

University of Nevada, Reno

**Potency and Process: A Naturalistic Defense of Mental Causation**

A thesis submitted in partial fulfillment of the  
requirements for the degree of Master of Arts in  
Philosophy

By

Devin J. Bray

Dr. Thomas Nickles/Thesis Advisor

May, 2016

**THE GRADUATE SCHOOL**

We recommend that the thesis  
prepared under our supervision by

**DEVIN JAMES BRAY**

Entitled

**Potency and Process: A Naturalistic Defense of Mental Causation**

be accepted in partial fulfillment of the  
requirements for the degree of

**MASTER OF ARTS**

Thomas Nickles, Ph.D., Advisor

Christopher Williams, Ph.D., Committee Member

Michael Webster, Ph.D., Graduate School Representative

David W. Zeh, Ph.D., Dean, Graduate School

May, 2016

## **Abstract**

It has been argued extensively in contemporary philosophy of mind and neuroscience that physicalism precludes the possibility of robust mental causation. “Exclusion arguments” of the sort offered by Jaegwon Kim present the physicalist with an apparent dilemma: either all putatively “mental” causation reduces to mere “bottom-up” causation, or else it is epiphenomenal. However, the purportedly “scientific” premises that bolster these arguments can be argued are metaphysically overladen, smuggling in assumptions about the fundamental nature of matter that are not warranted by our best current data. By examining the evidence against such assumptions and the feasibility of alternative metaphysical frameworks, serious doubt can be cast on the validity of Exclusion arguments. Accepting the plausibility of causal entanglements leading from the “top down,” so to speak, also turns out to be a better fit with experimental practice, especially in explaining psychosomatic medical phenomena.

**Table of Contents:**

~ Chapter One ~

What This Thesis is Not

*page 1*

~ Chapter Two ~

Hard Determinism and Incompatibilism:

Why Naturalism Seems To Motivate Skepticism About Self Control

*page 9*

~ Chapter Three ~

From Hume To Woodward:

What Instrumentalists About Causation Get Right

*page 20*

~ Chapter Four ~

Why Process? Towards A Thinner (and more permissive) Metaphysics of Causation

*page 38*

~ Conclusion ~

*Page 71*

## Chapter One: What This Thesis Is Not

In this thesis, I will be expounding positions that leave me open at any point to a myriad of misunderstandings on the part of the reader. This is for two primary reasons. Firstly, it's because I will be adopting an approach towards the philosophy of mind that is broadly in the tradition of pragmatic skeptics like Ludwig Wittgenstein, Richard Rorty, and Daniel Dennett. These philosophers are widely misunderstood in the literature. It seems to me the vast majority of papers that have been written in an attempt to refute their positions fundamentally misunderstand the views the targets of their critiques have been trying to articulate. Their critiques thus *miss* their intended targets, because the view at which they are taking aim is not the view expressed by the philosophers in question. I hope that when it comes time to spell out the differences between my position and those in the Wittgenstein-Rorty-Dennett tradition, my own thesis will not succumb to this same difficulty. Secondly, I risk being misunderstood because I will be attempting to provide an empirically responsible metaphysical basis for the kind of compatibilism each of these philosophers has endorsed, or at least gestured toward. As I begin to speak about "self control," and present my arguments against the epiphenomenalist, it is possible I could be misconstrued as arguing for a brand of property dualism, libertarian free will, or spooky mind-stuff that's capable of making molecules swerve in mid-career. Rather, I intend to show that with a rigorous commitment to both naturalism and empiricism, along with deference to the findings of contemporary physics, a view can be defended that although our consciousness arises from social practice and material events in the central nervous system, it still makes sense to talk about us being "in control of ourselves."

In an attempt to head off some of those misapprehensions at the pass, so to speak, it should be helpful to take some time, at the outset, to specify what positions I am *not* offering to

the reader, and to explain why. This will hopefully lessen the chance of confusion when I begin to make some of my constructive claims a bit later on.

### **a. A Denial of Subjective Experience**

This should clear up relatively quickly once I begin talking about what I will call “causal entanglement” and begin to advocate for the global causal relevance of the brain states denoted by folk-psychological terminology. However, at various points in this essay I will be casting serious doubts on some of our dearly-held intuitions, both about the nature of our subjective experience and about what kind of ontological commitments are implied by a naturalistic worldview. Later attacks on putative mentalistic entities like “qualia” (or what Richard Rorty calls “raw feels”) and on the idea of an “inner observer” (what Dennett refers to as the “Cartesian theater,”) should not then be construed as dismissing the idea that there is something it’s like to be you, as well as something it’s like to be me. I don’t wish to deny that there is a subjective dimension to the experience of the espresso I’m now drinking, or the warmth of the mug I feel in my hands, or the sound of the wind blowing through the trees outside my window. One of the aims of this project will be to argue that this subjective dimension to our lives as organisms is not an accidental, epiphenomenal, or eliminable aspect of our neuronal function. Rather, it is the result of a crucial, and perhaps inevitable development in the evolution of life on earth. It is central to the ideas contained here that multicellular life, *especially as it grows more complex*, cannot react intelligently without being capable of generating states that are experienced globally. These states give the organism a “feel” for its environment and internal milieu that allows it to make predictions about the future of itself and its environment, and strive towards those behaviors that best favor its ongoing homeostasis, survival and replication.

Part of the task of philosophy of mind, then, is to use the findings of neuroscience and psychology (and, one hopes, in some cases, to inspire new lines of investigation) to reveal how certain kinds of informational states associated with “subjectivity” and “agency” are realized by neuronal processes. We will need to understand how certain subsets of information gleaned by the central-nervous system become “value-laden,” so to speak, and how they make themselves globally available for control of behavior, introspective awareness, and (in some cases) communication or verbal report.

**b. An Argument that a Complete Physical Description Leaves something “Left Over”**

Some of the preliminary remarks in the previous section about the causal relevance of the psychological may give the impression that this thesis will turn out to defend some version of traditional, libertarian “free will,” or that there is some irreducible “mind-level” to reality that is not encompassed by a complete physical description of the environmentally embedded organism. This is not the case. In fact, the view I will eventually be expounding will cast doubt on the idea that talk of discrete “levels” to physical reality is ontologically well-motivated. My account is, at its core, a physicalist one. My aim will be to show that it is possible to remain responsibly naturalistic by adhering to the physicalist principle of causal closure (PCC) while still avoiding the pitfalls of eliminativism, epiphenomenalism, or strong reductionism.

The idea that naturalistic philosophers must choose between one of these last three options is best exemplified by “exclusion” arguments of the type developed in its clearest form by philosopher Jaegwon Kim, and, as such, Kim will be a sustained target of criticism in an upcoming chapter. By denying Kim’s fork, however, in which we must choose either to accept the ontological primacy of theoretical entities described by fundamental physics *or* consign

ourselves to epiphenomenalism, I will not be defending a view of mental “entites” which exist either apart from, or over and beyond the physical world. Rather, I will be dissecting the intuitions that motivate Kim’s version of the exclusion argument in an attempt to convince the reader that the choice it presents us with is a false one, based on an unjustifiably limited conception of the physical world. Like many of the arguments presented in this thesis, it will be my contention that the axioms Kim brings to the table are, although ostensibly motivated by the findings of empirical science, not a particularly good match with the best current data. Far from it – on closer inspection, I will argue, Kim’s premises turn out to rest primarily on dubious intuitions about the way the metaphysics of reality *simply must* work. But if there’s anything we’ve learned by way of the history of naturalism, it’s that intuitions about how things “simply must work” should be regarded with suspicion. By undermining the intuitional basis on which his argument rests, I hope to shift the burden of proof back onto the causal exclusionists, and to suggest empirical routes by which we might confirm or disconfirm the existence of patterns of activity that correspond to our subjective acts of “willing” and “intending.” With any luck, this will demonstrate that there is another way to understand the physical processes that underlie our use of folk psychological terms that does not divest them of their causal relevance.

### **c. A Case of “Quantum Flapdoodle<sup>1</sup>”**

In a later section of this paper meant to cast doubt on the empirical validity of a substance ontological approach to philosophy of mind, I will be presenting arguments to the effect that a contemporary survey of the findings of particle physics should urge us towards a process ontology, and a form of ontic structural realism. Many responsible, naturalistically minded

---

<sup>1</sup> If I may borrow a phrase from the late, great Murray Gell-Mann.

thinkers may be inspired to dismiss my conclusions here, being all too familiar with the spurious (and sometimes downright mystical) theories of consciousness usually offered up by those who believe quantum mechanics should play a role in our framing the problem. Perhaps chief among those proffering such notions is Roger Penrose, whose theory of “Orchestrated Objective Reduction” (Orch-OR) has penetrated deep into the public discourse. This is almost certainly due to his renown as a mathematician, and to his frequent collaboration with no less than superstar scientist Stephen Hawking. It is also highly likely that it owes some of its popularity to its connection with “new age” ideology. The notion that “consciousness” somehow permeates the cosmos as the realization of a kind of “universal mind” lends comfort to the Deepak Chopra types who would want to find out that their personal, subjective awareness can somehow come untethered from their body, opening the door to the possibility that they may survive their own death.

I have no such goals in mind. I will argue that our subjective awareness arises from organic processes, as well as from some of our “mind-talk” arising from the kinds of language games we play to describe that experience. While I will at one point defer to quantum mechanics in order to cast doubt on some of the fundamental assumptions of theorists like Kim, this is *not* to say that the indeterminism implied by quantum mechanics will rescue mental causation. Whether or not there are any quantum processes involved is a matter we can remain agnostic about for the moment, even with respect to the question of “free will” – for as others have extensively noted, random quantum fluctuations do not make for any kind of “freedom” worth wanting, in any case. For the purposes of this thesis I will be treating the biological processes that underlie subjective awareness as deterministic. Though it may turn out through further empirical enquiry that this is not strictly true, the arguments I am making should hold, even if we

should find that certain neuronal firings depend, for instance, on pairs of entangled electrons in dendritic tissue. Simply put, I do not believe that the question of universal determinism has any bearing on the question of agency or subjective awareness. It is a questionable leap indeed to conclude from the fact that some brain processes might utilize quantum mechanical processes in their functional architecture to the idea that there is somehow, then, a ghostly inner “observer” in control of the outcome of these randomized events. Even if the brain operates in a wholly probabilistic fashion, we would be left in exactly the same place with regard to the question as to how “we are in control” of our thoughts and actions. As Robert Kane put it: “If conscious willing is illusory or epiphenomenalism is true, *all* accounts of free will go down, *compatibilist and incompatibilist.*” (2005)

Why, then, invoke quantum mechanics, if not to rescue a notion of human specialness, or to re-inject some quasi-mystical sense of the divine into the question of agency and subjective awareness? This subject will be explored more deeply in a forthcoming chapter, but for now, it is enough to say that such arguments are designed to cast serious doubt on the substance-ontological intuitions upon which the arguments of eliminativists, epiphenomenalists, and exclusionists rest. In short, the premise that Jaegwon Kim calls “causal fundamentalism” loses its force *if* we can convincingly argue that three conclusions follow from taking the findings of fundamental physics seriously: (1) There is no sufficient empirical basis for treating the putative entities of fundamental physics as the only “real” ones, and the entities of the special sciences as somehow illusory, unnecessary, or simple aggregates of these “real entities;” (2) A thoroughgoing commitment to a process ontology (which I maintain is motivated in part by taking the findings of quantum mechanics seriously) commits us to a version of structural realism that undermines the notion that we should regard *any* of the putative entities involved as

objects bearing properties (such as “causal power,”); and (3) Jaegwon Kim’s supposed deference to fundamental physics is exposed as sloppy from a naturalistic point of view, as he imports into the world of the atoms and molecules a set of inappropriate intuitions about the way entities must interact. More specifically, he seems to treat entities on the level of fundamental physics as if they still followed the laws of Newtonian mechanics. Not only does he privilege the entities at the fundamental level as the only ones being “really real,” but he still treats them as though they were little billiard balls clacking into each other, reducing all causation to a set of tiny “microbangings.” As we will see, this stowaway intuition doesn’t accord well with our current understanding of physical science at all – and, therefore, conclusions drawn with its aid should be regarded with extreme skepticism.

We can see, then, that the endgame is not to show that somehow, globally effective complexes of central nervous system processes can somehow go “this way or that” willy nilly, like the photon emission that leaves the fate of Schrödinger’s cat indeterminate. Rather, it is to argue that the exclusion arguments which purport to show these complexes have *no* causal relevance rest on a set of insufficiently checked intuitions. Though these intuitions *seem* to be drawn from a scientific worldview, such intuitions have no place on the scale Kim is discussing, and which he privileges as causally “fundamental.” These are “stowaway intuitions” – drawn from the scale of everyday objects and applied in a way which is inappropriate to the putative entities being discussed. As we shall see, all of the standard arguments meant to support hard determinism are similarly reliant on such *seemingly* naturalistic principles which turn out, on closer inspection, to be little more than hunches drawn from everyday experience. If we want to gain an appreciation for how our nervous systems realize our conscious deliberations, and how those deliberations come to determine the behaviors we engage in, we will have to leave such a

priori metaphysics aside. To be *real* naturalists, we're going to have to take a much more empirical approach – one which is much more careful about what lessons we can responsibly draw from contemporary physical science. Before we can begin building our positive account, however, we're going to have to take a look at a distressingly common way the problem of free will is framed among naturalists – an issue we turn to in the next section.

## **Chapter Two: Hard Determinism and Incompatibilism – Why Naturalism Seems to Motivate Skepticism about Self-Control**

How do living brains generate behavior? Parts of the story have been clarified over the last few decades by progress in the neurosciences. We now understand a great deal about the molecular structure of the cells that make up our brains. We understand many of the chemical pathways by which they communicate, how certain signals propagate, and even in some cases how this can be used to explain how sensory information is processed, or how muscle movements are actuated. We have learned that certain areas of the brain seem differentially dedicated to certain tasks, with some areas doing the heavy lifting during speech generation, for instance, and others working hardest when we're trying to make fine discriminations between visually similar objects. We've also learned the role that certain chemicals play in effecting more globalized changes, such as gamma-aminobutyric acid and dopamine.

For materialists, the idea that mental activity is underpinned by complex physical mechanisms should have come as no surprise. Even Descartes and Hobbes both understood well that the body is full of mechanisms of one sort or another, and that sufficiently intricate mechanisms can give rise to surprisingly sophisticated behaviors. While Descartes may not have been prepared to accept that his mind was being generated by a mechanism, his descriptions of the working of the heart, his intense scrutiny of the pineal gland, and his talk of automata in Part V of the *Discourse on Method* belie his great respect for what a mechanism can accomplish – and he was a dualist! Naturally, then, for those philosophers throughout history who have found dualistic arguments too flawed and implausible to countenance, the idea that conscious experience is generated with the aid of mechanisms within the physical organism has been the

most plausible speculative explanation - one that got a big boost, understandably, as we came to better understand how physical mechanisms could carry out computations.

