suppose I have some a priori reason to assign a credence of 0.9 that w_1 is metaphysically possible and a credence of 0.1 that w_2 is metaphysically possible, even though both worlds are epistemic possibilities for me. Perhaps, for example, I have good reason to suspect that there is some subtle logical inconsistency in the description of world w_2 , even though I'm not *sure* that there is. Clearly, I should not be *neutral* with respect to w_1 or w_2 ! This is a straightforward counterexample to Epistemic Indifference. Moreover, because the ultimate truth about modal space will turn on the ineffable intrinsic natures of things, even *ideal* reasoners should not be certain about which epistemic possibilities are genuine metaphysical possibilities. Metaphysical Indifference does not have this flaw. Once one supposes that both possibilities are genuine metaphysical possibilities, there doesn't seem to be any reason to favor one possibility over another in the absence of any empirical evidence, and so it is natural to apply the principle of indifference.

According to this Metaphysical Indifference approach to the problem of induction, not only does our metaphysics help our epistemology, but epistemology greatly helps our metaphysics! In our toy case above, the empirical evidence of regularity constituted an *enormous* amount of evidence that the true metaphysical space of possibilities is given by REGULAR. In fact, the probability of REGULAR jumps from one in a million to ~0.999999999! If the most plausible explanation for REGULAR involves the claim that there are ineffable intrinsic natures of physical properties that are all powerful in non-gerrymandered ways, then we also have an extraordinary amount of purely empirical evidence for this highly abstract, metaphysical claim! One might have thought that arcane debates in modal metaphysics are entirely insensitive to ordinary empirical findings. But on the Metaphysical Indifference approach, ordinary empirical observations like an apple falling at 9.8 m/s² might have direct bearing on the ultimate nature of modal space.

Now of course, it should be emphasized that *far* more work has to be done to recover anything like induction in the real world. My discussion has been extremely abstract and has only concerned a very simple toy case. Nowhere have I talked about falling apples or rising suns. However, I hope to have motivated a distinctively *metaphysical* approach to the problem of induction, which is the only hope we have of avoiding the twin horns of Dogmatism and Skepticism.

References

Armstrong, David. 1978. Universals and Scientific Realism. Cambridge University Press.
Armstrong, David. 1983. What Is a Law of Nature? Cambridge University Press.
Armstrong, David. 1997. A World of States of Affairs. Cambridge University Press.
Bird, Alexander. 2007a. The Regress of Pure Powers? Philosophical Quarterly 57, no. 229:

- Bird, Alexander. 2007b. *Nature's Metaphysics: Laws and Properties*. Oxford University Press.Blackburn, Simon. 1990. Filling in Space. *Analysis* 50, no. 2: 62–5.
- Bourget, David, and David J. Chalmers. 2014. What Do Philosophers Believe? *Philosophical Studies* 170, no. 3: 465–500.

Bradley, Darren. Forthcoming. Naturalness as a Constraint on Priors. Mind.

Carroll, John. 1994. Laws of Nature. Cambridge University Press.

- Chalmers, David. 2002. Does Conceivability Entail Possibility? In *Conceivability and Possibility*, edited by Tamar S. Gendler and John Hawthorne, 145–200. Oxford University Press.
- Colyvan, Mark. 1998. Can the Eleatic Principle be Justified? *Canadian Journal of Philosophy* 28: 313–336.
- Cowling, Sam. 2015. Advice for Eleatics. In *The Palgrave Handbook of Philosophicl Methods*, edited by Chris Daly. Palgrave Macmillan.
- Darrigol, Olivier. 2014. Physics and Necessity: Rationalist Pursuits From the Cartesian Past to the Quantum Present. Oxford University Press.
- Ellis, Brian. 2001. Scientific Essentialism. Cambridge University Press.
- Field, Hartry. 1980. Science without Numbers. Oxford University Press.
- Foster, John. 2004. The Divine Lawmaker: Lectures on Induction, Laws of Nature, and the Existence of God. Oxford University Press.

Garber, Daniel. 1992. Descartes' Metaphysical Physics. University of Chicago Press.

Garber, Daniel. 1995. Leibniz: Physics and Philosophy. In The Cambridge Companion to

Leibniz, edited by Nicholas Jolley, 270–352. Cambridge University Press.

- Goff, Philip. 2017. *Consciousness and Fundamental Reality*. New York, USA: Oxford University Press.
- Goff, Philip. 2018. Revelation, Consciousness+ and the Phenomenal Powers View. *Topoi*: 1–4.

Grahek, Nikola. 2007. Feeling Pain and Being in Pain. Cambridge, MA: MIT Press.

Griffin, Michael V. 2012. Leibniz, God and Necessity. Cambridge University Press.

Hawking, Stephen. 1988. A Brief History of Time. Toronto: Bantam Books.

Hedden, Brian. 2015. A Defense of Objectivism About Evidential Support. *Canadian Journal of Philosophy* 45, no. 5: 716–743.

Heil, John. 2003. From an Ontological Point of View. Oxford University Press, 2003.

- Hofweber, Thomas. 2017. Are There Ineffable Aspects of Reality? Oxford Studies in Metaphysics, vol. 10, edited by Karen Bennett and Dean Zimmerman, 124–170. Oxford University Press.
- Huemer, Michael. 2009. Explanationist Aid for the Theory of Inductive Logic. *British Journal for the Philosophy of Science* 60, no. 2: 345–375.
- Ladyman, James. 2016. Structural Realism. In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta. URL =

<https://plato.stanford.edu/archives/win2016/entries/structural-realism/>.

Lange, Marc. 2007. Laws and Meta-Laws of Nature: Conservation Laws and Symmetries. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 38, no. 3: 457–481.

- Lange, Marc. 2012. There Sweep Great General Principles Which All the Laws Seem to Follow. In Oxford Studies in Metaphysics, Volume 7, edited by Karen Bennett and Dean Zimmerman, 154–185. Oxford University Press.
- Langsam, Harold. 2011. The Wonder of Consciousness: Understanding the Mind Through Philosophical Reflection. MIT Press, 2011.
- Langton, Rae. 2001. Kantian Humility: Our Ignorance of Things in Themselves. Oxford University Press.
- Langton, Rae. 2004. Elusive Knowledge of Things in Themselves. *Australasian Journal of Philosophy* 82, no. 1: 129–136.
- Leibniz, Gottfried. *NE. New Essays on Human Understanding*. Translated and edited by Peter Remnant and Jonathan Bennett. Cambridg University Press, 1981.
- Lewis, David. 2008. Ramseyan Humility. In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, 203–222. MIT Press.

Lipton, Peter. 2003. Inference to the Best Explanation. Routledge.

Lowe, E. J. 2006. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Clarendon Press.

Maudlin, Tim. 2007. The Metaphysics Within Physics. Oxford University Press.

Maudlin, Tim. 2019. Philosophy of Physics: Quantum Theory. Princeton University Press.

Mittelstaedt, Peter. Rational Reconstructions of Modern Physics. Springer.

Mørch, Hedda Hassel. 2014. Panpsychism and Causation: A New Argument and a Solution to

the Combination Problem. PhD Thesis, Oslo.

Mørch, Hedda Hassel. 2018a. The Evolutionary Argument for Phenomenal Powers. *Philosophical Perspectives*, 293–316.

Mørch, Hedda Hassel. 2018b. Does Dispositionalism Entail Panpsychism? Topoi: 1–16.

- Mørch, Hedda Hassel. 2019. Phenomenal Knowledge Why: The Explanatory Knowledge Argument Against Physicalism, in Coleman, S. (ed.) *The Knowledge Argument*, Cambridge University Press.
- Mørch, Hedda Hassel. Forthcoming. The Argument for Panpsychism From Experience of Causation. In *The Routledge Handbook of Panpsychism*, edited by William Seager. Routledge.
- Robinson, Howard. 1982. Matter and Sense: A Critique of Contemporary Materialism. Cambridge University Press.

Russell, Bertrand. 1927. The Analysis of Matter. London: Kegan Paul.

Sober, Elliott. 2015. Ockham's Razors: A User's Manual. Cambridge University Press.

- Strawson, Galen. 2006. Realistic Monism Why Physicalism Entails Panpsychism. Journal of Consciousness Studies 13, no. 10–11: 3–31.
- Van Inwagen, Peter. 1996. Why Is There Anything At All? *Aristotelian Society Supplementary Volume* 70, no. 1: 95–120.
- White, Roger. 2010. Evidential Symmetry and Mushy Credence. In Oxford Studies in Epistemology, edited by T. Szabo Gendler and J. Hawthorne, 161–186. Oxford University Press.

White, Roger. 2015. The Problem of the Problem of Induction. Episteme 12, no. 2: 275–290.

Williamson, Timothy. 2000. Knowledge and Its Limits. Oxford University Press.

Wilson, Alastair. 2013. Schaffer on Laws of Nature. *Philosophical Studies* 164, no. 3: 653–667.

PART TWO

The Problem of Self-Locating Belief

Chapter Three: Self-Locating Evidence and the Metaphysics of Time

For most metaphysical debates, ordinary observational evidence seems to be irrelevant. Intuitively, ordinary observational evidence can neither confirm nor disconfirm the existence of abstract objects, the existence of tropes or universals, the existence of arbitrarily scattered mereological fusions, etc. Arguments for these views tend to be at a very abstract, theoretical level. They are rarely of the following form: conditional on metaphysical theory X, we should expect to observe O_1 . Conditional on ~X, we should expect to observe O_2 , with $O_1 \neq O_2$. We have in fact observed O_1 , so we should raise our credence in X. The goal of this paper is to argue that ordinary observational evidence *can* directly confirm or disconfirm the two major positions in the metaphysics of time, Presentism and Eternalism, in exactly this way. There are cases where we should expect to observe different things depending on whether Eternalism or Presentism is true.

Presentism is the view that only present things exist, and so reality is three-dimensional. Eternalism is the view that objects from both the past and the future exist just as much as present objects, and so reality is four-dimensional. I will be considering an orthodox version of Eternalism according to which no time is objectively privileged as having the primitive property of *being present*. I will be setting aside the Growing Block theory, according to which only past and present objects exist, and the Moving Spotlight theory, according to which all past, present, and future objects exist but a specific time, the present, is objectively privileged.

If my thesis is right, that Presentism and Eternalism make different observational predictions, it puts pressure on a certain skeptical view about the entire debate between Eternalists and Presentists. According to one kind of skeptic, the debate between Eternalists and Presentists is 'merely verbal'.⁴⁷ Note that if 'exists' means 'exists now', then it's trivial that, say, 'Dinosaurs exist' is false, contra the Eternalist. If 'exists' means 'once existed or now exists or will exist', then it's trivial that 'Dinosaurs exist' is true, contra the Presentist. The skeptic then argues that it's very unclear what 'exists' could possibly mean that makes the dispute over 'Dinosaurs exist'

⁴⁷ Sider (2006) aptly characterizes this skeptical view and argues against it.

interesting and non-trivial.⁴⁸ However, if I am right that both views make different observational predictions, then even hard-core logical positivists will accept that the debate is substantive!

There are two main existing arguments regarding Presentism and Eternalism that are roughly empirical in character (although they are not at all as direct as the above template), so it is worth saying why the arguments in this paper are of an entirely different character than these. First, some authors have argued that general facts about our conscious experiences spell trouble for Eternalism, because our conscious experience can only be accounted for by a theory that recognizes 'objective becoming'. These sorts of arguments are highly contentious, and my arguments will not appeal to any controversial claims regarding conscious experience.⁴⁹ Another very popular argument in favor of Eternalism appeals to Special and General Relativity, according to which, on the standard interpretation, there are no objective facts about which events are simultaneous with which others. There are two worries about this style of argument that my arguments will avoid. First, there is some controversy about whether Quantum Gravity, which is supposed to supersede Special and General Relativity, will posit objective facts about which events are simultaneous with which others.⁵⁰ Second of all, one can make Special and General Relativity consistent with Presentism by positing a privileged reference frame (for example, see Prior (1972), Crisp (2007), and Markosian (2004)).⁵¹ Call this version of Presentism, which takes on board Special and General Relativity by positing a privileged foliation of space-time, Presentism*. There doesn't seem to be any ordinary observational evidence that can directly confirm Presentism* over Eternalism. My arguments avoid the first problem because they don't presuppose any controversial assumptions about a finalized physics. My arguments avoid the second problem because I will provide an explicit case where ordinary observational evidence can confirm or disconfirm Presentism* over Eternalism (section III).

Here is a roadmap for the paper. In Section 1, I will review some necessary basic distinctions (de dicto/de se and centered/uncentered worlds), and I will introduce the three basic ways of dealing

⁴⁸ For a recent defense of the view that there is no fact of the matter between Eternalism and Presentism, see Balaguer (2014). The genuineness of the presentism/eternalism debate is also discussed in Crisp (2003, 2004) and Ludlow (2004).

⁴⁹For a rebuttal of some of these arguments, see Skow (2011).

⁵⁰ For some reason to think that it might posit such facts, see Monton (2006).

⁵¹ It is worth noting, however, that there are some models of General Relativity for which this trick won't work (e.g. models with closed time-like curves can't have a privileged foliation of Cauchy slices).

with self-locating evidence by considering the three main responses to the classic Sleeping Beauty problem. In Section 2, I give the intuitive reason why self-locating evidence should be relevant to the debate between Presentism and Eternalism. In Sections 3-5, I present three distinct cases which cumulatively show that everyone should believe that self-locating evidence can be brought to bear on Presentism* vs Eternalism (in a relativistic setting) and Presentism vs Eternalism (in a classical setting). Lastly, in Section 6, I conclude with some morals of my argument.

1. Centered and Uncentered Worlds

Since my arguments will be appealing to self-locating evidence, we must first recall some basic distinctions. Following Lewis (1979), distinguish between *de dicto* beliefs and *de se* beliefs. De dicto beliefs characterize what reality is like in non-indexical terms. They are the sort of information that you would find in a complete 'map' of reality. Examples of de dicto beliefs include the belief that the earth orbits the sun, the belief that smoking causes cancer, and the belief that there are more than a billion humans. If two agents have the same de dicto belief, either both are right or both are wrong. However, we don't just have beliefs about which 'map' of reality is the correct one. We also have beliefs about our own location on the map. These are our *de se* beliefs. Examples of de se beliefs include the belief that I am now in California or that I am near a bank. They are the sort of information characterized by the 'you are here' sign on a map. Contrary to de dicto beliefs, if two people believe the de se content expressed by 'I am in Australia', one might be right and one might be wrong.

The content of de dicto beliefs may be represented using uncentered propositions, or sets of uncentered worlds. Uncentered worlds are just ordinary possible worlds. The content of de se beliefs may be represented using centered propositions, which are sets of centered worlds. I will take centered worlds to be ordered pairs consisting of an uncentered world w together with the spatiotemporal location of an individual in w.

There are three main updating rules that have been proposed for how one's credences ought to change in response to self-locating evidence, which correspond to the three most popular responses to the classic Sleeping Beauty problem. First, recall the setup of the case:

<u>Sleeping Beauty</u>: Beauty is a perfectly rational agent who is told that the following events will occur. On Sunday, she will be put to sleep. A fair coin will then be tossed. If it lands Heads, she will be awakened on Monday morning. Later, in the evening, she will be told that it is Monday, and then she will be let go. If the coin lands Tails, as before, Beauty will be awakened on Monday morning, and then she will be told that it is Monday later that evening. However, instead of being let go, she will be given a memory-loss drug that will make her forget all of her memories of Monday, and she will be put back to sleep. She will then be awakened on Tuesday, and then she will be let go. Her wakings on Monday and Tuesday will be indistinguishable. When she first awakens on Monday morning, what should her credence be that the coin landed Heads? When she is subsequently told that it is Monday, on Monday evening, what should her credence be that the coin landed Heads?

According to Thirders, Beauty's credence should be 1/3 in the morning and 1/2 in the evening (e.g. Dorr 2002, Elga 2000, and Titelbaum 2008). According to Lewisian Halfers, Beauty's credences should be 1/2 then 2/3 (e.g. Lewis 2001 and Bradley 2011). According to Double Halfers, Beauty's credences should be 1/2 in the morning and evening (e.g. Bostrom 2007 and Meacham 2008). For details about how to generalize these three positions to three generalized updating rules, see Briggs (2010), Meacham (2008), and Pettigrew (MS).

These three different views will respond differently to the cases I will present below. For each case, I will make sure to explain why each of the three positions responds in the way that it does. In my explanations, I will crucially be appealing to the bolded principles below.

First, consider the Double Halfer. The Double Halfer treats centered and uncentered information very differently. For the Double Halfer, one's credence in an uncentered proposition should only change when one receives evidence that rules out uncentered worlds. In other words, the Double Halfer endorses what Titelbaum (2008) calls the Relevance Limiting Thesis:

Relevance Limiting Thesis: Evidence that does not rule out any uncentered possibilities should not change one's credence in any uncentered proposition.

However, Double-Halfers will endorse the following principle:

Uncentered Elimination: For any two incompatible and jointly exhaustive uncentered propositions H_1 and H_2 , if E eliminates some uncentered possibilities in which H_1 is true (which are assigned some non-zero credence) but does not eliminate any uncentered possibilities in which H_2 is true, then upon learning E one's credence in H_2 should increase and one's credence in H_1 should decrease.^{52 53}

Second, consider the Thirder and Lewisian Halfer. The updating rules that generalize the Lewisian Halfer and Thirder position treat centered and uncentered information similarly. For my purposes, I only need the following claim:

Standard Conditionalization: For any two propositions H_1 and H_2 that are mutually exclusive and jointly exhaustive, and any other proposition E (which may be a centered proposition), if $Cr(E | H_1) < Cr(E | H_2)$, then updating on E should raise one's credence in H_2 and lower one's credence in H_1 . ⁵⁴

2. The Intuitive Case

I will be arguing that self-locating evidence can be brought to bear on the debate between Eternalists and Presentists. At first, this seems like an extremely strange result. Surely self-locating evidence can't be brought to bear on other metaphysical issues like the existence of abstract objects or the correct view of mereology. Why should the metaphysics of time be any different? The crucial observation is that *the Presentism/Eternalism debate has consequences for how to divide up the world into its centered and uncentered parts*. Given that different updating procedures handle information about centered worlds and uncentered worlds differently, it should come as no

⁵² Both **Relevance Limiting Thesis** and **Uncentered Elimination** are consequences of the updating rule that generalizes the Double-Halfer position. For details, see Briggs (2010) and Meacham (2008). Briggs calls the generalized rule the 'Halfer Rule' and Meacham calls the generalized rule 'Compartmentalized Conditionalization'. Both rules are equivalent.

⁵³ Purely for the sake of simplicity, in the cases below I will be assuming that Presentism and Eternalism are jointly exhaustive.

⁵⁴ **Standard Conditionalization** is a consequence of the updating rules that generalize the Thirder and Lewisian Halfer positions. In Meacham's (2008) terminology, this is because Thirders and Lewisian Halfers endorse 'Centered Conditionalization'. In Sleeping Beauty, Thirders and Lewisian Halfers use **Standard**

Conditionalization when they increase their credence that the coin will land Heads when they are told that 'It is Monday' on Monday evening.
73

surprise that some updating procedures can tease Presentism and Eternalism apart by exploiting their differences in how they divide the world up into centered and uncentered parts.

Here is an example. Suppose I know that there are three agents that are subjectively indistinguishable from each other that are located at different times and places in the world. I know that I am one of the three agents, but I don't know which I am. Two of them are simultaneous with each other and live in the year 2000 (perhaps I know one lives on Earth and the other lives on Twin Earth), and a third will live a thousand years after the first two, in the year 3000. For the Eternalist, this situation is described by having one uncentered four-dimensional possibility with three different epistemically possible centers of me. For the Presentist, this description is inadequate. For the Presentist, reality is three-dimensional, not four-dimensional. So, there are two distinct three-dimensional ways reality can be. I can either be in the three-dimensional uncentered possibility in which the year 2000 is present, or I can be in the three-dimensional uncentered possibility in which the year 3000 is present is actual. The first three-dimensional uncentered possibility has two epistemically possible centers of me, and the second uncentered world has one epistemically possible center of me. There just is no three-dimensional reality with three existing centers in it. In sum, for the Presentist uncentered possibilities are three-dimensional (since reality itself is three-dimensional), and for the Eternalist uncentered possibilities are four-dimensional (since reality itself is four-dimensional). Alternatively, we can put the difference as follows. For the Eternalist, facts about what time it is are purely indexical facts (and hence part of our de se beliefs). Facts about what time it is are just facts about what time the agent is located. However, for the Presentist, there is an *objective, non-indexical* distinction between what time *exists* and which times don't exist. Facts about what exists are paradigmatic de dicto facts.

Since Presentists and Eternalists differ on how to divide the world into its centered and uncentered parts and different updating procedures treat centered and uncentered information differently, it should come as no surprise that different updating procedures can tease apart Presentism and Eternalism. Even if this very abstract general point is currently unclear, hopefully it will become clearer when considering the three concrete cases below.

3. A Relativistic Case for Everyone

The following case distinguishes Presentism* and Eternalism for everyone: Thirders, Lewisian Halfers, and Double Halfers. It is set in the context of Special Relativity:

<u>Three Relativistic Rooms</u>: Suppose there are three indistinguishable rooms. You know that you are in one of the rooms, and the other two rooms have subjective duplicates of you inside them. Two of the rooms have a '1' on the outside and the other has a '2' on the outside. You know that the rooms are each space-like separated from each other (so that any pair of rooms exists at the same time according to some inertial frame of reference), but there is no inertial frame of reference in which all three rooms exist at the same time. From your perceptual evidence, you know that you are in one of these rooms, but you don't know which room you are in.



In the figure above, the four black lines correspond to the four different kinds of candidates for the 'objective present' for the Presentist*, while the yellow lines correspond to light-cones centered on each of the rooms. Intuitively, the left-right direction corresponds to space, and the up-down direction corresponds to time. So, the vertical lengths of the rooms correspond to how long in time the rooms exist for. For the Eternalist, the case involves one four-dimensional uncentered possibility with three centers in it. For the Presentist*, there are four different kinds of three-dimensional uncentered possibilities, corresponding to the four different kinds of candidates for the objective present. These four candidates are as follows:

(x) Only one '2' room exists.

- (x') Only one '1' room exists.
- (x'') Only two '1' rooms exist.
- (x''') One '2' room exists and one '1' room exists.

First, let us consider the case from the point of view of the Double-Halfer. Our crucial question for all of these cases will be the following: how should your credences change in Eternalism versus Presentism when you walk outside your room and observe which number is outside your room? Note that when you walk outside your room and observe your corresponding number, no uncentered Eternalist possibilities are eliminated. You will only eliminate either the '1' centers or the '2' center in the single uncentered Eternalist possibility. However, no matter what number you observe on your door, you *will* eliminate uncentered three-dimensional Presentist* possibilities. If you observe a '1', you will eliminate possibility (x). If you observe a '2', you will eliminate any uncentered Eternalist possibilities, and you will eliminate uncentered Presentist* possibilities. So, by **Uncentered Elimination**, no matter what you observe, your credence in Eternalism should increase and your credence in Presentism* will decrease, as desired.⁵⁵

Next, we will be able to consider the case from the point of view of Thirders and Lewisian Halfers together. Given Eternalism, there is just one four-dimensional world with three centers in it. Almost everyone in the Sleeping Beauty literature accepts the following 'highly restricted' principle of indifference (Lewis 2001: 172):

Center Indifference: If a single uncentered possibility has finitely many epistemically possible centers, one ought to divide one's credence equally among those centers.⁵⁶

⁵⁵ Yes, this is a reflection violation. Reflection violations are common place in self-locating cases. In the original Sleeping Beauty case, Thirders also face a reflection violation since on Sunday they know their credence that the coin will land Heads will drop when they wake up. Lewisian Halfers also know on Sunday that on Monday evening their credence that the coin will land Heads will and Heads will rise to 2/3.

⁵⁶ The restriction to the finite case is to avoid conflicting with Countable Additivity. I will not be needing the countable version of the principle anywhere in this paper.

So, letting Cr be your credence function before you see the number outside your room, you should set Cr(There's a '1' on your door | Eternalism) = 2/3.

Next, what should the value of Cr(There's a '1' on your door | Presentism*) be? Conditional on Presentism*, you are ignorant about which kind of three-dimensional reality you are in. Suppose you have credences $c_x, c_x', c_{x''}, c_{x'''}$ in the possibilities (x)-(x''') above. So, $0 \le c_x, c_x', c_{x''}, c_{x'''} \le 1$ and $c_x + c_{x'} + c_{x'''} + c_{x'''} = 1$. Then, the likelihood that there is a '1' on your door given these credences is $c_x' + c_{x''} + c_{x'''} + (c_{x'''/2})$. This is true since given (x), there are no rooms with a '1' on them. So, the likelihood that your room has a '1' is 0. Given (x') or (x''), there are only rooms with a '1' on them, so the likelihood your room has a '1' is 1. Lastly, given (x'''), the likelihood is 1/2 that your room has a '1' on it since there is one room with a '1' and one room with a '2'. Since, in almost every case, $c_{x'} + c_{x''} + (c_{x'''/2}) \ne 2/3$, it follows that in almost every case Cr(There's a '1' on your door | Presentism*) \ne Cr(There's a '2' on your door | Presentism*) \ne Cr(There's a '2' on your door | Eternalism). Similarly, in almost every case, Cr(There's a '2' on your door | Presentism*) \ne Cr(There's a '2' on your door | Eternalism). So, when you step outside your room and observe a particular number, after conditionalizing on you observing that number, Thirders and Lewisian Halfers agree that, in almost every case, your credences in Eternalism and Presentism* should change, by **Standard Conditionalization**.

