
HIM 1990-2015

2014

The Incompatibility of Freedom of the Will and Anthropological Physicalism

Ariel Gonzalez
University of Central Florida



Part of the [Philosophy Commons](#)

Find similar works at: <https://stars.library.ucf.edu/honorstheses1990-2015>

University of Central Florida Libraries <http://library.ucf.edu>

This Open Access is brought to you for free and open access by STARS. It has been accepted for inclusion in HIM 1990-2015 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

Recommended Citation

Gonzalez, Ariel, "The Incompatibility of Freedom of the Will and Anthropological Physicalism" (2014). *HIM 1990-2015*. 1570.

<https://stars.library.ucf.edu/honorstheses1990-2015/1570>



THE INCOMPATIBILITY OF FREEDOM OF THE WILL AND ANTHROPOLOGICAL
PHYSICALISM

by

ARIEL J. GONZALEZ

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Philosophy
in the College of Humanities and
in the Burnett Honors College
at the University of Central Florida
Orlando, Florida

Spring Term 2014

Thesis Chair: Travis Rodgers

ABSTRACT

Many contemporary naturalistic philosophers have taken it for granted that a robust theory of free will, one which would afford us with an agency substantial enough to render us morally responsible for our actions, is itself not conceptually compatible with the philosophical theory of naturalism. I attempt to account for why it is that free will (in its most substantial form) cannot be plausibly located within a naturalistic understanding of the world. I consider the issues surrounding an acceptance of a robust theory of free will within a naturalistic framework. Timothy O'Connor's reconciliatory effort in maintaining both a scientifically naturalist understanding of the human person and a full-blooded theory of agent-causal libertarian free will is considered. I conclude that Timothy O'Connor's reconciliatory model cannot be maintained and I reference several conceptual difficulties surrounding the reconciliation of agent-causal libertarian properties with physical properties that haunt the naturalistic libertarian.

DEDICATION

Debido al amor de mi madre

ACKNOWLEDGEMENTS

I'd like to begin by thanking my friends for the lively yet solemn conversations we've had on issues of great philosophical importance. I'd like to acknowledge Jordan Bell, Doug Zimmerman, Phil Maikkula, Michael Gaeta, Matthew Gilman, Kim Le, Sawyer Frescoln, Brad Rahr and Ben Cook (not listed in order of importance). I'd also like to express my deepest gratitude to my committee—Travis Rodgers, Mason Cash, and Valerie Sims—for their patience with me during my drafting of this thesis and their many suggestions and critiques which forced me to engage in more philosophical precision during the writing of this thesis.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION	1
The Challenge of Free Will	6
The Incompatibility of Event Causation and Free Will.....	8
John Searle’s Libertarianism	17
A Hierarchical Conception of Reality	19
CHAPTER 2: DEFINING “FREE WILL”	23
Necessary Conditions for Freedom.....	24
Choice and Contrastive Explanation	31
The Supervenience Relation and Emergence	36
Structural Properties and Emergence.....	41
CHAPTER 3: NATURALIZING FREE WILL	45
Causal Unity and Non-Reductionism	47
Creatio Ex Nihilo and Emergence.....	50
Naturalistic Libertarianism vis-à-vis Panprotopsyhism	53
Free Will and Supernaturalism (Conclusion)	58
BIBLIOGRAPHY	61

CHAPTER ONE: INTRODUCTION

Many contemporary naturalistic philosophers have taken it for granted that a robust theory of free will, one which would afford us with an agency substantial enough to render us morally responsible for our actions, is not itself conceptually compatible with the philosophical theory of naturalism¹. I will attempt to explain why it is that free will (in its most substantial form) is not plausibly located within a naturalistic understanding of the world. Free will (of the agent-causal variety) will be argued to be conceptually irreconcilable to a naturalistic picture, given the supernaturalistic elements required for this variety of free will to exist. In what follows, I will define the central philosophical concepts relevant to the thesis.