The framework for understanding conscious experience in terms of mechanisms of some sort, then, both precedes neuroscience and is bolstered by it. We understand with great enough clarity how some of these mechanisms work, and the kinds of experience they generate, to conclude that the neural activity thus facilitated is at the very least a necessary condition for generating the kind of experience we have. Given that our brains seem to be beholden to the same laws of physics that hold everywhere else in the universe, it is parsimonious to conclude that the tangle of electrical and chemical activity in our nervous systems is also sufficient for generating consciousness. The role and potency of mental mechanisms is too well-established to provide warrant for additional metaphysical presuppositions, such as the positing of new fundamental forces of nature, or of immaterial entities intervening in the physical order through some unknown means. All indications are that the tools for providing a satisfactory account of consciousness are at our fingertips.

When I decided that taking a neuroscience course would be a helpful cross-disciplinary exercise in the course of getting my masters degree in philosophy, I was overjoyed to find myself in the classroom with a group of relatively sober thinkers about the mind. Quite unlike some of the philosophy classrooms I was used to, *every* student in my neuroscience course was a thoroughgoing physicalist – they all believed that our conscious experience was generated by material processes in the brain. Some of them were even savvy enough to attribute some of what we call our “consciousness” to environmental and cultural factors – an idea to which I am sympathetic, and to which we will return in due course. None of them believed it was necessary to invoke strange, non-material substances to explain the subjective, first-person aspect of our

lives, nor did they believe that it would require the discovery of new physical forces and interactions that up to now have remained undetected by science. I felt I was among like-minded peers: no one was invoking any kind of magic or spookiness to account for this last bastion of mystery for philosophers of science – “the Mind.”

As I came to know many of them better, however, a distressing trend emerged. All of them, at least in *my* class, seemed to believe some version of the following propositions:

(1) Since consciousness is realized by material processes, and these processes are realized “bottom up” by matter adhering to the deterministic laws of chemistry and physics, our experience of free will must be an illusion. Our conscious “deliberations” and “desires” don’t determine our behavior – particles in motion do.

(2) Since the physical world is causally closed and dualism is false, our behaviors are connected by an unbroken causal chain back to the big bang itself. Our behaviors aren’t just determined by particles in motion – they were in a sense already “destined to happen” by causal processes operating before we were born.

I suspect that the slightly different lines of reasoning expressed by (1) and (2) were not at all unique to the particular neuroscience department of which I was briefly a part, but rather express a belief held by a large number of scientists in this area – whether or not they choose to say so to the world at large. Both positions are expressions of what philosophers call *incompatibilism* – the idea that if the physical world does indeed adhere to law-like, deterministic principles –and if that world is “causally closed” (which, for the moment at least allows us to remain agnostic about spooky non-material forces, as it simply says no such entity or force would be able to push particles around) then free will must be an illusion. In other

words, if there's nothing more to our "consciousness" than matter in motion, then we're really nothing more than puppets, destined to behave in ways that were fated long before we were born. The idea of determinism can be put more precisely than the two "common-sense" formulations above:

Determinism: The world is governed by (or is under the sway of) determinism if and only if, given a specified way things are at a time  $t$ , the way things go thereafter is fixed as a matter of natural law. (Butterfield, 1998)

This formulation brings out the issue clearly. We can put aside for a moment the fact that our current best theories tell us that matter on very small scales behaves indeterministically, for reasons we explore in depth later. For now, it suffices to say that on the scale of ordinary matter, the laws of physics *do* appear to operate according to laws that are deterministic, meaning that at any given moment in the causal history of the universe, there is exactly one and only one possible future. Even if quantum processes are involved, then, in the causal chains in our central nervous systems that lead to expressions of behavior, this merely adds *randomness* to the system on its smallest scale – and it seems hard to understand how *random* or *probabilistic* physical events could salvage a notion of our causal autonomy any better than deterministic physical events.

Worries about determinism have a long and storied history within the philosophical tradition. Perhaps the most famous formulation comes from Pierre-Simon Laplace, whose worries clearly stem from the idea that everything has a discrete, determinate cause:

We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all

things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes. The perfection that the human mind has been able to give to astronomy affords but a feeble outline of such an intelligence. (1820)

Laplace, reasoning about the nature of causality and ruling out *a priori* the incoherent notion of an uncaused cause or a *causa sui*, a thing which causes itself, arrived at what seems a natural conclusion: a being given complete knowledge about the positions of every particle in the universe at a particular time, as well as the causal interconnections that governed their motions, would be capable of deducing both all of their former states, as well as all of their future states. If this were true, then our formal definition of determinism presented above would hold – only one future would be possible, thus ruling out the idea that the “choices” we apparently make are real. They would only *seem* to be real, since the line of reasoning pursued above seems to rule out the idea that actual alternative possibilities can be open to us – after all, how can we have a choice between genuine alternatives in a universe that moves headlong in a career towards a single, inevitable future?

While Laplace is sometimes given the credit for being the first to articulate this line of reasoning concerning the potentially illusory nature of our volitional “freedom,” Spinoza in fact came to much the same conclusion almost two centuries earlier. In his *Ethics*, Spinoza dismisses the idea of free volition using a very similar argument about the necessity of causes: “...each volition can neither exist nor be determined to produce an effect unless it is determined by another cause, and this cause again by another, and so on, to infinity.”(1677) Like Laplace, Spinoza came to the conclusion that any apparent “contingency” (or, it follows, any volitional choice,) is an illusion produced by our lack of complete knowledge about the causal order. We

only think we are “free” to act in the ways dictated by our deliberations, desires, and intentions because “the order of causes is hidden from us,” and this “defect of our knowledge” leads us to declare something to be contingent.

Many, many philosophers have followed Spinoza and Laplace’s lead since, and we will eventually examine some of the more contemporary arguments in detail. The problem should at least, for the time being, be coming into focus: all effects necessarily have a cause, and the suspension of the causal order would entail some kind of miracle. Then, it seems, we are determined to act as we do by causes which stretch back through history to the very beginning of the causal order itself (the Big Bang, presumably) and, as Laplace and Spinoza concluded, volition is an illusion. Either that, or acting as little Gods unto ourselves, we’re capable of popping ourselves out of the causal order through a wave of our Magic Will – each little decision point in our lives constituting an unexplainable miracle that violates the principles of causality itself. The former of these incompatibilist positions (referring to the incompatibility between the deterministic story outlined above and our subjectively experienced acts of willing,) is often referred to as Hard Determinism. The latter, by which organic beings like ourselves somehow exempt ourselves from the causal order, is called Libertarianism by philosophers. “Take your pick!” the incompatibilist invites us. “Either the causal order holds, or your feelings of liberty are an artifact of your incomplete knowledge of the causal forces that move you.” In order to have genuine “freedom,” so the incompatibilist story goes, we would need some Principle of Alternative Possibilities, (hereafter PAP,) as Harry Frankfurt put it in 1969. We would need some mechanism by which *real* alternative possibilities were *really* open to us – and some way by which we could choose to go either this way or that by an act of will.

It should now be unsurprising that the Hard Determinist prong of this incompatibilist fork is the one most commonly grasped by neuroscientists, if they bother thinking about the problem of the Magic Will at all. After all, their entire profession is devoted to studying the underlying causes of our subjective experience at various levels of granularity. The very purpose of the science is to understand, to as great an extent as is possible, the underlying causes of mental and behavioral episodes in terms of physiochemical interactions. The very act of giving a neuroscientific explanation forces us to leave talk of irreducible mental entities behind, so we can talk about their underpinnings in deterministic physical processes. Arriving at the incompatibilist conclusion is a natural result of pursuing, in much the same way curious children do, a chain of “why?” questions back to the point at which explanation comes to an end. One begins with a simple question such as: “What caused me to raise my arm?” One can provide an answer to this behavioral question on the level of psychology, by saying, for instance, “Well, you had an intention to raise your arm, so you raised it.” I’ll argue later on that there is an important sense in which this answer is correct, but somehow it doesn’t feel very satisfying. Our hypothetical curious child would surely furrow his brow and stamp his feet. “Well, I know that,” she might say. “But HOW did it happen?” (We’ll imagine for the time being that we have a very philosophical youngster on our hands.)

It’s easy to see that both from the perspective of sheer curiosity into the nature of things, as well as from the explanatory perspective of the sciences, we’re going to need to descend down a “level of organization” (a concept which has its epistemological uses but about which I’ll express ontological skepticism in due time,) to provide a more satisfying answer. “Well,” we might respond, “neural signals to release acetylcholine were sent from your brain down your arm in a cascade of reactions involving many complexes of neurons in between. When it finally

reached the axon end-plates of your motor neurons, acetylcholine was released. This caused the neurons to fire that are responsible for contracting the muscles that raise your arm.”

This answer seems a bit more satisfying. It sounds very official and sciencey, and we have something like a mechanical story explaining to us how we got from activity in the brain to an arm raising. Anyone who’s spent an appreciable time among young children, however, knows that this isn’t where the Chain of Whys ends. “Okay,” our neuroscience-savvy but endlessly curious tyke might continue, “*why* were signals to release acetylcholine sent from my brain down my arm? How did that happen?” If the previous step in our regression back down the explanatory chain didn’t begin to squeeze our intuitions about what we’re talking about when we speak of freedom of the will, this next step surely does. Once again, to provide a satisfactory explanation, we have to progress down another “level” of organization, crossing the boundary now from talk about neurons and the tiny lock-and-key molecular structures that activate them to the level of physical particles: ions of sodium and potassium, the diffusion of which down gradients within the neuron determine where action potentials in the brain begin to fire. Populations of neurons fire due to voltage spikes generated when sodium and potassium “gates” in their cell walls allow a critical threshold of sodium ions to enter the cell. Once this critical threshold is reached, a sudden burst of electrical activity occurs reflexively, as a consequence of the mechanical architecture of the cell itself.

“But *why?*”

Could it be the case that what we call “agency” then, through some sort of quantum process, or through the kind of particle “swerving” effect imagined by Lucretius, amounts to

some mind-stuff in our brains directing the flows of sodium and potassium across cell walls in appropriate areas of cortex?

Unfortunately for a certain notion of what free will would entail, there's no evidence that such an answer can be given. The diffusion of sodium and potassium ions across gradients in the cell appear to operate on fully deterministic physical laws – mostly those of electromagnetism. Of course, since we want our account of mental causation to be naturalistic, it wouldn't be right to invoke such a "Swervy" account of agency, as this would violate the Principle of Causal Closure (PCC.) Such a result would take us out of the realm of material sciences and into the realm of mysticism, and so should be avoided at all costs. Our Chain of Whys, then, finally reaches a dead-end, and one that motivates naturalistic denials of free will: Why do sodium and potassium diffuse in the way they do, at the times they do? Due to the action of Fundamental Forces of nature.<sup>2</sup> Who's really in charge, then? Is it some entity called "me" which vanishes as we regress down the explanatory chain of Whys? Or is it nature itself, as Spinoza and LaPlace imagined?

Our actions, then, appear to be explainable in terms of fundamental forces of nature. Via the Principle of Causal Closure, we're not permitted to imagine some magical, quasi-divine intervention on the part of a ghost inside the machine. No wonder then, that neuroscientists conclude that there is no magic, and hence, no will at all. After all, if our thoughts and behaviors are realized by deterministic mechanisms, wouldn't this make us into a kind of biological robot? If my deliberation over an important decision is really the outcome of a computational procedure, set in motion by the opening and closing of sodium and potassium gates in my

---

<sup>2</sup> This, incidentally, appears to be where all explanation in science comes to an end.

neurons and propagated by a deterministic electrical cascade, how can I claim that “I” am in control of what decision I come to? Isn’t the electrochemical process simply going to do its thing, *bypassing* what I experience as “me?” But in order for there to be a sensible naturalistic story about self-control and agency, perhaps we’re making a mistake by thinking of the will in terms of the Magic Will, which can Swerve its way into suspending the laws of causality.

The position I have been describing as the “Magic Will” is often called, within the confines of the traditional debate about free will, the “libertarian” position. It is, like the position advocated by the hard determinist, an incompatibilist position. It simply grasps the other horn of the dilemma: instead of affirming the truth of determinism and concluding this precludes the possibility of free action, it embraces the supposedly certain, subjective knowledge we have of our own moment-to-moment freedom to do what we please, and concludes on this basis that determinism must be false. It must be – for how else would it be possible for me to choose either to raise my arm at this moment or to not raise it? In order for there to be a PAP, my choices must constitute legitimate forks in the causal history of the universe. Whether my material body goes this way or that is truly *up to me*, up until the last moment, at which time my act of volition intervenes into the causal order, presumably causing particles to swerve in their career as a result.

It’s clear why this strategy is rejected by neuroscientists, and should be rejected by naturalistic philosophers as well. Such violation of causal closure would amount to a miracle, and such “God of the gaps” reasoning is anathema to the entire project of science. If we’re going to mount a defense for the idea of self-control, against the epiphenomenalists, whatever we mean by “will” mustn’t rely on magic. Our task is now to develop a framework for thinking about causality that avoids the conclusion urged by following back the Chain of Whys, and provides a

naturalistic account of what it means for us to have self-control that doesn't rely on a notion of the Magic Will or violate the PCC.

Before we can set about developing such a framework, however, we have to lay some groundwork. What is in question is how it could be possible that what we call our mental states could have causal efficacy. Having already ousted the libertarian option as not scientifically tenable (due to its violation of the PCC,) we must make due with a notion of the causal relevance of our mental states that works *within* the limitations set by deterministic concerns. There mustn't be any Magic Wills, nor any swerving particles. As romantic as Sartre's notion of freedom is, we must accept that whatever agency and self-control involve, it is *not* a break with our entire causal history. Naturalism demands that we recognize we are beings who are thoroughly causally enmeshed – we are organisms in dynamic interaction with our surrounding environment, causally determined not only by the fundamental laws of physics, but by our genetics, our experiential history, environmental cues, and synaptic makeup. The fact that we are so enmeshed, however need not entail that what we experience as our mental states are somehow *bypassed* in the series of events that lead to our behavior. To see how this could be so, some talk about causation itself is in order. Before we can begin laying a preliminary physical and metaphysical account of how what I will be calling “self-control” is possible, it will be worth our time to remind ourselves of the limitations of the empirical project, and the danger of engaging in a-priori, analytic metaphysical speculation under the banner of “naturalism.”

### 3. From Hume to Woodward – What Instrumentalists about Causation Get Right

While this thesis will eventually engage in some physical and metaphysical speculation about the material underpinnings of mental causation, my position has a surprising amount of overlap with broadly anti-metaphysical accounts of causation of the sort advocated by Hume. This may sound paradoxical at first. How can one point to Hume as a forebear and then proceed to make claims about the nature of the causal “tie” that he argued every responsible empiricist should abjure?