4. A Finite Classical Case for Lewisian Halfers and Double Halfers

Next, consider the following Sleeping Beauty-like case in a classical, non-relativistic setting:

<u>Three Rooms</u>: Suppose you are told that the following events will occur. On Sunday, you will be put to sleep and will never be woken up again. While you are asleep, three perfect sleeping duplicates will be made of you. These three duplicates will be placed in three different rooms, each of which are indistinguishable from the inside. Two of the duplicates will be woken up at noon on Monday, and then let go. The third duplicate will be woken up at noon on Tuesday, and then let go. The room which contains the duplicate that will be woken up on Tuesday has the number '2' written on the outside of the room. One of the rooms which contains a duplicate that will be woken up on Monday has the number '1' written on the outside, and the other room has the number '2' written on the outside.

Since all these rooms are indistinguishable, the duplicates will not know which room they are in when they wake up. For the Eternalist, there is only one four-dimensional uncentered possibility with three possible centers in it. For the Presentist, there are two three-dimensional uncentered possibilities, one with two centers and one with one center.



First, let us consider the situation from the perspective of the Double Halfer. Much of what was said in the last section still holds true. No matter what you observe when you walk outside your room and observe your corresponding number, you will not eliminate any Eternalist uncentered possibilities. However, if you observe a '1', you will eliminate one Presentist possibility. In particular, you will eliminate the three-dimensional uncentered possibility in which there is only one '2' room. If you observe a '2' on the other hand, you will eliminate no uncentered possibilites. So, our verdict is as follows. For the Double-Halfer, if you observe a '1', your credence in Eternalism should increase and your credence in Presentism should decrease (by **Uncentered Elimination**). If you observe a '2', your credences in Eternalism and Presentism should remain the same, since in either case you don't eliminate any uncentered possibilites (by **Relevance-Limiting Thesis**)

Next, let us analyze the case from the perspective of Thirders and Lewisian Halfers together. As before, everyone agrees that you should set Cr(There's a '1' on the outside of your room | Eternalism) = 1/3 and Cr(There's a '2' on the outside of your room | Eternalism) = 2/3, by Center Indifference.

The case under Presentism is trickier. For the Presentist, one uncentered possibility T_1 is the threedimensional world in which it is Monday. Another uncentered possibility T_2 is the threedimensional world in which it is Tuesday. In T_1 , there are two relevant centers, and in T_2 , there is only one relevant center. *Note that this is structurally identical to the Sleeping Beauty problem*. In <u>Sleeping Beauty</u>, there are also two relevant uncentered possibilities (corresponding to Heads and Tails), where one of the uncentered possibilities has one center and the other uncentered possibility has two centers. T_2 is similar to the possible world where Heads is flipped, since it only has one center, and T_1 is similar to the possible world where Tails is flipped, since it has two centers. Because of this structural similarity, I believe that Thirders and Lewisian Halfers should treat <u>Three Rooms</u> (given Presentism) similarly to <u>Sleeping Beauty</u>.

The one disanalogy between <u>Sleeping Beauty</u> and <u>Three Rooms</u> is that in <u>Sleeping Beauty</u>, there is an objective chance of 1/2 that the Heads (or Tails) world is actual because of the fair coin, but there are no objective chances at work in <u>Three Rooms</u>. In response, it should be noted that <u>Sleeping Beauty</u> need not involve any objective chances. For example, suppose on Sunday Beauty had credence 1/2 that it rained in Boston on Saturday, and the experimenter told her that he would wake her up either once or twice depending on if it rained in Boston on Saturday. The case should be treated similarly even though there is no objective chance involved. Thirders should think that on Monday Morning Beauty's credence in the proposition that it rained in Boston on Saturday should deviate from 1/2, and Lewisian Halfers should think it should stay at 1/2.

According to the generalized updating procedure for Thirding developed in Briggs (2010), Pettigrew (MS), and Meacham (2008), what really matters in <u>Sleeping Beauty</u> is that Beauty assigns 1/2 to Heads *conditional on the uncentered portion of her evidence* on Monday Morning. An agent's 'uncentered portion of her evidence' only consists in all of the *uncentered* worlds that the agent has ruled out. For example, Beauty's uncentered portion of her evidence is the same on Sunday night and on Monday morning, since Beauty does not rule out any uncentered worlds on Monday morning. Beauty only learns the indexical fact 'it is now either Monday or Tuesday' when she wakes up, which does not rule out any uncentered worlds. Thirders then think that this additional indexical fact should *decrease* Beauty's credence in Heads given only the uncentered portion of her evidence. Lewisian Halfers believe that this additional indexical fact should *not* change Beauty's credence in Heads given only the uncentered.

So, to make the <u>Sleeping Beauty</u> case analogous to our <u>Three Rooms</u> case, I will stipulate that when the duplicates wake up their credence in T_1 conditional on the uncentered portion of their evidence is 1/2 (analogously, when Beauty wakes up her credence in Tails conditional on the uncentered portion of her evidence is also 1/2). It's worth mentioning that this stipulation is extremely natural. Since the uncentered portion of the duplicate's evidence is roughly *there is a room with an agent having experiences like such-and such*, this evidence is entirely neutral between T_1 and T_2 . Given that this uncentered evidence doesn't favor either T_1 or T_2 , natural indifference reasoning would suggest that the agent's credence in T_1 conditional on this evidence should be 1/2 anyway.

So, I have argued that, given this stipulation, the <u>Sleeping Beauty</u> case and our <u>Three Rooms</u> case should be treated similarly. In particular, the Thirder should believe that the Presentist should have credence 1/3 in T₂ and 2/3 in T₁, since, roughly speaking, the Thirder 'weighs' uncentered possibilities by how many relevant centers it has (similarly, the Thirder assigns 1/3 to Heads and 2/3 to Tails in <u>Sleeping Beauty</u>). This would have the consequence that the Thirder should think that Cr(There's a '1' on the outside of your room | Presentism) = 1/3 (using **Center Indifference**), which is the same as the Eternalist conditional credence 1/2 in T₁ and 1/2 in T₂ (similarly, the Lewisian Halfer should believe that the Presentist should have credence 1/2 in T₁ and 1/2 in T₂ (similarly, the Lewisian Halfer assigns 1/2 to Heads and 1/2 to Tails in <u>Sleeping Beauty</u>). This will have the consequence that the Lewisian Halfer should think that Cr(There's a '1' on the outside of your room | Presentism) = 1/4. So, since Cr(There's a '1' on the outside of your room | Beauty). This will have the consequence that the Lewisian Halfer should think that Cr(There's a '1' on the outside of your room | Presentism) = 1/4. So, since Cr(There's a '1' on the outside of your room | Beauty) > Cr(There's a '1' on the outside of your room | Presentism), if you see a '1' on the outside of your room when you exit you should increase your credence in Eternalism over Presentism. If you see a '2' on the outside of your room when you exit, you should increase your credence in Presentism over Eternalism (by **Standard Conditionalization**).

5. An Infinite Classical Case for Thirders and Lewisian Halfers

The following is a case that, I believe, distinguishes Presentism and Eternalism for both Thirders and Lewisian Halfers:

<u>Infinite Rooms</u>: Suppose you know that the world is such that there are infinitely many times $t_1 < t_2 < t_3 < ...$, where for each t_i there are three indistinguishable rooms, one with a

'1' on the outside and two with a '2' on the outside. All of the rooms across different times are also indistinguishable, and they each have subjective duplicates of you. You don't know what time it is or what room you are in.



From the perspective of a Double-Halfer, you will not change your credences in Presentism or Eternalism no matter what you see, because no matter what you see you will not eliminate any uncentered possibilities (by **Relevance-Limiting Thesis**).

Next, let us consider the case from the perspective of Thirders and Lewisian Halfers. This time, Thirders and Lewisian Halfers will agree on all of the conditional probabilities. We will start with the case of Presentism since it is easier. What should the value of Cr(There's a '1' on the outside of your room | Presentism) be? Well, conditional on Presentism, *you know for certain that there are exactly three rooms*, and that one of them has a '1' on it and the other two have a '2' on them. So, it seems natural to set Cr(There's a '1' on the outside of your room | Presentism) = 1/3 and Cr(There's a '2' on the outside of your room | Presentism) = 2/3. This natural thought also follows from **Center Indifference** together with the standard axioms of probability (including the axiom of Countable Additivity). Let T_i be the uncentered proposition that time t_i is the objective present. Then, in the context of this setup, Presentism is equivalent to T₁ \vee T₂ \vee T₃ \vee ... Note that the T_i are mutually exclusive. By **Center Indifference**, Cr(There's a '1' on the outside of your room | T_i) = 1/3 for any *i*. From the probability axioms, it therefore follows that Cr(There's a '1' on the

outside of my room | Presentism) = Cr(There's a '1' on the outside of your room | $T_1 \lor T_2 \lor T_3 \lor \dots$) = 1/3, as desired.

What should the value of Cr(There's a '1' on the outside of your room | Eternalism) be? It's hard to say. You have credence 1 that there exists infinitely many rooms that you could be in that have a '1' on them, and you have credence 1 that there exists infinitely many rooms that you could be in that have a '2' on them. If one were to naively apply **Center Indifference** (ignoring its restriction to finite cases), then one would get the result that your credence should be undefined, since 'infinity / (infinity + infinity)' doesn't make any sense. In the absence of any such norm to follow, one natural thought is that 'anything goes' in a situation like this. In particular, perhaps it is permissible to assign any credence at all to Cr(There's a '1' on the outside of your room | Eternalism). If this is correct, then the difference between Presentism and Eternalism is secured, since the likelihood of observing a '1' conditional on Presentism is determinately 1/3, while the likelihoods will differ.

However, you might think that you should have credence 1/3 even under Eternalism, because it's most 'natural' or 'simple' to group up the rooms in terms of the time slices they are at. Using this grouping, one can use a limiting frequency argument to justify staying at 1/3. This would probably be the position endorsed by Cian Dorr and Frank Arntzenius (forthcoming), according to their 'compromising approach'. Roughly speaking, the compromising approach says that reasonable credences will be 'generated by taking a weighted average of the credences that result from the different simple [groupings], in which the simpler ones get weighted more heavily' (29). This view, together with the claim that the above grouping is the only simple grouping there is for the case, would entail that your credence should be 1/3. The proposal raises obvious follow up questions about the nature of naturalness and simplicity. However, rather than following this line, one can concoct a case where even the compromising approach ought to say that your credence should not be 1/3.

Let us modify the case to account for the *spatial* positions of the rooms as follows. Suppose there are countably many planets p_1 , p_2 , p_3 , ... that are 'lined up' in a row. Stipulate that the case is such that each planet only ever has one of the rooms in question, where p_i has a room with a '2' on it if

and only if *i* is divisible by 10. If the case is modified like this, everyone should admit that one natural credence function for the Eternalist to have is 9/10, for the very same reasons that 1/3 is natural. Namely, we can group the rooms up in terms of what planet they are on, and then we can just do a limiting frequency argument. Surely the compromising approach will give *some* non-zero weight to this credence. In which case, the Eternalist ought to have a credence strictly between 1/3 and 9/10 that they will observe a '1'. Therefore, they will have a different likelihood of observing a '1' than the Presentist, as desired.

6. Some Morals

Let us take a step back from these cases and review the upshot of all of this. I have argued that, unlike most metaphysical debates, Eternalism and Presentism make different observational predictions. In other words, there are cases where ordinary observational evidence (like looking at a number on a door) should make one increase (or decrease) one's credence in Eternalism over Presentism. I have done this by considering three different cases which cumulatively show that Thirders, Lewisian Halfers, and Double Halfers should all agree both that Presentism and Eternalism make different observational predictions and that Presentism* and Eternalism make different observational predictions (note that all three of the cases are needed for this conclusion). I take this result on its own to be intrinsically interesting and counter-intuitive. However, I believe it also has some important morals.

First, if the above arguments are sound, suspicions that the debate between Eternalism and Presentism is 'merely verbal' are not justified. Theories that are mere notational variants cannot be directly confirmed or disconfirmed in different ways by ordinary empirical, observational evidence. Therefore, the debate in the metaphysics of time is substantive.

Second, I have argued that there is a quite strong connection between the metaphysics of time and puzzles about self-locating credences. This has the result that puzzle cases in the literature on self-location are often under-described. The three cases above show that one's credences in a particular self-locating case are sensitive to one's credences in Presentism and Eternalism. In effect, it seems that most of the self-location literature has implicitly assumed that the agents all have credence 1 that Eternalism is true. For example, in <u>Sleeping Beauty</u>, authors typically describe the case as having two uncentered possibilities (Heads and Tails), with one uncentered possibility having two

epistemically possible centers and the other uncentered possibility having one epistemically possible center. This is simply false if Presentism is true. If Presentism is true, there are three uncentered possibilities (Heads and Monday, Tails and Monday, and Tails and Tuesday), each of which has one epistemically possible center.

Lastly, I suspect that many readers will be suspicious of the argument in this paper because of its counterintuitive conclusion. It just seems wrong that looking at numbers on doors should be able to give one evidence about the metaphysics of time! Perhaps, since authors in the Sleeping Beauty literature seem to be presupposing Eternalism, the three updating procedures considered here should only be considered valid under Eternalism.⁵⁷ If this is right, then we are in need of a completely general self-locating updating procedure that is valid under different hypotheses about the metaphysics of time. One might then hope that the true general updating procedure, whatever it is, will not have the consequence that Presentism and Eternalism make different observational predictions. At the very least, then, I hope that the arguments in this paper motivate the development of such an updating procedure.⁵⁸

References

Arntzenius, Frank & Dorr, Cian. Forthcoming. Self-locating Priors and Cosmological Measures.

In Khalil Chamcham, John Barrow, Simon Saunders & Joe Silk (eds.), *The Philosophy of Cosmology*. Cambridge University Press.

Balaguer, Mark. 2014. Anti-Metaphysicalism, Necessity, and Temporal Ontology. Philosophy

and Phenomenological Research 89, no. 1: 145–167.

Bostrom, Nick. 2007. Sleeping Beauty and Self-Location: A Hybrid Model. Synthese 157, no. 1:

⁵⁷ Thanks to an anonymous referee for this suggestion.

⁵⁸ Thanks to David Balcarras, Caspar Hare, Agustin Rayo, Haley Schilling, Miriam Schoenfield, Ginger Schultheis, Brad Skow, Jack Spencer, Roger White, an anonymous referee, and the members of MIT's MATTI Graduate Student Workshop.

- Bradley, Darren. 2011. Self-Location Is No Problem for Conditionalization. Synthese 182, no. 3: 393–411.
- Briggs, Rachael. 2010. Putting a Value on Beauty. In Tamar Szabo Gendler and JohnHawthorne (Eds.), Oxford Studies in Epistemology, Volume 3. Oxford University Press, 3–34.
- Crisp, Thomas. 2003. Presentism. In *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman. Oxford University Press.
- Crisp, Thomas. 2004. Reply to Ludlow. Oxford Studies in Metaphysics 1: 37–46.
- Crisp, Thomas. 2007. Presentism, Eternalism and Relativity Physics. In W. Craig and Q. Smith (Eds.), Einstein, Relativity and Absolute Simultaneity, New York: Routledge: 262-278

Dorr, Cian. 2002. Sleeping Beauty: In Defence of Elga. Analysis 62, no. 4: 292–296.

Elga, Adam. 2000. Self-Locating Belief and the Sleeping Beauty Problem. Analysis 60, no. 2:

143–147.

- Lewis, David. 1979. Attitudes de Dicto and de Se. Philosophical Review 88, no. 4: 513–543.
- Lewis, David. 2001. Sleeping Beauty: Reply to Elga. Analysis 61, no. 3: 171–76.
- Ludlow, Peter. 2004. Presentism, Triviality, and the Varieties of Tensism. Oxford Studies in

Metaphysics 1: 21–36.

Markosian, Ned. 2004. A Defence of Presentism. Oxford Studies in Metaphysics 1, no. 3: 47-

84

82.

Meacham, Christopher J. G. 2008. Sleeping Beauty and the Dynamics of de Se Beliefs.

Philosophical Studies 138, no. 2: 245–269.

Monton, Bradley. 2006. Presentism and Quantum Gravity. In The Ontology of Spacetime, edited by Dennis Dieks.

Pettigrew, Richard. Unpublished Manuscript. Self-Locating Belief and the Goal of Accuracy.

Prior, Arthur N. 1972. The Notion of the Present. In *The Study of Time*, edited by J. T. Fraser, F. Haber, and G. Muller, 320–323. Springer Verlag.

Sider, Theodore. 2006. Quantifiers and Temporal Ontology. Mind 115, no. 457: 75–97.

- Skow, Bradford. 2011. Experience and the Passage of Time. Philosophical Perspectives 25, no. 1: 359–387.
- Titelbaum, Michael G. 2008. The Relevance of Self-Locating Beliefs. Philosophical Review 117, no. 4: 555–606.

Chapter Four: Time-Slice Rationality and Self-Locating Belief

The epistemology of self-locating belief concerns itself with how rational agents ought to respond to indexical information concerning where they are located in the world. Over the past couple of decades, it has been discovered that this issue is deeply interconnected with a wide variety of unresolved questions in formal epistemology. As Titelbaum (2013) has noted, the epistemology of self-locating belief is bound up with questions about relative frequencies, objective chances, Dutch book arguments, accuracy arguments, reflection principles, and indifference principles. Debates about self-locating belief are even interconnected with many of the central questions in the metaphysics of science, including debates about the correct interpretation of quantum mechanics (Bradley (2011a), Lewis (2006), Sebens and Carroll (2018)), the existence of multiple universes (Bradley (2012)), the measure problem in cosmology (Arntzenius and Dorr (2017)), and even classic questions in the metaphysics of time (Builes (2019)).

The goal of this paper is to argue for a novel way of resolving these debates about the epistemology of self-locating belief, by appealing to the recently popular thesis of *Time-Slice Rationality*, defended by Hedden (2015a, 2015b) and Moss (2015).⁵⁹ Since much of the literature on self-locating belief has revolved around Elga's (2000) Sleeping Beauty problem, my discussion will follow suit.

Before we start, some introductory remarks on Time-Slice Rationality and Sleeping Beauty are in order. Following Hedden (2015b), Time-Slice Rationality is the conjunction of two theses:

Synchronicity: What attitudes you ought to have at a time does not directly depend on what attitudes you have at other times.

Impartiality: In determining what attitudes you ought to have at a time, your *beliefs* about what attitudes you have at other times play the same role as your beliefs about what attitudes other people have. (p. 452)

⁵⁹ For some more recent criticisms and defenses of Time-Slice Rationality, see Podgorski (2016), Dori Döring and Eker (2017), Snedegar (2017), and Hedden (2016, 2017).

There are two main motivations for adopting Time-Slice Rationality. The first stems from the thought that the requirements of rationality should not make reference to the relation of personal identity over time, which notoriously gets very murky in certain puzzling cases (e.g. teletransportation, fission, partial brain transplants, etc.). As Hedden (2015b) writes:

Determining what an agent ought to believe does not require first figuring out the correct theory of personal identity over time. This means that requirements of rationality should not make reference to the relation of personal identity over time; what you ought to believe does not depend on who *you* are. That is, the requirements of rationality should be *impersonal*. (p. 452)

Let us call this intuition, that the requirements of rationality should not make reference to the relation of personal identity over time, the *No Reference* intuition. This intuition will bear much of the argumentative weight in the arguments below, and since the *No Reference* intuition is so central to Time-Slice Rationality, I will be taking it as an implicit commitment of the view from now on. Both *Synchronicity* and *Impartiality* are meant to be precisifications of the idea that the requirements of rationality should be 'impersonal' in the way that *No Reference* demands. The second motivation stems from the internalist intuition that what it is rational for an agent to believe should supervene on that agent's perspective. Since, according to the internalist, an agent's perspective on the world at a time is constituted by their mental life at that time, *Synchronicity* follows.

Next, let us recall the setup of the Sleeping Beauty case:

<u>Sleeping Beauty</u>: Beauty is a perfectly rational agent who is told that the following events will occur. On Sunday, she will be put to sleep. A fair coin will then be tossed. If it lands Heads, she will be awakened on Monday morning. Later, in the evening, she will be told that it is Monday, and then she will be let go. If the coin lands Tails, as before, Beauty will be awakened on Monday morning, and then she will be told that it is Monday later that evening. However, instead of being let go, she will be given a memory-loss drug that will make her forget all of her memories of Monday, and she will be put back to sleep. She will then be awakened on Tuesday, and then she will be let go. Her wakings on Monday and Tuesday will be indistinguishable. When she first awakens on Monday morning, what

should her credence be that the coin landed Heads? When she is subsequently told that it is Monday, on Monday evening, what should her credence be that the coin landed Heads?

There are three main responses to the problem. Currently, the most popular answer to the problem seems to be the Thirder position, according to which Beauty's credence should be 1/3 in the morning and 1/2 in the evening.⁶⁰ There are two other less popular answers. According to Lewisian Halfers, Beauty's credences should be 1/2 in the morning and 2/3 in the evening (e.g. Lewis (2001) and Bradley (2011a)). According to Double Halfers, Beauty's credences should be 1/2 in the morning and evening (e.g. Bostrom (2007) and Meacham (2008)).

I will be arguing that the connection between Time-Slice Rationality and Double Halfing runs very deep, and in fact the connection is already implicit in much of the literature. To show this, I will be presenting three independent arguments for why Time-Slicers should be Double Halfers. The first argument is a positive argument in favor of Double Halfing: if one takes an accuracy-first approach to <u>Sleeping Beauty</u>, then Time-Slicers should be Double Halfers. This first argument will be inspired by the discussion in Kierland and Monton (2005). The second argument is an argument against the popular Thirder view: perhaps the best way for Thirders to avoid certain highly implausible consequences of their view is to reject Time-Slice Rationality. This second argument will be inspired by discussions in Leitgeb (2010), Meacham (2008), and Bradley (2011a, 2015). The third argument is a defense of Double Halfing: perhaps the main objection to Double Halfing fails if Time-Slice Rationality is true. This third argument will be inspired by the initial exchange between Elga (2000) and Lewis (2001).

I. The First Argument: Accuracy and Sleeping Beauty

It would be nice if <u>Sleeping Beauty</u> could be settled using accuracy arguments. After all, it's natural to think that what matters at the end of the day is getting at the truth, not avoiding clever Dutch books or best satisfying our intuitions.⁶¹ Unfortunately, several authors have argued that accuracy arguments alone fail to settle <u>Sleeping Beauty</u> (e.g. Briggs (2010), Kierland and Monton (2005),

⁶⁰ Authors who have defended the Thirder solution include Arntzenius (2003), Dorr (2002), Draper and Pust (2008), Elga (2000), Hitchcock (2004), Horgan (2004), Monton (2002), Stalnaker (2008), Titelbaum (2008), and Weintraub (2004). For a useful survey of the arguments for the Thirder position, see Ross (2010).

⁶¹ For a book length defense of an accuracy-first approach to epistemology, see Pettigrew (2016).

89

and Pettigrew (MS)). I agree with these authors. However, I will argue that accuracy arguments *do* settle <u>Sleeping Beauty</u> when combined with the thesis of Time-Slice Rationality.

The reason why an accuracy-first approach to <u>Sleeping Beauty</u> fails to settle the problem is because on one precisification of 'minimizing inaccuracy', we should be Thirders, and on another precisification we should be Double Halfers. In the rest of this section, I will briefly outline why this is so.⁶²

Following Lewis (1979), we will be taking the objects of credence to be sets of *centered worlds*, which are ordered pairs consisting of a world *w* together with the location of a particular time-slice of an agent within *w*. A set *X* of centered worlds is called a *de dicto* or *uncentered* proposition iff whenever it contains some centered world, it also contains all centered worlds with the same world-coordinate. A set of centered worlds is called a *de se* or *centered* proposition otherwise. Intuitively, de dicto propositions are entirely about what the world is like, while de se propositions are also about where you are in the world. Examples of de dicto beliefs include the belief that space-time is curved, the belief that laptops exist, and the belief that a Republican won the 1992 US presidential election. If two agents have the same de dicto belief, either both are right or both are wrong. Examples of de se beliefs include the belief that I am in Minnesota. If two people believe the de se content expressed by 'It is Monday', one might be right and one might be wrong.

Next, some remarks about accuracy. Just as beliefs may be true or false, credences can be more or less accurate. Intuitively, we would like a measure of how accurate a credence is based on how far away it is from the credence function that assigns 1 to all truths and 0 to all falsehoods. As is common, the particular measure I will be using to measure the inaccuracy of a credal state will be the Brier score.⁶³ Here's how the Brier score works. First, for any world *w* and any uncentered proposition *X*, we will let w(X) = 1 iff *X* is true at *w* and w(X) = 0 iff *X* is false at *w*.⁶⁴ Then, given some credence function *c* and some uncentered proposition *X*, we define the *Brier score* of *c* at *w* with respect to *X* as:

⁶² The following discussion will largely follow Kierland and Monton (2005).