After doing so, I will discuss the issues surrounding an acceptance of a robust theory of free will within a naturalistic framework. I will discuss Timothy O'Connor's reconciliatory effort in maintaining both a scientifically naturalist understanding of the human person and a full-blooded theory of agent-causal libertarian free will. I conclude that Timothy O'Connor's reconciliatory model cannot be maintained by reference to several conceptual difficulties surrounding physical properties that haunt the naturalistic libertarian.

¹ John Bishop acknowledges this traditional view and defends it in John Bishop, "Prospects for a naturalist libertarianism: O'Connor's persons and causes," *Philosophy and Phenomenological Research*, 66, no. 1 (2003): 228-243.

As was mentioned, contemporary naturalism is normally understood to have eschewed the notion of free will within its ontology² (by ontology, I refer specifically to the entities postulated as constituting naturalism's exclusive domain of discourse). According to naturalism, the spatiotemporal universe(s) which consists of physical entities such as quarks, molecules, trees, planets, galaxies and other spatially and temporally situated entities (all entities which are essentially publicly-accessible) are all that exists³. The thought here is that every entity and every feature (property) belonging to any entity must be describable within the languages of the physical sciences (or at least 'reducible' to languages within the physical sciences). Different strategies of reduction are possible, but the essential claim is made that all entities in existence are physically describable using the theoretical predicates found in the physical sciences.

This ontological list of entities is derived from a strict adherence to the scientific method, a methodology used as the principal means by which one knows what exists at the fundamental level (it is an epistemological stance). Merely assenting to the scientific method does not make one a naturalist, but rather allegiance to the principle that the scientific method is the most authoritative epistemic route is what makes one a faithful naturalist. This epistemic stance explains why naturalists are usually reluctant in accepting entities such as numbers and sets (which are paradigmatically non-physical and non-causal entities) into their ontology. They will usually reduce them to physical features within this universe, or they might just eliminate them from existence. Some naturalists have formulated arguments to the effect that there is no

² Ibid.

³D.M. Armstrong proposes this characterization of Naturalism in D.M. Armstrong, "Naturalism, materialism and first philosophy," *Philosophia*, 8, no. 2-3 (1978): 261-276,

reason to admit of their existence due to their inefficacy in causally impacting the physical events of the universe.⁴

There are similar reductive and eliminative maneuvers which can be found within the literature on the philosophy of mind, particularly regarding the ontological status of mental states (thoughts, pains, smells etc.), and specifically whether mental states are appropriately physical or non-physical features of the brain.⁵ The naturalistic maneuver is to either eliminate any first-person points of view which are normally found in conscious experience (my private experience with pain, or my private thoughts), or to reduce and situate them alongside entities which enjoy public-accessibility.⁶ For example, there is no first-person perspective to chairs since the features of chairs are publicly accessible to any external observer and are not observer-dependent. The naturalist wants states of consciousness to enjoy a similar status as that of the physical states of chairs, desks, planets, and galaxies.

The hard problem of consciousness, however, is that we understand the sensation of pain to be privately accessible (only I can feel my pain), and yet this is an actual feature which I possess. Is this feature physical? If so, why is it not publicly accessible to any observer? Scientific entities and features are normally considered to be publicly observable. My brain is observable by more than one person, and the “physical” features of the brain (neuron-firings, electrical activity etc.) are accessible to anyone. However, when it comes to the mental features

⁴See Jaegwon Kim, *Supervenience and Mind*, (New York: Cambridge University Press, 1993), 269-270.

⁵ *Ibid*, 267.