In the section that follows, my aim will be to show that Hume, and those who take a broadly Humean approach to causation, have gotten a number of things right. Namely, I will be praising them for advocating the view that:

- (1) No a-priori argument from analytic metaphysics should be taken seriously as overturning inferences about causal relationships drawn from empirically observed regularities.
- (2) Talk about causes is *relational* – there is no empirical warrant for talk about entities as “endowed” with “causal properties.”
- (3) We should take seriously a counterfactual or manipulationist account of causation, in which only empirical regularities can establish warrant for assenting to causal claims.
- (4) Nothing in the physical sciences gives us empirical warrant for reducing all “real” causation to the level of fundamental physics.

Where I depart (much to the Humean’s dismay, I’m sure) from the instrumentalist picture can be summed up simply for the time being. Far from being in conflict with a consistent interpretation

of the natural sciences, I believe that a careful, up-to-date understanding of those sciences actually provides a plausible underpinning for the Humean conviction that, from a strictly empirical point of view, “any thing may [possibly] cause any thing” – that is, we have no basis for making *a priori* claims about what kinds of things are precluded from causing other things. As I will argue in section 4, arguments from analytic metaphysics that attempt to establish the opposite conclusion, such as Jaegwon Kim’s “exclusion argument,” rely on intuitional premises that fall apart under careful scrutiny by the thoroughgoing naturalist. Before I venture into that metaphysical thicket, however, it’s important to establish exactly why it is that *no* such intuitions have a legitimate place in metaphysics, and observed empirical regularities should be the primary arbiters of any ontological or nomological claim. While I will take some time to discuss modern takes on an instrumentalist approach to a broadly naturalistic philosophy, and to causation in particular, Hume provides the best place to begin analyzing such an approach. His simple arguments about the futility of searching for *a-priori* epistemological “grounding” provide an excellent in-road to understanding why we have no good reason to grant the traditional arguments meant to support epiphenomenalism about the mental, and what responsible naturalistic philosophy entails.

*a. Hume on Induction and Causal Powers*

One of the greatest interpretive difficulties in Hume’s philosophy consists in the attempt to reconcile two main threads of his philosophy that seem to run in opposite directions: the broadly naturalistic thrust of his accounts of mind, morality, and society on one hand, and his highly destructive arguments concerning induction, causal power, and external objects on the other. How are we to understand Hume’s commitment to scientific methodology in light of the arguments he advances that seem to undercut its justification?

There is ample evidence in Hume's *Treatise* and *Enquiries* to show that he feels his skepticism and naturalism can live peacefully together. In fact, he seems to think they mutually support one another. Some of the best evidence for this can be found in those passages where Hume speaks of "true skepticism." It is here, I contend, that we discover the real purpose of Hume's skeptical arguments. Far from undercutting the merits of the naturalist/empiricist project, these arguments are meant to encourage us to hew as closely as possible to the results of controlled experiments and careful, disinterested observation. Hume, then, is not a "skeptic" of the sort Descartes had in mind when he framed his representational account of the mind, and thereby opened the door to skepticism about the existence of the external world. Rather, Hume is what I will call a "proto-skeptic" of another species altogether, anticipating the so-called "skeptical" movement championed by public figures like James Randi, Michael Shermer, and Daniel Dennett. The message advanced by Hume is of a piece with theirs: beware of conclusions arrived at by the light of one's own reason. Distrust your metaphysical intuitions. Defer, whenever possible, to the evidence, and draw responsible conclusions from those things that can be measured, tested, and observed. Where that evidence contradicts our most deeply held intuitions and rigorously demonstrated philosophical convictions, it's the intuitions and convictions that are obliged to bend first.

Before we can understand how Hume's skeptical arguments are meant to support a thoroughgoing commitment to experimentally informed naturalistic reasoning, we need to examine the arguments themselves, and see what conclusions Hume purports to draw from them. Let's begin with the supposed "problem of induction" – his account of causal reasoning which denies we can ever have "ultimate" justification for making causal inferences. In section seven of his *Enquiry Concerning Human Understanding*, Hume introduces his copy principle: "All our

simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly represent" (E 11). Continuing in the empiricist tradition, Hume claims that anything we claim to have knowledge of whatsoever must first be derived from some interaction with the world, or with our bodies. This allows for the idea that we can derive concepts from our dealings with objects in our environment, (what he calls "impressions of sensation") as well as from the exercise of our bodies and familiarity with our internal states (what he calls "impressions of reflection.") What it rules out, accordingly, is the idea that we can derive ideas from a mysterious intuitional wellspring called "reason" that would somehow afford us direct insight into the nature of things unaided by anything in our experience. (The analytic metaphysician's perennially ill-explained allergy to what is called "overdetermination" provides an excellent modern-day example of an idea of dubious origin.)

Wielding his "copy principle," Hume launches an investigation into the possible origins of our notion of "force, power, or necessity." It stands to reason that if we are justified in making claims such as "every effect must have a single efficient cause," or "all *real* causation must take place at the micro-level in the form of particle collisions," then we must have some direct empirical acquaintance with causes. After an exhaustive search, however, Hume famously concludes that we do not: all we experience is *constant conjunction* or one event following reliably from another, and from this we are accustomed to call the first event the "cause" and the latter an "effect." When we see one billiard ball hit another, for instance, we do not actually *see* this thing called the "causing," nor could we have guessed, prior to all experience with the behavior of objects, what the result of such a collision would have been.<sup>3</sup> Nor do we gain a

---

<sup>3</sup> It bears mentioning here that despite 300 years of scientific progress, Hume's argument still stands. Even with such high-powered instruments such as the Large Hadron collider, we still have no way of "seeing causation" itself, such that we could assign entities "causal powers" or make sweeping nomological generalizations. By means of data analysis, we can come to see patterns in events such that we become justified in saying what things cause

clear idea of causation through acquaintance with our own volition since, as the need for this thesis demonstrates, there is nothing at all transparent about how my desire to raise my arm causes my arm to raise – after all, my purpose for writing is to refute those who would say the desiring doesn't cause the raising at all!

So Hume searches high and low for the impression from which our notion of “cause” must be copied if it's to be a notion about which we're qualified to speak, and he comes up empty (T 89, 92, 97, 158-164). At this point, he concludes that, lacking a God-given faculty of Reason that would deliver metaphysical truth to us through some non-empirical means, we must not have a clear notion of force, power, or necessity at all. In fact, he claims the notion “entirely incomprehensible” (E 72). By the light of our experience, then, we can only discover that “anything may appear able to produce anything” (E 164) – we never gain insight into nature so as to perceive a necessary “tie” between distinct events in the world, and therefore understand the *nature* of any causal connection underlying our inference. We therefore have no justification for placing any a priori restraints on how causes must behave, what things are metaphysically capable of causing other things, and what entities do or do not contain “causal powers” (which calls back to points (1) and (2) introduced at the outset of this chapter.)

It apparently gets worse, however. Inductive inferences themselves, according to Hume, can't be grounded in the sense sought-after by the rationalist philosophers of his day (or by the analytic metaphysicians of today.) Making causal claims based on induction already assumes that events in the past are a meaningful guide to events in the future. Hume calls this assumption the “uniformity principle,” since it presumes the truth of the premise that events in the future will exhibit a uniformity with observed events in the past. While we may not get any deep

---

what – but this is not due to any deeper understanding about what “causality” itself is than there was when Hume wrote his *Treatise*.

metaphysical insight into nature's "ultimate springs and principles" by observing sequences of cause and effect, we can at least justify our causal inferences by recourse to such the principle of uniformity, can't we? Hume seems to suggest we can't. After all, nothing in logic seems to preclude the possibility of the laws of nature suddenly changing. The entire universe might undergo a violent upheaval, perhaps as a result of a collision with another universe in the quantum "foam" contemporary physicists find fashionable to invoke. The uniformity principle is just too "rock bottom" in our ideology. No logical principle can demonstrate its truth, and no amount of empirical data is ever sufficient to verify it.

It shouldn't come as a surprise that Hume denies the principle can be known to be true *a priori* – "there can be no demonstrative arguments to prove that those instances, of which we have had no experience, resemble those, of which we have had experience" (T 89). What may be more surprising is that Hume also denies we can justify the principle of uniformity using probabilistic reasoning. Probabilistic reasoning, after all, is based on causal inferences we make by way of recourse to past experience, so using this technique to justify the principle of uniformity itself would be circular – we'd be assuming the very thing we're trying to prove. No argument, then, can never provide a firm foundation for inductive reasoning, "tho' aided by experience, and the observation of their constant conjunction in all past inferences" (T 92). How is it possible for a naturalist to express such grave doubts about whether we're justified in making causal inferences, given that such inferences are the keystone of the scientific enterprise?

There are several clues within Hume's writing that can help us begin to answer this question. Firstly, it's important to note that after presenting this ostensibly devastating argument against causal reasoning, Hume goes on to make use of causal reasoning throughout the rest of the treatise. He even details "rules to judge of causes and effects," later in part three, which

appears to be a straightforward guide to determining when causal inferences are justified. If causal inferences are, at bottom, justifiable neither by reason nor by anything in our experience, why does Hume not abandon the practice?

The answer can be found in Hume's discussion of the *real* origins of our causal reasoning. There are many places where Hume alludes to the actual basis for our judgments concerning cause and effect, but perhaps one of the most revealing can be found in the section of the Treatise "Of the reason of animals." In the final passage of the section, Hume reveals that he believes "reason is nothing but a wonderful and unintelligible *instinct* in our souls" [emphasis mine]. He goes on: "This instinct, 'tis true, arises from past observation and experience; but can any one give the ultimate reason, why past experience and observation produces such an effect...nay, habit is nothing but one of the principles of nature..." (T 179). Here, we see that far from undermining the basis for his naturalism, Hume's skepticism *grows out of* his naturalism – in destroying the notion that we can rely on a faculty of reason to give us an a priori basis for metaphysics, Hume leaves naturalistic empiricism as the only method we have for making reliable statements about the world. We haven't been endowed with any transcendent faculty for gaining unmediated access to reality-in-itself, so we should expect that our epistemology terminates in ungrounded assumptions, rather than demonstrable truths. If reason were, for instance, a God-given ability, that transcendent standard might afford us a way to firmly ground its operations. If reason were perfect, we may be able to find that it operates on rock-solid logical foundations which could be discovered *a priori* – THIS could provide the kind of epistemological certainty Descartes so badly wanted.

If, however, it's merely a glorified animal faculty - if it's nothing more than an instinctual reflex organisms developed over time "for their own preservation, and for the propagation of

their species,” then we shouldn’t expect that the beliefs it engenders should be perfect, logically grounded, or even fully intelligible (T 177). We should only expect that we’ll hold beliefs that have proven useful in the past – and that there’s no additional justification for the operations of the understanding to be had. The beliefs our reason affords us are “more properly an act of the sensitive, than of the cogitative part of our natures” (T 183). This being the case, we cannot simply trust our reason to afford us direct insight into the nature of reality. Instead, we must allow the data itself to determine our conclusions, using the scientific tools of observation, experiment, and probabilistic reasoning. We have recourse to nothing more effective than these modest implements.

There turns out, then, to be a threefold basis for falling back on causal inferences drawn from the data of past experience, given the skeptical argument as laid out by Hume. *First*, we have no choice. The habit is reflexive and automatic, and cannot be subverted by philosophical arguments. Before long, “nature breaks the force” any such arguments may have over us (T 187). *Second*, while the principle of uniformity is not ultimately justifiable by reason alone, there is also nothing in our experience that gives us reason to doubt it. We assume the principle of uniformity every day, even to perform such basic tasks as driving a car, going to the store, and feeding ourselves. Hume doesn’t even seem to think the principle of uniformity *needs* to be rigorously grounded, since we’re bound by nature to reason by way of it in any case. In fact, the overall anti-rationalist thrust of the arguments in the Treatise should suggest to us that he does not consider grounding a requisite condition for making reasonable epistemological claims.

Assuming the principle of uniformity in our more abstract reflections, then, is only a matter of consistency, even though our most “strained metaphysical conviction” can’t seem to permanently fix its foundation (T 214). *Thirdly*, given that reason is an imperfect, merely animal

faculty, working from an empirical basis, *we shouldn't expect there to be philosophical certainty about much of anything*. Reason is no supernatural endowment that allows us to penetrate reality itself, and uncover its deep metaphysical secrets. Instead, it is an fallible animal reflex that, “when it acts alone, and according to its most general principles, entirely subverts itself, and leaves not the lowest degree of evidence in any proposition” (T 267). Given such unreliable mental apparatus, causal reasoning based on evidence is *the best and only investigative tool at our disposal*, uncertain as it may be. The fact that it cannot be propped up by an airtight philosophical argument should be unsurprising once we adopt this proto-evolutionary perspective – perhaps only a God-given faculty of reality-penetration would make such ultimate grounding possible. The lack of such ultimate ground, then, should only prove troubling to those who imagined that the source of human understanding was fundamentally different – more divine than down-to-earth.

Hume, then, is giving a naturalistic account of man's powers of understanding as those of a souped-up animal, not a miniature God. This puts him far ahead of his time, and understandably, led some of his first academic interpreters to completely miss his point. Once we understand, however, that Hume's intention is not to plunge us into doubt but to pull us down to earth, the purpose of the foregoing arguments snaps into focus. Their ultimate upshot is similar to the arguments presented by Wittgenstein in *On Certainty* and carried further by Rorty in *Philosophy and the Mirror of Nature*: it simply isn't possible to secure any philosophical notion on an “ultimate” ground. At one point, one simply has to commit to a set of assumptions and see where they lead – “investigation comes to an end.” Continued and focused reflection on any subject of knowledge, then, far from producing the grounds for certainty in any area of epistemology, will continually erode our certainty, reveal our fallibility, and expose the (for

some) painful fact that none of our knowledge claims can be placed on an unimpeachable, indubitable foundation. Such fallibility is no reason to fall into external-world skepticism and despair, however. It's a blow to our ego we must accept. We must dust ourselves off and get back to the business of investigating the natural world, always with one eye on our tendency to make mistakes. What's more, we can do so with a clearer idea of the *role* of abstruse philosophical arguments. Having abandoned the notion that they reveal deep metaphysical truths about the universe, we can see them instead as revealing the "whimsical condition of mankind" – the very same rational faculties that make the world intelligible to us also subvert themselves when trying to formalize and ground that intelligibility (E 160). Put more pithily, the more rigorously one reasons, the more one discovers the infirmity of reason. Ironic indeed.

Sentiments of just this sort are found throughout the last section of part four of the *Treatise*, and this is where he reveals to us what he thinks it means to be a "true" skeptic. Rather than privileging philosophical argumentation over the immediate evidence of our senses and instinctual inclinations, Hume champions exactly the opposite course of action. Given that such "strained metaphysical convictions" (T 214) always have their force eventually broken by our return to the course of everyday life, our best remedy is simply to return to everyday life, and turn our minds away from such "philosophical melancholy and delirium" (T 269). Skeptical arguments of the sort proffered by Hume in the preceding sections "admit of no answer" – but this is not reason to embrace them, for they also "produce no conviction" (E 160). Instead, returning from our sojourn into skepticism thoroughly convinced of the inadequacy of our mental apparatus, we are instead well advised to "yield to the current of nature, in submitting to [our] senses and understanding" (T 269).