⁶³ For much more on different inaccuracy measures and their justifications see Leitgeb and Pettigrew (2010).

 $^{^{64}}$ We can think of a world w as a function from uncentered propositions to the truth value of that proposition at w.

$$B_{c,w}(X) = (w(X) - c(X))^2$$

So, what credence should Beauty assign to the proposition that the coin landed Heads when she wakes up if she wants to minimize her expected inaccuracy? There are two plausible ways to do this. First, one may try to minimize the expected *total* inaccuracy that one will accrue across all wakings by setting one's credence in the proposition that *the coin will land Heads* to *h*. There's an objective chance of 1/2 that one's *total* inaccuracy for this proposition will be $(1-h)^2$ if the coin lands Heads (since there will only be one waking), and there's an objective chance of 1/2 that one's *total* in accuracy for this proposition will be $(1-h)^2$ if the coin lands Heads (since there will only be one waking), and there's an objective chance of 1/2 that one's total inaccuracy will be $(0 - h)^2 + (0 - h)^2$ if the coin lands Tails (one summand for each waking). Using a bit of elementary calculus, we find that the value of *h* that will minimize one's expected total inaccuracy, which is $(1/2)*(1-h)^2 + 1/2*((0 - h)^2 + (0 - h)^2)$, is 1/3, vindicating Thirders.

Second, one may try to minimize the expected *average* inaccuracy that one will accrue in the following way. There's an objective chance of 1/2 that one's average inaccuracy in the proposition that *the coin will land Heads* will be $(1-h)^2$ if the coin lands Heads, and there's an objective chance of 1/2 that one's average inaccuracy will be $((0 - h)^2 + (0 - h)^2)/2 = h^2$ if the coin lands Tails, which is the average inaccuracy of both wakings. The value of *h* that will minimize $(1/2)*(1-h)^2 + 1/2*h^2$ is 1/2, vindicating Halfers.

Lastly, it is easy to see that both methods recommend adopting a credence of 1/2 in h on Monday evening. On Monday evening, there are two uncentered possible worlds that might be actual, the world in which the coin lands Heads and the world in which the coin lands Tails. Both worlds have an objective chance of 1/2 of being the actual world. Furthermore, since one knows that it is Monday evening on Monday evening, there is only one center in each world that is compatible with one's current evidence. So, since both uncentered worlds have an objective chance of 1/2 and both uncentered worlds only have one center compatible with one's evidence, the method of minimizing one's expected total inaccuracy and the method of minimizing one's expected average inaccuracy agree that one should have a credence of 1/2 in h on Monday evening.

So, minimizing expected total inaccuracy entails the Thirder position in <u>Sleeping Beauty</u>, and minimizing expected average inaccuracy entails the Double Halfer position in <u>Sleeping Beauty</u>. Briggs (2010) has generalized this result to arrive at fully general updating procedures that handle

all de se cases. In Briggs terminology, minimizing expected total inaccuracy implies the 'Thirder Rule' and minimizing expected average inaccuracy implies the 'Halfer Rule'.⁶⁵

The Halfer Rule is easy to describe. Suppose an agent's total evidence is E, and their current rational credence function is Cr. Let Cr* be the agent's rational prior credence function, and let E^* be the strongest de dicto proposition entailed by E. Then, according to the Halfer Rule, for any uncentered proposition X:

Halfer Rule:
$$Cr(X) = Cr^*(X | E^*)$$

While I will not be going over the details of Briggs' derivations of the Halfer Rule and the Thirder Rule here, it should be stressed that a successful defense of either one of these accuracy goals, given Briggs' results, amounts to a successful defense of a completely general theory of how to handle all de se and de dicto evidence.

So, how are we to decide between the competing goals of minimizing our expected average inaccuracy and minimizing our expected total inaccuracy? Prima facie, either goal looks equally permissible. Pettigrew sums up the dialectical situation between these two epistemic goals as follows:

Are there any such arguments [that favor one goal over the other]? I haven't seen them, nor have I been able to formulate them. Moreover, I find it hard to imagine how such an argument might go. Both understandings of probability seem reasonable; both seem to give reasonable definitions of expected inaccuracy; and, most importantly, both give definitions of a quantity that one would hope, intuitively, to minimize. Thus, any such argument would have to favour one and explain why our intuitive attraction to the other is mistaken. This is a difficult task. But that is not to say that it cannot be done. (p. 16)

In the next section, I will try to take up this challenge.

⁶⁵ The Halfer Rule is also equivalent to the updating procedures defended by Meacham (2008) and Halpern (2006).

II. The First Argument: Solving the Problem

We have two epistemic goals on the table: the goal of minimizing one's total inaccuracy and the goal of minimizing one's average inaccuracy. I will now argue that the goal of minimizing one's total inaccuracy is inconsistent with Time-Slice Rationality.

This fact can be brought out by considering the following variant of <u>Sleeping Beauty</u>:

<u>Duplicating Beauty</u>: On Sunday, Beauty will be put to sleep. She will be woken up on Monday, and then let go. A coin will then be tossed on Monday night. If it lands Heads, nothing happens. If it lands Tails, a perfect subjective duplicate of Beauty, call her Tuesday Beauty, will be created, and this duplicate will be woken up on Tuesday morning. Tuesday Beauty will then be let go. Beauty is told that her Monday waking will be subjectively indistinguishable from the Tuesday waking of Tuesday Beauty.

Note that this case is structurally identical to <u>Sleeping Beauty</u>. There are two uncentered possibilities, the Heads possibility and the Tails possibility, which each have an objective chance of 1/2. In the Heads possibility, there is only one epistemically possible center of Beauty, and in the Tails possibility, there are two epistemically possible centers of Beauty. When Beauty wakes up, for all she knows, it might be Monday or Tuesday. The only relevant change in this case is the personal identity facts. Beauty is not Tuesday Beauty.

However, in this case, the goal of minimizing one's total inaccuracy, as Kierland and Monton (2005) themselves note, recommends that Beauty ought to have a credence of 1/2 when she wakes up.⁶⁶ This is because the goal is to minimize one's *own* total inaccuracy. When Beauty wakes up, she knows that, no matter which uncentered possibility is actual (Heads or Tails), there is only one epistemically possible center that is *her*. Even if the coin lands Tails, she knows that she is either Beauty (and not Tuesday Beauty) or that she is Tuesday Beauty (and not Beauty). It easy to see that the credence that will minimize her expected total inaccuracy in this case is therefore 1/2.⁶⁷

⁶⁶ Kierland and Monton's position is that one ought to be a Halfer in <u>Duplicating Beauty</u>, but either Thirding or Halfing is permissible in <u>Sleeping Beauty</u>.

⁶⁷ If her credence in the coin landing Heads is *h*, her total inaccuracy will be $(1-h)^2$ if the coin lands Heads, and her total inaccuracy will be $(0-h)^2$ if the coin lands Tails (because there will only be *one* center that is her). So, her

So, the goal of minimizing one's total inaccuracy gives different recommendations in <u>Sleeping</u> <u>Beauty</u> and <u>Duplicating Beauty</u>. Since the only relevant difference in <u>Sleeping Beauty</u> and <u>Duplicating Beauty</u> are the personal identity facts, the recommendations that the goal gives explicitly depend on the personal identity facts, contradicting the *No Reference* intuition. In particular, since the goal requires Beauty to treat the attitudes of *other* people (Tuesday Beauty) very differently than the attitudes of herself, it explicitly violates the requirement of *Impartiality*. The goal is therefore inconsistent with Time-Slice Rationality.

Next, let us turn to our second main goal: the goal of minimizing one's average inaccuracy. This goal is not only consistent with Time-Slice Rationality, but it is the goal that is best motivated by Time-Slice Rationality. Given Time-Slice Rationality, you should only care about minimizing the inaccuracy of your *current* time-slice. Since you may be unsure of which time-slice you are in the world in cases of self-locating uncertainty, your best guess for the expected inaccuracy of your current time-slice is the expected *average* inaccuracy of all the time-slices that may be you, for all you know.⁶⁸ So, the goal of minimizing one's expected average inaccuracy *just is* the goal of minimizing the expected inaccuracy of your current time-slice. For the averager, whether or not certain centers are identical to other centers is entirely irrelevant.

Here is an analogous ethical case to drive the point home. Consider the debate between total utilitarians, who want to maximize the total utility in the world, and average utilitarians, who want to maximize the average utility in the world. Let possible world w_1 contain three agents, each with utilities 10, 20, and 30 respectively, and let possible world w_2 contain two agents, each with 25 utility. The total utilitarian will think w_1 is better than w_2 , and the average utilitarian will think that w_2 is better than w_1 . The crucial point is that average utilitarianism can be motivated in the following way. Suppose you were behind a veil of ignorance, and you were unsure of which agent you were going to be in the world. If your goal was to selfishly maximize your own utility, you

expected total inaccuracy is $(1/2)^*(1-h)^2 + (1/2)^*(0-h)^2 = h^2 - h + 1/2$. The value that minimizes this quantity is h = 1/2.

⁶⁸ It is worth noting that the expected inaccuracy of your current time-slice should actually be the *weighted* average of the inaccuracies of all the time-slices you may be (weighted by your credence in each of the corresponding centers). However, if one appeals to Elga's (2000) 'highly restricted principle of indifference', which says that one should assign equal credence to all the epistemically possible centers within an uncentered word, one will be required to assign equal weights to all those time-slices within the same uncentered world. While Elga's principle has proven very popular, for some pushback see Weatherson (2005). For a brief response to Weatherson (2005), see Bradley (2011a: 338-339).

would prefer that w_2 be actual rather than w_1 , because your expected utility in w_1 is the *average* utility of the three agents in w_1 , namely 20, and your expected utility in w_2 is the *average* utility of the agents in w_2 , namely 25. So, in the ethical case, if you selfishly want to maximize your own utility when you don't know who you are in the world, you take averages of the utilities of all the agents you might be.⁶⁹ Similarly, in the epistemic case, if you selfishly want to minimize the inaccuracy of your current time-slice when you don't know which time-slice you are in the world, you take averages of the inaccuracies of all the time-slices you might be.

So, Time-Slicers have a principled way to resolve the disagreement between the two competing epistemic goals. They should minimize expected average inaccuracy, resulting in the Double Halfer position and the Halfer Rule.

III. The Second Argument: Thirding and Time-Slice Rationality

For many philosophers, Thirding is unattractive because its most natural generalizations lead to unacceptable consequences. Leitgeb (2010) has argued that the Thirder is committed to implausible cosmological consequences. Meacham (2008) has argued that the Thirder is committed to implausible skeptical consequences. Bradley (2011a, 2015) has argued that the Thirder is committed to implausible quantum mechanical consequences. However, *all* of these consequences may be defused if the Thirder is willing to generalize their position in a way that explicitly rejects Time-Slice Rationality (by, for example, adopting the goal of minimizing total inaccuracy discussed above). This is a powerful reason for Thirders to reject Time-Slice Rationality. In the absence of other ways to avoid these consequences, Time-Slicers who wish to avoid these implausible consequences must reject Thirding.

Leitgeb's cosmological case is as follows:

<u>Eternal Recurrence</u>: Suppose astrophysicists have pinned down the actual evolution of our universe to either of two models. According to the first model, our universe is going to expand indefinitely. On the second model, our universe is expanding and contracting indefinitely, so that history repeats itself over and over again. An indeterministic quantum event shortly after the Big Bang determined whether our universe would evolve according

⁶⁹ Here I am assuming Elga's highly restricted principle of indifference (see footnote 11).

to the first or second model. There is an objective chance of 1/2 that the quantum event went either way. What should your credence be in Eternal Recurrence?

While the prima facie obvious answer is '1/2', Leitgeb argues that Thirders should have credence 1 in Eternal Recurrence in this case. Moreover, Thirders should have credence 1 in Eternal Recurrence so long as the objective chance of Eternal Recurrence is *non-zero*! In fact, given that the objective chances are dispensable, Thirders seem to be forced to have credence 1 in Eternal Recurrence, so long as their prior probability in Eternal Recurrence is non-zero!⁷⁰

Meacham's skeptical case is as follows:

<u>Many Brains</u>: Consider the proposition that you're in a world where brains in vats are constantly being constructed in states subjectively indistinguishable from your own. Let your credence in this possibility be 0 , and your credence that there will be no multiplication of [your subjective states] be <math>1 - p.

Meacham argues that if you accept Thirding, then 'you should come to believe (if not yet, then in a little while) that these brains in vats are being created . . . as you become certain that these brains in vats are being created, you should become certain that you're a brain in a vat' (p. 260). This is an unwelcome result.

Bradley's (2011a, 2015) quantum-mechanical case is as follows, where 'MWI' refers to the Many Worlds interpretation of Quantum Mechanics:

<u>Quantum Measurement</u>: You are about to perform a spin measurement with possible outcomes Up and Down. Quantum mechanics says that Up and Down each has a chance of fifty percent. According to MWI, the universe will divide, so you will have two future successors, one of whom will observe Up, and one Down. According to a stochastic theory (ST), there will be only one future successor, who will observe either Up or Down, each

⁷⁰ To my knowledge, the only Thirder who has argued that the presence of objective chance makes a crucial difference in cases of self-locating uncertainty is Wilson (2014). Wilson also argues against the case of <u>Quantum</u> <u>Measurement</u> below on the grounds that the correct interpretation of Quantum Mechanics is not decided by a chancy process. For a response to this suggestion, see Bradley (2015: 689-692).

with fifty percent probability. You are unsure of whether MWI or ST is correct, and you assign each a credence of fifty percent.

Bradley argues that Thirders are committed to the view that you ought to increase your credence in MWI after you perform the experiment, regardless of what your observation will be! If he is right, Thirders are therefore committed to 'easy confirmation' of MWI. Every time the universe branches, we gain reason to believe in MWI over ST! As Bradley writes, '. . . our everyday observations are constantly confirming MWI. On this reasoning, MWI gets enormous confirmation without the need for modern physics. The Ancients could have worked out that they have overwhelming evidence for MWI merely by realizing it was a logical possibility and observing the weather' (2011a: p. 336).

It's not hard to see why these philosophers have seen connections between these cases and <u>Sleeping Beauty</u>. Each of the above three cases shares a common structure with <u>Sleeping Beauty</u>. In <u>Sleeping Beauty</u>, there are two relevant uncentered possibilities: the Heads possibility and the Tails possibility. Similarly, in <u>Eternal Recurrence</u> there are two relevant cosmological models; in <u>Many Brains</u> there is one skeptical possibility and one non-skeptical possibility; in <u>Quantum Measuerment</u> there are two different interpretations of Quantum Mechanics. In <u>Sleeping Beauty</u>, Thirders deviate from their initial credence in these two possibilities by giving more 'weight' to possibilities in proportion to how more epistemically possible centers it contains. Similarly, in each of our three cases, one of the two relevant possibilities only includes one epistemically possible center while the other includes more than one epistemically possible center. So, it seems like Thirders should assign more credence to the possibilities which includes more centers (i.e. the eternal recurrence possibility, the skeptical possibility, and the MWI possibility).

Moreover, many of the standard arguments for the Thirder position in <u>Sleeping Beauty</u> straightforwardly apply to these cases. For example, Elga's (2000) first argument for the Thirder position is a long-run frequency argument. If we imagined the Sleeping Beauty experiment being run many different times, in the long run roughly 1/3 of Beauty's wakings will correspond to a Heads flip and roughly 2/3 of the wakings would correspond to a Tails flip. Similarly, if we imagine a 'multiverse' in which many different universes come into existence, each with chance 1/2 of being one-history worlds or eternal recurrence worlds, then almost every agent would end

up being in an Eternal Recurrence world. Following the same line of reasoning leads one to have credence 1 that one is an Eternal Recurrence world!

In addition, the result that Thirders should adopt these implausible consequences is also a consequence of many of the formal, mathematically precise generalizations of the Thirder position that have been developed in the literature. While I will not rehearse these generalizations here, readers are encouraged to see Briggs (2010), Meacham (2008), and Pettigrew (MS) for generalizations of the Thirder position which all imply the counterintuitive consequences above.⁷¹

The Thirder who wishes to avoid these three consequences must identify some relevant difference between <u>Sleeping Beauty</u> and these cases. Fortunately, there is a clear difference: only in <u>Sleeping Beauty</u> do the relevant centers within each uncentered possibility correspond to the *same* agent.⁷² The Thirder also has a principled reason for thinking that this difference is a *relevant* difference, namely that the goal of minimizing *total* inaccuracy (discussed above) is a principled Thirder position that is sensitive to this sort of difference. Because the goal of minimizing total inaccuracy gives 'Halfer' verdicts in cases where there aren't multiple centers of the *same* agent in a single possible world, it avoids giving implausible verdicts in any of our three cases above. In the absence of some other relevant difference between <u>Sleeping Beauty</u> and these three cases, Time-Slicers who wish to avoid these three counterintuitive consequences should reject Thirding.

IV. The Third Argument: The Diachronic Argument Against Double Halfing

In the original exchange between Elga (2000) and Lewis (2001), which introduced the Sleeping Beauty problem to philosophers, it is striking that the Double Halfer view was never even

⁷¹ Interestingly, while Titelbaum's (2013) Thirder framework does have counterintuitive consequences in <u>Eternal Recurrence</u> and <u>Many Brains</u>, Titelbaum has argued that it avoids the charge of easy confirmation in <u>Quantum Measurement</u> (p. 273 – 282). However, Titelbaum only avoids the charge of easy confirmation *given a certain theory about personal identity*. If the experimenter before the experiment is not identical to *either* person after the experiment, or if the experimenter before the experiment is identical to *both* persons after the experiment, Titelbaum's framework does not generate the desired probabilities. However, given an account of personal identity defended by Lewis (1976) and supplemented by Saunders and Wallace (2008), Titelbaum's account does give the desired probabilities. On this account, there really are two observers *before* the experiment, each of which is identical to one of the observers after the experiment. So, there really isn't an increase in the number of observers before and after the experiment! While this interpretation of the personal identity facts can be questioned on independent grounds (e.g. see Tappenden (2008)), the Time-Slicer will also be wary of the fact that this framework rejects *No Reference*, since it gives different verdicts depending on the correct theory of personal identity.
⁷² This assumes that in cases of fission (like <u>Quantum Measurement</u>), neither of the post-fission agents is identical to the pre-fission agent.

mentioned as a possible solution to the problem. In fact, the debate over the problem continued for several years under the assumption that the Thirder and Lewisian Halfer positions were the only possible solutions. The reason for this widespread presumption against the Double Halfer position is that philosophers were relying on an essentially *diachronic* constraint on any possible solution to the problem. Namely, everyone agreed that Beauty's credence in Heads, whatever it is, should *increase* between Monday morning and Monday evening. The motivation for this diachronic constraint is straightforward. On Monday morning, Beauty has three options open to her corresponding to the two possible wakings on Tails and the one possible waking on Heads. Once Beauty is told that it is Monday on Monday evening, she is able to eliminate one of the possible Tails wakings. So, it seems to be a straightforward consequence of conditionalization that Beauty's credence in Heads should increase.

Given this diachronic constraint, Elga used as a (synchronic) premise that Beauty's credence in Heads should be 1/2 on Monday evening and retroactively inferred that her credence must have been 1/3 on Monday morning, and Lewis used as a (synchronic) premise that Beauty's credence in Heads should be 1/2 on Monday morning and inferred that her credence should increase to 2/3 in the evening. So, it seemed that one simply had to pick one's poison. One could either be a Halfer on Monday morning or a Halfer on Monday evening, but one couldn't be a Halfer at both times.

Many philosophers have questioned whether this diachronic constraint should really be adhered to in this case.⁷³ However, the Time-Slicer has an easy response here. Time-Slicers are free to choose neither poison and endorse both of Lewis' and Elga's synchronic premises. According to Time-Slicers there *just are* no essentially diachronic constraints on rationality; the fundamental norms of rationality are all synchronic. For any given case, the Time-Slicer should only be asking what credences an agent should assign at a particular time given the agent's evidence at that time, without any heed to what credences the agent assigned at any other time.

That being said, it might well be that there are diachronic norms of rationality which can be *derived* from purely synchronic norms of rationality. So one might naturally wonder whether the diachronic premise appealed to by Lewis and Elga can be derived in this way. Unfortunately,

⁷³ Some philosophers who have questioned this diachronic constraint include Bostrom (2007), Meacham (2008), Briggs (2010), and Cozic (2011).

perhaps the most promising way to attempt such a derivation, which roughly follows the way Hedden (2015a, 2015b) derives a diachronic version of conditionalization, runs into difficulties when dealing with diachronic updates on self-locating information.

Here is how one might attempt such a derivation. First, stipulate as part of the case that Beauty has the same prior probability function on Monday morning and Monday evening, which encodes what credences she would have in every (centered and uncentered) proposition in the absence of any evidence.⁷⁴ Next, propose the following synchronic norm:

Synchronic Conditionalization: If an agent A at time t has total evidence E and prior credence function C, then their credence at t in each proposition H should equal C(H | E).

Given Synchronic Conditionalization, then if one's evidence grows monotonically and one retains the same prior credence function, then one's credences should evolve in the standard way governed by diachronic versions of conditionalization.⁷⁵

In the context of <u>Sleeping Beauty</u>, however, Bostrom (2007) notes that Beauty's total evidence on Monday evening is *not* strictly greater than Beauty's total evidence on Monday morning. On Monday morning, it is part of Beauty's evidence that *I have not been told that it is Monday*, but on Monday evening, Beauty 'loses' this piece of evidence and learns its negation: *I have been told that it is Monday*! Consequently, there is no straightforward way to apply conditionalization in this case (as well as other cases where self-locating evidence is at issue).

In response to Bostrom, Titelbaum (2013) has developed a modeling framework, the 'Certainty-Loss Framework' (CLF), which entails that Beauty's credence in Heads should increase from Monday morning to Monday evening, even when accounting for Beauty's evidence that *I have been told that it is Monday* (p. 217-219). However, Titelbaum's framework invokes a crucial principle called (PEP), and as Titelbaum himself says, 'I categorize (PEP) as a diachronic

⁷⁴ Hedden himself endorses *Uniqueness*, which, in the context of Bayesian epistemology, is the claim that there is a uniquely rational prior probability function. While *Uniqueness* would guarantee that Beauty is rationally required to have the same prior on Monday morning and Monday evening, for our purposes we need not rely on it.

⁷⁵ For suppose that at t_1 your total evidence is E_1 , and at t_2 you gain evidence E_2 (which makes your total evidence $E_1 \land E_2$). Then, if you have prior credence function *C* at both times, by *Synchronic Conditionalization*, your credences at t_1 should be $C_1(-) = C(-|E_1 \land E_2|)$. C_2 is just the probability function that results from taking C_1 and conditionalizing on E_2 .

systematic constraint of CLF' (p. 194). Time-Slicers, however, will be skeptical of any such diachronic principles. Meacham (2008) also notes that both Thirders and Lewisian Halfers need to endorse certain (underived) diachronic Continuity principles to get the diachronic results that they want to have. Again, Time-Slicers will not find any such primitive diachronic principles attractive. While the diachronic premise invoked by Elga and Lewis seems uncontroversial at first glance, it turns out to be very difficult to justify by only appealing to uncontroversial synchronic principles.

In sum, given Time-Slice Rationality, the only objections to the Double Halfer position that have any hope of being successful are ones that rely on entirely synchronic premises. Consequently, perhaps the most powerful objection to the Double Halfer position fails to get off the ground given Time-Slice Rationality.⁷⁶

V. The Costs of Denying Time-Slice Rationality

So far, I have only argued for the conditional claim that if you're a Time-Slicer, you should be a Double Halfer. In response, one might think that this conditional claim isn't terribly interesting, given that Time-Slice Rationality is a relatively new thesis, and it's dubious whether most philosophers would endorse it anyway. In this section, I will try to push back on this natural thought by drawing out some implausible consequences of any epistemology of self-location that explicitly denied Time-Slice Rationality.

Consider the following variant of <u>Sleeping Beauty</u>:

<u>Sorites Beauty</u>_x: On Sunday, Beauty will be put to sleep. A fair coin will then be tossed. If it lands Heads, she will be woken up on Monday and then let go. If it lands Tails, she will be woken up on Monday and then put back to sleep. An evil neurosurgeon will then replace

⁷⁶ One important synchronic argument against Double Halfing is given in Titelbaum (2012). If one supplements <u>Sleeping Beauty</u> with the claim that a fair coin will *also* be flipped on Tuesday (which will have no effect on the rest of the experiment), then it turns out that Double Halfers must assign a credence greater than 1/2 in the proposition that 'today's coin flip will land Heads' on Monday morning. In response, I grant that this a strong objection to Double Halfers who are primarily motivated by aligning their credences with the objective chances. However, this is not the relevant motivation for the Time-Slicer. The Time-Slicer is primarily motivated by having an *impersonal* epistemology, which is encapsulated in principles like *No Reference, Synchronicity*, and *Impartiality*. Given that the proposition 'today's coin will land Heads' is a merely indexical proposition, the fact that the Double Halfer gives an unintuitive verdict on this indexical proposition shouldn't be much of an embarrassment for the Time-Slicer.