⁶ *Ibid*.

of the brain (pain, thoughts, beliefs, desires, and intentions), only the possessor of these mental states can access them and have an awareness of their existence. The notion of the privately-accessible is not easily reconcilable with a naturalistic epistemology since the properties or features of the mental life cannot be observed in the same way that, say, other features of the brain or of the central nervous system can be observed.⁷

At most, all one can do is to believe other individuals when they report to us their own mental experiences. The naturalist will have to find a way in which she can either eliminate the existence of mental states from her ontology, or she can reduce them to (render them nothing over and above) a third-person feature of the brain (as opposed to a first-person feature inaccessible to scientific investigation and thus not “naturalistic”). Every philosophical naturalist is going to be a physicalist, but not every physicalist is committed to the naturalistic project. Physicalism is the thesis that every event is physical (this definition will suffice for this section, though the thesis of physicalism will be elaborated upon in the next section). The physicalist is not committed to the denial of the ontological reality of numbers, for instance. The physicalist is simply committed to the claim that every substance (every entity possessing features or properties) is entirely describable using the language of the physical sciences. This is the philosophical vocation of the naturalist and in doing so she embraces the title of ‘physicalist’ regarding mental states.

⁷ Richard Swinburne, *Mind, Brain, & Free Will*, (Oxford: Oxford University Press, 2013), 87-93.

In this thesis, I will be discussing what I call “Anthropological Physicalism” which is to say that I will be focusing on the doctrine of physicalism only insofar as it relates to the human person as a substantial entity. I will reference naturalism only insofar as it serves as the backdrop or justification for O’Connor’s view of the human person as a physical organism (and substance), though he grants that consciousness is not a physical property of this physical organism. To be clear, O’Connor is not a physicalist in the category of property (though he holds that all human persons are physical in the category of substance).⁸ O’Connor personally grants the non-physicality of consciousness (this will be discussed in Chapter 2), but his theory of free will is intended to be a thoroughly naturalistic and even physicalistic picture of free will. In other words, his theory of free will is supposed to be acceptable to physicalists and naturalists, even if he personally does not accept some of the positions (regarding the ontology of mental states) held by both.

O’Connor advances an agent-causal theory of libertarian free will, which is a non-deterministic understanding of free will and he wants to wed this view to an ontologically naturalistic picture of the world. I will argue that while his agent-causal theory of free will does not suffer from any obvious deficiencies in its robustness, the theses of naturalism and physicalism cannot allow for his robust account to be naturalized in principle. In this thesis, I will argue that O’Connor’s view of free will not only renders his attempt to reconcile

⁸ O’Connor holds that human persons are emergent individuals; mental states are emergent upon the biological organisms that possess them. Humans are fundamentally physical substances, even in spite of having non-physical mental properties. See Timothy O’Connor, and Jonathan Jacobs, "Emergent Individuals," *The Philosophical Quarterly*, 53, no. 213 (2003): 540-555, <http://www.jonathandjacobs.com/resources/papers/OConnorJacobs-EmergentIndividuals.pdf> (accessed April 14, 2014).

physicalism and free will impossible, but it also creates insurmountable worries for a reconciliation between naturalism and O'Connor's view of free will, seeing that naturalism plays a pivotal role as the basis for O'Connor's reconciliatory project.

The Challenge of Free Will

To understand how free will poses a threat to naturalism (or perhaps, how naturalism poses a threat to free will) one must understand that naturalism is usually associated (in some version or other) with physicalism. By physicalism, what is meant is that everything is composed of fundamentally physical entities (e.g. atoms, quarks, fields) and all things (substances) act in accordance with the laws of nature.

The concept of "physical" is constrained by what we find within the naturalistic epistemic stance such that what we deem to be physical must be described in the language of physics, chemistry or in the language of any other physical science. It is noticeable how physicalism is in some sense or other parasitic upon the thesis of naturalism. Both theses betray a deep reverence for knowledge acquired through an appropriation of the scientific method. While physicalism is notoriously difficult to define, the most precise definition of physicalism in my estimation has been offered by David Chalmers who defines physicalism as:

“Physicalism is true of our world iff any world that is a physical duplicate of our world is either a duplicate of our world simpliciter or it contains a duplicate of our world as a proper part.”⁹

Another way to understand physicalism is by way of the causal closure thesis, which is 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*.”¹⁰