Hume states in no uncertain terms, more than once, that in his “blind submission” to the naturalist project, he “shew[s] most perfectly [his] skeptical disposition and principles” (T 269.) Elsewhere, he states that far from succumbing to Pyrrhonian doubts, the “true sceptic will be diffident of his philosophical doubts, as well as of his philosophical conviction” (T 273). The message here, I argue, is clear. The appropriate response to skeptical arguments of the sort offered by Hume throughout Book 1, Part 4 is not to renounce belief in the external world, or in the reliability of our causal inferences. Rather, the appropriate response is to renounce the arrogant pretense that our analytic philosophical musings constitute the final say on epistemological claims, and that we shan’t rest until we find grounds for unqualified certainty.

Understood this way, Hume’s purpose in introducing skeptical arguments is of a piece not only with Wittgenstein’s *On Certainty*, but with his forebear Francis Bacon’s discussion of the four “Idols.” Specifically, we can see Hume as delivering the same warning Bacon did when he spoke of “idols of the theater” – those impediments to naturalistic enquiry “which have immigrated into men’s minds from the various dogmas of philosophies, and also from wrong laws of demonstration.” Where philosophy and the naturalist project are at loggerheads, in other words, it’s the philosophy that must be cast aside. The same goes for intuitions, chains of reasoning, and any other proposition which simply appears as self-evident to the understanding. Being “truly” skeptical in Hume’s sense means not trusting your own mind – and deferring to the relevant evidence, whether or not the evidence is intuitively satisfying. “There is another species of skepticism,” Hume tells us, “*consequent* to science and enquiry, when men are supposed to have discovered, either the absolute fallaciousness of their mental faculties, or their unfitness to reach any fixed determination in all those curious subjects of speculation, about which they are commonly employed” (E 150). Stick to the facts, and draw only those conclusions which follow

reliably from them. Any attempt to claim deeper epistemic access is “mere sophistry and illusion” (E 163).

In this way, Hume can be seen as a progenitor of sorts for a cultural movement that wouldn't come to be for centuries after his death. Fittingly enough, this is the so-called “skeptical” movement, founded by magician James Randi, and boasting such members as Richard Dawkins, Lawrence Krauss, Daniel Dennett, and Steven Pinker. The essence of this movement is very much in the spirit of Hume's “true skepticism” – the message such Skeptics propound to the general public is one that Hume would almost certainly agree with. Demand evidence for extraordinary claims, learn to interpret scientific results, and defer to them whenever possible -a lesson which I shall be taking to heart in the next few sections in suggesting a naturalistic basis for belief in mental causation. Our minds are so beset by biases, cognitive frailties, prejudices, and illusions that we are almost certain to be led astray when we trust our own *a-priori judgments* rather than the results of a well-controlled trial.

Alternative medicine provides a perfect example of how this epistemological attitude gets applied, today. One might have an experience, for instance, of continually finding that the duration of one's colds seems shorter when taking vitamin C supplements. Trusting in one's own direct experience and mental faculties, one might conclude that vitamin C reduces the duration of the common cold. However, there have also been several large, randomized, double-blind placebo controlled trials which have shown that vitamin C has no such effect. This, combined with knowledge of what's sometimes called “confirmation bias,” might lead a skeptic in the contemporary sense to conclude that his own experience was mistaken – it only *seemed as though* the vitamin C was what was reducing the duration of his colds. (I shall be returning to the placebo effect in due course.)

Where a careful consideration of all of the evidence contradicts the conclusions of our own understanding, we would be foolish to privilege our own understanding. Our own understanding just isn't that good – left to our own devices, we succumb to all manner of delusions, but the tools of science are just the remedy to guide us in the right direction.

I think Hume would find this worldview most agreeable, and would proudly call himself a “skeptic” in this contemporary sense. He was the first to recognize the naturalistic project as what Sagan called a “candle in the dark” – a lone beacon for lighting the path towards truth in a world that is otherwise continually subject to the “universal perplexity and confusion, which is inherent in human nature” (E 161). It is in this spirit that he offers his counterfactual definition of causation, which calls back to point (3) I introduced at the start of this chapter as part of my list of things for which the instrumentalists are to be praised: “We may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed.*” But how are we to determine whether a sequence of events fits this counterfactual criterion? This brings us to Woodward, who presents us with an updated framework for making practical causal inferences he calls “manipulationism.”

*b. Woodward, Nahmias, and Manipulationism – Inching Toward Causal Realism*

Hume’s account of causation is entirely pragmatic in that it is mum on any metaphysical questions about the mind-independent reality of causal connections. Hume neither affirms nor denies them. He thinks it leads one into meaningless and errant speculation to think about necessary connections that somehow *underlie* the inferences we make based on regularity of precedence, contiguity, and counterfactual observation – so it seems to follow from his remarks

that he must remain agnostic about the existence of “ultimate springs and principles” underlying our experiences of constant conjunction. He might not want to call it *causal* agnosticism, though, because, as he’s defined them, he *does* believe in causes. For him, however, we cannot separate meaningful talk about it from the mental construct we use for interpreting them. For this reason, we can call Hume’s brand of instrumentalism about causes a “Whole-HOG” instrumentalism – the extent to which we can remark on causal connections in the world is through a subjective framework that is Human Observation Governed. Though his account is immensely useful for shedding unnecessary metaphysical baggage that would declare, a priori, the impossibility of more than one efficient cause, or epiphenomenalism for causes beyond A-level chemistry, it also constitutes a somewhat unsatisfying answer to the question of whether naturalism justifies a belief in mental causation. Sure, Hume says, we make predictively reliable inferences based on our understanding of someone’s mental states all the time. To that extent, we’re justified in calling those mental states causal, to the same extent we’re justified in calling the movement of the initial billiard ball the cause of the second billiard ball’s movement.

Surely, however, an account of causation that will convince the skeptics must go further than this complete shrugging off of metaphysics. James Woodward takes a page from Hume in defining our justification for positing really existing causal connections between events in nature, but his account, we might say, is only “half-HOG,” for he manages through careful articulation of his experimentalist paradigm to avoid falling headlong into anthropocentric subjectivism with respect to causal claims. While his account is still metaphysically quite modest and noncommittal as compared with, for instance, the Causal-Mechanical model proposed by Wesley Salmon (an account we must reject, as it embraces a form of local realism dependent on substance ontology) it nevertheless provides us a road towards a theory that establishes a basis

for asserting the mind-independence of causal relationships on the basis of counterfactual investigation. This does not commit us to a “monolithic” vision of causation in which manipulability is the sole criterion determining causal relevance (Cartwright 2003), but rather provides a pragmatic way forward in sorting out the presence or absence of causal relationships across different organizational levels in specific cases.

Woodward nicely embodies all four virtues of empirical approaches to causation mentioned at the beginning of this chapter. I will not get into the more technical aspects of his system, but, put briefly, some of his chief arguments can be summarized as follows:

Causal relationships exist independent of human perception – Causal relationships are “out there” in nature, but only can be explanatory insofar as the information is somehow accessible by us for recognition, appreciation, and surveyance. (23)

Manipulation and experimental intervention set the criteria for working out what causes what – Such potentially explanatory causal relationships are discoverable insofar as they are “exploitable for purposes of manipulation and control.” (25) No further speculation about “transmission of causal influence” or other metaphysical worries is necessary. Essentially, the fact that A causes B can be responsibly affirmed insofar as interventions can be performed which demonstrate counterfactuals. In its simplest form, “If A is prevented from happening, B does not happen either.” (28)

Anti-reductive manipulationism avoids circularity – Shedding the metaphysical baggage associated with reductionist theories of causation that appeal to Aristotelian notions of

“production” or “generation” allows us to avoid trading in one mystery for another (that of “causation” with that of “production.”) Adhering to strictly non-reductionist, manipulationist criteria based on experimental intervention allows us to understand causation empirically, without appealing to concepts “that lie outside of the circle of concepts to which ‘cause’ belongs.” (22) In this sense, Woodward’s manipulationism is both non-reductive and metaphysically deflationary – advantages which are broadly Humean in the sense we have defined, and which comport well with the version of process ontology we will endorse.

Reducing all cause-talk to talk about “laws of nature” is unilluminating – Woodward is rightly committed to the sorts of models and experimental procedures used to predict and manipulate outputs from within the experimental sciences. Causal claims as embedded in structural models in psychology, for example, would be adjudicated using rigorous manipulationist criteria from \*within\* said models. This avoids the overlaid and anti-Humean tack (the one Kim and his ilk still labor under) which aims to subsume our understanding of all causation under the banner of “laws of nature.” (Woodward 2003, p.343) Equations in the so-called “special” sciences, Woodward insists, “can express local causal truths without qualifying as laws.” (Woodward 2003, p. 349)

Once again, Woodward’s manipulationist models are too technical in their precise details to be of interest to our general discussion about the implications of such a system. In its basic outline, though, Woodward provides an excellent example of a more contemporary, more technically developed empirical system in the Humean tradition. It has the virtue of eschewing unjustified metaphysical speculation in favor of methodical interpretation of the data, as well as

generation of new data through newly conceived interventions. This allows us to adjudicate causal claims *on their own terms* from within the explanatory practices of a discipline, rather than trying to import dubious premises from “Reason” to tell us what is and is not possible. It inches us towards realism without sacrificing the wisdom of avoiding analytic metaphysics.

This line of reasoning has also been developed to some extent by Eddy Nahmias, who points out (correctly) that incompatibilists have yet to produce any compelling data to suggest that “brain processes associated with our conscious intentions are causally cut off from those that produce actions.”(2011) Determinism, he argues, is not at all intuitively obvious, nor is it clear that it would threaten free will in any meaningful way. Determinism does not preclude the causal efficacy of higher order mental states, downstream effects, or varieties of “self-programming” possible through conscious consideration of future alternatives. (2011) It is only a threat to free will if it can be shown that the causal forces at work in producing an action somehow *bypass* the agent in an important respect. That is, with inputs leading to outputs without any meaningful causal role for lived subjective experience – willing and desiring and thinking and believing and deliberating. Unless this is empirically demonstrated, however, Nahmias makes the perfectly fair point that a person could be *determined* through processes that are deeply entangled with that person’s subjective decision-making. In the meantime, Nahmias insists, we ought to defer to the explanatory practices of the involved disciplines.

This notion of bypassing vs. self-programming or self-determination is crucial to the remainder of our discussion. Epiphenomenalism is the contention that our conscious, lived experienced is “bypassed” in just the sort of way Nahmias has in mind: our behaviors would be the result of purely reflexive, unconscious processes, with our lived experience reduced to nothing more than an inert byproduct. Deliberations and expectations wouldn’t actually have

any causal role, but would only appear to. In contrast, Kim-style reductionists want us to believe that in order for deliberations and expectations to *have* any appreciable causal role, they must be fully reducible to the interactions of entities existing on the level of physics and chemistry.

In the next section, we will be moving from our discussion about responsible empiricism to a discussion of the undue metaphysical assumptions that come attached to arguments about causal exclusion, overdetermination, epiphenomenalism, and the like. By endorsing a very thin conception of process philosophy, the hope is that we can accomplish two goals: first, to bridge the gap between the virtues of Humean empiricism and some form of Realism; and second to suggest that there is no reason for a thoroughgoing naturalist to endorse arguments of the sort the exclusionist or the epiphenomenalist makes. We will not be able to rescue free will of the libertarian sort. However, we can certainly urge the conclusion that the choice some have offered us between crude reductionism and epiphenomenalism is a false one.

#### 4. Why Process? Towards a Thinner (and more permissive) Metaphysics of Causation

Are we justified in admitting folk psychological entities like beliefs, desires and intentions into our ontology? Or are they mere “*abstracta*” – conceptual entities that are of value in explanatory theories, but which don’t strictly exist as “concrete” objects? (Reichenbach 1938) What if we are justified in admitting beliefs, desires and intentions into our metaphysical view because *all* objects reduce, ultimately, to “abstracta?”

Jettisoning all commitment to the *being of entities* involved in causal relationships may be our best strategy in disentangling the problem of “mental causation.” This program of analysis is largely deflationary, disallowing for any sort of “substance” talk about “real” ontological “entities” and their “properties” or “causal powers,” taking this to be an inappropriate way of framing the discussion – both metaphysically ungrounded and unscientific. A process-based account of causation, then, strongly disallows for any sort of reification – not just of processes on higher levels of organization, but also of the so-called “basic” constituents. With these obstacles removed, the groundwork for a plausible theory of causal relationships can be laid.

It may seem a strange strategy to begin by talking about Hume and the virtues of not getting too entangled with analytic metaphysics, only to conclude by advocating for a complete shift in the way we think about fundamental ontology, matter, and causality. However, I believe the following points can be convincingly argued that, if correct, show that these argumentative tacks do not in fact pull in opposite directions. Rather, the initial goal of providing an empirically responsible, more metaphysically neutral account of causal relations would remain intact through an endorsement of Process Metaphysics *if*:

- A.) There exists evidence within physics itself to lend credence to a view that would prioritize the notion of process over the notion of substantial, self-same entities over time,
- B.) Adopting a processual view as opposed to one committed to the metaphysical existence of selfsame entities over time is *deflationary*, allowing one to systematically *remove* premises from traditional arguments about causation, rather than introducing new ones, and
- C.) The resulting loosening of metaphysical constraints helps us account for data that is otherwise unaccountable.

I contend that all three of these can be convincingly supported. We shall tackle them in order, turning first to why our best current theories about the nature of the universe actually strongly suggest that substance ontology is incorrect, and something like a commitment to Dennettian “Real Patterns” (1991) is in fact a more accurate way of describing nature – that is, the view that a process or event can be considered “Real” insofar as invoking the process or event allows for a certain kind of informational compressibility of description.

Next, we shall examine some of the traditional arguments within the free will debate that seem to preclude or limit the possibility of free will, and see that some of the premises at work in such arguments lose their justification within a thoroughgoing process metaphysics. Finally, we shall see that by defeating these arguments through casting serious doubt on some of their key premises, there are empirical phenomena we become able to capture which remain mysterious or inexplicable within the discussion as commonly framed. By showing how process metaphysics in turn enables us to square our theory better with a wider and more inclusive data set, we can hopefully make clear how we can simultaneously endorse quasi-Humean empirical boundaries for the philosophy of science *and* a theory which turns our common-sense notions about matter upside down.

a. *The Evidence from Physics*

The reasons most people, philosophers of science or not, have for embracing a view rooted in substance ontology are practical. It makes psychological sense. It enables us to stick labels onto things, and by virtue of sticking such labels, (the nature of which persists, for the most part,) our conviction that “things” in themselves similarly persist is deepened. While Newtonian mechanics explicitly takes on board a conception of the universe as carved up into “entities,” the frontiers of science have been hinting at grave difficulties in such an ontological view for quite some time. In their bracing, polemical book *Everything Must Go: Metaphysics Naturalized*, James Ladyman and Don Ross provide what may be the definitive refutation of the traditional ways of thinking about the matter. Mustering evidence primarily from quantum mechanics and relativity theory, they have already demonstrated, I think conclusively, that the traditional conception of matter as being made up of discrete, self-same entities persisting over time is forlorn, having been superseded and contradicted by the balance of the evidence over the past century.