X% of Beauty's body with qualitative duplicate parts while she sleeps, so that the resulting body will be qualitatively just like the original body. The resulting agent, call her Tuesday Beauty, will then be woken up on Tuesday in the usual room.

Again, this case is structurally similar to <u>Sleeping Beauty</u>. Given Heads, there is one waking, and given Tails, there are two wakings, and all three possible wakings are indistinguishable. However, in this case the personal identity facts are being distorted. When X=0, the case should clearly be treated the same as <u>Sleeping Beauty</u>. When X=100, Tuesday Beauty is only a qualitative duplicate of Beauty, so Beauty \neq Tuesday Beauty.⁷⁷

There are two different ways one can react to <u>Sorites Beauty</u>, given that one treats <u>Sleeping Beauty</u> and <u>Duplicating Beauty</u> differently (presumably by being a Thirder in <u>Sleeping Beauty</u> and a Halfer in <u>Duplicating Beauty</u>, as per the goal of minimizing total inaccuracy). First, there is an externalist approach. On this externalist approach, if Tuesday Beauty is *in fact* identical to Beauty (regardless of what beliefs Beauty has about whether Tuesday Beauty is Beauty), then Beauty ought to assign credence 1/3 on Monday morning, just as in <u>Sleeping Beauty</u>. If Tuesday Beauty is *in fact* not identical to Beauty (regardless of what beliefs Beauty (regardless of what beliefs Beauty (regardless of what beliefs Beauty as about whether Tuesday Beauty). If Tuesday Beauty is *in fact* not identical to Beauty (regardless of what beliefs Beauty (regardless of what beliefs Beauty as about whether Tuesday Beauty has about whether Tuesday Beauty is *in fact* not identical to Beauty (regardless of what beliefs Beauty has about whether Tuesday Beauty has about whether Tuesday Beauty is Beauty), then Beauty ought to assign credence 1/2 on Monday morning, just as in <u>Duplicate Beauty</u>.

This externalist approach clearly conflicts with *No Reference*. What credence Beauty ought to assign on Monday morning crucially depends on whether Beauty = Tuesday Beauty. Moreover, it also seems to conflict with *Synchronicity*. *Synchronicity* states that what attitudes you ought to have at a time does not directly depend on what attitudes you have at other times. However, on the externalist approach, what credence Beauty ought to have on Monday morning explicitly depends on whether it is *Beauty* who is adopting a credal state about the coin on Tuesday morning. If she is the one adopting attitudes on Tuesday morning, Beauty should have a 1/2 credence on Monday morning, and if she is not adopting any attitudes on Tuesday morning, Beauty should have a 1/3 credence on Monday morning. So, the credence Beauty ought to have on Monday morning depends on what Beauty's attitudes are at other times.

⁷⁷ If you do not believe that Beauty \neq Tuesday Beauty in <u>Sorites Beauty₁₀₀</u>, then pick some other analogous soritical case where Beauty = Tuesday Beauty in <u>Sorites Beauty₀</u>, and Beauty \neq Tuesday Beauty in <u>Sorites Beauty₁₀₀</u>

Even bracketing *Synchronicity*, the externalist approach seems unattractive. It faces the familiar internalist worry that it does not give followable advice, and an agent would be epistemically blameless if they failed to follow the externalist advice.⁷⁸

A second approach is an internalist one. According to this approach, if Beauty *believes* that Tuesday Beauty is identical to Beauty (regardless of whether Beauty is in fact identical to Tuesday Beauty), then Beauty ought to assign credence 1/3 on Monday morning just as in <u>Sleeping Beauty</u>. If Beauty *believes* that Tuesday Beauty is not identical to Beauty (regardless of whether Tuesday Beauty is in fact identical to Beauty), then Beauty ought to assign credence 1/2 on Monday morning, just as in <u>Duplicate Beauty</u>.

This internalist approach explicitly goes against *Impartiality*, which is the thesis that your *beliefs* about what attitudes you have at other times should play the same role as your beliefs about what attitudes other people have. According to the internalist approach, your beliefs about what attitudes *you* have at other times plays an entirely different role than your beliefs about what attitudes *other* people have (e.g. when Tuesday Beauty is not Beauty).

It is worth pausing to draw out some implausible consequences of the internalist version of the goal, even bracketing any commitment to the thesis of Time-Slice Rationality. What should the internalist version of the goal say when Beauty is unsure about whether Tuesday Beauty is identical to Beauty? For example, if we let X=50 in <u>Sorites Beauty</u>, it is natural to be unsure whether Tuesday Beauty is identical to Beauty.⁷⁹ Suppose, for example, that Beauty has a credence of 1/2 that Beauty = Tuesday Beauty in this case. We know that if Beauty = Tuesday Beauty, <u>Sorites Beauty</u>₅₀ should be treated like <u>Sleeping Beauty</u>, and if Beauty \neq Tuesday Beauty, <u>Sorites Beauty</u>₅₀ should be treated like <u>Duplicating Beauty</u>. So, it looks like if Beauty has credence 1/2 that Beauty = Tuesday Beauty should assign a credence of (1/2)*(1/2) + (1/2)*(1/3) = 5/12 to the coin landing Heads! So, in this case, neither Thirders nor Halfers are correct, but 5/12-ers are correct!

⁷⁸ Of course, there are familiar externalist reasons for thinking that *no* epistemic theory can always give followable advice having to do with Williamson's (2000) Luminosity argument. For a response to the Luminosity argument, see Berker (2008). For motivations to be an internalist that do not appeal to Luminosity, see Schoenfield (2015). ⁷⁹ If you have determinate intuitions for the case of X=50, pick some other value of X for which you lack a determinate intuition.

It is telling that no author has ever offered any answer other than 1/3 or 1/2 in any structurally similar variation to the <u>Sleeping Beauty</u> case.

A second implausible consequence has to do with what sort of evidence this internalist version of the goal deems relevant to the proposition that the coin landed Heads. Suppose Beauty wakes up on Monday morning and assigns credence 1/2 that Beauty = Tuesday Beauty, and hence assigns a credence of 5/12 to the proposition that the coin lands Heads. Suppose Beauty really wants to assign the best credence she can to the coin's landing Heads, so she spends all of Monday morning reading literature on the metaphysics of personal identity (before she is told that it is Monday). After reading many journal articles, her credence that Beauty = Tuesday Beauty drops to 1/4, and hence her credence that the coin will land Heads goes from 5/12 to (1/4)*(1/2)+(3/4)*(1/3) = 3/8. Intuitively, however, all these abstract arguments in the metaphysics journals should be treated as evidentially irrelevant to the coin's landing Heads or Tails. One shouldn't be able to pursue one's curiosity about the state of a fair coin by reading metaphysics journal articles on personal identity!

VI. Concluding Remarks

Apart from these three more specific arguments, it is worth noting that the general philosophical spirit behind the Time-Slice view and the Double Halfer view are remarkably similar. Both the Double Halfer and the Time-Slicer adopt a deflationary stance on the importance of 'I' in epistemology. Time-Slicers think that facts about which time-slices across the universe count as 'I' are simply irrelevant to one's theorizing. Time-slices that count as other people should be treated in the same way as time-slices that count as oneself. Similarly, the natural generalization of the Double Halfer position, in the form of the Halfer Rule, also adopts a deflationary take on the importance of 'I' in epistemology. The Halfer Rule entails the so-called 'Relevance Limiting Thesis' introduced by Titelbaum (2008), which says that one should only revise one's credences in uncentered propositions when one learns new uncentered propositions as evidence. In other words, essentially indexical information is *always irrelevant* to one's theorizing about non-indexical matters. For example, in <u>Sleeping Beauty</u>, since Beauty learns no new non-indexical information on Monday morning or Monday evening, she ought to retain her credence of 1/2 in Heads. Given the Time-Slicers deflationism about the role of 'I' in epistemology, a principle like the Relevance Limiting Thesis is a natural one to adopt. Combining the Double Halfer View and

the Time-Slice view results in an epistemology that takes a sort of 'view from nowhere', which deflates the importance of the subject in one's inquiry into the objective, non-indexical world.

Meacham (2010) has argued that de se puzzle cases form a 'tangled web' with other issues that are independent of self-locating beliefs *per se*, such as the internalism vs externalism debate in epistemology and debates about how to handle identity over time. I believe my discussion has supported this contention that these de se cases form such a tangled web, but at the same time I hope to have offered a principled way to untangle this web for those who endorse Time-Slice Rationality. Time-Slicers should be Double Halfers.⁸⁰

References

- Arntzenius, Frank. 2003. Some Problems for Conditionalization and Reflection. *Journal of Philosophy* 100, no. 7: 356–370.
- Arntzenius, Frank, and Cian Dorr. 2017. Self-Locating Priors and Cosmological Measures. In *The Philosophy of Cosmology*, edited by Khalil Chamcham, John Barrow, Simon Saunders, and Joe Silk, 396–428. Cambridge: Cambridge University Press.

Berker, Selim. 2008. Luminosity Regained. Philosophers' Imprint 8: 1-22.

- Bostrom, Nick. 2007. Sleeping Beauty and Self-Location: A Hybrid Model. *Synthese* 157, no. 1: 59–78.
- Bradley, Darren. 2011a. Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty. *British Journal for the Philosophy of Science* 62, no. 2: 323–342.

Bradley, Darren. 2011b. Self-Location Is No Problem for Conditionalization. *Synthese* 182, no. 3: 393–411.

⁸⁰ Thanks to Miriam Schoenfield, Jack Spencer, Roger White, and an anonymous referee for helpful feedback.

- Bradley, Darren. 2012. Four Problems About Self-Locating Belief. *Philosophical Review* 121, no. 2: 149-177.
- Bradley, Darren. 2015. Everettian Confirmation and Sleeping Beauty: Reply to Wilson. *British Journal for the Philosophy of Science* 66, no. 3: 683–693.
- Briggs, Rachael. 2010. Putting a Value on Beauty. In *Oxford Studies in Epistemology, Volume 3*, edited by Tamar Szabo Gendler and John Hawthorne, 3-34. Oxford University Press.
- Builes, David. 2019. Self-Locating Evidence and the Metaphysics of Time. *Philosophy and Phenomenological Research* 99, no. 2: 478–490.
- Chalmers, David. 2011. Frege's Puzzle and the Objects of Credence. Mind 120, no. 479: 587-635.
- Cozic, Mikael. 2011. Imagining and Sleeping Beauty: A Case for Double-Halfers.' *International Journal of Approximate Reasoning* 52, no. 2: 137–143.
- Draper, Kai, and Joel Pust. 2008. Diachronic Dutch Books and Sleeping Beauty. *Synthese* 164, no. 2: 281–87.
- Döring, Sabine, and Bahadir Eker. 2017. Rationality, Time and Normativity: On Hedden's Time-Slice Rationality. *Analysis* 77, no. 3: 571–585.
- Dorr, Cian. 2002. Sleeping Beauty: In Defence of Elga. Analysis 62, no. 4: 292–296.
- Elga, Adam. 2000. Self-Locating Belief and the Sleeping Beauty Problem. *Analysis* 60, no. 2: 143–147.
- Halpern, Joseph. 2006. Sleeping Beauty Reconsidered: Conditioning and Reflection inAsychronous Systems. In Oxford Studies in Epistemology Vol. 1, edited by TamarGendler and John Hawthorne, 111-142. Oxford University Press.
- Hedden, Brian. 2015a. *Reasons Without Persons: Rationality, Identity, and Time*. Oxford University Press UK.

- Hedden, Brian. 2015b. Time-Slice Rationality. Mind 124, no. 494: 449–491.
- Hedden, Brian. 2016. Mental Processes and Synchronicity. Mind 125, no. 499 (2016): 873-888.
- Hedden, Brian. 2017. Replies to Döring and Eker, Snedegar and Lenman. *Analysis* 77, no. 3: 607–618.

Hitchcock, Christopher. 2004. Beauty and the Bets. Synthese 139, no. 3: 405-420.

- Horgan, Terry. 2004. Sleeping Beauty Awakened: New Odds at the Dawn of the New Day. *Analysis* 64, no. 1: 10–21.
- Kierland, Brian, and Bradley Monton. 2005. Minimizing Inaccuracy for Self-Locating Beliefs. *Philosophy and Phenomenological Research* 70, no. 2: 384–395.
- Leitgeb, Hannes, and Richard Pettigrew. 2010. An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science* 77, no. 2: 201–35.
- Leitgeb, Hannes. 2010. Sleeping Beauty and Eternal Recurrence. Analysis 70, no. 2: 203–205.

Lenman, James. 2017. Reasons Without Humans. Analysis 77, no. 3: 586–595.

Lewis, David. 1976. Survival and Identity. In *The Identities of Persons*, edited by Amelie Oksenberg Rorty, 17–40. University of California Press.

Lewis, David. 1979. Attitudes de Dicto and de Se. Philosophical Review 88, no. 4: 513-543.

Lewis, David. 2001. Sleeping Beauty: Reply to Elga. Analysis 61, no. 3: 171-76.

Lewis, Peter J. 2006. Quantum Sleeping Beauty. Analysis 67, no. 1: 59-65.

- Meacham, Christopher J. G. 2008. Sleeping Beauty and the Dynamics of de Se Beliefs. *Philosophical Studies* 138, no. 2: 245–269.
- Meacham, Christopher J. G. 2010. Unravelling the Tangled Web: Continuity, Internalism, Non-Uniqueness and Self-Locating Beliefs. In *Oxford Studies in Epistemology, Volume 3*, edited by Tamar Szabo Gendler and John Hawthorne, 86. Oxford University Press, 2010.

Monton, Bradley. 2002. Sleeping Beauty and the Forgetful Bayesian. Analysis 62, no. 1: 47–53.

Moss, Sarah. 2015. Time-Slice Epistemology and Action Under Indeterminacy. In *Oxford Studies in Epistemology*, edited by Tamar Szabó Gendler and John Hawthorne, 172–94. Oxford University Press.

Pettigrew, Richard. 2016. Accuracy and the Laws of Credence. Oxford University Press Uk.

Pettigrew, Richard. Unpublished Manuscript. Self-Locating Belief and the Goal of Accuracy.

Podgorski, Abelard. 2016. A Reply to the Synchronist. Mind 125, no. 499: 859-871.

Ross, Jacob. 2010. Sleeping Beauty, Countable Additivity, and Rational Dilemmas.

Philosophical Review 119: 411-447.

- Saunders, Simon, and David Wallace. 2008. Branching and Uncertainty. *British Journal for the Philosophy of Science* 59, no. 3: 293–305.
- Schoenfield, Miriam. 2015. Internalism Without Luminosity. *Philosophical Issues* 25, no. 1: 252–272.
- Sebens, Charles T., and Sean M. Carroll. 2018. Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. *The British Journal for the Philosophy of Science* 69, no. 1: 25–74.

Snedegar, Justin. 2017. Time-Slice Rationality and Filling in Plans. Analysis 77, no. 3 : 595–607.

Stalnaker, Robert. 2008. Our Knowledge of the Internal World. Oxford: Oxford University Press.

- Tappenden, Paul. 2008. Saunders and Wallace on Everett and Lewis. *British Journal for the Philosophy of Science* 59, no. 3: 307–314.
- Titelbaum, Michael G. 2008. The Relevance of Self-Locating Beliefs. *Philosophical Review* 117, no. 4: 555–606.

Titelbaum, Michael G. 2012. An Embarrassment for Double-Halfers. Thought: A Journal of

Philosophy 1, no. 2: 146–151.

- Titelbaum, Michael G. 2013. Ten Reasons to Care About the Sleeping Beauty Problem. *Philosophy Compass* 8, no. 11: 1003–1017.
- Titelbaum, Michael G. 2013. *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief.* Oxford University Press.
- Weatherson, Brian. 2005. Should We Respond to Evil with Indifference? *Philosophy and Phenomenological Research* 70, no. 3: 613–635.
- Weintraub, R. 2004. Sleeping Beauty: A Simple Solution. Analysis 64, no. 1: 8–10.
- Williamson, Timothy. 2000. Knowledge and Its Limits. Oxford University Press.
- Wilson, Alastair. 2014. Everettian Confirmation and Sleeping Beauty. *British Journal for the Philosophy of Science*, no. 3: 573–598.
Chapter 5: No-Self Location

Imagine an ideal reasoner, Alice, who is enclosed in a windowless room. Alice knows everything there is to know about fundamental physics: the initial conditions of the universe, the laws of physics, the position and momentum of every particle in the universe, etc. From this information, Alice comes to know that there are exactly two agents in the universe, both of which are exact duplicates who are enclosed in rooms that are also exact duplicates. One of the rooms is located next to a lake, and the other is not. From this information, Alice comes to know that she must be one of the agents. However, even though Alice knows everything about the physical features of the universe, it still seems like she is ignorant about certain facts. For example, she doesn't know if there's a lake outside her room.

As has been noted before, this story bears a striking resemblance to Jackson's (1982) famous 'knowledge' argument regarding phenomenal consciousness.⁸¹ Mary, a brilliant color scientist (and also an ideal reasoner), is locked in a black and white room her whole life. Inside the room, she learns about all the relevant physical facts regarding color, including all of the relevant physics and all of the relevant neuroscience. In fact, we can imagine that she also knows everything there is to know about the initial conditions of the universe, the laws of physics, the position and momentum of every particle, etc. Still, it seems like she is ignorant about certain facts. For example, Mary does not know what it's like to see red.

These epistemic gaps are surprising. It is surprising that an ideal reasoner can know everything there is to know about what physics says (and what any other science says), yet still be ignorant about a certain class of facts.

In response to the knowledge argument, some physicalists have denied that Mary is ignorant about the relevant kind of facts. One radical version of this view endorses *illusionism about consciousness*.⁸² According to illusionists about consciousness, there are only facts concerning what David Chalmers (1996) has called the 'easy' problems of consciousness. There are facts

⁸¹ Stalnaker (2008) and Kwon (2017) also discuss this analogy.

⁸² For defenses of illusionism, see Frankish (2016), Dennett (2016), and Humphrey (2012).

about behavior, verbal reports, highly complicated brain functioning, and perhaps facts about what the brain represents (understood in a physicalist-friendly way). Crucially, however there are no further facts about 'phenomenal consciousness' that are unanalyzable in functional terms. Since Mary could certainly know everything there is to know about behavior, brain functioning, and (physicalist-friendly) representation, she thereby knows everything there is to know about the mind. Any lingering feeling that there are further facts for Mary to know, concerning the *redness of red*, is an illusion.

I myself am not an illusionist about consciousness. However, I am attracted to *illusionism about the self*. Just as illusionists about consciousness deny that there are any further facts for Mary to know, illusionists about the self deny that there are any further facts for Alice to know. According to illusionism about the self, *de se* facts, which are essentially indexical facts about where I am located in the world, or how I relate to various other items in the worlds, are illusory.⁸³

Illusionism about the self is, of course, a radical view, and I will not be arguing for its truth here. My first goal will be to explain how it *could* be true. After all, given that Alice *exists*, it seems like she must have some location or other. It (surely) can't be that she's in *neither* location, or in *both* locations, or somehow *indeterminately* located! However, I will be arguing that in fact *all three of these possibilities are realized by different metaphysical theories*. Consequently, a variety of different metaphysical theories vindicate illusionism about the self (sections 1-3). My second goal is to argue that illusionism about the self *makes empirical predictions*, because it entails a certain epistemology of self-locating belief that itself licenses different metaphysical debates can indirectly have empirical consequences. It thereby also puts pressure on certain popular meta-metaphysical views, according to which these abstract metaphysical debates are somehow verbal, or somehow a matter of convention, or otherwise not substantive. If different positions on these abstract metaphysical recognize that these debates are substantive.

⁸³ Kwon (2017) extends a corresponding 'ability hypothesis' that has been defended in the case of the knowledge argument to the case of de se facts. Crucially, however, Kwon's account implies that Alice *is* ignorant about whether there is a lake next to her room. He categorizes this ignorance as essentially *de re* ignorance rather than *de se* ignorance (see section 5 of Kwon 2017). However, the account that I will be developing below implies that Alice is not ignorant of any facts, whether such facts are categorized as de re or de se facts.

1. Alice is in *Neither* Location

Metaphysicians disagree about which material objects exist. We may categorize many of these different views as disagreeing about how to 'carve up' the world into objects. *Mereological Nihilists* carve out only its smallest components (such as space-time points or point particles) (e.g. Sider (2013)). *Organicists* also carve out living organisms (e.g. van Inwagen (1990) and Merricks (2001)). Other metaphysicians, call them *Mooreans*, also choose to carve out those objects that common-sense recognizes (e.g. Korman (2015)). *Mereological Universalists* also choose to carve out every possible fusion of the mereological simples that the Nihilist recognizes (e.g. van Cleve (2008)).

Some metaphysicians, however, disagree that the world should be 'carved up' at all. According to *Monists*, the only material object is the Cosmos as a whole. The Cosmos does not come 'pre-carved' into objects like tables and chairs.

The Monist position is perhaps the least familiar position, however in recent years a flurry of considerations have been mounted in its favor. Horgan and Potrč (2008) have written a booklength defense of the thesis, and Schaffer (2007) has argued that the view is strictly superior than mereological nihilism. Moreover, a number of arguments have been given that the Cosmos is at least the only *fundamental* object. This view has been motivated by considerations of physical law (Schaffer (2013)), quantum entanglement (Schaffer (2010a), Ismael and Schaffer (2016)), causal essentialism (Segal (2014)), space-time substantivalism (Perry (2017)), mereology (Schaffer (2007, 2010a)), the denial of external relations (Vallicella (2002)), failures of free modal recombination (Schaffer (2010b)), and concerns about truthmakers (e.g. Schaffer (2010c)). The view has even been appealed to in various radical solutions to the hard problem of consciousness (Goff (2017), Kastrup (2018), Chalmers (2015, 2019))!

For our purposes, we will be focusing on versions of Nihilism, Organicism, Mooreanism, Universalism, and Monism that are formulated as comprehensive theses about what there is. I will *not* be understanding these theses as merely making claims about which things are 'fundamental' (e.g. Schaffer (2009)), or which things exist in the most fundamental sense of 'exists' (e.g. Sider (2011)), or which things exist 'in reality' (e.g. Fine (2009)). Although these 'fundamentalist' views have become more popular in recent times, a number of philosophers have criticized these views

on a variety of different grounds (e.g. van Inwagen (2014) and Miller, Duncan, and Norton (forthcoming)). For example, Korman (forthcoming) and Lipman (2018) have argued that formulating these revisionary ontological theories in these more heavyweight metaphysical terms does nothing to address the problems that these theories were meant to solve in the first place. So, in what follows, I will have in mind a Nihilist (or Monist or Organicist) who thinks that there is no sense in which 'tables exist' is literally true, just as there is no sense in which 'unicorns exist' is literally true.

With these different views in mind, let us return to Alice, who does not know whether there is a lake outside her room. Mooreans, Universalists, and Organicists agree that Alice exists, so of course she must have some particular location. However, Nihilists and Monists explicitly *deny* the existence of Alice. So, for them, of course there will be no further de se facts concerning the location of *Alice*, just as there are no facts concerning the location of unicorns.

At this point, however, both Nihilism and Monism seem to face refutation due to Descartes' *cogito*. Surely I think, therefore I am! If we are not to revert to illusionism about consciousness, both Nihilism and Monism must somehow be able to recover facts about our own consciousness without appealing to what we ordinarily think of as the *subjects* or *bearers* of phenomenal properties (whether such subjects are bodies, brains, or extra non-physical souls). In response, Nihilists and Monists should admit that they must deny that 'I think', but they need not deny that 'Thought is occurring'.⁸⁴

In order to account for mental phenomenon, the Nihilist should say that phenomenal properties like *being in pain* must somehow be predicated *plurally* to one's simples-arranged-brain-wise (or simples-arranged-body-wise). There is no single bearer of the property of *being in pain*, but rather *being in pain* should be understood as an activity that many things do *together*. Analogously, if various people are surrounding an object, no *single* person 'surrounds' the object. Rather, surrounding an object is an activity that many people do together. Similar things should be said about the other mental states of Alice. Alice can be said to have beliefs, desires, and other mental

⁸⁴ This response to Descartes' *cogito* goes back to Georg Lichtenburg in the 18th century, and it is also defended by Russell (1945: 567).

states only if this is understood as various simples jointly having beliefs, desires, and other mental states.

Once this move is made, it seems that the Nihilist is able to recover the de se facts that Alice was ignorant of. Instead of wondering where the single subject of her mental states is located, Alice can instead wonder where the many simples whose activity is responsible for her mental states are located. For all the simples-arranged-Alice-wise know, those simples could be located next to a lake, or they may not be.

However, while the Nihilist can recover Alice's de se ignorance, the Monist cannot.