The purpose behind the causal closure principle is that in order to have an ideally complete physical theory, it must be able to explain every physical event in terms of a physical cause (which turns out to be another physical event). This finalized physical theory cannot leave out an explanation for any particular event. If there is no physical event which causally explains a particular physical event, then this “theory of everything” is incomplete and physicalism must be false since there are non-physical causal explanations for certain physical events. In order to have a causally unified scientific picture, one must assume that the universe is causally closed from non-physical causal powers. Naturalism is similarly committed to such a notion. Moreover, the causal closure of the universe requires the conservation of energy and therefore precludes the injection of energy into the universe by non-physical events. The causal closure thesis is relevant to the present examination of the incompatibility between a robust theory of free will

⁹ This simple formulation of Chalmers’ definition is provided in Howell, Robert. *Consciousness and the Limits of Objectivity: The Case for Subjective Physicalism*. Oxford: Oxford University Press, 2013.

¹⁰ Jaegwon Kim has formulated this causal closure principle consistent with physicalism in Jaegwon Kim, *Physicalism, or Something Near Enough*, (Princeton: Princeton University Press, 2005), 15.

and both naturalism and physicalism, and so the causal closure thesis' impact upon this discussion will be elucidated after a discussion on the nature of free will, which I turn to now.

The Incompatibility of Event Causation and Free Will

In order to understand the conceptual dissimilarity between event causation and what O'Connor's robust notion of free will requires (namely, agent causation), event causation must be adequately defined. The agent is understood to act freely as a constituent of an event by determinists and event-causal libertarians ("libertarian" referring to a non-deterministic or indeterministic understanding of human action). The various theories of free will which require deterministic (or indeterministic) events in order for humans to act freely will not be exhaustively discussed, but a broad description will be offered (though the deterministic account of free will shall be discussed further in Chapter 2).¹¹ The notion of event causation (deterministic and otherwise) will be the focus of the discussion in this section.

The thesis of determinism states that every present event is entailed by a full description of all past events conjoined with a description of the laws of nature. A full description of both every past event (or a full description of the Universe at a particular time t[x]) and the relevant laws of nature would be sufficient to accurately foreknow a complete

¹¹ A comprehensive treatment of the possible compatibility between free will and determinism can be found in John Martin Fischer, "Compatibilism," *Four Views on Free Will*, (Malden: Blackwell Publishing, 2007), 44-84.

description of any future event (or any event whether past or future at any other time $t[y]$).

This would require every future event to be an inevitable consequence of past events in conjunction with the laws of nature.¹²

David Lewis expressed the thesis of determinism by asking us to imagine a possible world (namely ours) that has a specific collection of past events. If one were to survey the infinite collection of possible worlds and fail to find a possible world that has the same past as ours and yet has a different future than our world, then our possible world is a deterministic world.¹³ In other words, (for the determinist) every possible world with the same past as our world will necessarily have the same future (and that is what would render our world deterministic). This would be due to the fact that the past causally necessitates the present and future. There could not be two possible deterministic worlds with the same past up to the year 2014 which causally diverge in the year 2014.

Another way to express this thesis would be in terms of the occurrence of an event. According to determinism, events necessitate subsequent events to occur. Every possible world with the same series of past events as a deterministic world will have (by necessity) the same type of present and future events in that world. An event is the instantiation (or the realization)

¹² John Bishop formulates the thesis of determinism in this fashion in John Bishop, "Prospects for a Naturalist Libertarianism: O'Connor's Persons and Causes," *Philosophy and Phenomenological Research*, 66, no. 1 (2003): 230-231.