The way they describe the fundamental picture in need of replacing is in terms of a metaphor of “*containment*” – that is the doctrine that “the world is a container bearing objects that change location and properties over time. These objects cause things to happen by interacting directly with one another. Prototypically, they move each other about by banging into one another.” (2007) This is an intriguing and revealing way to describe the world-picture offered by substance ontology – that is, one which at least appears to comport with a certain aspect of our daily experience and with common sense. But there is little reason, as we shall see, to afford this kind of “common sense” conception any empirical weight, especially if it is contradicted by our best available evidence. Ladyman and Ross give a name to this “explaining

[physics] in terms of things that common sense thinks it comprehends” – they call it “domesticating metaphysics,” and we shall adopt this term as a way of describing one type of abuse committed by those who approach metaphysical speculation with undue confidence in the analyticity of certain propositions. (2007) Ladyman and Ross go on to describe the consequences of the containment metaphor, not just for ontology, but for causation itself, the primary topic of this thesis. Their analysis is worth quoting at length, at least in its introductory form, as it will allow me to define several terms going forward:

Causation to the modern domesticating metaphysician is, by contrast, typically identified with what Aristotle called ‘efficient causation.’ A characteristic of efficient causation, in the context of the containment metaphor, is that it is imagined to ‘flow’ always from ‘inside out.’ Thus the ultimate constituents of the world that halt the regress of containment are also taken to be the ultimate bearers of causal powers, which somehow support and determine the whole edifice of (often complex) causal relations that constitute the domain of observable dynamics.

The metaphysics of domestication tends to consist of attempts to render pieces of contemporary science – and, at least as often, simplified, mythical interpretations of contemporary science – into terms that can be made sense of by reference to the containment metaphor. That is, it seeks to account for the world as ‘made of’ myriad ‘little things’ in roughly the way that (some) walls are made of bricks. Unlike bricks in walls, however, the little things are often in motion. Their causal powers are usually understood as manifest in the effects they have on each other when they collide. Thus, the causal structure of the world is decomposed by domesticating metaphysics into reverberating networks of what we will call ‘microbangings’ – the types of ultimate causal relations that prevail amongst the basic types of little things, whatever exactly those turn out to be. Metaphysicians, especially recently, are heavily preoccupied with the search for ‘genuine causal oomph,’ particularly in relation to what they perceive to be the competition between different levels of reality. We will argue that this is profoundly unscientific, and we will reject both the conception of causation and levels of reality upon which it is based. (2007)

Several important things are worth drawing attention to in these two paragraphs, all three of which we shall return to in more detail shortly. First, we have the correlation between analytic

(or “domesticated”) substance metaphysics and the notion of “*causal powers*.” This of course makes sense. If the most accurate way of thinking about nature were to think of it as made up of little selfsame entities, then it would make sense to impute reality to an entity only insofar as it contained the causal “oomph” necessary to take part in physical systems. No causal powers, it is sometimes reasoned, no reality. Of course, if our empirical metaphysics does not impute reality to any “entities,” strictly speaking, then there is nothing for so-called causal “powers” to inhere in – and many of the assumptions we make fall by the wayside once we dispense with this dubious concept.

Second, it’s worth noting before delving into Jaegwon Kim that the very notion that causality flows from lower so-called “levels of organization” to higher ones, or “inside out” as Ladyman and Ross put it, is a result of holdovers peculiar to the history of Western metaphysics and a kind of realism about entities at the “bottom” level that follows naturally from the substance-ontological containment doctrine. Neither one, as we shall see, is well-founded. Unfortunately for common sense, and for analytic metaphysicians in general, particles (which would be the “ultimate” constituents of matter and therefore the best candidates for having causal “oomph”) do not seem to behave at all like localized entities. *Causal fundamentalism*, therefore, a fundamental premise lurking behind arguments of the sort made by epiphenomenalists and others who might deny free will, is unfounded. Contemporary physics provides little justification for regarding elementary particles as the truly “real” ontological bedrock of reality. Quantum mechanics reconceptualizes “particles” in terms of field processes and interactions – a stance which moves away from reification and toward the anti-substance view I have been advocating. In addition, phenomena like the decay of protons and neutrons suggest that “entities” on this scale, too, are transient – there is a certain time-scale anthropomorphism in

declaring them to be the uniquely solid foundations of our ontology if their existence amounts to quasi-stable patterns of behavior in intersecting fields.

If elementary particles themselves are emergent phenomena relative to certain measurement procedures, then the sense in which they are causally relevant is not in a “brute” sense that trumps all other forms of causality, but rather in a configurational sense, as with emergent phenomena on higher levels. Contemporary physical theories are forcing us toward a picture of elementary particles that characterizes their being in terms of interrelationships between observed behaviors. If this picture is correct, then the “temperature” of a gas and the “atomic particles” underlying it are both “abstracta” in a meaningful sense. On this view, there is no good reason to assign “causal primacy” to the interactions of such so-called elementary particles – causal relevance in any case would be described in terms of the entangled relationships between types of patterns best described as “configurations” or “topologies” – regardless of which or how many “levels” are supposedly involved. Physicalists spend a great deal of time arguing about the ontological status of higher-level entities, but this is matched by a strangely uncritical acceptance regarding the ontological status of lower-level entities. The truth of the matter may be, however, that the realm of quantum mechanics and elementary particles is just as inhospitable a place for talk of “entities” and “properties” as the realm of the mental.

Third, it is worth drawing attention once more to this notion of causality-as-microbangings. This narrow, efficient-causation-inspired view of what constitutes a cause is profoundly lacking, as we shall see. Of course it is bound to be false from the outset if we are not making a metaphysical commitment to the existence of entities or their “causal powers.” We will see further problems for this view when discussing whether there’s good reason to believe all causation must supervene locally on the properties of the microverse. If instead, we are

simply mapping the entangled causal systems that arise from Real Patterns or topologies organized across multiple scales, much of the metaphysical baggage that is dragged along with notions of efficient or “Productive” causation falls away.

A good deal of the evidence that Ladyman and Ross present from within physics to support their view is highly technical and mathematical, which is only to be expected as it is very closely tied to actual empirical findings within the discipline of experimental physics.<sup>4</sup> Rather than present this evidence at length here, I will latch onto two of the issues they address, namely the notion of “individuality” in quantum mechanics (which ties directly to the notion of locally real, selfsame “entities”) and the notion of “objectivity” in particle physics. My aim will be to provide, rather than the kind of technically precise analysis Ladyman & Ross provide, a kind of pop-science counterpart. This will hopefully still retain some argumentative force while being considerably easier to understand and digest. For the sake of brevity and by way of providing a kind of “update” to the kinds of arguments offered by Ladyman & Ross, I have chosen a few examples from well-regarded peer-reviewed studies published in the last two years.

The first is a paper published in the October edition of *Nature* by a team of researchers at Delft university in the Netherlands. The paper on the one hand represents quite a timely finding for the publication of this thesis, but on the other, it only really serves to confirm a number of existing theories about the fundamental nature of reality. The experiment is a confirmation of Bell’s theorem, or the notion that the behaviors observed in the quantum realm cannot be explained by recourse to “hidden variables.” That is, strange effects like quantum superposition,

---

<sup>4</sup> Attempting to summarize these arguments in too much detail will occupy too much of our time and space here – suffice it to say that those wanting to know more about the empirical warrant for process theories of metaphysics should read sections 3.1 and 3.2 of *Every Thing Must Go*.

entanglement, tunneling, probabilistic distribution and the like must be accepted as fundamental metaphysical truths about the nature of the phenomena we are studying, rather than being somehow “artifacts” of epistemic poverty. Another way of saying this is that it appears that *matter itself at very small scales does not behave like persistent entities which are selfsame over time*. Perhaps even more importantly for our present considerations, *causality in the quantum realm appears to not be fundamentally localized, casting serious doubt on “supervenience” hypotheses about reduction and part-whole relations*.

What makes this experiment unique is that it is designed in such a way, for the first time, so as to close a number of “loopholes” which rendered the results of previous such experiments ambiguous or indeterminate. These mostly consisted of additional premises or assumptions the scientists had to import into their data sets in order to soundly disconfirm the possibility of hidden variable theories. By starting with no such assumptions, (thereby conforming nicely with our distaste for analytic metaphysics and our commitment to empirical methods,) the team at Delft was nonetheless able to robustly demonstrate that two entangled electrons, separated by 1.3 kilometers, were able to effect one another’s state *instantaneously* when turned either spin up or spin down. All errors in equipment and other measurement artifacts were controlled for, and the group was able to show beyond a doubt that the “communication” between the electrons took place with no delay whatsoever, seemingly allowing for the possibility of instantaneous transfer of information. (Henson et. al 2015)

The group at Delft, however, resists this interpretation of the results. There is good reason, they reckon, for rejecting the notion that what we are witnessing in the experiment is faster-than-light communication. Other experiments seem to have demonstrated rather definitively that it is still “against the law” in our universe, so to speak, to travel faster than light.

Rather, the physicists contend, the experiment undermines (if not outright disproves) theories of nature that rely on *local realism* – that is, the idea that the universe is made up of locally positioned selfsame entities that persist over time. In other words, whatever electrons are, they are *not* like little points, and their causal relations bear no resemblance to the common-sense Newtonian picture that would have us reduce all “real” causality to particulate microbangings. “A null-hypothesis test,” their findings state, “yields a probability of at most  $P = 0.039$  that a local-realist model for space-like separated sites could produce data with a violation at least as large as we observe, even when allowing for memory in the devices.”

They conclude, therefore: “Our data hence imply statistically significant rejection of the local-realist null hypothesis.”

This is a rather striking result for a peer-reviewed study published in *Nature*. Like many of the more technical arguments offered by Ladyman & Ross, this finding seriously undercuts the premise of “individuality” – that is, that nature is composed, at bottom, of discrete, individual units. Instead of thinking about some mysterious interaction occurring between “two electrons” associated by some mysterious quality called “entanglement,” such that the two communicate instantaneously by some unknown means, we are better off thinking of a *single* “electron process” that has been stretched out over 1.3 kilometers. In abandoning local realism, we no longer have to ponder the mysterious interaction between two apparently discrete entities that effect each other by “spooky action at a distance.” Rather, we must simply give up on the idea that an electron is fundamentally like a discrete object located at a discrete point in spacetime. As philosopher Ernst Cassirer put it in 1936, examining quantum field theory:

The field is not a ‘thing,’ it is a system of effects (Wirkungen), and from this system no individual element can be isolated and retained as permanent, as being ‘identical

with itself' through the course of time. The individual electron no longer has any substantiality in the sense that it per se est et per se concipitur; it 'exists' only in its relation to the field, as a 'singular location' in it. (Cassirer 2014)

As I have suggested, evidence suggesting the need for this gestalt shift from “things” to “processes and configurations” has been piling up for some time now. Our common-sense attachment to the metaphysics of substance and to the local realism it implies, it would seem, is not easily dislodged by disconfirming evidence, especially in the case of most philosophers of science. As we shall see, however, allowing ourselves to embrace the idea that the universe at bottom is not very “entity-like” allows us to solve a number of otherwise quite vexing puzzles, not least of which some of those related to mental causation.

A second paper, published May 2014 in *Nature Photonics*, undermines another critical piece of the substance-metaphysical world picture by encouraging a picture of “fundamental particles” that, rather than seeming to demarcate fundamentally different categories of entities, seems to paint a picture better aligned with combinatoric processes capable of interleaving and reconstituting, resulting in entirely different sets of relational features. The paper details an experimental mechanism it is within our present capabilities to craft that, by colliding photons together, would be able to produce matter. Essentially, this would be a collider that would be able to turn light into matter, on the order of  $10^5$  “Breit–Wheeler” electron-positron pairs in a single shot. (*Nature Photonics 2014*) This is difficult to wrap one’s mind around in terms of the traditional, common-sense substance-ontological world picture previously described in terms of the containment metaphor. If all putative “entity” talk, however, is really talk about quasi-stable “Real Patterns” or structural/behavioral relations, with no rock-bottom entities endowed with essential properties to appeal to, it is easier to see how such radical configurational shifts could occur, along with the associated changes in relational properties. (This has been discussed more

detail by Campbell and Bickhard in the context of their discussion of causation in terms of multifarious interactions within a physical system.) (2011) If putative “photons” are a semantic label for the dynamic propagation of a certain influence through fields over time, it is easier to stomach how reconfiguration of said influence could result in the process taking on new dynamical features. Here, the language of another abstract analyzing this Breit-Wheeler process is revealing in its avoidance of traditional substance-ontological language:

The generalized Breit-Wheeler process, i.e., the emission of  $e^+e^-$  pairs off a probe photon propagating through a polarized short-pulsed electromagnetic (e.g., laser) wave field, is analyzed. We show that the production probability is determined by the interplay of two dynamical effects. The first one is related to the shape and duration of the pulse and the second one is the nonlinear dynamics of the interaction of  $e^\pm$  with the strong electromagnetic field. The first effect manifests itself most clearly in the weak-field regime, where the small field intensity is compensated by the rapid variation of the electromagnetic field in a limited space-time region, which intensifies the few-photon **events** [emphasis mine] and can enhance the production probability by orders of magnitude compared to an infinitely long pulse. Therefore, short pulses may be considered as a powerful amplifier. The nonlinear dynamics in the multiphoton Breit-Wheeler regime plays a decisive role at large field intensities, where effects of the pulse shape and duration are less important. In the transition regime, both effects must be taken into account simultaneously. (*Physics Review 2013*)

It is less important to understand all of the details in this rather technical passage and more important to note that, in describing such a process, a great deal of focus is justly put on the dynamicity of the processes involved, and of the time-sensitive nature of their propagation. Note the description of the involved putative photons as “photon events,” as well as talk of a “transition regime” rather than a strictly demarcated moment of transformation.

Relatedly, the containment metaphor necessitates making a sharp distinction between those things that exist at a given time, and those that do not. This satisfying cleavage arises quite naturally from the containment metaphor - thinking of the universe as being populated by

fundamentally by entities with a definite ontological status at a given time – a view which, as we have been arguing, our current best theories suggest is deeply incorrect.