Start with the question of how the Monist recovers facts about consciousness. The Monist cannot say that properties like *being in pain* are attributed to bodies, or brains, or immaterial souls, nor can they say that such properties should be treated as disguised plural predications concerning many simples working together. The Monist also cannot say that the property of *being in pain* is had, in any straightforward way, by the Cosmos itself. For suppose that the world (as described by the Moorean) contains two subjects of experience, one of which is in pain and the other which is not in pain. The Monist should not say the Cosmos bears both of these properties, so that the Cosmos is both *in-pain-and-not-in-pain*.

The way that Monists typically deal with properties that we would ordinarily ascribe to things within the Cosmos (colors, temperatures, masses, etc.) is by invoking *distributional* properties. Distributional properties are ways of 'filling in' some spatially extended object with some quality. For example, *being polka-dotted* is a color distributional property and *being hot at one end and being cold at the other* is an example of a heat-distributional property.⁸⁵ In the case of our best physical theories, perhaps the most plausible candidate for the basic properties of fundamental physics are distributional field-theoretic properties, which specify the distribution and magnitude of various quantum fields across space-time.⁸⁶ The Monist could give a similar treatment to phenomenal properties. The property that specifies the distribution of quantum-fields across space-

⁸⁵ For much more on distributional properties, see Parsons (2004).

⁸⁶ For a defense of this sort of picture, see Schaffer (2009).

114

time also serves to specify (perhaps together with psycho-physical laws) a distributional phenomenal property, that specifies the distribution of phenomenal qualities across the Cosmos.

The trouble is that Alice *already* knows the distribution of phenomenal qualities across the Cosmos, since she *already* knows that there are phenomenal qualities in (what the Moorean would describe as) two windowless rooms, one of which is next to a lake. In fact, Alice already knows everything there is to know about how various properties are distributed across the Cosmos. However, according to the Monist *these are all the facts there are*. It is an illusion to think that there must be a *further* fact, concerning the location of Alice, or of the location of simples-arranged-Alice-wise within the Cosmos. Our deep-seated intuition that there must be such further facts is incompatible with the fact that *there is only the Cosmos and its various distributional properties*.

Before moving on, let us illustrate the Monist's point of view with a concrete, toy example. Consider a world that seems to contain four squares, two of which are grey and two of which are white:



Suppose you take yourself to be one of the grey squares (e.g. suppose your visual field is uniformly grey). You may know exactly how the colors are distributed among the world, but you may still be ignorant about where *your* square is located in the world. Are you the square on the top left? Or maybe the one on the bottom right?

Now, suppose you abandoned your Mooreanism and were instead convinced that none of the four squares you thought existed actually does exist. Instead, you come to think that there are four smaller squares that compose each of the original squares, and *those* are the squares that actually exist. Then, while you can no longer be curious about where your medium sized square is located

(because there are no medium-sized squares), you can still be curious about where the smallersized squares that 'compose' you are located:

Now, suppose you finally become a Monist. Then, given that you still already know how the colors are distributed among the world, *there is nothing left for you to be curious about*. All there is to the world is the distribution of colors! There *just are* no individual squares inside the world that have any locations. There is just the colored world:



From the perspective of the Monist, these extra de se questions only arise when we start carving up the world into separate objects. When we start thinking that I'm *here* and you're *there*, and we are all separate and distinct entities that are the bearers of distinct properties with our own individual locations in the world, then we commit ourselves to this extra realm of facts. However, if we instead come to think of the world as genuinely *one* – where there is only a distribution of qualities had by the *one* - then we are able to avoid this extra realm of facts that is outside the reach of Alice.

2. Alice is in Both Locations

Let us turn to a different, albeit similarly abstract, metaphysical dispute. Metaphysicians disagree about how objects relate to properties. On the standard way of thinking about things, there are two kinds of entities in the world. First, there are things that 'have' or 'instantiate' various properties, which philosophers sometimes call 'particulars'. Particulars include such things as tables, apples, and laptops. Second, there are properties that are had *by* these particulars, such as mass, charge, color, etc. This sort of theory is sometimes called the 'substratum theory'.

For various reasons, many philosophers have been wary of this sort of picture. For one, if this sort of picture is right, why couldn't there be a particular – a 'bare' particular - that had no properties whatsoever?⁸⁷ To many philosophers, however, the idea of a bare particular has seemed unintelligible. Other philosophers have argued that we can only detect the distribution of qualities in the world, and not the particulars 'behind' these various qualities (e.g. Dasgupta 2009, 2017).⁸⁸ After all, things would appear the very same if the same distribution of qualities was instantiated by numerically distinct particulars. So, if we can construct theories that don't posit extra undetectable structure in the form of particulars that 'bear' qualities, we should.

On an alternative way of thinking of things, contrary to common sense, there are only properties. What we think of 'objects' or 'particulars' are really just properties that are 'bundled' together by a particular relation, which bundle theorists sometimes call 'compresence'. Instead of saying that there is some particular *a* that is both red and round, we should instead say that redness and roundness are *compresent*. On this sort of picture, the only facts are facts that specify which qualities are compresent with which others and the relations that hold between these various compresent qualities.

According to the bundle theory, the typical motivation for thinking that there must be further de se facts is absent. Alice, at least when thought of as a particular that is the bearer of various properties, does not exist. So, *a fortiori*, there is no question of where Alice is located.

⁸⁷ As many are quick to point out (e.g. Sider (2006)), there is of course *a* sense in which bare particulars have properties. After all, bare particulars have the properties of being colorless, massless, shapeless, odorless, etc. They also have the property of *being such that* 2+2 = 4! Given an abundant conception of properties, of course bare particulars have many properties. However, if we restrict ourselves to the fundamental, perfectly natural properties that objects are supposed to instantiate, bare particulars instantiate no such properties. They simply *lack* all such properties. It is this sense in which they lack any positive, fundamental, intrinsic nature, and it is in this sense that bare particulars strike so many philosophers as deeply mysterious.

⁸⁸ Locke (1689) described these hidden particulars as an 'unknown support of those qualities', and Russell (1940) argued that 'We experience qualities, but not the subject in which they are supposed to inhere' (122).

One might try to recover a sort of de se ignorance as follows. Suppose Alice is in pain. Since properties do exist according to the bundle theorist, Alice can still wonder: where is *this* pain located? Is it next to the lake or not?

Whether Alice knows the answer to this fact turns on what metaphysics of properties the bundle theorist adopts. Some bundle theorists think that properties should be understood as *universals* and others think they should be understood as *tropes*. Universals are repeatable: they are wholly located wherever they are instantiated. Tropes, however, are not repeatable. Consider, for example, two duplicate red apples. The Universals theorist will think that the universal *redness* is wholly located in both of the apples, whereas the Trope theorist will think that the particular redness of the first apple is distinct from the particular redness of the second apple.

On the trope bundle theory, Alice is ignorant about whether her pain is next to the lake. This is because she does not know whether her particular pain trope is located next to the lake or not. After all, one such trope is next to a lake and a distinct (but indiscernible) trope is not next to a lake. On this view, de se ignorance can be rethought of as ignorance about the location of certain tropes.

However, on the universal bundle theory, Alice is not ignorant about anything. She *already* knows that 'her' pain is a universal that is wholly located in *both* places at once, one place that is next to a lake and one place which is not. De se ignorance, therefore, cannot be rethought of as ignorance about the location of any universal, since Alice already knows the locations of every universal – they are wholly located wherever they are instantiated. Unlike with tropes, the location of every universal can be fully determined just by a purely qualitative description of reality.

On the universal bundle theory, there is a sense in which Alice already knows, inside of her room, that she is located in *both* places at once! This is exactly analogous to the usual 'counterexample' that is often levied against the universal bundle theorist. Consider Black's (1952) case of a symmetric universe consisting of nothing but two duplicate iron spheres. The universal bundle theorist has a hard time recognizing that this possibility contains two *distinct* spheres, since both spheres are composed of the exact same universals that are wholly present in both locations. In response, the bundle theorist typically replies that there is a *single* bundle of universals that is multi-located (e.g. Hawthorne (1995), Curtis (2014)). The exact same response is operative in the

case of Alice: the bundle of universals that is Alice is multi-located, and Alice already knows this. Any lingering feeling that there is something that Alice is ignorant of is therefore an illusion.

3. Alice is Indeterminately Located

The bundle theory with universals is an example of a theory where there are only *qualitative* facts, as opposed to facts concerning any particular individuals. Dasgupta (2009, 2017) has developed two other metaphysical theories, Quantifier Generalism and Algebraic Generalism, according to which all of the fundamental facts are purely qualitative.

Consider a world which we would ordinarily describe as having two individuals, a and b, such that a is blue, b is green, and a and b stand in some relation R. Ordinarily, we would describe the facts in this world as follows:

Ba Gb Rab

These facts are clearly not qualitative facts, since they concern particular individuals. According to Quantifier Generalism, the fundamental facts should not make any reference to particular individuals. Instead, the fundamental facts should be *quantified* facts that only say that something is blue, some other distinct thing is green, and the blue thing stands in relation R to the green thing. More formally, the fundamental fact should be as follows:

 $\exists x \exists y (Bx \land Gy \land Rxy \land x \neq y)$

This quantified fact is entirely neutral, for example, as to whether a or b (or some other possible individual) is blue. For the Quantifier Generalist, fundamentally there is only a quantified fact concerning the number of objects in the world, together with their qualitative properties and relations to each other.

The workings of Algebraic Generalism are somewhat more complicated, and the details may be found in Dasgupta (2009). Briefly, however, Algebraic Generalism is like the bundle theory in that

it countenances a range of qualitative properties, however it is unlike the bundle theory in that it allows for more complex ways for properties to be 'stitched together'. These more complex ways of forming properties allow the theory to mimic the resources of first-order logic. For example, for any property F, there will be the property '~F' which applies to non-F's. For any properties F and G, there will be the property 'F&G' applying to things that are both F and G. For any two-place relation R, there will be a cropping operator 'c' that forms the one-place property cR, which we would normally describe as the property of *bearing R to something*. Applying the cropping operator again, ccR would be the zero-place 'property' we'd normally describe as *something bearing R to something*. This cropping operator is a surrogate for the existential quantifier in firstorder logic. For example, instead of saying '∃x∃yRxy', the Algebraic Generalists says that 'ccR obtains'.

The crucial thing to note about these theories is simply that the fundamental facts are purely qualitative. According to Dasgupta, non-qualitative facts concerning particular individuals hold *in virtue of*, or are *grounded in*, these fundamental qualitative facts.

Given this metaphysics, let us turn back to Alice. Inside of her windowless room, Alice already knows the fundamental qualitative facts that are posited by the Quantifier Generalist and the Algebraic Generalist. If Alice is wondering about her own location, this is clearly a non-qualitative fact concerning some particular individual, Alice. Prima facie, it seems like there are two possibilities that may obtain: either (i) the qualitative facts are exactly as they are and Alice is located next to the lake or (ii) the qualitative facts are exactly as they are and Alice is not located next to the lake or (ii) the qualitative facts are exactly as they are and Alice is not located next to the lake. Since every non-fundamental fact must supervene on the fundamental facts, it must be that the fundamental qualitative facts necessitate either (i) or (ii). So, which one is it? Here, I think the Qualitativist should follow Russell (2017), who has argued that, conditional on some plausible assumptions, the Qualitativist should think that non-qualitative facts like these are simply *indeterminate*. After all, it would seem wholly arbitrary if the fundamental qualitative facts necessitate (i) or (ii). Given that Russell is right, there is no *further* fact for Alice to know concerning her own location, since such facts are simply indeterminate. The Qualitativist should therefore think that such extra de se facts are illusory and endorse illusionism about the self.

4. The Epistemology of Self-Locating Belief

Unsurprisingly, illusionism about the self has important consequences for the *epistemology* of self-locating belief, which concerns itself with how rational agents ought to respond to essentially indexical information concerning where they are located in the world. Over the past couple of decades, it has been discovered that this issue is deeply interconnected with a wide variety of unresolved questions in formal epistemology. As Titelbaum (2013) has noted, the epistemology of self-locating belief is bound up with questions about relative frequencies, objective chances, Dutch Book arguments, accuracy arguments, reflection principles, and indifference principles. Debates about self-locating belief are even interconnected with many of the central questions in the metaphysics of science, including debates about the correct interpretation of quantum mechanics (Bradley (2011a), Lewis (2006), Sebens and Carroll (2018)), the existence of multiple universes (Bradley (2012)), the measure problem in cosmology (Arntzenius and Dorr (2017)), and even classic questions in the metaphysics of time (Builes (2019)).

Typically, discussions of the epistemology of self-locating belief follow Lewis (1979) in distinguishing between *de dicto* beliefs and *de se* beliefs. Intuitively, de dicto beliefs are wholly about what the world is like, whereas de se beliefs are partly about an agent's location within a world. De dicto beliefs give you a 'map' of the world, whereas de se beliefs provide the 'you are here' sign on the map. The content of de dicto beliefs may be represented using uncentered propositions, or sets of uncentered worlds. The content of de se beliefs may be represented using centered propositions, which are sets of centered worlds. Centered worlds may be modeled as ordered pairs consisting of an uncentered world *w* together with a particular time-slice of an agent within *w*. In our original thought experiment, Alice knew the truth-value of every uncentered proposition, but was still unsure of the truth-value of the centered proposition that *I am next to a lake*.

According to illusionism about the self, *there are only uncentered propositions*. There are no *extra* facts concerning the location of a particular individual within a world. In the case of Monism, the uncentered propositions are facts about the distributional properties of the Cosmos. In the case of the bundle theory with universals, the uncentered facts are facts about how various universals are

bundled together and related to one another. In the case of other qualitative theories, the uncentered facts are purely qualitative facts.

Illusionism about the self provides an entirely *metaphysical* solution to the famous 'Sleeping Beauty' problem, which has been the subject of much of the literature on the epistemology of self-locating belief:

<u>Sleeping Beauty</u>: Beauty is a perfectly rational agent who is told that the following events will occur. On Sunday, she will be put to sleep. A fair coin will then be tossed. If it lands Heads, she will be awakened on Monday morning. Later, in the evening, she will be told that it is Monday, and then she will be let go. If the coin lands Tails, as before, Beauty will be awakened on Monday morning, and then she will be told that it is Monday later that evening. However, instead of being let go, she will be given a memory-loss drug that will make her forget all of her memories of Monday, and she will be put back to sleep. She will then be awakened on Tuesday, and then she will be let go. Her wakings on Monday and Tuesday will be qualitatively indistinguishable. When she first awakens on Monday morning, what should her credence be that the coin landed Heads? When she is subsequently told that it is Monday, on Monday evening, what should her credence be that the coin landed Heads?

There are three main responses to the problem. Currently, the most popular answer to the problem seems to be the Thirder position, according to which Beauty's credence should be 1/3 in the morning and 1/2 in the evening.⁸⁹ There are two other less popular answers. According to Lewisian Halfers, Beauty's credences should be 1/2 in the morning and 2/3 in the evening (e.g. Lewis (2001) and Bradley (2011a)). According to Double Halfers, Beauty's credences should be 1/2 in the morning and evening (e.g. Bostrom (2007), Meacham (2008), and Builes (forthcoming)).

It is agreed by all parties that Sleeping Beauty does not learn any uncentered proposition throughout the experiment. She *already* knew before hand that the world was going to be in exactly one of two states: either (i) the coin will land Heads, she will be woken up, told it's Monday, and

⁸⁹ Authors who have defended the Thirder solution include Arntzenius (2003), Dorr (2002), Draper and Pust (2008), Elga (2000), Hitchcock (2004), Horgan (2004), Monton (2002), Stalnaker (2008), Titelbaum (2008), and Weintraub (2004). For a useful survey of the arguments for the Thirder position, see Ross (2010).

let go or (ii) the coin will lands Tails, she will be woken up, told it's Monday, put back to sleep, woken up again on Tuesday, and then let go. When Beauty wakes up, she does not learn which of these two uncentered possibilities is actual.⁹⁰

According to illusionism about the self, Double Halfing is the correct answer to the Sleeping Beauty Problem. This is because, according to illusionism about the self, Beauty does not learn any new information *whatsoever* throughout the experiment. So, of course, Beauty's credence of 1/2 on Sunday evening should not change throughout the experiment. Both Thirders and Lewisian Halfers believe that Beauty's credence should change when she learns certain *centered* propositions, but this sort of account is not available to the illusionist.

In general, illusionists about the self endorse the following principle discussed by Titelbaum (2008):

The Relevance Limiting Thesis: It is never rational for an agent who learns only centered propositions to revise their credences in uncentered propositions.

It is easy to see that The Relevance Limiting Thesis is rendered trivial by illusionism, since it is impossible for an agent to learn about 'centered propositions' in the first place!

The correct epistemology for illusionists about the self is simple. One's credence about any uncentered proposition should simply be one's prior probability function conditional on one's total uncentered evidence. This epistemology is explicitly defended in Meacham (2008), who defends the Double Halfer solution to Sleeping Beauty on independent grounds. For the illusionist, one needn't worry about one's credences in centered propositions, or about what happens when one's evidence includes centered propositions, since there simply *are* no centered propositions in the first place.

This sort of epistemology is very controversial. Titelbaum (2016) brings this out by considering the following case:

⁹⁰ This is assuming that Beauty is always woken up in qualitative duplicate rooms. If this requirement is not met (as in Titelbaum's (2008) 'Technicolor Beauty' case), then more will have to be said. This reliance on qualitative duplicate rooms is also needed for the standard generalization of the 'Double Halfer' position, as developed in Meacham (2008) and Briggs (2010). Such generalizations are exactly what the illusionist about the self should endorse.

<u>Mystery Bag</u>: You are one of ten people arranged in a circle in a room. A fair coin has been flipped to determine the contents of a bag: on heads the bag contains nine black balls and one white ball; on tails it's nine white and one black. The bag is passed around the room. Each person draws one ball, holds onto it, and passes the bag until it's empty. You can't see anyone else's ball, but the one you've drawn is black.

In this case, everyone should agree that one's credence that the coin came up heads should increase from 1/2 to 9/10. However, Titelbaum then asks us to imagine the following:

[N]ow let's add the wrinkle that in this story you have no qualitative way of discriminating between yourself and the other subjects in the room. We'll have to imagine that you don't know your own name (perhaps your memory has been erased, or perhaps you were raised by scientists and never given a name); you look exactly like everyone else in the room; the room is cylindrical so you can't describe yourself as, say, 'The guy in the corner'; etc. Whatever science-fiction elements are needed to make this work, consider them added to the story. (673)

Given this stipulation, you do not learn any uncentered proposition when you draw a black ball from the bag. The only thing you learn is the centered proposition that 'I drew a black ball'. You already knew that *someone* would draw a black ball, and in fact you already knew that someone with exactly your qualitative properties would draw a black ball. Because of this, the Relevance-Limiting Thesis implies that your credence that the coin came up heads should stay at 1/2. Titelbaum presents this case as a straightforward counterexample to the Relevance-Limiting Thesis. However, given illusionism about the self, one is required to retain one's credence of 1/2 that the coin came up heads. If this really is a 'counterexample' to the Relevance-Limiting thesis, then, surprisingly, it will also be a counterexample to Monism, bundle theory with universals, and other qualitativist theories in metaphysics! What Mystery Bag brings out is that we all have a strong intuition that there *are* additional, irreducible de se facts that we might learn, and it is precisely this intuition that these radical metaphysical views are denying.

5. Meta-metaphysical Consequences

I have argued that certain metaphysical theories imply illusionism about the self. For those who find these metaphysical theories to be utterly implausible, however, these claims might not be terribly interesting. Why care about what follows from obviously false metaphysical theories?

In response, I would like to suggest that the interest that these claims have far outstrips whatever consequences they have for those who hold these radical metaphysical views. In particular, they support the following disjunction: either (i) *everyone* should endorse the Relevance-Limiting Thesis (regardless of whether they endorse illusionism about the self) or (ii) a widespread version of metaphysical deflationism is untenable. I will argue for this disjunction by arguing that if (i) is false, then (ii) must be true.

Following in the footsteps of Carnap (1950), many philosophers have felt that debates about how we should 'carve up' the world into objects are, in some way or other, not substantive (e.g. Hirsch (2010), Balaguer (2018), Chalmers (2009), and Rayo (2017)). We may imagine various linguistic communities that are otherwise like English, except each of them espouses a different view about how to carve up the world. There might be a Nihilist community, an Organicist community, a Universalist community, and a Monist community. According to these philosophers, metaphysicians are wrong in thinking that one community is 'right' and the rest are 'wrong'. Perhaps there is no fact of the matter about who is right, or perhaps they just each mean different thing by 'exist'.

While the dispute over the bundle theory and other qualitative theories has gotten less attention from metaphysical deflationists, the same anti-metaphysical spirit that seeks to deflate debates in mereology seems to also recommend deflating these debates. We could similarly imagine a linguistic community of bundle theorists and a linguistic community of substratum theorists, and it easy to be skeptical about whether one community is right and the other is wrong.

If what I have argued for is correct, however, these different metaphysical theories make different *empirical predictions*. Consider the case of Mystery Bag. Conditional on Monism, we should expect the coin to land heads with probability 1/2, and conditional on Organicism, we should expect the coin to land heads with probability 9/10!

Moreover, the empirical consequences of these theories far outstrip the highly idealized cases of Sleeping Beauty and Mystery Bag. They also make important predictions in our most fundamental physical theories.

Consider, for example, the case of large-scale cosmology. Many have wondered how we can get evidence for one cosmological theory over another when both imply that (more or less) every observation (or series of observations) that *can* happen actually *does* happen somewhere in the universe. This is related to the so-called 'measure problem' of cosmology (e.g. Freigovel (2011), Salem (2012), Vilenkin (2012)). For example, perhaps for any possible experience, there is a 'Boltzmann Brain' (Albrecht and Sorbo 2004) – a short-lived, isolated observer that spontaneously fluctuates into existence out of thermal equilibrium – that has that experience. According to the Relevance-Limiting Thesis, no amount of evidence could favor one such theory over another. All our uncentered evidence tells us is that somewhere-or-other there is an observer who has the evidence we have, which is entailed by both theories. However, those who reject the Relevance-Limiting Thesis are free to say that if one hypothesis makes our observations more 'typical' (among the class of all observers) than the other hypothesis, then that hypothesis should be favored (e.g. Dorr and Arntzenius (2017)). In this way, theories that are committed to the Relevance-Limiting Thesis (e.g. illusionists about the self) make different probabilistic predictions about the large-scale cosmological structure of the universe. In general, those who accept the Relevance Limiting Thesis will think that our evidence does not distinguish among possibilities according to which there is some-observer-or-other that has our evidence, but those who deny the Relevance Limiting Thesis *will* make probabilistic distinctions among such possibilities.

The epistemology of self-locating belief is also crucially important to the empirical adequacy of the Many Worlds Interpretation of quantum mechanics.⁹¹ Although the Many Worlds interpretation (MWI) is a fully deterministic theory, it must be able to make sense of the probabilities associated with quantum mechanics to be empirically adequate. Suppose than an experiment is about to be performed that has a 90% chance of having outcome O_1 and a 10% chance of having outcome O_2 . According to MWI, once the experiment is performed, there are two resulting 'branches' of the universe, one in which outcome O_1 occurred and one in which

⁹¹ For an overview and defense of the Many Worlds interpretation, see Wallace (2012).

outcome O_2 occurred. One way of making sense of these probabilities in a deterministic MWI universe appeals to self-locating ignorance. Once the experiment is performed, there are two copies of the experimenter, one on the branch where O_1 occurred and one on the branch where O_2 occurred, and one way one might attempt to recover these probabilities is by saying that the experimenter rationally ought to assign a 0.9 credence to their being on the O_1 branch and a 0.1 credence to their being on the O_2 branch (before the outcome is observed). Recently, Vaidman (2011) and Carroll and Sebens (2018) have defended this rational requirement, however their justification explicitly denies the Relevance-Limiting Thesis, since it commits them to the Thirder position in Sleeping Beauty. If the probabilities cannot be justified given the Relevance-Limiting Thesis, then this is another way in which illusionism about the self has scientific consequences.

The fact that Monism, the bundle theory, and other qualitative theories make different empirical predictions about coins, large-scale cosmology, and quantum mechanics (among other things) casts doubt on metaphysical deflationism. One might think that disputes about the nature of objects and properties are merely verbal, but our different linguistic communities will also disagree about the status of flipped coins, large-scale cosmology, and quantum mechanics. And surely these debates are not verbal! If what we should think about these paradigmatic non-verbal questions *crucially turns on whether certain metaphysical theories are true*, then it seems to me that even hard-core logical positivists should think that these abstract metaphysical debates are substantive. After all, in principle we are able to get *empirical evidence* for these different metaphysical theories. For example, discovering that the correct cosmological theory is one that is made more likely conditional on Monism constitutes empirical evidence for the truth of Monism.

6. No-Self

Just as many physicalists have questioned whether Mary is ignorant about what it's like to see red in her black and white room, I have also tried to question whether Alice is ignorant about whether she is next to a lake. I have done this by arguing that certain metaphysical theories, which some philosophers have found plausible for independent reasons, have the consequence that Alice is not ignorant about any such fact. Moreover, I have argued that the fact that these metaphysical theories imply a sort of illusionism about the self should be of wide interest, not just to the proponents of those metaphysical theories. I have done this by arguing that either *everyone* should endorse the (unpopular) Relevance-Limiting Thesis or otherwise a popular sort of metaphysical deflationism is untenable.