¹³ David Lewis, "New work for a theory of universals," *Australasian Journal of Philosophy*, 61, no. 4 (1983): 359-361.

of a substance's (or thing's) possession of a property (or feature).¹⁴ More specifically, an event is a change of state within one or more substances.¹⁵ By instantiation, I merely refer to the coming to be of a feature (I use "feature" synonymously with the term "property") possessed by a substance. For example, a brown chair that is painted red will be involved in an event (as it will possess a new feature at some time or other at some time in the future). In this example, the future event will be caused by the past event of the painting of the chair. The painting of the chair is an event which consists in the painter's possession of many features (e.g. possessing a brush, intention to paint). This event is itself a change of state owing its existence to yet another temporally prior event.

In these instances, the person qua agent (or substance) is not responsible for the painting of the chair but she is a constituent of an event along with the chair, her brush, and even her mental intention (and other occurrent mental states). The complete event is what contributes to the occurrence of the painted chair. The human agent qua agent is causally inefficacious and is a mere constituent of a causally efficacious event. It is not the agent who causes the chair to be red, but the agent's intending and carrying out her intention to paint with her brush (this change in state is an event). In this view, the states of the agent do all of the explaining. Physical events operate according to the laws of nature (whether deterministic or probabilistic) and agents themselves do not violate these laws of nature in this event-causal

¹⁴Here I follow Jaegwon Kim's classical defense of the property-exemplification account of events found in Jaegwon Kim, "Events as Property Exemplifications," *Action Theory*, 97 (1976): 159-177.

¹⁵ See Richard Swinburne, *Mind, Brain, & Free Will*, (Oxford: Oxford University Press, 2013), 6.

understanding of action. Agents qua agent do not make a difference to the occurrence of an event; only agents insofar as they are parts of events make a difference on this view.¹⁶

While events are connected by laws of nature, but some models of free will allow for an agent qua agent (in the absence of being part of an event) to execute intentions which cause events (the carrying out of an action). In the standard agent-causal libertarian model of free will, an agent is not causally influenced by prior events to cause further events via some connecting law between both events (as this would make the agent a constituent of the initial event). Rather, the agent possesses a sui generis (of its own kind) ability to execute certain intentions in the absence of any event of which the agent may be a part. Therefore, one has fundamentally two kinds of causation in reality: Event-causation and agent-causation. The free agent cannot be subsumed under certain laws which describe the regular causes and effects between events. The agent in itself must be the appropriate source for her own control.¹⁷

The model of free will which will be discussed eschews a deterministic (or even probabilistic) nomological mechanism for the exercise of the will. It also disavows the notion that even probabilistic laws can ground freedom. If the action of an agent is contingent upon an external non-deterministic (probabilistic) antecedent event which originated outside of the agent's control, it would still not be sufficient to be classified as a free action.¹⁸ Likewise, if the

¹⁶ Timothy O'Connor competently explains the differences between an agent as cause and an event as cause in Timothy O'Connor, *Persons and Causes*, (New York: Oxford University Press, 2000), 67-73.

¹⁷ An exhaustive account of this model (the agent-causal libertarian model) can be found in Timothy O'Connor, "Agent Causation," *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. Timothy O'Connor (New York: Oxford University Press, 1995), 173-200.

¹⁸ This non-deterministic event-causal view will be further elucidated in the next section.

event of an agent's formation and execution of an intention (in the form of an action) has been probabilistically fixed by prior events, then the agent is still a mere constituent of an event and makes no significant causal contribution as an agent over and above the event of which she is a part of causally contributes.

The agent would still be part of an event whose causal connection with prior and subsequent events is probabilistic as opposed to being merely deterministic. For the agent to have a final say in the matter, or to settle the action, she must act spontaneously as a first cause of her action in the absence of past events that determine any act of the will. The agent must exercise some intrinsic faculty (or power) and this exertion must be such that it occurs in the absence of any prior causally sufficient conditions. This absence of prior causally sufficient conditions requires the agent to possess an intrinsic capacity to exert active causal influence.