One of my favorite examples of a metaphysical confusion arising from this intuitively substance-based conception of matter is the notion that, at the quantum level in space, “virtual particles” are in a constant state of “popping in and out of existence.” (Choi 2013) We know such “virtual particles” represent real structural features of the universe, and that the processes they represent are capable of having measurable effects on their surroundings. The cumulative result of this “Casimir effect” is actually so significant that it drives cosmic inflation. (Tegmark 2006) However, from a metaphysical standpoint, this notion should seem puzzling, at the very least. If the fundamental metaphysical reality was that the universe was composed of little selfsame units, how could one possibly account for them coming into existence, in pairs, and then going out of existence, all within a fraction of a second?

Dissatisfaction with such head-scratchingly strange “something from nothing” metaphysical language has led some physicists to try to distance themselves from entity-focused language and speak instead of “fluctuations of the quantum vacuum” or “resonances.” (Meulenber 2011) This seems altogether more metaphysically appealing, as it dispenses with the idea that something is literally coming in and out of existence, replacing it with the idea that we \*say\* a virtual particle comes into existence at the moment fluctuations of energy resonances within the quantum vacuum produce a process that is capable of perturbing other nearby processes. Nothing actually need enter our ontological realm and then exit it, however – rather, fundamental processes within the fabric of material reality reach a certain point of coherent activity, which then decoheres back into random noise. We could say, then, that a quasi-stable “Real Pattern” or configuration is instantiated for a brief period of time. The resulting transient

properties, rather than inhering in some object which is their bearer, are instead emergent properties of the dynamic quasi-stable configuration insofar as it retains certain structural features. Again, this seems straightforwardly metaphysically preferable to the statement that “things” are going in and out of existence.

“Virtual particles,” then, suggest a reading of nature which will drive us towards a version of *ontic structural realism* (OSR) – the view that what the success of the empirical sciences commits us to metaphysically is not an ontology of things, but rather one of *events* taking place in the physical manifold. Configurations or relational structure is all that’s real. Versions of structural realism and associated event or process ontologies have been defended in respect to quantum field theory in Auyang (1995) and Dieks, (2011) as well as Seibt (2002) and Hättich, (2004) but the version that will concern us chiefly here is the version championed by Ladyman & Ross:

Ontic Structural Realism is the view that the world has an objective modal structure that is ontologically fundamental, in the sense of not supervening on the intrinsic properties of a set of individuals. According to OSR, even the identity and individuality of objects depends on the relational structure of the world. (2007)

In order to clarify this notion of objective modal structure, Ladyman & Ross introduce a more technically robust definition of “Real Pattern” to serve as their fundamental ontological concept. The objective existence of such patterns boils down ultimately, much as it did for Dennett, to the facility of the concept in making better-than-chance predictions in physically possible computer models, as well as carrying *information* in a crucial sense. Specifically, a pattern, to be considered Real, must *compress* information in a meaningful sense, such that it tracks

relationships and makes predictions both more effectively than a one-to-one, zero-compression bit-map encoding of a scenario and predictions which such an encoding would effectively miss.<sup>5</sup>

As Dennett noted in his original paper on Real Patterns, the predictive capacity and allowance for compressibility provided by folk psychology is “stupendous.” (1991) By adopting what Dennett calls “the intentional stance” towards organisms – that is, treating them as if they have beliefs, desires, intentions, etc. – we achieve a previously impossible degree of predictive power and understanding with respect to them. The enormous computational leverage provided by conceiving of an organism as moved by what it “wants” far outstrips the leverage provided by a description at the level of chemistry. If, in the future, this turns out not to be the case as we learn more about chemistry, so much the worse for intentional patterns. This, however, is still an empirical question, and cannot be settled *a priori*. In the meantime, we can take the success of the intentional stance as an indication that the patterns it proposes are at least *plausible candidates* for having objective reality and meaningful causal bearing.<sup>6</sup>

Such an ontological system based off the notion of Real Patterns brings us to the kind of responsible, empirically-minded realism we have been seeking in trying to make sense of causal narratives and causal relations. The account is still broadly deflationary compared to most kinds of Realisms that would force us to say something about our commitment (or lack thereof) to putative unobservables, and other theoretical entities. It also eschews some of the traditional

---

<sup>5</sup> The authors here offer a more technically precise definition: “To be is to be a real pattern; and a pattern  $x \rightarrow y$  is real iff

(i) it is projectible [able to make better than chance predictions]; and  
(ii) it has a model that carries information about at least one pattern P in an encoding that has logical depth less than the bit-map encoding of P, and where P is not projectible by a physically possible device computing information about another real pattern of lower logical depth than  $x \rightarrow y$ .”

<sup>6</sup> Dennett has clarified in *Intuition Pumps and Other Tools for Thinking* (2013) that physical underpinnings for a mechanism (what he calls a “Rube Golberg machine”) do not lead to epiphenomenalism about content, as patterns on higher levels of description are still Real insofar as they afford compressibility and projectibility.

problems problems and paradoxes that arise from naturalistic worldviews that give undue ontological and nomological preference to the *smallest* theoretical entities currently postulated.<sup>7</sup> What's more, it allows us to embrace the scale-relativity of ontology, which takes into account a wider swath of data regarding potential causal relationships. It also carves a path towards unity of science without reduction, since there are a multitude of ways for real patterns to carry information about one another without inter-reducing.

We also emerge with a robustly defined, metaphysically neutral way of defining what kinds of events can serve as links in a causal chain: no longer strictly Human Observation Governed (Hume's Whole-HOG instrumentalism,) Ladyman & Ross also go one step closer than the Half-HOG manipulationism of a Woodward by specifying ways in which the data of one realm of scientific inquiry can trump another. Once again, the critical concept is information: rather than throwing out causes on the basis of shaky analytic arguments about "overdetermination" and "exclusion" that tend to rule out so-called "high-level" patterns from the outset, redundant patterns are rejected only insofar as they do not increase projectibility while decreasing logical depth. Thus, intentional patterns of the sort that interest us in this thesis would be rejected by a mature theory insofar as "all projectible patterns tracked by intentional psychology along with others it doesn't track are projected by mature cognitive neuroscience." That is, if the individuation of the pattern is essentially a crude approximation motivated by epistemic limitations, and whose invocation in a simulation would increase the logical depth of the encoding without leading to any increase in projectibility.

---

<sup>7</sup> More of this will be said in the next subsection.

This section has argued that there is strong evidence from the frontiers of physics that casts doubt on substance metaphysics' metaphor of *containment*, by breaking down notions of individuality and self-similarity over time. Instead, we have argued that OSR is a neater fit to the actual data, and retains many of the positive features of instrumentalist readings of causation visited earlier by being metaphysically deflationary. Through Ladyman & Ross' definition of "Real Patterns" and unity without intertheoretic reduction, we are poised to utilize this empirically-minded approach to systematically remove some premises from traditional arguments about inter-level causation.

*b. Unjustified Premises in Exclusion Arguments – Shifting the Burden of Proof*

As I stated in the beginning of this chapter, dispensing with the strong realist assumptions of substance metaphysics in favor of OSR is metaphysically deflationary. Although it does not commit us to the kind of instrumentalism discussed in the second chapter, it retains a strong empirical bent which rules out metaphysical speculation about the relationship between different "levels" of organization as distinct from questions about projectibility and manipulability. What's more, by refusing to domesticate our ontological world-picture to the metaphor of containment, we avoid thinking of causal relationships in terms of the causal "powers" of the entities populating the container, and of their causal influence flowing from the "inside out" or "bottom up" after the fashion of traditional Western metaphysics (see page 38.) If it's configurations, topologies, and Real Patterns all the way down, then there is no justification for granting nomological privilege those patterns picked out by the "hard sciences" as opposed to the "special sciences."

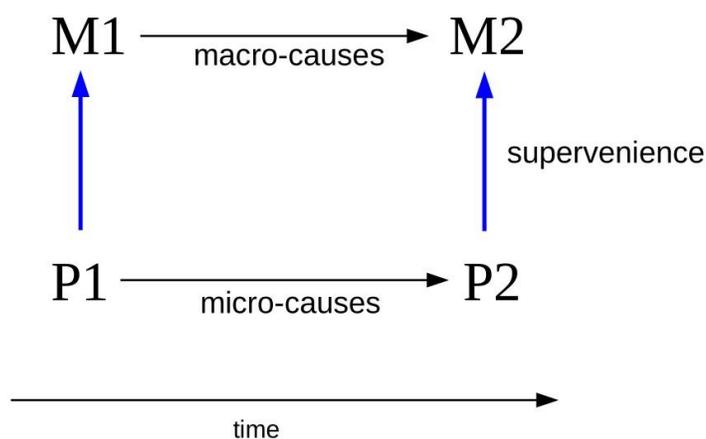
As a result, many analytic metaphysical arguments widely considered to be formidable within conversations about physicalism and mental causation, simply lack sure enough footing in their premises to get off the ground. Let's take a look at one such argument, perhaps one of the strongest and most convincing formulations, offered by Jaegwon Kim in his book *Physicalism, or Something Near Enough*.

Kim, like us, is interested in using a responsible reading of the naturalistic sciences to vindicate the casual efficacy of certain mental states, resisting epiphenomenalism. Unlike us, however, Kim believes that such mental states can only have "causal powers" insofar as they are physically reducible to the causal powers of the involved microprocesses. This is just the analytic metaphysical view of causation we have been challenging – the one which arises from a view of causation as a power that inheres in individuated entities on the microscale, presumably through "microbangings." Kim makes clear how strong his commitment to such causal fundamentalism is: "mental residue," he says, "insofar as it resists physical reduction, remains epiphenomenal." Elsewhere, he states: "If anything is to exercise causal powers in the physical domain, it must be an element in the physical world or be reducible to it." (2005)

This statement would not be so problematic were it not for what Jaegwon Kim means by the "physical world." Kim does not mean those physical processes with which we can become acquainted through empirical investigation, across multiple scales of reality. Nor does he mean those things that can be singled out as "Real Patterns" in the rigorously empirical and revisable sense detailed earlier. Rather, Kim means those putative theoretical entities which are meant to populate the micro world, and only those. Since, for Kim, this "level" is one that possesses hard-line causal "powers," any other process which claims to have an influence on the material world must do so exclusively by means of such processes. Kim refines his argument:

It is only when we reach the fundamental level of microphysics that we are likely to get to a causally closed domain. As I take it, the so-called Standard Model is currently taken to represent the bottom level. Assuming that this level is causally closed; the supervenience argument, if it works, shows that mental causal relations give way to causal relations at this microlevel...the bottom level of fundamental particles (assuming that this is the only level that is causally closed) is always the reference physical domain.

Kim often presents his exclusion argument in the form of a diagram:



Since, as he takes it, it is not just the physical world that is causally closed, but the *lowest level* of the physical world that is causally closed, he takes this lowest level itself to be sufficient for explaining the causal entanglements present in the physical domain. Thus, the arrow  $P1 \rightarrow P2$  is meant to indicate that the changes in any physical system over time should be fully explainable in terms of the causal interactions between microparticles at the “bottom” level. The physical state at P1 causes the physical state at P2, thereby pre-empting any putative causation at higher levels. This is the substance-ontological “microbangings” world picture we have been criticizing at its most direct. Kim rightly points out that he is simply taking it as an assumption that the bottom level in physics is the only one that is causally closed. Why should we assume this? Is this an assumption drawn responsibly from the empirical sciences, or is it merely a metaphysical prejudice ingrained in us by the history of our discipline?

Kim makes an additional mistake in assuming that putative mental causes must supervene locally on features of the microphysical world. The arrows pointing from the lower-level physical states to the macrostates at the top are meant to indicate that the features of said macrostates are always dependent upon local features of the microphysical world. But again, why should we assume this? Causation at the level of concepts and experience is likely to be entangled not only with subpersonal physical processes within the body, but with features of the psychosocial milieu in which the organism is embedded. Causal entanglements that run through symbolic manipulations occurring over conscious operands within the agent will likely involve Real Patterns occurring at interpersonal and cultural levels, as well. For instance, the efficacy of a symbolic ritual in providing placebo relief from pain symptoms will depend strongly on conceptions of illness and wellness, histories of herbal remedies within the community, and the role of the ritual in the agent's belief system. While all of this takes place, of course, within a physical domain, the causal factors at work are unlikely to be in a supervenience relation with local arrangements of microprocesses.

All of this boils down to a defense of a certain flavor of functionalist reductionism that a thoroughgoing process-based account need not accept. While we are not permitted to simply take empirically suggested putative causal connections "at face value," so to speak (as the Real Patterns involved would also have to be verified to have projectibility,) we need not join Kim in believing that to be a cause is to be a productive cause in the fundamentalist sense, or a concatenation of such discrete causes.

This is made clearer still analyzing Kim's version of the doctrine he calls "emergentism." As Batterman (2002) and Ladyman & Ross (2007) have observed, several of the "central doctrines" Kim invokes in setting up his arguments against "emergentism" (1999) express

commitment to a substance metaphysical view in which, as Ladyman & Ross put it, the world “decomposes uniquely into non-overlapping components,” as well as “dividing naturally into levels related at each step by a composition relation – commitments, we argue, that current physics does not support.” If Kim’s assumptions about what it would mean to reject his version of ontological reductionism assume, from the outset, a substance metaphysical worldview which is highly doubtful given our current best empirical research, then we should regard these particular arguments as no serious “discovery” about the limits of physical possibility.

From the point of view of OSR, the Exclusion Argument itself is fraught with problems. Similarly to Kim’s reductios against a view which would accommodate the scale-relativity of ontology by granting existence to Real Patterns, these arguments again rest on premises which tacitly endorse substance ontology. Kim enumerates them as follows:

The causal closure of the physical domain. If a physical event has a cause at  $t$ , then it has a physical cause at  $t$ .

Principle of causal exclusion. If an event  $e$  has a sufficient cause  $c$  at  $t$ , no event at  $t$  distinct from  $c$  can be a cause of  $e$  (unless this is a genuine case of causal overdetermination)

Principle of determinative/generative exclusion. If the occurrence of an event  $e$ , or an instantiation of a property  $P$ , is determined/generated by an event  $c$  – causally or otherwise – then  $e$ ’s occurrence is not determined/generated by any event wholly distinct from or independent of  $c$  – unless this is a genuine case of overdetermination.

As we have previously noted, by almost unquestioningly defining the “physical domain” as meaning “the level of fundamental particles and molecules,” Kim as already tacitly endorsed a view which is more metaphysically heavy-handed than anyone should be at the outset of such an investigation. As always, one of our chief goals is to expose

the role of unjustified *a priori* premises smuggled into naturalistic philosophy – and here, why is it that we should determine what entities are “real” by simplistic recourse to which are most “fundamental,” rather than which constitute meaningful variables that can be manipulated in experimental scenarios, or those which afford projectibility and compressibility? If we are being responsible naturalists, why is our account so laden from the outset as to presume that the domain which is meaningfully causally “closed” is the so-called “bottom level,” reducing all causation in the universe to microbangings?