In closing, I would like to briefly address the question of whether illusionism about the self is at all plausible, independently of any of the metaphysical theories I have been considering.

I suspect that for many philosophers, conscious introspection clearly reveals that there is an 'I' that is the subject of our conscious experiences. For these philosophers, Descartes was absolutely right in insisting that he can know with certainty that '*I* think', not merely that 'thought is occurring'. If there is a (determinate) 'I' that is the subject of our experiences and has some location or other inside the universe, then illusionism about the self is untenable. If illusionism about the self is to be tenable, it must deny this core intuition we have about our conscious experiences.

However, there is a long history in philosophy, in both eastern and western traditions, of denying exactly this intuition, *precisely on the basis of conscious introspection*. As Hume famously remarked in *A Treatise of Human Nature*:

There are some philosophers who imagine we are every moment intimately conscious of what we call our *self*; that we feel its existence and its continuance in existence; and are certain, beyond the evidence of a demonstration, both of its perfect identity and simplicity. . . For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, or heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception and never can observe anything but the perception . . .[minds] are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement.

Hume's bundle theory of the self is exactly what we should expect given the bundle theory in section 3. More recently, Flanagan (1992), echoing James (1892: 83), also writes:

The illusion is that there are two things: on one side, a self, an ego, an 'I', that organizes experience, originates action, and accounts for our unchanging identity as persons and, on the other side, the stream of experience. If this view is misleading, what is the better view? The better vies is that what there is, and all there is, is the stream of experience.

'Preposterous! What then does the thinking?' comes the response. The answer is that 'the thoughts themselves are the thinkers'...We are egoless. (178)

Many other western philosophers have also expressed skepticism about the idea of a self, including Metzinger (2010), Hood (2012), and Dennett (1992).

Turning to Eastern thought, Buddhism teaches that one is able to see through the illusion of the self by following various Buddhist practices. According to Alabahari (2014):

A central tenet of *nibbana*, as described in early Buddhist tradition, is that it yields a complete understanding of the truth of no-self. Sometimes called 'awakening', the realization is said to permanently free the aspirant from the affective, behavioral, and motivational drives that stem from having a sense of self... There is no identification with elements of the mind and body, which would give rise to such thoughts as 'this is *me*, this is *mine*, this is *my* action'. The process of understanding the reality of selflessness is thus said, in the Buddhist tradition, to be not merely intellectual, but deeply transformative. (2-3)

Perhaps the most radical vision of the self is supplied by a school of Indian philosophy known as Advaita Vedānta. By maintaining that there is a single entity, 'Brahman', that is the sole ground of all of reality, this school of thought fully embraces the Monistic picture of section 2. Rather than being eliminativists about the self, however, Advaita Vedanta teaches that our 'true' self 'Ātman' is actually *identical to* Brahman. On this picture, we are not entities that have any location 'within' the universe, but rather we literally *are* the (ground of the) universe itself! The illusion that we are all separate and distinct subjects is known as *Māyā* in the Advaita Vedānta tradition.

Albahari (2019b) has recently defended a 'mystical' solution to the mind-body problem along the lines of the Advaita Vedānta tradition. In describing the core tenets of Advaita Vedānta, Albahari (2019a) writes:

All that appears as concrete reality – the spatio-temporal universe with its furniture of tables, pebbles, stars, atoms, cats, human beings – is fundamentally grounded in 'universal consciousness', which is beyond the subject/object division, beyond the bounds of space and time, completely self-subsistent . . .What inhibits full direct realization of our conscious

nature as identical to universal consciousness as ground of all being is the sense of being a separate self, or 'I'. The self is a cognitive illusion that makes the nature of consciousness appear intrinsically confined to the focal perspective of a personal subject. It is possible, via meditation techniques, to eliminate the illusion of self, such that one fully, directly and non-discursively realizes one's fundamental grounding as universal consciousness. (120)

Of course, these tenets are all extremely controversial. What I hope to point out is simply that illusionism about the self need not be seen as directly contradicting some datum of our conscious experience. In fact, there are schools of thought that directly argue for a sort of illusionism about the self precisely on the basis of our conscious experience.

Already in philosophy, many feel comfortable denying that there are further facts about the *diachronic* self. Following in the footsteps of Parfit (1984, 2001), it seems to many that once one knows all of the relevant physical and psychological facts, there is no *further* question about whether two selves at different times are really one and the same self. I have tried to develop a similar picture for the *synchronic* self. Once Alice knows everything there is to know about the initial conditions of the universe, the laws of physics, and the distribution of quantum fields across the Cosmos, there is no *further* question to be asked about where any 'self' is at any particular time. Just as it is an illusion that there must be further facts concerning the diachronic self.

References

Albahari, Miri. 2014. Insight Knowledge of No Self in Buddhism: An Epistemic Analysis. *Philosophers' Imprint* 14, no. 21: 1–30.

Albahari, Miri. 2019a. Beyond cosmopsychism and the great I am: How the world might be grounded in universal 'advaitic' consciousness. In *The Routledge Handbook of Panpsychism*, edited by William Seager. Routledge.

Albahari, Miri. 2019b. Perennial Idealism: A Mystical Solution to the Mind-Body Problem.

Philosophers' Imprint 19, no. 40: 1–37.

- Albrecht, Andreas, and Sorbo, L. 2004. Can the Universe Afford Inflation? *Physical Review D* 70: 63528
- Arntzenius, Frank. 2003. Some Problems for Conditionalization and Reflection. Journal of Philosophy 100, no. 7: 356–370.
- Arntzenius, Frank, and Cian Dorr. 2017. Self-Locating Priors and Cosmological Measures. In The Philosophy of Cosmology, edited by Khalil Chamcham, John Barrow, Simon Saunders, and Joe Silk, 396–428. Cambridge: Cambridge University Press.
- Balaguer, Mark. 2018. Why the Debate About Composition Is Factually Empty. *Synthese* 195, no. 9: 3975–4008.

Black, Max. 1952. The Identity of Indiscernibles. Mind 61, no. 242: 153–164.

- Bostrom, Nick. 2007. Sleeping Beauty and Self-Location: A Hybrid Model. Synthese 157, no. 1: 59–78.
- Bradley, Darren. 2011a. Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty. *British Journal for the Philosophy of Science* 62, no. 2: 323–342.
- Bradley, Darren. 2011b. Self-Location Is No Problem for Conditionalization. *Synthese* 182, no. 3: 393–411.
- Bradley, Darren. 2012. Four Problems About Self-Locating Belief. *Philosophical Review* 121, no. 2: 149-177.
- Briggs, Ray. 2010. Putting a Value on Beauty. In *Oxford Studies in Epistemology*, *Volume 3*, edited by Tamar Szabo Gendler and John Hawthorne, 3–34. Oxford University Press.

Builes, David. 2019. Self-Locating Evidence and the Metaphysics of Time. Philosophy and

Phenomenological Research 99, no. 2: 478–490.

- Builes, David. Forthcoming. Time-Slice Rationality and Self-Locating Belief. *Philosophical Studies*.
- Carnap, Rudolf. 1950. Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie* 4, no. 11: 20–40.
- Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, David. 2009. Ontological Anti-Realism. In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David John Chalmers, David Manley, and Ryan Wasserman. Oxford University Press.
- Chalmers, David. 2015. Panpsychism and Panprotopsychism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, edited by T. Alter and Y. Nagasawa, 246-76.
 Oxford University Press.
- Chalmers, David. 2019. Idealism and the Mind-Body Problem. In *The Routledge Handbook of Panpsychism*, edited by William Seager. Routledge.
- Curtis, Benjamin L. 2014. The Rumble in the Bundle. Noûs 48, no. 2: 298–313.
- Dasgupta, Shamik. 2009. Individuals: An Essay in Revisionary Metaphysics. *Philosophical Studies* 145, no. 1: 35–67.
- Dasgupta, Shamik. 2017. Can We Do Without Fundamental Individuals? Yes. In *Current Controversies in Metaphysics*, edited by Elizabeth Barnes. Routledge.
- Dennett, Daniel. 2016. Illusionism as the Obvious Default Theory of Consciousness. *Journal of Consciousness Studies* 23, no. 11–12: 65–72.

Dorr, Cian. 2002. Sleeping Beauty: In Defence of Elga. Analysis 62, no. 4: 292–296.

- Draper, Kai, and Joel Pust. 2008. Diachronic Dutch Books and Sleeping Beauty. *Synthese* 164, no. 2: 281–87.
- Elga, Adam. 2000. Self-Locating Belief and the Sleeping Beauty Problem. *Analysis* 60, no. 2: 143–147.
- Fine, Kit. 2009. The Question of Ontology. In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David John Chalmers, David Manley, and Ryan Wasserman, 157– 177. Oxford University Press.

Flanagan, Owen J. 1992. Consciousness Reconsidered. MIT Press.

- Frankish, Keith. 2016. Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies* 23, no. 11–12: 11–39.
- Freivogel, Ben. 2011. Making Predictions in the Multiverse. *Classical and Quantum Gravity* 28: 204007.
- Goff, Philip. 2017. Consciousness and Fundamental Reality. New York, USA: Oxford University Press.
- Hawthorne, John. 1995. The Bundle Theory of Substance and the Identity of Indiscernibles. *Analysis* 55, no. 3: 191–196.
- Hirsch, Eli. 2010. *Quantifier Variance and Realism: Essays in Metaontology*. Oxford University Press.
- Hitchcock, Christopher. 2004. Beauty and the Bets. Synthese 139, no. 3: 405–420.

Hood, Bruce. 2013. The Self Illusion: How the Social Brain Creates Identity. Oxford University

Press.

- Horgan, Terry. 2004. Sleeping Beauty Awakened: New Odds at the Dawn of the New Day. Analysis 64, no. 1: 10–21.
- Horgan, Terence, and Matjaž Potrč. 2008. Austere Realism: Contextual Semantics Meets Minimal Ontology. MIT Press.

Humphrey, Nicholas. 2012. Soul Dust: The Magic of Consciousness. Princeton University Press.

Ismael, Jenann, and Jonathan Schaffer. 2016. Quantum Holism: Nonseparability as

Common Ground. Synthese. https://doi.org/10.1007/s11229-016-1201-2

Jackson, Frank. 1982. Epiphenomenal Qualia. Philosophical Quarterly 32: 127–136.

- James, Williams. 1892. The Stream of Consciousness. In *Psychology*, edited by William James, 71–82. MIT Press.
- Kastrup, Bernardo. 2018. The Universe in Consciousness. *Journal of Consciousness Studies* 25, no. 5–6: 125–155.

Korman, Daniel Z. 2015. Objects: Nothing Out of the Ordinary. Oxford University Press UK.

Korman, Daniel Z. Forthcoming. Why Care About What There Is? In *The Question of Ontology: The Contemporary Debate*, edited by Javier Cumpa. Oxford University Press.

Kwon, Hongwoo. 2017. Mary and the Two Gods: Trying Out an Ability Hypothesis.

Philosophical Review 126, no. 2: 191–217.

Lewis, David. 1979. Attitudes De Dicto and De Se. Philosophical Review 88, no. 4: 513–543.

- Lewis, David. 2001. Sleeping Beauty: Reply to Elga. Analysis 61, no. 3: 171-76.
- Lewis, Peter J. 2006. Quantum Sleeping Beauty. Analysis 67, no. 1: 59-65.

Lipman, Martin A. 2018. Against Fundamentality-Based Metaphysics. Noûs 52, no. 3: 587-610.

Locke, John. 1689. An Essay Concerning Human Understanding. Oxford University Press.

Meacham, Christopher J. G. 2008. Sleeping Beauty and the Dynamics of de Se Beliefs.

Philosophical Studies 138, no. 2: 245–269.

Merricks, Trenton. 2001. Objects and Persons. New York: Oxford University Press.

Metzinger, Thomas. 2010. The Ego Tunnel. Basic Books.

- Miller, Kristie, Michael J. Duncan, and James Norton. Forthcoming. Ontology Without Hierarchy. In *The Question of Ontology: The Contemporary Debate*, edited by Javier Cumpa. Oxford University Press.
- Monton, Bradley. 2002. Sleeping Beauty and the Forgetful Bayesian. Analysis 62, no. 1: 47–53.

Parfit, Derek. 1984. Reasons and Persons. Oxford University Press.

- Parfit, Derek. 2001. The Unimportance of Identity. In *Identity*, edited by Henry Harris. Oxford University Press.
- Parsons, Josh. 2004. Distributional Properties. In *Lewisian Themes: The Philosophy of David K. Lewis*, edited by Frank Jackson and Graham Priest. Clarendon Press.
- Perry, Zee R. 2017. How to Be a Substantivalist Without Getting Shifty About It. *Philosophical Issues* 27, no. 1: 223–249.
- Rayo, Agustín. 2017. The World Is the Totality of Facts, Not of Things. *Philosophical Issues* 27, no. 1: 250–278.
- Ross, Jacob. 2010. Sleeping Beauty, Countable Additivity, and Rational Dilemmas. *Philosophical Review* 119: 411-447.
- Russell, Bertrand. 1940. An Inquiry Into Meaning and Truth. Routledge.
- Russell, Bertrand, 1945. A History of Western Philosophy, New York: Simon and Schuster.

- Russell, Jeffrey Sanford. 2017. Qualitative Grounds. *Philosophical Perspectives* 30, no. 1: 309–348.
- Salem, Michael. 2012. Bubble Collisions and Measures of the Multiverse. *Journal of Cosmology and Astroparticle Physics* 1201: 021.
- Sartre, Jean-Paul. 1957. *The Transcendence of the Ego*, translated by Forrest Williams and Robert. Kirkpatrick. New York : The Noonday Press.
- Schaffer, Jonathan. 2007. From Nihilism to Monism. *Australasian Journal of Philosophy* 85, no. 2: 175–191.
- Schaffer, Jonathan. 2009. On What Grounds What. In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Manley, David J. Chalmers, and Ryan Wasserman, 347–383. Oxford University Press.
- Schaffer, Jonathan. 2010a. Monism: The Priority of the Whole. *Philosophical Review* 119, no. 1: 31–76.
- Schaffer, Jonathan. 2010b. The Internal Relatedness of All Things. *Mind* 119, no. 474: 341–376.
- Schaffer, Jonathan. 2010c. The Least Discerning and Most Promiscuous Truthmaker. *Philosophical Quarterly* 60, no. 239: 307–324.
- Schaffer, Jonathan. 2013. The Action of the Whole. *Aristotelian Society Supplementary Volume* 87, no. 1: 67–87.
- Sebens, Charles T., and Sean M. Carroll. 2018. Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. The British Journal for the Philosophy of

Science 69, no. 1: 25–74.

- Segal, Aaron. 2014. Causal Essentialism and Mereological Monism. *Philosophical Studies* 169, no. 2: 227–255.
- Sider, Theodore. 2006. Bare Particulars. Philosophical Perspectives 20, no. 1: 387–397.
- Sider, Theodore. 2013. Against Parthood. Oxford Studies in Metaphysics 8: 237–293.
- Sider, Theodore. 2011. Writing the Book of the World. Oxford University Press.
- Stalnaker, Robert C. 2008. Our Knowledge of the Internal World. Oxford University Press.
- Titelbaum, Michael. 2008. The Relevance of Self-Locating Beliefs. *Philosophical Review* 117, no. 4: 555–606.
- Titelbaum, Michael G. 2013. Ten Reasons to Care About the Sleeping Beauty Problem. Philosophy Compass 8, no. 11: 1003–1017.
- Titelbaum, Michael G. 2016. Self-Locating Credences. In *The Oxford Handbook of Probability and Philosophy*, edited by Alan Hajek Christopher Hitchcock. Oxford University Press.
- Cleve, James Van. 2008. The Moon and Sixpence : A Defense of Mereological Universalism. In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean W. Zimmerman. Blackwell.
- Vaidman, L. 2011. Probability in the Many-Worlds Interpretation of Quantum Mechanics. In *Probability in Physics*, edited by Yemima Ben-Menahem and Meir Hemmo, 299-311. Berlin: Springer.
- Vallicella, William F. 2002. Relations, Monism, and the Vindication of Bradley's Regress. *Dialectica* 56, no. 1: 3–35.

Van Inwagen, Peter. 1990. Material Beings. Cornell University Press.

Vilenkin, Alexander. 2012. Global Structure of the Multiverse and the Measure Problem. *AIP Conference Proceedings*, 1514: 7–13.

Wallace, David. 2012. The Emergent Multiverse: Quantum Theory According to the Everett Interpretation. Oxford University Press, 2012.

Weintraub, R. 2004. Sleeping Beauty: A Simple Solution. Analysis 64, no. 1: 8-10.

Chapter 6: The World Just Is The Way It Is

We ordinarily think of the world as composed of a number of *individuals* or *particulars*. These individuals have various properties, and they are related to each other in a number of ways. It's not hard to see why this view should come so naturally to us: this sharp division between individuals and their properties is directly built into the language we use to think about the world. On the one hand there are *subjects*, and on the other hand there are *predicates*.

Philosophers who believe in the existence of individuals, *individualists*, disagree over many things. *Nominalists* think that properties should not be reified: while we can speak of individuals being various ways, strictly speaking there are no properties; there are only individuals.⁹² Among those who think there are such things as properties, some think that properties are *universals*, and some think that properties are *tropes*. Universals are repeatable: two individuals can have the *very same* universal. Tropes, on the other hand, are not supposed to be repeatable. Consider two duplicate red apples. The Universal theorist will think that the very same universal of *redness* is present in both apples, while the Trope theorist will think that the particular redness of the first apple is distinct from the particular redness of the second apple.

Despite these significant variations, however, there are compelling philosophical reasons for rejecting any theory which countenances individuals. At the same time, however, there are compelling reasons for rejecting the most popular alternative to the standard individualist picture, the so-called 'bundle' theory. In the first half of this paper, I present the case against both theories, with the goal of clearly distinguishing and making explicit the theoretical challenges that both types of theories face (sections 1 and 2). In the second half of this paper, I turn to developing my own positive theory of objects and properties, whose ultimate aim is to collapse the object-property distinction altogether. The main *argument* for the theory will be that it allows us to overcome all

⁹² 'Nominalism' is used in slightly different ways by different philosophers, so to a certain extent my use of the word is stipulative. 'Class Nominalists', for example, believe that properties are classes of their corresponding extensions, and sometimes trope theorists get categorized as 'Trope Nominalists'. My use of the word 'Nominalism' corresponds to what some philosophers call 'Austere Nominalism'. Since I will be giving very general objections towards individualist theories in section I, these more fine-grained distinctions will not be relevant for our purposes.

of the challenges facing standard individualists and bundle theorists discussed in the first half of the paper (sections 3 and 4).

Before beginning, I should say upfront that the view I will end up defending deeply disagrees with common sense. In particular, it will be committed to a particularly radical version of *Monism*, according to which the whole of reality is (in some yet to be precisified sense) 'one'. For some philosophers, this radical departure from commonsense will be sufficient reason to reject the theory, irrespective of whatever theoretical virtues it may have to offer. I count myself among those philosophers who find it puzzling why common sense should have any bearing on questions about the fundamental nature of reality. While issues in the methodology of metaphysics are certainly deep and important, for the remainder of this paper I will be setting the issue of Mooreanism to one side.

I. Individuals

There are at least four different sorts of considerations that have persuaded many philosophers to be skeptical of individuals. The first two are arguments found in Paul (2013, 2017) and Dasgupta (2009, 2017), and the third is an argument that has been widely influential throughout the history of philosophy. Each of these three arguments has considerable force. However, I also provide a fourth argument which targets the *modal relationship* between individuals and their properties. Going forward, a central desideratum of anti-individualist theories will be to comply with these four considerations.

I.I. Paul's Argument: A One-Category Ontology

L.A. Paul's primary motivation for developing her own distinctive kind of *mereological bundle theory*, which dispenses with the thought that there are particulars that 'have' properties, is to have a *one-category ontology*. Metaphysicians have distinguished many different sorts of ontological categories: particulars, events, states of affairs, universals, tropes, etc. According to a one-category ontology, however, the world is only built from one fundamental ontological category. The appeal of such an ontology is clear. As Paul (2017) writes, 'One category ontologies are deeply appealing, because their ontological simplicity gives them an unmatched elegance and sparseness' (32).

While this sort of appeal to theoretical elegance and simplicity certainly has its attractions, an important limitation of this argument is that it does not rule out *every* theory of individuals. It rules out theories according to which there are individuals *and* properties (whether they are conceived of universals or tropes), but it fails to rule out theories that only countenance individuals, such as Nominalism (which Paul herself rejects on independent grounds).

I.II. Dasgupta's Argument: Individuals as Danglers

Dasgupta argues against individuals on the grounds that they are explanatory 'danglers', which is to say that they are *empirically undetectable* and *physically redundant*.

As an example, Dasgupta considers the case of absolute velocity. If material objects had an absolute velocity, independent of any reference frame, then these properties would be undetectable to us. We are only able to detect the *relative* velocities of material objects. Moreover, these facts would be physically redundant, in the sense that any two closed physical systems that *only* differed with respect to their absolute velocity, would behave in indistinguishable ways. The physical dynamics of a system are completely insensitive to the absolute velocity of material objects. It is for these reasons that the orthodox view in the philosophy of physics is that there is no such thing as absolute velocity.

Regarding individuals, first consider undetectability. Dasgupta (2017) writes:

[T]he idea is that a primitive individual is 'hidden' behind its qualities. We can detect those qualities and come to know that *something or other* has them, but according to individualism there is a *further fact of the matter* as to which individual it is, and I claim that this further fact is epistemically inaccessible. After all, if two situations were qualitative duplicates – that is, if they contained the same number of primitive individuals propertied and related in *exactly* the same way, so that they differed *only* in which primitive individuals lie behind the qualities – you would never tell them apart. (10)

This idea that individuals are 'hidden' behind the qualities that they instantiate has had a long history in philosophy. Locke (1690/1975) described these hidden individuals as an 'unknown support of those qualities', and Russell (1940) argued that 'We experience qualities, but not the subject in which they are supposed to inhere' (122).

Second, consider physical redundancy. Given any closed physical system, its evolution across time will only depend on purely qualitative facts about it - facts concerning the distribution of qualitative properties and relations in the system. Any closed physical system that only differed with respect to *which* individuals were instantiating these qualities and relations would behave in the very same way. Physics is entirely indifferent with respect to the identities of the primitive individuals present in any closed system.

Given that we have strong reason to eliminate undetectable and physically redundant structure from our best theories, we have strong reason to eliminate individuals.

I.III. The Possibility of Bare Particulars

One of the most historically influential objections against individualism is that it naturally leads to the possibility of *bare particulars*.⁹³ If individuals occupy their own distinct ontological category independent from properties, what is there to stop the existence of an individual that instantiates no property whatsoever?⁹⁴ To many philosophers, bare particulars seem obviously unintelligible. Everything has to have some positive, intrinsic nature. To be *just is* to be someway or other!

As many are quick to point out (e.g. Sider (2006)), there is of course *a* sense in which bare particulars have properties. After all, bare particulars have the properties of being colorless, massless, shapeless, odorless, etc. They also have the property of *being such that* 2+2=4! These sorts of properties, however, are not the sort of properties that philosophers have in mind when they object that bare particulars 'have no properties'. Given an abundant conception of properties, of course bare particulars have many properties. However, if we restrict ourselves to the fundamental, perfectly natural properties that objects are supposed to instantiate, bare particulars instantiate no such properties. They simply *lack* all such properties. It is this sense in which they lack any positive, fundamental, intrinsic nature, and it is in this sense that bare particulars strike so many philosophers as deeply mysterious.

⁹³ Philosophers who have worried about bare particulars in some form or other include Plato (*Timaeus* 48c-53c), Aristotle (*Metaphysics* 1029a20-33), Locke (1690/1975), and Russell (1940). For more recent criticisms, see, for example, Mertz (2001, 2003), Bailey (2012), and Gilberman (2012).

⁹⁴ As Sider (2006) points out, some philosophers use the term 'bare particular' to refer to the underlying individual or 'substratum' that instantiates properties, even when the individual *does* instantiate properties. My focus will be on what Sider calls 'truly' bare particulars, which instantiate no fundamental, intrinsic properties.

It is important to note that Nominalists, who only endorse the existence of individuals, face the same problem. While Nominalists don't admit properties into their ontology, in order to describe the world adequately they must pay the ideological cost of having a stock of predicates that are meant to describe the ways that individuals are. Even given Nominalism, it still seems like nothing is stopping the possibility of an individual x such that it is not the case that 'x is F' for every perfectly natural predicate F.

This objection to individualist theories is *not* the objection that all individualist theories *entail* that bare particulars are possible. The individualist can always simply supplement their theory with the claim that bare particulars are impossible. However, this addition would simply be an unexplained constraint on modal space.⁹⁵ What one would like from a theory is an *explanation* as to why bare particulars are impossible. Ideally, we should be able to *derive* the impossibility of bare particulars from the correct theory of objects and properties, rather than simply barring their existence by fiat. Fortunately, as we will see, there are anti-individualist theories that do exactly this.