I will sketch what I consider to be a full-blooded conception of free will, but a defense of this view will be delayed for the next chapter. Actions exercised by the person's will are explained by the reasons the agent had for acting in order to fulfill the content of those reasons. A choice (or decision) is a settled intention to act (immediately or at some point in the future), and the will is the faculty which brings about the formation of an intention. Choices may or may not result in an external action (such as the choice to raise one's hand while it is severely damaged). Since one is intending as an agent to directly execute an action, this qualifies as an exercise of agency (more precisely, an exercise of rational agency due to the reasons involved in forming the intention to execute an act). The freedom of the will refers to

the fact that the agent is not restrained in her forming of certain intentions by her faculty (the will) and the execution of those intentions is similarly unrestrained (though the resulting external act may be unsuccessful due to some external impediments). The phrases “freedom of choice” and “freedom of action” similarly convey this familiar point (in my view, choices are always made for reasons, there are no unconscious choices).

One may contrast this with merely being acted upon, wherein we are merely passive participants in events which are casually produced by prior events which subsequently cause us to enter into a state which causes us to act. As stated earlier, agent-causal power (the capacity or power to exercise one’s rational agency in forming intentions to act) must be exercised in the absence of prior causally sufficient conditions¹⁹. In order for there to be free will, the state of the brain at a time $t(1)$ does not determine the state of the brain at $t(2)$. In other words, prior to forming an intention at $t(2)$, the state of my brain at $t(1)$ in conjunction with all of the other facts of the universe in conjunction with the laws of nature which may mutually determine future facts about me, do not (in concert with one another) determine my own forming of an intention (which may culminate in a decision) to act at $t(2)$.

There is a noticeable gap between the state of the brain at $t(1)$ and the state of the brain at $t(2)$, due to the fact that a free action originates in the absence of prior determining

¹⁹ A comprehensive account of such a power or capacity is found in Timothy O'Connor, "Agent Causation," *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. Timothy O'Connor (New York: Oxford University Press, 1995), 173-200.

conditions.²⁰ Nothing externally acts upon the individual's brain states so as to cause the brain to enter into an intention-forming state (which consequently executes the intention in the form of an action). For the agent-causalist, the agent injects causal influence upon events from above (so to speak) without herself being involved in an event (or caused to be in an event by a prior event) which is ultimately responsible for this causal influence. By contrast, an event-causal understanding of an agent's action would render the brain a passive participant within the event, which relies upon the energy-transfer of the previous state to cause the action. Matter is, as it seems, intrinsically passive or inert in that it relies upon external causal conditions (being in such and such a state or event) in concert with the laws of nature for it to make any causal contributions to any event.²¹

Physical objects do not seem to possess any intrinsically active causal powers (powers that are exercisable in the absence of any powers external to them which would cause or determine them to act). Free agents, on the other hand, must be able to exercise an agency that is by its very nature spontaneous and not determined (or even probabilistically fixed) by prior internal or external states of the agent. In a naturalistic universe, an individual with knowledge of every relevant physical state and feature combined with knowledge of the operating laws of nature would be able to predict what will occur in the future (assuming that the universe is fundamentally deterministic).

²⁰ Searle makes this astute observation in John Searle, *Mind: A Brief Introduction*, (New York: Oxford University Press, 2004), 232.

²¹ J.P. Moreland makes this very point in James Porter Moreland, "Naturalism and Libertarian Agency," *Philosophy & Theology*, 10, no. 2 (1997): 353-383.

An omniscient individual would be able to predict what would probably occur given certain probabilistically fixed laws of nature (if we allow for the possibility of a fundamentally indeterministic universe). This knowledge would only be possible if this individual had knowledge of a completed ideal theory of physics of some sort that we currently do not possess. If we assume a naturalistic ontology, whereby the only entities and features (or states) that exist are those that are characterized in the physical sciences, then this individual would have knowledge of the future universe.²²