This same problem is echoed in Kim’s second and third premises, which adopt the microbanging approach once more by invoking discrete, localized, efficient causes. Kim rules out any other such causation as “overdetermination” which, it seems to me, is rejected mostly as a matter of metaphysical prejudice. Kim makes special allowance for supposedly “legitimate” cases of causal determination, such as “two bullets hitting the victim’s heart at the same time, the short circuit and the overturned lantern causing a house fire, and so on,” where “each overdetermining cause plays a distinct and distinctive causal role.” Given our commitment to OSR, we need not claim that mental causes play a “distinct and distinctive” causal role from those identifiable at other scales of empirical investigation – in fact, such causes are likely deeply intertwined or entangled with one another.<sup>8</sup>

We prefer to let the question of mental causation be adjudicated by manipulationist experimental designs and rigorous assessment of the projectibility of the Real Patterns invoked by psychology. Not *a priori* metaphysical commitments. However, it has been pointed out that Kim’s premises, besides being a bit overweening

---

<sup>8</sup> One way to imagine this would be to imagine that structural features of experienced mental states *constrain* or place physical limitations on the possible behaviors of entangled processes at lower scales.

in their reliance on intuitions, are problematic in additional ways. Philosopher Ned Block has pointed out, so long as the argument is valid, causal powers seem to “drain away” to the bottom level. If, as has been proposed by some respected physicists, physics has no fundamental “bottom level,” with process infinitely decomposing into other processes, then it seems no causal claim can get any traction, since it is always pre-empted by causes at the level beneath it (Block, 2003). Kim attempts to head off this objection by insisting the causal oomph resides in the “*union* of all of the microlevels” in this proposed infinite chain, but this only raises anew the question of why such a union would not include Real Patterns at larger scales. If it is true that causation at higher levels is always pre-empted insofar as lower levels are not explanatorily complete, what justifies our taking the current level of fundamental physics as “closed,” given the incompleteness of the standard model?

As before, our account here is deflationary in a classically empiricist sense. Our arguments are not meant to conclusively demonstrate the existence of mental causation, but rather to show that it can be commensurate with a responsible reading of science that is not overburdened by unjustified presuppositions. We are shifting the burden of proof back to those who would deny its possibility, epiphenomenalists and Kim-style reductionists alike. If the assumptions of substance metaphysics are incorrect, is there still a basis for embracing exclusion arguments? Is there any reason analytic metaphysics should lead us to the conclusion that our conscious deliberations are not the causal locus of our decision-making?

*c. The Role of Expectations in Neuroscience – Placebo and Nocebo Effects*

Of course, in returning the causal locus of decision-making to the agent, we are not suggesting that these are not realized physically and mechanistically within the brain – we have merely been advocating for a certain form of *causal ensnarling*, whereby processes occurring at multiple “levels” of scientific investigation effect each other in complex and entangled ways. In other words, we argue, it is possible for the “Real Pattern” of physical activity represented by an expectation, a belief, or a desire to have meaningful causal bearing (perhaps through constraint) on the summary behavior of its constituent molecular mechanisms within the brain.

Given that our brains seem to be beholden to the same laws of physics that hold everywhere else in the universe, it is parsimonious to conclude that the tangle of electrical and chemical activity in our nervous systems is also sufficient (given the right environmental and social conditions<sup>9</sup>) for generating the kind of consciousness we enjoy. The role and potency of mental mechanisms is too well-established to provide warrant for additional metaphysical presuppositions, such as the positing of new fundamental forces of nature, or of immaterial entities intervening in the physical order through some unknown backdoor. All indications are that the tools for providing a satisfactory account of consciousness are at our fingertips.

Of course, as we have discussed, such a mechanistic understanding of consciousness generally goes hand in hand with worries about epiphenomenalism, or the notion that beliefs, desires, expectations, and deliberations are causally inert, merely “piggybacking” on the causal effects generated at atomic and molecular level by the “real,” “rock bottom” furniture of the universe.

Hopefully we have already raised considerable skepticism about this sort of view by showing how its central assumptions, including those of causal fundamentalism, supervenience,

---

<sup>9</sup> This kind of externalist bent is important for establishing the kind of non-local mental causation that rejects “supervenience” arguments.

and distaste regarding overdeterminism are inextricably linked to an unjustified commitment to substance metaphysics. In this final section, however, we will take a look at some contemporary candidates for cases like those Nahmias and Woodward discuss – namely, cases in which the empirical sciences seem to be making a non-trivial commitment to the existence of mental causation as a matter of explanatory practice.

Psychosomatic medical phenomena in particular present difficulties for epiphenomenalist theories of consciousness, which would seek to explain the phenomena of human experience without recourse to folk psychological processes, or which would seek to explain their causal role entirely in terms of neural processes happening “beneath” the level of subjectivity. This is the choice we are forced to make if we want to be responsible physicalists according to Kim – a notion we have argued is the dubious result of an overinflated and empirically ungrounded metaphysics. Either we accept that all mental causation is reducible to “bottom up” causation, Kim tells us, or else abandon our belief in it altogether. Once a functional property of a system has been identified, “scientific work can begin in search of the ‘realizers’ of the functional property – that is, the mechanisms or properties that *actually* perform the specified causal work – in the population of interest to us.” (2005) [Emphasis mine.]

Of course, I am equally hopeful that a physicalist program will uncover how the causal work behind human thought and behavior is accomplished, and that those answers will be found, largely (but not entirely<sup>10</sup>) within the neural architecture of the brain. I also grant that whatever the mechanism is behind mental causation, it will be underwritten by neural “realizers.” I am not proposing that mental states are somehow divorced from, or realized by other means than electrochemical signals within the brain. The “actually” in the above quote from Kim, however,

---

<sup>10</sup> My sympathy towards “extended mind” theories, as well as externalist theories of meaning, necessitate the qualification here.

is revealing, and points once again to his causal fundamentalism – his belief that realizers would constitute “real” or “legitimate” causes, while the states they realize somehow cannot enter the causal chain.

We have already discussed evidence from the world of physics that casts doubt on causal fundamentalism, but the question that currently concerns us is this: can certain psychosomatic medical phenomena ever be made sense of within a view that assumes causal fundamentalism? In other words, can we make sense of an epiphenomenalist or hardline, Kim-style reductionist view in light of explanatory practice within the sciences?

To explore this question, we’ll need an example of a *psychosomatic phenomenon*. These are an excellent way to explore the question of mental causation and the role of folk psychological entities, as such processes as the patient’s “beliefs” and “expectations” are frequently referenced in the medical literature investigating such phenomena. They also seem to be paradigm cases of at least putative “downward” causation - a term we have rejected in favor of causal entanglement or ensnarling. After all, it’s right there in the name. It goes from the *psycho* to the *soma* - mental causation! Case closed, right?

Well, things won’t be quite so easy. Nevertheless, I hope to make a convincing case that explanatory practice in psychosomatic medical phenomena suggests we need to take seriously the idea that processes operating across multiple scales are capable of becoming deeply causally entangled. For the purposes of this discussion, we’ll choose a phenomena which should be particularly puzzling to the causal fundamentalist. *Medical student’s disease*, also called “hypochondriasis of medical students,” is a well-documented phenomenon whereby aspiring medical professionals, during the course of their study, seem to mysteriously contract the symptoms of whatever disease they happen to be studying at the time. From a common-sense

perspective, it's easy to see why this might happen. The students are ruminating about a certain kind of malady for hours a day, sometimes for days or weeks at a time. Their thoughts are consumed by reflection on some condition or other, and so they have a certain kind of heightened awareness to any sensation that might be a sign or symptom. Any sensation they \*do\* experience, surely, gets blown out of proportion in their minds, and before you know it, they're showing up to their own physician, full of anxiety, convinced they've come down with the very thing they've recently been thinking so much about. Of course, as it turns out, they're perfectly healthy.

From the perspective of a philosopher of mind, mechanistic explanations for such a phenomena are not so easy to come by. How *could* such a process work? We don't currently have a completed theory in medicine (though we will discuss some leading research shortly,) and so as of right now, no one can definitively answer the "how" question. As philosophers concerned with the reality of mental causation, however, we would be remiss in not acknowledging that it *does* happen, and not trying to ensure that our theories can accommodate this reality. To borrow a Kantian phrase, what is the condition of the possibility for medical students' syndrome? What *has* to be the case in order for such a thing to come about? The answer, I argue, is threefold. In order to bring about the phenomenon in question:

- 1) Sensory information drawn from the environment *must* be presented to the agent as a conscious perception.
- 2) That conscious perception fits into a both a phenomenological and conceptual framework such that it can become the subject of conscious reflection, deliberation, and other forms of cognitive manipulation.

- 3) The patterns of neural behavior that follow 1) and 2) (such as those that generate behavior) are generated *as a result of associations made at the conceptual level*. In other words, they are as they are *only because* the information they contain has been made globally available for cognitive operations.

If this set of conditions of possibility is correct, then the existence of a phenomenon like medical students syndrome provides compelling evidence that mental causation exists – that is, that epiphenomenalism cannot be true. On this account, the brain generates behavior *\*by\** presenting it to the agent – this is its’ strategy for generating flexible, intelligent responses that transcend mere reflex. For if epiphenomenalism were the case, no such condition as medical students’ syndrome could ever arise. It is difficult to imagine how someone could make themselves sick by thinking about something too much if thoughts have no causal efficacy. Let us imagine what an epiphenomenalist, or eliminativist account might have to assert.

Let’s come up with a hypothetical medical student. We’ll call him David. At the moment, he’s studying celiac disease – allergy to gluten. David does most of his current research through reading, but that he also internalizes some of the information he’s presented with by listening intently to his Professor’s lectures. After hearing lectures on celiac disease for a week straight, reading medical texts on the subject, and preparing to give a presentation on the condition, David begins to notice odd symptoms. He feels unusually fatigued. One night after eating pasta for dinner, he develops a rash on his skin and has severe diarrhea. The symptoms subside for a few days, only to flare up again the next time he eats a meal that is high in gluten. Symptoms persist, including bloating, cramping, rashes and diarrhea, all associated with his intake of gluten. Finally, panicked after two and a half weeks of continuous symptoms, David

submits himself to testing for a gluten allergy. The tests come up negative. David is puzzled but relieved, and never experiences problems associated with eating gluten again.

Cases like David's are not unheard of. Medical student's disease can produce genuine symptoms that mimic the course of an actual disorder, including measurable effects such as rashing. This manifestation of such real and measurable symptoms in the body has to be accounted for, one way or another. I am proposing that the way of doing so which is both commensurate with our manifest experience and our explanatory practices is to look for a cause which is psychological – somehow, sustained, voluntary cognitive focus on celiac disease and its symptoms was responsible for setting off other non-voluntary effects within David's body. These effects would not have been set off if not for David's voluntary sustaining of cognitive focus, and therefore would not have been set off if David's cognitive activity was not capable of altering David's brain on a neurochemical level *in virtue of its content*.

Some rather hefty and implausible assumptions would have to be made in order to assert that the following sequence of events involves only causes at the “fundamental,” “bottom-up” level without invoking the content of David's thoughts, and therefore assigning causal efficacy at the level of concepts and mental operations. Obviously, nothing has changed about the structure of the gluten molecules in the products David buys. Nor has his body's internal chemistry begun to interact differently with those molecules, as evidenced by his negative test results and subsequent recovery. No exogenous trigger, it would seem, has changed, except for those materials David is interacting with *conceptually* in the form of textbooks and lectures. We would have to assume, then, that the sequence of events which led to the sensation of pain in David's abdomen, and the rashing on his face, was triggered by more basic, atomistic elements of David's perception that make no references to the conceptual framework of which they

become a part. We would have to assume, in other words, a complex mechanism which connects exogenous triggers into motor outputs of a sort, as well as into commands that can be acted upon by the immune system. Perhaps, then, certain sequences of black marks on white paper are capable of triggering a pain response in the abdomen? Of causing diarrhea? Perhaps it just so happens that certain vibrational frequencies in the air corresponding to his professor's words are capable of starting a neural cascade that leads the body to mimic the symptoms of celiac disease?

Of course, I don't mean to be dismissive. I present the above puzzle as an honest challenge to those philosophers who believe that empirical evidence has already shut the door on the possibility of any causation besides the "bottom up" sort conceived in terms of microbangings. If we are being epiphenomenalists about consciousness, we can't make reference to the *concepts* David is entertaining and declare those causally relevant to the physical symptoms he is experiencing. That would be assigning causal efficacy to the mental, since it would be tantamount to granting that mental content and how it is consciously manipulated has measurable and irreducible causal relevance. Even if we do grant that the words in David's textbook, and that his professor speaks, (1) are realized as conscious perceptions in David's experience, this is not enough to explain the effect – for hypochondria is not a simple reflexive response, such as the one that would be triggered by hearing someone yelling "duck!" It also *must* be the case that (2) – the conscious perception fits into a larger framework for later rumination and cognitive manipulation. David's symptoms weren't triggered the instant he learned about celiac disease, but rather as the result of prolonged reflection on the concepts. It seems highly problematic to imagine that sensory data might eventually make its way down to causally influence the nerves surrounding the intestines without realizing a causally potent

mental state. Such a string of sensory data would, one imagines, have to be highly specific to the point of being arbitrary – vibrations in the air and scribbles on paper are causally efficacious because they invoke words that fit into a context for us. They can effect our behavior because they effect what we think.

Sense data may constitute the beginning of a causal narrative, but only *through* invoking concepts – which then must be manipulated on the level of conscious thought in order for them to produce their wide range of effects. In this case, *it is the thinking about celiac and its classic presentation that caused David's symptoms*. In other words, (3) – the effects our brains produce in the form of thoughts and behavior would not be as they were if mental causation was impossible. Or, to put it another way, *hypochondria is impossible to explain without recourse to ensnarled (mental) causation*. This is another causal scenario that seems difficult to account for on theories of reality that would make all causation supervene locally on the actions of particles – while there are of course black squiggles on the page that correspond to the words David is reading, and acoustic blasts coming out of his professor's mouths that correspond to the words he's hearing, it is not the Real Pattern represented by the shape of the acoustic waves themselves that triggers a complex series of reflexes in David's head, like an infinitely complex key entering an infinitely complex lock. Rather, it is the Real Patterns represented by the *ideas* he is absorbing as they exist in his cultural/psychological milieu, and how those come to interact with other ideas David has already taken on board.

While there has not been enough research on hypochondria itself to confidently propose mechanisms for its operation, there is good cause for believing that one can extrapolate responsibly from data on placebo effects. After all, our argument has been that responsible empiricists should read conclusions about causal connections off of data and explanatory

practices, rather than analytic metaphysical arguments. What are some common explanatory practices from within placebo effect research?