I.IV. The Modal Relationship between Individuals and their Properties

Given a metaphysics of individuals and properties, there are two natural modal questions one could ask about how they relate to each other. First, given some distribution of qualitative properties, could the *very same* qualitative distribution of properties have been instantiated by *different* individuals? Second, given some individual bearing some properties, could the *very same* individual have had *different* properties? If so, which ones?

The first question concerns *haecceitism*, the doctrine that the world could differ non-qualitatively without differing qualitatively. Suppose the world only contained two individuals, *a* and *b*, such that *a* is F and *b* is G. Could it have been that some *other* individuals, other than *a* and *b*, instantiate F and G (respectively)? For example, could it have been that *a* is G and *b* is F? On an individualist theory, the natural answer is *yes*. After all, given a metaphysics of individuals, it seems perfectly *conceivable* that some other individuals instantiate F and G.⁹⁶ Claiming that it is *impossible* for any other individuals to instantiate F and G in this way, without any further justification, seems to

⁹⁵ Moreover, this constraint would violate a natural Humean-inspired doctrine that entities of one fundamental kind (e.g. individuals) may exist independently of entities of other fundamental kinds (e.g. properties). See Gilberman (2015: 315).

⁹⁶ For a survey of conceivability arguments for haecceitism, see Cowling (2017).

amount to a *brute necessity*, which many philosophers wish to avoid.⁹⁷ If we want to avoid these merely haecceitisitcally different possibilities, then we should look for an *explanation* as to why they can't obtain. Ideally, we should be able to *derive* the doctrine of anti-haecceitism (i.e. the doctrine that it's impossible for the world to differ without differing qualitatively) from our underlying metaphysics.⁹⁸

Why should one want to endorse anti-haecceitism in the first place? Different philosophers have different motivations. To some, the strangeness of certain haeccaeitistic possibilities is reason enough. Could there really be a world qualitatively indiscernible from the actual world except that the individual that is my iPhone switches places with the individual that is my water bottle? Such feelings of strangeness are dialectically ineffective though. Here are two different kinds of arguments for anti-haecceitism.

Many philosophers of physics have wanted to endorse anti-haecceitism as a way of securing that the laws of physics are *deterministic*. The famous *hole argument* in the context of general relativity purports to show that the theory of general relativity fails to be deterministic because of certain haeccaeitistically different possibilities that are left open by the laws of general relativity.⁹⁹ Hawthorne (2006) also gives a priori arguments, independent of any particular physical theory, against determinism, which only work given haecceitism. For those philosophers who wish to salvage at least the empirical *viability* of determinism, there is strong reason to endorse anti-haecceitism.

Haecceitism also commits one to a realm of inexplicable, brute facts that it would be better to avoid. Again, consider the world where a is F and b is G. Why is it that a is the individual that is F and b is the individual that is G, rather than vice versa? Why did F 'attach' to a rather than to b? For the haecceitist, who recognizes that these are both genuine possibilities, there will simply be no explanation. It is simply brute. We can also ask other somewhat embarrassing questions. Just *how many* distinct possibilities are there with exactly two individuals, one of which is F and one

⁹⁷ For some philosophers who oppose brute necessities, see Dorr (2004, 2008), Cameron (2008), Chalmers (2009), and Goswick (2018). See Van Cleve (2018) for an overview.

⁹⁸ Dasgupta (2015) has also argued that we should be able to derive anti-haecceitism from our underlying non-modal metaphysics, rather than simply putting it into our theory 'by hand'.

⁹⁹ See Pooley (2013, Section 7) and Norton (2015) for recent overviews of the hole argument. Perhaps the most popular way to secure determinism while holding on to haecceitism appeals to facts involving the essences of space-time points. See Teitel (forthcoming) for arguments against these essentialist approaches.

of which is G? Finitely many? Infinitely many? If there are infinitely many, exactly what cardinality of infinity? Once one opens the door to haecceitism, many such seemingly unanswerable questions raise their head. For example, for which two individuals in the actual world is it possible for everything to be exactly as it is except those two individuals swap places? Could it be that everything is as it is except I swap places with an electron? My laptop with my phone? One blade of grass with another? An infinity of such awkward questions need to be answered given haecceitism.

While the issue of haecceitism is concerned with whether qualitative possibilities can differ with respect to its individuals, one can also ask how an individual could differ with respect to its qualities. This is the general issue of de re modality. Consider Socrates. Could he have had different parents? Could he have been a female? A non-biological robot? A poached egg? A nonconcrete entity (e.g. Williamson (2003))? Exactly how tall could Socrates have grown? 10 feet? 10¹⁰ feet? Exactly what is the cut-off? Given a metaphysics of individuals which instantiate various properties, these seem to be perfectly sensible questions. After all, there doesn't seem to be any reason why Socrates had to have all the properties he in fact has. So it seems like there must be a realm of facts concerning which properties he could have had and which properties he couldn't have had. It is easy to be skeptical of such facts, however. How could we possibly come to have knowledge of these facts? Won't such facts inevitably be vague, or involve entirely arbitrary cutoffs (such as Socrates' height)? Moreover, taking de re modal properties seriously results in the infamous puzzles of constitution. Famously, if a lump of clay was molded into a statue, it seems like the lump could survive being squashed, but the statue could not. Should we be error theorists about our conflicting intuitions? Should we believe there are multiple co-located objects? If so, exactly how many co-located objects are there coincident with the statue?¹⁰⁰

Perhaps the most prominent way that some philosophers have tried to avoid these issues, while still endorsing a metaphysics of individuals, involves the apparatus of counterpart theory. However, the counterpart theorist's analysis of de re modal claims remains highly controversial, and even if successful, it is a matter of controversy whether it really secures anti-haecceitism.¹⁰¹

¹⁰⁰ For an overview of the problem of material constitution, see Wasserman (2018).

¹⁰¹ Perhaps the most famous objection facing counterpart theory is the so-called 'Humphrey Objection', originally put forward by Plantinga (1974: 116) and Kripke (1980: 45) and more recently advanced by Merricks (2003). Skow
Again, rather than simply taking on board a package of controversial modal views in the form of counterpart theory, we should instead try to formulate an underlying non-modal metaphysics from which we can *derive* the fact that these questions are bad questions. Fortunately, as we will see, there are theories that straightforwardly dissolve these questions concerning the modal relationship between individuals and their properties, without any controversial de re modal analyses.

II. Bundles

The most popular alternative view is the bundle theory. Just as individualist views come in many different flavors, so do bundle theories. As I shall be using the term 'bundle theory', all bundle theories are committed to the existence of properties (either understood as tropes or universals) and some sort of relation C, which we will call 'compresence', that holds between properties that serves to 'bundle' properties together. Primitive individuals are then reduced to mere 'bundles' of properties. For example, consider a world which we would ordinarily describe as only having two particulars, *a* and *b*, such that *a* has properties F and G and *b* has properties G and H. One traditional kind of bundle theorist will describe this world as one according to which there are three universals F, G, and H such that C(F,G) and C(G,H):

F-G G-H

For our purposes, bundle theories may be divided into two broad camps. Those that *identify* bundles with the same properties and those that do not. Let us call the first type *Extensional Bundle Theories* and the second type *Non-Extensional Bundle Theories*.

II.I. Non-Extensional Bundle Theories

There are many different ways one could be a Non-Extensional Bundle Theorist. Consider a world which we would intuitively think of as having two different individuals, each with exactly two properties, F-ness and G-ness. How could we make sense of this situation having *two* bundles if bundles are entirely constituted by their properties and both bundles have the very same properties?

^(2008, 2011) argues, for example, that Lewis' version of counterpart theory is best interpreted as a version of haecceitism.

Typically, bundles are thought to be analogous to sets: both are supposed to be uniquely determined by their members.¹⁰²

Here's one way.¹⁰³ Suppose the properties being bundled together are *tropes* instead of universals. Then, there could be two distinct bundles of F-ness and G-ness because the particular F-ness of the first bundle is just primitively distinct (i.e. distinct, but not by virtue of its qualitative nature) from the particular F-ness of the second bundle (and same with G-ness). According to this view, there would be four tropes, F_1 , F_2 , G_1 , and G_2 , such that $C(F_1, G_1)$ and $C(F_2, G_2)$:

Here's another way. Suppose that the compresence relation itself is an item in one's ontology that should be conceived as a trope. According to this view, there would be two universals F and G (or perhaps four tropes), and two primitively distinct compresence relations C_1 and C_2 , such that $C_1(F, G)$ and $C_2(F, G)$:

There are many other non-extensional bundle theories that philosophers have developed, but each of them crucially involves a similar idea: positing some kind of entity that can be *primitively distinct* from a corresponding qualitatively indiscernible entity (e.g. Rodriguez-Pereyra (2004), Zhang (2018), and Paul (2017)).

While these sorts of theories are perfectly coherent, because they are committed to primitively distinct entities, they do not live up to the motivations behind many anti-individualist theories in the first place. For the sake of concreteness, let us focus on the trope-theoretic version of the bundle theory. Consider the following two situations:

$$F_1-G_1$$
 F_2-G_2
 F_1-G_2 F_2-G_1

¹⁰² In fact, if bundles were not uniquely determined by their members, then bundle theorists would face the same modal questions that the individualist did. Why did these properties give rise to *this* bundle rather than some other one? Could this very same bundle have been composed of different properties? If so, which ones?

¹⁰³ The following two strategies are explicitly discussed in Benovsky (2008).

Because these situations are qualitatively indiscernible, they differ in an undetectable and physically redundant way. After all, physics isn't sensitive to the particular identities of the topes that are instantiated in some particular situation. This runs afoul of Dasgupta's (2009, 2017) argument against danglers. Furthermore, because both situations differ without differing qualitatively, such a theory implies haecceitism. If the first situation obtained, we could wonder why F_1 was compresent with G_1 rather than with G_2 (and why F_2 was compresent with G_2 rather than G_1), but such curiosity could only be met with an entirely inexplicable, brute fact.

II.II. Extensional Bundle Theories

Extensional Bundle theories have universals (rather than tropes) as part of their ontology, together with a primitive relation of *compresence* that serves to bundle universals together. Unsurprisingly, the main objection to such theories targets the fact that they *identify* bundles containing the very same universals. Because of this, such theories have a hard time accounting for certain possibilities. Consider, for example, a world that we would ordinarily describe as only containing two duplicate spheres that are ten feet apart. The natural thing to say is that there are two distinct bundles of universals that are ten feet apart. However, this violates the extensionality of the theory. In response, extensional theorists typically say that there is a *single* bundle of universals that is ten feet apart from itself (e.g. Hawthorne (1995) and Curtis (2014)). Just as universals may be multiply located, why can't 'bundles' of them be multiply located as well? Many objections have been raised towards this sort of move, and it is still a matter of much controversy whether it is successful. Below, I will quickly canvass three different sorts of objections that such a theory faces, which I think cumulatively make a strong case that the anti-individualist should look elsewhere.

First, there is *The Multi-Location Objection*. It is certainly counterintuitive that a bundle of universals can be 'ten feet apart from itself', but Vallicella (1997) has supplemented this intuition with an argument, conditional on certain prima facie plausible claims about bundles, locations, and instantiation. If we are to say that bundle B is the sort of thing that can be multiply-located, then, Vallicella argues, B must a universal, rather than a particular. Particulars are not the sort of thing that can be wholly located at different places at once, whereas universals clearly are. But a universal is multiply-located only if it is multiply instantiated, and so if B is to be multiply-located it must also be multiply instantiated. But, according to the bundle theory, what it is for a universal to be instantiated is for it to be compresent with other universals. But B is clearly not compresent

with any other universals, so it is not instantiated. So, it is not multiply instantiated, and so it is not multiply located.

Second, there is *The Modal Objection*. We initially objected that the extensional theory could not account for the existence of two duplicate spheres that are ten feet apart. In response, the extensionalist could redescribe the possibility as one where there is one sphere that is ten feet apart from itself. However, it can be argued that this is besides the point: what was intuitive was that there could be *two* spheres that are ten feet apart, not one sphere ten feet apart from itself. Rodriguez-Pereyera (2004) defends this modal judgement on the basis of continuity concerns. Everyone should agree that there can be two *almost* duplicate spheres that are ten feet apart. The first sphere can have temperature T and the second sphere can have an infinitesimally higher different temperature, T*. But if this is possible, then it should be possible for the second sphere to decrease its temperature ever so slightly, to temperature T. This establishes the possibility of *two* duplicate spheres that are ten feet apart, contrary to the extensional theory.¹⁰⁴

Third, there is *The Relation Objection*, which I take to be the strongest objection to the theory. Sider and Hawthorne (2002) note that, even if we grant the extensionalist that they can deal with symmetric cases, the extensionalist has a hard time distinguishing certain sorts of relational possibilities. Consider, for example, the following two distinct possibilities. Each consists of three duplicate particulars, such that adjacent particulars are ten feet apart, together with two different sorts of relations, R_1 (red) and R_2 (blue):

<i>B</i> —	B	B
B	B	В

On simple versions of the extensionalist theory, these two possibilities are indistinguishable. In both possibilities, the very same facts obtain, namely:

B is ten feet from itself

¹⁰⁴ In his recent defense of extensional bundle theory, Curtis (2014) simply denies the modal intuition that the second sphere could have decreased its temperature very slightly. In addition, to respond to other sorts of modal objections, Curtis (2014) resorts to the (controversial) apparatus of counterpart theory. The anti-individualist theory that I will come to endorse is not committed to counterpart theory.

B is twenty feet from itself

B bears R_1 to itself

B bears R_2 to itself

What the extensionalist needs to distinguish these possibilities is some way to distinguish 'which' particulars bear the relations to each other, even when it is the *same* bundle bearing these relations. Hawthorne and Sider (2002) have developed the most detailed account of how to accommodate these relational possibilities, but their theory leaves much to be desired, even by their own lights. As they write, 'Rather than containing a single notion of compresence, the primitive ideology of the bundle theorist now contains infinitely many locutions, each of which can be used to make a different sort of statement about universals' (64). Swallowing an infinite number of primitive notions to make sense of these sorts of possibilities is a rather large bullet to bite for the extensionalist.

III. Object Monism

Our goal is to develop an anti-individualist theory which accommodates our four desiderata in section I while avoiding the pitfalls of standard bundle theories canvassed in section II. The key to developing such a theory is to modify extensional bundle theories in two crucial ways.

The first modification involves rethinking the mereological structure of the world. On the standard picture, the world is ultimately built up out of the smallest elements of reality, perhaps point particles or space-time points that are related in various ways. On an alternative Monist picture, the world isn't 'built up' out of any tiny pieces, rather the world as a whole is what everything derives from. While 'bottom-up' views remain popular, in recent years there has been a surge of interest in the Monist picture, according to which, at least fundamentally speaking, there is only the one Cosmos. The view has been motivated by considerations of physical law (Schaffer (2013)), quantum entanglement (Schaffer (2010a), Ismael and Schaffer (2016)), causal essentialism (Segal (2014)), space-time substantivalism (Perry (2017)), vagueness and parsimony (Horgan and Potrč (2008)), mereology (Schaffer (2007, 2010a)), the denial of external relations (Vallicella (2002)), and failures of free modal recombination (Schaffer (2010b)). The view has even been appealed to

in various radical solutions to the hard problem of consciousness (Goff (2017), Kastrup (2018), Chalmers (2015, forthcoming))!

Specifically with respect to anti-individualist theories, there have already been some signs that point to Monism. The most detailed development of the extensional bundle theory, developed by Hawthorne and Sider (2002), entails a certain kind of metaphysical holism, according to which 'whenever there is a network of interrelated things, the facts cannot be captured by anything simpler than a single statement describing the entire organic whole' (62). Similarly, the sort of anti-individualist theory developed by Dasgupta (2009, 2017) *also* implies a kind of holism. Dasgupta describes his view as one in which 'one can state something fundamental only by characterizing the fundamental nature of the world as a whole...this is what might be called a 'many-from-one' metaphysics, on which many elements (the individualistic facts) flow together from the one source (the World Fact)' (2017: 20). On a bottom-up metaphysics, this is very surprising. If the world is really built up out of a plurality of tiny components that are related in various ways, why should one expect there to be indispensable, fundamental facts that concern the entire world as a whole? On a Monist metaphysics, however, this isn't at all surprising. In fact, this sort of holism is *entailed* by the Monist's view!

Combining Monism with the bundle theory gives us a theory with only one Cosmic Bundle from which everything else derives. The properties contained within the Cosmic Bundle would be certain sorts of global, fundamental *distributional* properties. Distributional properties are ways of 'filling in' some spatially extended object with some quality. For example, *being polka-dotted* is a color distributional property and *being hot at one end and being cold at the other* is an example of a heat-distributional property.¹⁰⁵ In the case of modern physics, perhaps the most plausible candidate for such distributional properties are distributional field-theoretic properties, which specify the distribution and magnitude of various fields across space-time. This kind of theory, I claim, has the resources to avoid all of the problems facing extensional bundle theories.

The initial problem facing the extensional theory is its failure to recognize certain sorts of symmetric possibilities. The reason why this spells trouble for extensionalist theories is because, at least on a 'bottom-up' metaphysics, such symmetrical possibilities are interpreted as having

¹⁰⁵ For much more on distributional properties, see Parsons (2004).

multiple distinct indiscernible individuals (e.g. two electrons or two spheres). The existence of multiple qualitatively indiscernible individuals, however, is incompatible with extensionalist theories. In contrast, such symmetric possibilities are *not* interpreted as containing distinct qualitatively indiscernible individuals given Monism. Rather, they are simply interpreted as the single Cosmic Bundle having a certain *symmetric distributional property*, which in no way clashes with extensionalist theories!

Furthermore, the Monist does not face *The Multi-Location Objection*, since it does not posit bundles that are multi-located. The Cosmic bundle is not 'ten feet away from itself'!

Neither does the Monist face *The Modal Objection*. According to this objection, the extensionalist needs to accommodate the existence of a world containing only two indiscernible spheres, because such a possibility is 'continuous' with possibilities containing only two *almost* indiscernible spheres. However, the Monist will simply reinterpret the possibility of two almost indiscernible spheres as one in which the Cosmic bundle contains an *almost* symmetric field-theoretic distributional property. The modal space that the Monist bundle theory recognizes does *not* have any abrupt discontinuities, unlike bottom-up extensional bundle theorists.

Turning to the last objection, the Monist easily avoids *The Relation Objection*. In order to distinguish certain sorts of relational possibilities, Hawthorne and Sider (2002) have argued that standard extensionalist theories need to adopt infinitely many distinct primitive notions of *compresence*. However, the Monist has no need for any relations between bundles. For the Monist, there is only the one Cosmic bundle. Therefore, the Monist bundle theorist is free to adopt a single primitive relation of compresence that serves the function of binding together the various fundamental distributional qualities of the Cosmos.

Having dealt with these objections, it is helpful to have an intuitive picture for how the Monist is thinking. To this end, consider the following analogy. Imagine that the world was one large *carpet*. Suppose that the carpet wasn't a flat carpet, but rather was bumpy and wrinkly throughout. In the ordinary business of life, small inhabitants in this one large carpet might find it convenient to *reify* these bumps and wrinkles. They might start ascribing various properties to them and talking about how they relate to other bumps and wrinkles. The metaphysicians among them might start asking questions about the persistence conditions of various bumps (if we made this bump smaller, would

it be the same bump?) and their de re modal properties (could this very bump have been a wrinkle instead?). In fact, since the inhabitants are so small, they would find little need to ever talk about the global properties of the carpet as a whole. Rather, they would spend all of their time talking about the various indentations, crevices, bumps, and wrinkles in the carpet. Because of this, the metaphysicians among them might start to think that the carpet was *composed* of these various tiny aspects of the carpet. The bundle theorists among these metaphysicians would then face various conundrums: couldn't there be a carpet with exactly two perfectly symmetrical bumps, which is otherwise flat? Unfortunately, these inhabitants would be making a mistake. The various bumps, wrinkles, indentations, and crevices in the carpet should *not* be reified in this way. In reality, there is only the carpet, albeit a bumpy and wrinkly one.

From the Monist perspective, our situation is similar. On a popular field-theoretic interpretation of our best quantum field theories, ordinary objects like tables and chairs, and even the tiniest of 'particles' such as electrons, are just *excitations* in these global quantum fields, analogous to 'bumps' or 'wrinkles' in the one carpet.¹⁰⁶ As the philosopher of physics d'Espagnat concludes:

In quantum field theory, reality lies at a deeper level than could be imagined by common sense or even elementary quantum mechanics. A particle is not itself 'a reality'; it is simply a more or less transient property of reality, a level of excitation (to speak as physicists do)... of reality, excited in a fashion corresponding to the field in question (1983: 85).

IV. Property Monism

The Monist bundle theory already avoids all of the problems that we wanted to avoid in section II. At this point, the anti-individualist bundle theorist could simply endorse Monism and call it a day. However, a much more theoretically satisfying anti-individualist theory waits around the corner, as long as we are willing to make one more (fairly radical) change to our Monist bundle theory.

IV.I. Against Compresence

It would be better if the bundle theorist could somehow do without compresence. As Benovsky (2008) writes, 'the compresence relation is usually taken as unanalyzable and ontologically

¹⁰⁶ For defenses of the field-theoretic interpretation of our best quantum field theories, see Halvorson and Clifton (2002), Redhead (1995), and Zeh (2003). For an overview of different interpretations, see Kuhlman (2018).

primitive. It is thus defined and individuated not by its nature or intrinsic features of which we are not told much by [the bundle theory], but rather by its theoretical role: it is a *unifying device*, a device that takes properties to make up objects' (177). Benovsky goes on to note that the compresence relation in fact plays the very same role as individuals do. Just as many properties are unified by a single compresence relation to make a single object, individuals are the singular item in which multiple properties are 'instantiated' to make a propertied object. Benovksy writes:

Both [the bundle theory] and [the individuals theory] thus have a unifying device, a primitive and under-defined one, an entity whose purpose is to tie or glue together properties of a single object. Paraphrasing Locke, in *both* cases this unifying device is a 'we-know-not-what'...but it is a 'we-know-what-it-does', that is, we know its theoretical role. (177)

On reflection, it's not at all clear how the compresence relation could manage to accomplish its intended role: how could *redness* and *sphericity* be combined to make a red sphere? Because the compresence relation is usually left as primitive, we are left in the dark. Call this problem, that the compresence relation is a primitive (we-know-not-what) relation that somehow serves to unify properties (we-know-not-how), the *Unity Problem*.

Also because of the compresence relation, bundle theorists violate the spirit of a 'one-category ontology', if not the letter. On the one hand, the bundle theorist is committed to fundamental, intrinsic *qualities*, but on the other hand it is also committed to an entirely different kind of thing, a primitive relation of *compresence*. Do both of these kinds of things belong to the same category? Partly because of the unclarity of 'ontological category', it's not so clear. Moreover, when various properties are related by the compresence relation, this is supposed to generate a new entity, a 'bundle' of properties. For straightforward reasons, such a bundle cannot be identified with the qualities that compose it; the bundle is *one* while the qualities are *many*.¹⁰⁷ Since a bundle isn't itself a quality, again this seem to pose a threat to a one-category ontology. At the very least, the considerations in favor of having a one-category ontology (i.e. theoretical elegance and

¹⁰⁷ There is a similar problem in the metaphysics of mereology: one cannot identify a whole with its parts, since this violates the fact that identity is a one-to-one relation. While this remains the orthodox position, some philosophers have denied that identity is a one-to-one relation and have defended the thesis of *composition as identity*. For a recent collection of articles on the subject, see Cotnoir and Baxter (2014).

parsimony), give us reason to try to do without the compresence relation and the bundle it generates. Call this problem the *One-Category Problem*.

Lastly, there is independent pressure for the Monist in particular to be wary of fundamental relations. One of the most historically influential arguments for Monism, Bradley's regress, is incompatible with the existence of ungrounded, fundamental relations.¹⁰⁸ There is therefore extra dialectical pressure for the Monist to get rid of the bundling relation. Call this problem the *Relation Problem*.

IV.II. Instantiation as Identity

The reason why the Monist bundle theory needs a compresence relation is to unify the many qualities of the Cosmos into one. After all, the many qualities are supposed to be all *of* the one Cosmos. How are we to capture this important unity among the qualities of the Cosmos without some sort of unifying device? The way out is to deny that the Cosmos has 'many' qualities at all. There is really only one fundamental quality that the Cosmos has, its *maximal* quality. Call this view, that the Cosmos only has one fundamental quality, *Property Monism*. Call the view that the Cosmos has many distinct fundamental qualities, *Property Pluralism*.

Once this move is made, the Monist bundle theory becomes the view that there is a single maximal quality that is 'bundle' together with itself using a compresence relation, which generates a 'bundle' distinct from the one maximal quality. At this point, the redundancy of the compresence relation is clear. Rather than identifying the Cosmos with a 'bundle' consisting of only one quality, we should dispense with the compresence relation altogether and identity the Cosmos with the one maximal Cosmic quality instead. This view has the resources to collapse the object-property distinction altogether. The World *just is* the Way it is.

This view easily avoids the *Unity Problem*. There is no mystery of how distinct qualities can somehow combine into one, since there is only one quality in the first place.

The view also avoids the *One Category Problem*. The view has no need of a primitive compresence relation, nor does it have any need for positing a numerically distinct 'bundle'. The reason why Property Pluralists needed to endorse the existence of a 'bundle' is because the Cosmos is a single

¹⁰⁸ For an overview of Bradley's Regress, see Perovic (2017).

thing, while its qualities are many. So, the Cosmos cannot be identified with its qualities. This objection is absent for Property Monists, so the Property Monist is free to identify the Cosmos with its maximal quality. In slogan form, Property Monism lets one endorse the thesis of *Instantiation as Identity*. In ordinary English, we have two uses of the word 'is'. There is the 'is' of identity (e.g. Hesperus is Phosophorus), and there is the 'is' of predication or instantiation (e.g. the apple is red). Instead of taking on board the second 'is' of predication as *sui generis*, as most philosophers do, the Property Monist can identify the 'is' of predication with the 'is' of identity!¹⁰⁹ Instead of saying that the Cosmos *is* F (in the sense of predication), we should instead say that the Cosmos is F-ness itself (in the sense of identity).

Lastly, by rejecting the compresence relation, the view does away with the Relation Problem.¹¹⁰

IV.III. In Defense of Property Monism

Property Monism certainly contradicts common sense. To many, I suspect it will just seem obviously false.

Property Monism shares this feature with Object Monism. Just as to many it seems obvious that there are plurality of fundamental objects (e.g. particles or space-time points), to many it seems obvious that there are a plurality of fundamental properties (e.g. mass, spin, charge). The problem that each theory faces is structurally analogous. In the case of Object Monism, there is the general mereological question of how we ought to 'carve up' the world into objects. Some only carve out its smallest components; some also choose to carve out ordinary objects; some also choose to carve out objects that no one would ever dream of. In contrast, the Object Monism rejects the very idea of carving up the world into objects. The world is fine without any carvings. Similarly, the Property Monist faces the question of how the *nature* of the world should be carved up. If we do choose to

¹⁰⁹ Orilia and Swoyer (2017, Section 2) provide an overview of different attempts to analyze the 'is' of instantiation, and they conclude by saying, 'Despite these proposals, it seems safe to say that most philosophers, whether tacitly or overtly, take exemplification to be a primitive and unanalyzable phenomenon'.

¹¹⁰ While I have mainly been addressing the problems with bundle theory, Property Monism *also* addresses some of the main problems with Dasgupta's (2009, 2017) anti-individualist theory, *Algebraic Generalism*. First, Turner (2017) criticizes Algebraic Generalism because it uses a primitive piece of ideology ('&') that serves to 'stitch' properties together, similar to the compresence relation. Just as Property Monism dissolves the need for compresence, it dissolves the need for this primitive operator. Second, the holism that Algebraic Generalism is committed to is *explained* by a Monist metaphysics, but it is a surprise given a bottom-up metaphysics. Third, it is controversial whether individuals can be adequately grounded in the one 'World Fact' of Algebraic Generalism (e.g. Russell (2017)), whereas by endorsing *Instantiation as Identity*, there just is no difference between an individual and the corresponding quality that it 'has'.

156

carve up the nature of the world, exactly *how many* qualities 'combine' to make up how the world is? The Property Monist rejects the thought that the nature of the world should be carved up in this way, just as the Object Monist rejects the thought that the world should be carved up into objects.

In response to this sort of worry, the Object Monist needs to explain how the immense structure and complexity of the world can be accommodated by a single object (e.g. by positing global distributional properties of the Cosmos that account for the fundamental fields of physics). Similarly, the Property Monist needs to explain how the structure and complexity of a single object, the Cosmos, can be accommodated by a single quality. Property Monism can have all the theoretical virtues in the world, but if it cannot account for the complexity of the Cosmos, then it does not merit our belief.

In response, I want to suggest that instead of thinking of the world as having different fundamental qualities, we should instead think of these different qualities as mere *aspects* or *modes* of a single underlying fundamental quality. What is this 'aspect' talk supposed to mean? The easiest way to get a grip on this is by way of example. Consider a maximally determinate color property C. C has a rich internal structure to it, owing to its different aspects. C has a certain hue, a certain saturation, and a certain *brightness*. It is sometimes fruitful to distinguish these different aspects of colors and study them separately. Nonetheless, a natural thought is that these three aspects are all *derivative* off C. The reason why an object counts as having a particular hue, or a particular saturation, or a particular brightness, is because of what color it is! It is incorrect to say that the object really has three wholly distinct qualities -a hue, saturation, a brightness -and in virtue of those three qualities somehow 'combining' it counts as having a certain color. At the very least, the view that the more 'global' property C is the more fundamental one, from which the other three are mere aspects, is an intelligible view. And if it is an intelligible view, then this establishes the possibility of qualities that have a rich internal structure. The mere epistemological fact that it might be fruitful to distinguish different aspects of a quality and study them independently is perfectly consistent with the Property Monist's claim that ultimately there is only one maximal quality.¹¹¹

¹¹¹ In indulging in this 'aspect' talk, I do not mean to be *reifying* aspects as numerically distinct entities apart from the one maximal quality that is the Cosmos. Strictly speaking, there is only the one maximal quality, but it has a rich, complex structure that is helpfully conveyed by speaking of its various 'aspects'.

Here is a deeper, and perhaps more controversial, example. Consider the category of *phenomenal qualities*. Phenomenal qualities specify *what it's like to be* something. A sizeable fraction of philosophers think that phenomenal qualities are among the fundamental qualities of the universe, including dualists, panpsychists, and idealists.¹¹² In fact in recent years, several philosophers have defended the idea that the universe as a whole has fundamental phenomenal qualities, and that our own phenomenal states are ultimately grounded in them.¹¹³ At the very least, there is more reason to think that phenomenal qualities are fundamental qualities of the universe than color qualities! I'd like to suggest that we treat phenomenal qualities analogously to color qualities. Fundamentally, there is only my *total experiential state*. This quality fully specifies exactly what it's like to be me. This quality has various *aspects* to it – visual aspects, auditory aspects, olfactory aspects, etc. However, the reason why I count as having these aspects is because of my total phenomenal state, not vice versa. Perhaps the most powerful argument in favor of this thought is the much-discussed phenomenon of the *unity of consciousness*. Brook and Raymont (2017) introduce the phenomenon as follows:

Human consciousness usually displays a striking unity. When one experiences a noise and, say, a pain, one is not conscious of the noise and then, separately, of the pain. One is conscious of the noise and pain together, as aspects of a single conscious experience. Since at least the time of Immanuel Kant, this phenomenon has been called the *unity of consciousness*. More generally, it is consciousness not of *A* and, separately, of *B* and, separately, of *C*, but of *A*-and-*B*-and-*C* together, as the contents of a single conscious state.

According to Bayne and Chalmers (2003), we should say that experiences are 'unified' just in case they are 'aspects of a single encompassing state of consciousness'.¹¹⁴ Searle (2002) argues that we have a 'single, unified conscious field containing visual, auditory, and other aspects' and that 'there is no such thing as a separate visual consciousness' (54). Tye (2003) also endorses what he calls 'the one-experience view'. In considering the different aspects of our experience he writes, 'There are not five different . . . experiences somehow combined together to produce a new unified

¹¹² For recent articulations and defenses of panpsyschsim and idealism, see Chalmers (2015, forthcoming). For an overview of dualism, see Robinson (2017).

¹¹³ See, for example, Jaskolla and Buck (2012), Nagasawa and Wager (2017), Goff (2017), Kastrup (2018), and Albahari (forthcoming).

¹¹⁴ Also see Bayne (2010: 20, 31).

experience.' Rather, 'there is just one experience here' (27). If one rejects Property Monism with respect to consciousness, one faces the awkward question of saying *exactly how many* basic phenomenal qualities somehow combine to make one's total experiential state. The prospects for answering this question, however, don't seem very promising (e.g. is there supposed to be a basic phenomenal quality for every 'pixel' of one's visual field?).

Even besides these two natural examples, the Object Monist is *already* committed to there being immensely richly structured qualities in the form of fundamental distributional qualities, that serve to specify (say) the distribution and magnitude of quantum fields throughout all of space-time. There doesn't seem, then, to be any *in principle* worry to the idea that the structural complexity of the Cosmos can be captured by a single immensely rich and structured maximal quality.

Given that there are no in principle objections to Property Monism, one might still wonder: does contemporary physical theory give us any reason to favor Property Monism over Property Pluralism, or vice versa? Many philosophers have defended claims that suggest that such evidence would be very hard to come by, due to the fact that physics can only ever tell us about the causal/nomic/mathematical/structural features of the world, rather than the intrinsic qualities underneath this structure.¹¹⁵ In the absence of any direct evidence about these intrinsic qualities, it seems like the structural complexity that the world exhibits could equally well be attributed to a single quality with many different aspects, or a plurality of wholly distinct qualities that somehow combine together.

This being said, perhaps we could look for indirect evidence. In particular, one might speculate that, given Property Pluralism, one should find it harder to formulate the laws of physics elegantly and compactly with reference to the 'maximal' state of the Cosmos, since the maximal state of the Cosmos would presumably be a gerrymandered, merely 'conjunctive' state. Given Property Monism, however, we should find this possibility to be more likely. In our earlier, classical physical theories, this sort of evidence seemed to point towards Property Pluralism. The property of mass was governed by one set of laws (Newton's Laws), and electromagnetic properties were

¹¹⁵ Contemporary Russellian Monists, following Russell (1927), endorse this sort of claim. Langton (2001, 2004) defends an interpretation of Kant in which we can never have knowledge of things 'in themselves', and Lewis (2008) also defends this conclusion. Many who identify as 'structural realists' in the philosophy of science are also sympathetic to this claim. For an overview of Structural Realism, see Ladyman (2016).

governed by an entirely different set of laws (Maxwell's Equations). This makes it seem more plausible that mass and charge are wholly distinct qualities. As we look to more contemporary physics, the evidence seems to point more towards Property Monism. In the case of non-relativistic Quantum Mechanics, perhaps the most basic equation is Schrodinger's equation:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle$$

This law directly evolves the entire global state of the cosmos at a time, represented by the wave function $\psi(t)$, into the future. The wavefunction is a central object of study in physics, and it isn't at all considered to be a merely gerrymandered object of study. That being said, the exact philosophical interpretation of the wave function is a hotly contested topic.¹¹⁶ It should also be noted that $\psi(t)$ doesn't represent a single global quality of the entire four-dimensional world, it only represents the intrinsic state of the world at a single time. If non-relativistic quantum mechanics were the final physical theory, the Property Monist should say that the single distributional quality across space-time is represented by the *evolution* of the wave function, where for each time t, $\psi(t)$ represents a certain temporal aspect of the single distributional quality. Looking forward to more speculative areas of physics, there is the Wheeler-DeWitt equation, which is an important equation in the field of quantum gravity. In its most simple, compressed form, the equation reads:

$\widehat{H}(x)|\psi>=0$

The interpretation of the symbols in this equation differs substantially from Schrodinger's equation, but the important thing to note is that this time the wave function ψ represents the global state of the entire spatiotemporal universe. Although the true theory of quantum gravity is still far off, an elegant equation like this one is exactly what one would expect, given Property Monism.

Moreover, if one looks across the history of physics, much of the progress that has been made consists of *unifying* seemingly different properties into different 'aspects' of a single underlying property. Mass and energy were unified into mass-energy utilizing Einstein's famous equation, $E = mc^2$. Electricity and magnetism were unified into electromagnetism in Maxwell's equations.

¹¹⁶ For a recent collection of essays on the metaphysics of the wave function, see Ney and Albert (2013).

Electromagnetic forces were unified with the so-called 'weak' interactions in the electroweak theory, resulting in the 1979 Nobel Prize in Physics. In modern times, several so-called Grand Unified Theories (GUTs) have been proposed that attempt to unify electromagnetism with both the strong and weak nuclear forces (e.g. Ross (1985)). Perhaps the holy grail of physics is to unify all four fundamental forces - gravitation, electromagnetism, and the strong and weak nuclear forces - into one 'theory of everything'.

Of course, there is much more to say about the relation between fundamental physics and Property Monism, and a serious assessment of these issues will inevitably be sensitive to one's interpretation of quantum theory. At the very least, however, there doesn't seem to be a clear case against Property Monism from fundamental physics; in fact, fundamental physics may end up providing additional support for Property Monism.

V. Taking Stock

Let us step back and see why we have ended up with the theory we have.

Our first motivation was to adopt a One-Category Ontology. Our theory does exactly this. It collapses the distinction between object and property, and only posits a single enormously structured Cosmic Quality (that *just is* the Cosmos itself). By doing this, the theory has the extra benefit of doing away with a *sui generis* notion of predication or instantiation. The 'is' of predication is simply identified with the 'is' of identity. In the words of Galen Strawson (2008), 'the being of an object is literally identical with the being of its propertiedness' (281). The distinction between objects and properties is a mere distinction of language.

Our second motivation was to remove undetectable and physically redundant structure from the world. We have accomplished this by getting rid of primitive individuals that 'have' properties, since physics is completely insensitive to *which* individuals underlie the qualities in the world.

Our third motivation was to explain the impossibility of bare particulars. On our theory, an object *just is* a kind of maximal quality. If there were no maximal quality, there would be no object. The notion of a bare particular is explained away as a confusion that results from distinguishing objects and properties.

Our fourth motivation was to deflate questions concerning the modal relationship between objects and properties. Could the *very same* individual have different qualities? Could the *very same* qualities be had by different individuals? These questions don't make any sense if an object is identical to a quality. It makes no sense to fix an object and vary its qualities (or fix some quality and vary its object) if an object and its quality are one and the same.

Moreover, we have avoided positing primitively distinct entities (unlike non-extensional bundle theories), and we have accommodated symmetric possibilities while avoiding *The Multi-Location Objection*, *The Modal Objection*, and *The Relation Objection* (unlike extensional bundle theories). We have also gotten rid of the two most worrying elements of bundle theories in general, the 'compresence' relation and the existence of a single 'bundle' distinct from its qualities. By doing this, we have avoided the *Unity Problem*, the *One Category Problem*, and the *Relation Problem*.

The thesis that the world just is a certain quality is counterintuitive. But its theoretical fruits are abundant. I think we should follow the arguments where they lead.

References

Albahari, Miri. Forthcoming. Beyond cosmopsychism and the great I am: How the world might

be grounded in universal 'advaitic' consciousness. In *The Routledge Handbook of Panpsychism*, edited by William Seager. Routledge.

Bayne, Tim. 2010. The Unity of Consciousness. Oxford University Press UK.

Bayne, Timothy J., and David J. Chalmers. 2003. What Is the Unity of Consciousness? In *The Unity of Consciousness*, edited by Axel Cleeremans. Oxford University Press.

Benovsky, Jiri. 2008. The Bundle Theory and the Substratum Theory: Deadly Enemies or Twin Brothers? *Philosophical Studies* 141, no. 2: 175–190.

Brook, Andrew and Raymont, Paul. 2017. The Unity of Consciousness. In The Stanford

Bailey, Andrew M. 2012. No Bare Particulars. Philosophical Studies 158, no. 1: 31-41.

Encyclopedia of Philosophy (Summer 2017 Edition), edited by Edward N. Zalta.

- Cameron, Ross P. 2008. Truthmakers and Ontological Commitment: Or How to Deal with Complex Objects and Mathematical Ontology Without Getting Into Trouble. *Philosophical Studies* 140, no. 1: 1–18.
- Chalmers, David. 2009. The Two-Dimensional Argument Against Materialism. In Oxford
 Handbook to the Philosophy of Mind, edited by Brian P. McLaughlin and Sven Walter.
 Oxford University Press.
- Chalmers, David. 2015. Panpsychism and Panprotopsychism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, edited by T. Alter and Y. Nagasawa, 246-76.
 Oxford University Press.
- Chalmers, David. Forthcoming. Idealism and the Mind-Body Problem. In *The Routledge Handbook of Panpsychism*, edited by William Seager. Routledge.
- Cotnoir, A. J., and Donald L. M. Baxter. *Composition as Identity*. Oxford: Oxford University Press, 2014.
- Cowling, Sam. 2017. Conceivability Arguments for Haecceitism. *Synthese* 194, no. 10: 4171–4190.
- Curtis, Benjamin L. 2014. The Rumble in the Bundle. Noûs 48, no. 2: 298–313.
- d'Espagnat, Bernard. 1983. In Search of Reality. New York. Springer-Verlag.
- Dasgupta, Shamik. 2009. Individuals: An Essay in Revisionary Metaphysics. *Philosophical Studies* 145, no. 1: 35–67.

Dasgupta, Shamik. 2015. Substantivalism Vs Relationalism About Space in Classical Physics.

Philosophy Compass 10, no. 9: 601–624.

Dasgupta, Shamik. 2017. Quality and Structure. In *Current Controversies in Metaphysics*, edited by Elizabeth Barnes. New York: Routledge, 7-23.

Dorr, Cian. 2004. Non-Symmetric Relations. Oxford Studies in Metaphysics 1: 155–92.

Dorr, Cian. 2008. There Are No Abstract Objects. In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean W. Zimmerman. Blackwell.

Giberman, Daniel. 2012. Against Zero-Dimensional Material Objects (and Other Bare Particulars). *Philosophical Studies* 160, no. 2: 305–321.

- Goff, Philip. 2017. Consciousness and Fundamental Reality. New York, USA: Oxford University Press.
- Goswick, Dana. 2018. Are Modal Facts Brute Facts? In *Brute Facts*, edited by Elly Vintiadis and Constantinos Mekios. Oxford University Press.
- Halvorson, Hans, and Rob Clifton. 2001. No Place for Particles in Relativistic Quantum Theories? *Philosophy of Science* 69, no. 1: 1–28.
- Hawthorne, John. 1995. The Bundle Theory of Substance and the Identity of Indiscernibles. *Analysis* 55, no. 3: 191–196.
- Hawthorne, John. 2006. Determinism De Re. In *Metaphysical Essays*, 239–243. Oxford: Oxford University Press
- Hawthorne, John, and Theodore Sider. 2002. Locations. *Philosophical Topics* 30, no. 1: 53–76.
- Horgan, Terence, and Matjaž Potrč. 2008. Austere Realism: Contextual Semantics Meets Minimal Ontology. MIT Press.

- Ismael, Jenann, and Jonathan Schaffer. 2016. Quantum Holism: Nonseparability as Common Ground. *Synthese*. https://doi.org/10.1007/s11229-016-1201-2
- Jaskolla, Ludwig, and Alexander J. Buck. 2012. Does Panexperiential Holism Solve the Combination Problem? *Journal of Consciousness Studies* 19, no. 9–10: 9–10.
- Kastrup, Bernardo. 2018. The Universe in Consciousness. *Journal of Consciousness Studies* 25, no. 5–6: 125–155.
- Kripke, Saul A. Naming and Necessity. Harvard University Press, 1980.
- Kuhlmann, Meinard. 2018. Quantum Field Theory. In *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), edited by Edward N. Zalta.
- Ladyman, James. 2016. Structural Realism. In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta.
- Langton, Rae. 2001. Kantian Humility: Our Ignorance of Things in Themselves. Oxford University Press.
- Langton, Rae. 2004. Elusive Knowledge of Things in Themselves. *Australasian Journal of Philosophy* 82, no. 1: 129–136.
- Lewis, David. 2009. Ramseyan Humility. In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, 203–222. MIT Press.

Locke, John. 1690/1975. An Essay Concerning Human Understanding. Oxford University Press.

- Merricks, Trenton. 2003. The End of Counterpart Theory. *Journal of Philosophy* 100, no. 10: 521–549.
- Mertz, D. W. 2003. Against Bare Particulars A Response to Moreland and Pickavance. *Australasian Journal of Philosophy* 81, no. 1: 14–20.

- Mertz, D. W. 2001. Individuation and Instance Ontology. *Australasian Journal of Philosophy* 79, no. 1: 45–61.
- Nagasawa, Yujin, and Khai Wager. 2016. Panpsychism and Priority Cosmopsychism. In *Panpsychism: Contemporary Perspectives*, edited by Godehard Bruntrup and Ludwig Jaskolla. Oxford: Oxford University Press.
- Ney, Alyssa, and David Z. Albert. 2013. *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*. Oxford University Press USA.
- Norton, John D. 2015. The Hole Argument. In *The Stanford Encyclopedia of Philosophy* (Fall 2015 edition), edited by Edward N. Zalta.
- Orilia, Francesco, and Chris Swoyer. Properties. In *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), edited by Edward N. Zalta.
- Parsons, Josh. 2004. Distributional Properties. In *Lewisian Themes: The Philosophy of David K. Lewis*, edited by Frank Jackson and Graham Priest. Clarendon Press.
- Paul, L.A. 2013. Categorical Priority and Categorical Collapse. Aristotelian Society Supplementary Volume 87, no. 1: 89-113.
- Paul, L. A. 2017. A One Category Ontology. In *Being, Freedom, and Method: Themes From the Philosophy of Peter van Inwagen*, edited by John A. Keller. Oxford University Press.
- Perovic, Katarina. 2017. Bradley's Regress. *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), edited by Edward N. Zalta.
- Perry, Zee R. 2017. How to Be a Substantivalist Without Getting Shifty About It. *Philosophical Issues* 27, no. 1: 223–249.
- Plantinga, Alvin. 1974. The Nature of Necessity. Clarendon Press.

Pooley, Oliver. 2013. Substantivalist and Relationalist Approaches to Spacetime. In *The Oxford Handbook of Philosophy of Physics*, edited by Robert Batterman, 522–586. Oxford:
 Oxford University Press.

Redhead, Michael. 1995. More Ado About Nothing. Foundations of Physics 25, no. 1: 123–137.

- Robinson, Howard. Dualism. 2017. In *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), edited by Edward N. Zalta.
- Rodriguez-Pereyra, Gonzalo. 2004. The Bundle Theory Is Compatible with Distinct but Indiscernible Particulars. *Analysis* 64, no. 1: 72–81.
- Rodriguez-Pereyra, Gonzalo. 2017. Indiscernible Universals. *Inquiry: An Interdisciplinary Journal of Philosophy* 60, no. 6: 604–624.

Ross, Graham. 1985. Grand Unified Theories. Benjamin & Cummings.

Russell, Bertrand. 1927. The Analysis of Matter. London: Kegan Paul.

Russell, Bertrand. 1940. An Inquiry Into Meaning and Truth. Routledge.

- Russell, Jeffrey Sanford. 2017. Qualitative Grounds. *Philosophical Perspectives* 30, no. 1: 309–348.
- Schaffer, Jonathan. 2007. From Nihilism to Monism. *Australasian Journal of Philosophy* 85, no. 2: 175–191.
- Schaffer, Jonathan. 2010a. Monism: The Priority of the Whole. *Philosophical Review* 119, no. 1:

31–76.

Schaffer, Jonathan. 2010b. The Internal Relatedness of All Things. *Mind* 119, no. 474: 341–376.

Schaffer, Jonathan. 2013. The Action of the Whole. Aristotelian Society Supplementary Volume

87, no. 1: 67–87.

- Searle, John R., [2000] 2002, "Consciousness", in his *Consciousness and Language*, Cambridge: Cambridge University Press.
- Segal, Aaron. 2014. Causal Essentialism and Mereological Monism. *Philosophical Studies* 169, no. 2: 227–255.

Sider, Theodore. 2013. Against Parthood. Oxford Studies in Metaphysics 8: 237-293.

- Skow, Brad. 2008. Haecceitism, Anti-Haecceitism, and Possible Worlds: A Case Study. *Philosophical Quarterly* 58, no. 230: 97–107.
- Skow, Bradford. 2011. More on Haecceitism and Possible Worlds. Analytic Philosophy 52, no. 4: 267–269.
- Strawson, Galen. 2008. The Identity of the Categorical and the Dispositional. *Analysis* 68, no. 300: 271–282.
- Teitel, Trevor. Forthcoming. Holes in Spacetime: Some Neglected Essentials. *The Journal of Philosophy*.
- Turner, Jason. 2017. Can We Do Without Fundamental Individuals? In *Current Controversies in Metaphysics*, edited by Elizabeth Barnes. New York: Routledge, 24-42.

Tye, Michael. 2003. Consciousness and Persons: Unity and Identity. MIT Press.

- Vallicella, William F. 1997. Bundles and Indiscernibility: A Reply to o'Leary-Hawthorne. *Analysis* 57, no. 1: 91–94.
- Vallicella, William F. 2002. Relations, Monism, and the Vindication of Bradley's Regress. *Dialectica* 56, no. 1: 3–35.

Van Cleve, James. 2018. Brute Necessity. Philosophy Compass 13, no. 9: e12516.

Wasserman, Ryan. 2018. Material Constitution. In The Stanford Encyclopedia of

Philosophy (Fall 2018 Edition), edited by Edward N. Zalta.

Williamson, Timothy. 2013. Modal Logic as Metaphysics. Oxford University Press.

Zeh, H. D. 2003. There Is No 'First' Quantization. Physics Letters A309: 329-34

Zhang, Ruoyu. 2018. A New Universal Bundle Theory. Philosophia 46, no. 2: 473–486.