Some of the leading work in this area has been done by Fabrizio Benedetti, who has published a number of articles and books on the neurobiological mechanisms that underlie such effects. His research has found that the operation of placebo effects is highly dependent upon what he calls the “psychosocial context” the patient finds herself in. (2003) He also makes considerable reference to the “conscious expectations” engendered by this psychosocial context. He has even stated that “the placebo effect appears to be a very good model to understand how a complex mental activity, such as expectancy, interacts with different neuronal systems.” (2005)

His research has uncovered several mechanisms by which expectations and beliefs seem to be irreducibly capable of triggering molecular activity within the brain. One fascinating discovery is that when patients are administered an inactive compound and told that the substance has pain-relieving properties, endogenous opioids are released in the patient’s nervous system that provide a very real analgesic effect. In fact, the placebo analgesia also induces respiratory depression, as well as its effects being totally nullified by an anti-opioid compound, naloxone. (1999) These are essentially the same effects we would expect from the administration of a real opioid. Yet, nothing at the level of physics and chemistry seems to have been meaningfully triggered – no “microbangings” traveled into the endogenous opioid receptors from the mouth of the physician. I own that this suggests that the critical ingredient transforming the humble, inactive sugar pill into a dose of Codeine is something being represented in subjective experience in the form of a Real Pattern we would call an “expectation” – a priming of some set of cortical processes that constrains, and ultimately has downstream effects on, the molecular machinery helping to realize it.

What's more, what kinds of chemicals the brain releases are even further malleable by the patient's conscious expectation of what the drug should do in their body. Benedetti has also been able to induce the endogenous production of cannabinoids, as well as inducing either improvement or worsening in parkinsonian patients through placebo/nocebo regulation of endogenous dopamine. In the latter case, it was found that "verbally induced expectations of analgesia/hyperalgesia and of motor improvement/worsening antagonized completely the effects of a conditioning procedure." (2005) In other words, the effect that the supposed "drug" had on the patients was entirely dependent, biochemically, on what they had been induced to believe beforehand. This was shown again in other experiments which showed that a procedure designed to isolate and manipulate the strength of the belief in the placebo could manipulate the strength of the analgesia. Stronger belief, stronger pain relief. (1999) Clearly, if there were ever a place to point in experimental practices to say that we shouldn't count out "downward" varieties of causation based on analytic arguments, here is the place. How else shall we draw the causal arrow but from the consciously experienced verbal suggestion/acute belief state to certain molecular subprocesses? (Recall, as well, that in the case of medical students' syndrome, the patient needn't ever before have experienced the symptoms that befall her.) What better candidate could we have, furthermore, for a mental cause which is both sufficiently projectable (for OSR's ontological standards) and manipulable through interventions (for Woodward's nomological ones?)

While Benedetti owns that placebo effects are sometimes "mediated by conditioning when unconscious physiological functions, such as hormonal secretion, are involved" (by which he means classical, Pavlovian conditioning that arguably makes a poorer case for the role of higher-level patterns causally influencing lower-level ones,) he also affirms that the release of

certain molecules in the nervous system is “mediated by expectation when conscious physiological processes, such as pain and motor performance, come into play.” (2005) This appears to work in virtue of activity in multiple prefrontal processing regions which are known to be involved in representing aspects of the situational context to the agent and generating expectations about reality, such as the ventrolateral prefrontal cortex and the orbitofrontal cortex. (Handy et al. 2001) In accordance with (3,) Benedetti seems to confirm that in these cases, it is crucially important not just that brain activity is going on in a certain region, but that said brain activity *represents* certain belief states which have an identifiable causal effect, confirmable through experimental interventions. He entertains two competing hypotheses at one point in his publication on the neurobiological mechanisms of placebo effects – that they are driven by executive attention, and that they work by reducing anxiety. He concludes that the latter hypothesis is insufficient to account for the data in a number of ways, and that the evidence points towards forms of self-regulation which involve top-down control of molecular mechanisms by emotion and belief states. (2005)

Other studies have reached similar conclusions about the irreducible explanatory role of conscious expectations in psychosomatic phenomena. Lieberman et. al. in 2003, studying placebo effects in patients with irritable bowel syndrome, concluded that “top down, belief-related placebo effects modulate the activity of brain regions ordinarily affected by other treatments that presumably operate through bottom-up mechanisms that are not belief-related.” (Neuroimage 2004) Their study supported what is sometimes called a “Disruption Theory” of placebo mechanisms, in which, (presumably in addition to the release of endogenous opioids,) regions that consciously represent pain unpleasantness (such as the dorsal anterior cingulate or dACC and the right ventrolateral prefrontal cortex or RVL PFC,) have downstream analgesic

effects on negative affective processes. These authors even go so far as to implicate the involved regions in “reflective conscious responses to negative affect.” As a result, the authors concluded that their results suggest “placebos may operate, in part, by increasing thoughts about the affective aspects of the pain (i.e. ‘I believe I am going to be less bothered by pain now.’)” (Neuroimage 2004)

Once again, all of these results seem to strongly affirm premises (1) and (2) and (3) at the beginning of this chapter. The neurobiological mechanisms involved in psychosomatic medical phenomena operate as they do *in virtue of* the belief processes made globally available for conscious consideration. The psychosocial context surrounding the administration of the placebo, along with the cultural milieu that provides a backdrop of semantic meanings and associations, engender a consciously-held expectation in the patient. It is *\*this expectation\** that figures into the explanatory chain, at least insofar as relevant explanatory practices and empirical considerations are concerned. The strong reductionist will find himself unable to account for psychosomatic medical phenomena so long as the Real Patterns that bear causally are only usefully projectable at the level of expectations and beliefs. It seems clear, then, that we should take this evidence seriously. For now, then, it seems irresponsible to allow specious analytic arguments about what kinds of “overdetermination” are permitted to trump such evidence. We have already enumerated, my means of manipulation and OSR, responsible empirical tools capable of adjudicating between cases of legitimate causation and mere correlation. It is those tools, and the methods of empirical science in general, that should be relied upon in such cases – not logical constructions which require smuggling in metaphysical premises at the outset.

## Conclusion

At the beginning of this thesis, it was our aim to show that questions about mental causation are generally misframed, and rest on premises that only appear superficially to be in line with empirical naturalism. The choice we are offered between a crude, Kim-style reductionism and epiphenomenalism, both of which are inhospitable environments in which to construct a meaningful account of free will, is a false one. This choice is only arrived at, however, due to metaphysical baggage – the baggage of substance ontology and the picture of causation that follows naturally from it. As Hume showed us centuries ago, however, we need not take up this baggage ourselves. If OSR allows us to account for the data while making fewer metaphysical commitments to conclusions that lie beyond it, we should be encouraged to accept it – if there is data to suggest it indeed represents a more accurate conception of the physical world so much the better.

We have argued, along with others, that such data does exist, and it is compelling. If this is true, the consequences for the debate about mental causation and “exclusion” seem to be enormous, for as we have shown, such arguments are fundamentally constructed within the gestalt of the containment metaphor, with all “honest-to-goodness” causation reducing to microbangings. In addition to providing evidence that this view is groundless, we take ourselves to have shown that there is also meaningful evidence for mental causation from within existing neuroscientific explanatory practices. While this doesn’t at all prove that indeterminism is correct, what it *does* show is that a responsible commitment to science and to the PNC does not mean that our decisions are made at the layer of physics and chemistry. It certainly doesn’t mean that what we refer to as our “selves” are somehow meaningfully “bypassed” by the forces of the universe, as Nahmias puts it.

It is quite likely that as we come to better understand the causal entanglements that characterize an organism embedded in an environment, including the relevant neurobiological processes, that dynamic systems or “complexity” theory will play a crucial role. The causal entanglements that lead to the aforementioned “snarl” in the case of patterns involving living organisms are likely to involve feedback mechanisms both intrinsic (contained within the body) and extrinsic (involving psychosocial context, cultural milieu, the influence of other agents in a group.) While all of the involved processes *could* of course be described in terms of physics and A-level chemistry, our contention has been that such a description may not be *explanatory* insofar as it fails to capture causal relationships between processes at larger, conceptual/psychological scales of organization. The ability of a person to be swayed by the mood of a room, for instance, or to change his mind on a political issue based on a book he has read, means all causation is unlikely to supervene locally on fundamental physics, *even though a fundamental physical description is always available in principle*. This does not mean that the putative causes are *non-physical*, but rather that what is physically instantiated has causal relevance *in virtue of its conceptual content and its conceptual content alone* – a view crude reductionisms are unable to capture. It may be possible in the future to describe patterns projectable at the mental/conceptual level through the use of strange attractors (Neil 2007,) another way in which complexity theory may pave the way toward a more nuanced understanding of how mental causation works.

Some neuroscientists have been hard at work trying to understand the role of mental patterns, as well. Peter Ulrich Tse, in his book *The Neural Basis of Free Will*, outlines a number of extremely plausible mechanisms by which large-scale patterns in the nervous system could, in virtue of their global properties, alter the properties of component processes projectable at the

cellular or molecular level. (2013) While he takes these processes to be a vindication of traditional intuitions about libertarian free will, I'm not entirely convinced he supports this case adequately. We can safely disagree with Tse, however, that any meaningful conception of free will must incorporate a principle of alternative possibilities. As we have contended in this thesis, it is sufficient to show that the deterministic processes that lead to our behaviors encompass those we experience as our conscious decision-making, deliberating, and reflecting. We are not bypassed by the physical forces that produce our thoughts and actions in the world, but rather we are an integral part of them.

As physics continues to push ahead into realms that challenge our intuitions about the fundamental nature of reality, the duty will press upon us ever stronger to resist smuggling them back into our theories and interpretations through the back door. Working out how consciousness arises from physical systems is bound to involve a lot more hard work, and much of it will surely be worked out using neuroscience. We cannot afford to entrust such a vital endeavor to the *a priori* convictions of analytic metaphysics. There's simply no good reason for us to believe that answering the question naturalistically will divest us of our agency, and of the power of our mental lives.

## References:

Auyang, S. Y., 1995, *How is Quantum Field Theory Possible?*, Oxford-New York: Oxford University Press.

Barnett MW & Larkman PM. "The action potential," *Practical Neurology* 7 (3): 192–7. June 2007.

Batterman, R. W. (2002). *The Devil is in the Details*. Oxford: Oxford University Press.

Benedetti F, Pollo A, Lopiano L, Lanotte M, Vighetti S, Rainero I. Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/nocebo responses. *J Neurosci* 2003;23:4315-23.

Benedetti F, Mayberg HS, Wager TD, Stohler CS, Zubieta J-K. Neurobiological mechanisms of the placebo effect. *J Neurosci* 2005;25:10390-402.

Benedetti F, Amanzio M, Rosato R, Blanchard C. (2011) Nonopioid placebo analgesia is mediated by CB1 cannabinoid receptors. *Nat. Med.* 2011 Oct 2;17(10):1228-30.

Block, N., 2003, "Do Causal Powers Drain Away?" in *Philosophy and Phenomenological Research*, LXVII, 1 July.

Butterfield, J., 1998, "Determinism and Indeterminism," in *Routledge Encyclopedia of Philosophy*, E. Craig (ed.), London: Routledge.

Campbell, R. J., Bickhard, M. H. (2011). Physicalism, Emergence and Downward Causation. *Axiomathes*, 21: 33-56.

Cartwright, N. “Two Theorems on Invariance and Causality”, *Philosophy of Science*, 70 203-24, 2003.

Cassirer, E. *Determinism and Indeterminism in Modern Physics*. New Haven: Yale University Press 1936.

Choi, Charles Q. “A Vacuum can yield flashes of light” *Scientific American*. February 2013.

Dennett, Daniel C. “Real Patterns.” *The Journal of Philosophy*, Vol. 88, No. 1. (Jan., 1991), pp. 27-51.

Dennett, Daniel C. *Intuition Pumps and Other Tools for Thinking*. W. W. Norton Inc, New York 2013.

Dieks, D. and A. Lubberdink, 2011, “How classical particles emerge from the quantum world”, *Foundations of Physics*, 41: 1051–1064.

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy*, 66/23 (1969), 829-39

Handy TC, Green V, Klein RM, Mangum GR (2001) Combined expectancies: event-related potentials reveal the early benefits of spatial attention that are obscured by reaction time measures. *J exp Psychol Hum Percept Perform* 27: 303-317

Hättich, F., 2004, *Quantum Processes — A Whiteheadian Interpretation of Quantum Field Theory*, Münster: agenda Verlag.

Henson, B. et al. “Experimental loophole-free violation of a Bell inequality using entangled electron spins separated by 1.3km,” *Nature* 526, 682-686. 2015

Kane, R. "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism."

Journal of Philosophy 96: 217-240. 2005.

Kim, J. Physicalism, or Something Near Enough, Princeton: Princeton University Press 2005.

Kim, J. "Making Sense of Emergence." Philosophical Studies 95: 3-36 1999.

Ladyman, J. & Ross, D. et al. Everything Must Go: Metaphysics Naturalized, New York:

Oxford University Press 2007.

Laplace, P., 1820, Essai Philosophique sur les Probabilités forming the introduction to his

Théorie Analytique des Probabilités, Paris: V Courcier; repr. F.W. Truscott and F.L. Emory

(trans.), A Philosophical Essay on Probabilities, New York: Dover, 1951 .

Lieberman M, Jarcho J, Berman S, Naliboff B, Suyenobu B, Mandelkern M, Mayer E (2003)

The neural correlates of placebo effects: a disruption account. NeuroImage 22 (2004) 447-455

Meulenberg, A. Jr. "Virtual and Real Photons," Proc. SPIE 8121, The Nature of Light: What are

Photons? IV, 812114 (September 28, 2011)

Nahmias, Eddy. "Intuitions about Free Will, Determinism, and Bypassing." The Oxford

Handbook of Free Will, 2nd Ed. Oxford University Press (2011)

Pike, O.J. Mackenroth, F. Hill, E.G. and Rose, S.J. (18 May 2014). "A photon–photon collider

in a vacuum hohlraum". Nature Photonics. doi:10.1038/nphoton.2014.95.

Price DD, Milling LS, Kirsch I, Duff A, Montgomery GH, Nicholls SS (1999) An analysis of

factors that contribute to the magnitude of placebo analgesia in an experimental paradigm. Pain

83: 147-156

Reichenbach, Hans. *Experience and prediction: an analysis of the foundations and the structure of knowledge*. University of Chicago Press, 1938.

Seibt, J., 2002, "The matrix of ontological thinking: Heuristic preliminaries for an ontology of QFT," in Kuhlmann et al. 2002, pp. 53-97.

Spinoza, Benedict. *Ethics*, repr. Edwin Curley (trans.), Princeton: Princeton University Press, 1994 101-106

Tegmark, M et al. [SDSS Collaboration], *Phys. Rev. D* 69, 103501 (2004); D.J. Eisenstein et al., *Astrophys. J.* 633, 560 (2005); J.K. Adelman-McCarthy et al., *Astrophys. J. Suppl.* 162, 38 (2006).

Titov, A.I. Kämpfer, B. Takabe, H. and Hosaka, A. (2013) *Physics Review A* 87, 042106 April 2013

Tse, Peter Ulrich (2013). *The Neural Basis of Free Will: Criterial Causation*. MIT Press.

Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.