A total understanding of all physical facts is sufficient to inform us of what the future will be like only because present (or past events) events determine (or entail) future events. Due to this naturalistic stance, the naturalist will hold that the physical universe is causally closed in the sense that there are no causal gaps. A complete knowledge of the state of the universe at time $t(1)$ will yield a state of the universe at $t(2)$ that can be reasonably entailed by the state of the universe at $t(1)$. There are not supposed to be any mind-boggling surprises if we have an omniscient understanding of all of the physical states and laws of nature at work at some time $t(x)$ in the universe. However, allowing for the existence of a rational individual with an agent-causal capacity (the ability to act as a cause of actions [for reasons] not caused by prior events and the ability to refrain from such actions [for reasons]) threatens this naturalistic outlook of the physical causal closure. No longer would knowledge of the complete universe at some prior time $t(x)$ entail certain results regarding future states of the universe. As was

²² Carl Ginet makes a similar argument regarding causal closure in Carl Ginet, *On Action*, (New York: Cambridge University Press, 1990), 92.

previously stated, the agent would inject causal influence (and thus new energy not derivable from previous physical events) upon events without being caused to do so (or being part of an event).

What if one wants to maintain a non-deterministic view of free action but desires to uphold an event-causal view? In this view, one must propose that the universe is not solely deterministic but is also indeterministic. In this case, a physical event x at $t(1)$ will not always necessitate physical event y at $t(2)$, but rather x might render it probabilistically likely (any probability under 1 may be assigned) that event y will occur at $t(2)$. In this case, the causal power of x simply fixes the probability assigned to the possible occurrence of y . This, however, is not to be taken as an example of intrinsic agent-causal power. Physical event x has a (let us stipulate) 0.5% chance of resulting in physical event y only due to the fact that x itself has a prior physical cause or state which necessitates or determines the probability it currently has (namely 0.5%) of producing physical event y . There is no causal acting “from above”; all causal energy is derived from previous events.

In other words, it has a derived causal efficacy from prior events as opposed to intrinsically possessing active agency to spontaneously produce events.²³ In a critique of event-causal libertarian Mark Balaguer, philosopher Derk Pereboom argued that if the state an agent is in at $t(1)$ is simply due to prior external events or even prior internal states, and that state is

²³ This indeterministic view of free will which does not require a causal relation between agent and event but consists in causation between probabilistic events is known as event-causal libertarianism. Robert Kane defends this view in Robert Kane, *The Significance of Free Will*, (New York: Oxford University Press, 1998).

what solely contributes to what occurs at t(2), then the agent itself (qua agent) makes no causal contribution. In other words, for an event-causalist, a probabilistic agent cause is just as passive and intrinsically inert as a deterministic agent cause (in the absence of any states or events).²⁴ Both of their properties are subsumed under laws of nature (property F affects property G due to a law of nature, but property F qua property F is not causally efficacious without being connected to another property via laws of nature). Due to the genuinely spontaneous power entailed by freedom of the will, rational agents have a dual-ability²⁵ to either form an intention or to refrain from forming an intention. Both are expressions of this sui generis capacity. I will now turn to consider a test-case in which naturalist John Searle espouses both libertarianism and an adherence to naturalistic principles.

John Searle's Libertarianism

The previous section illustrates the tough issue of achieving a commensuration between agent-causal free will and naturalism, but some philosophers have faced this challenge head-on and have attempted to elucidate their own naturalistic theories of free will which incorporate purely physicalistic entities. For instance, John Searle has recently advocated a version of naturalistic libertarianism (libertarianism referring to the indeterministic notion of free will

²⁴ I owe this insight to Derk Pereboom. Pereboom has a discussion of this issue in Derk Pereboom, *Free Will, Agency, and Meaning in Life*, (New York: Oxford University Press, 2014), 32-39.

²⁵ To borrow the terminology from Moreland in James Porter Moreland, "Naturalism and Libertarian Agency," *Philosophy & Theology*, 10, no. 2 (1997): 353-383.

currently in discussion). Despite his ardent naturalism, Searle has attempted to reconcile his naturalistic view of the human person with our own suspicion of freedom.²⁶ In reality, his view has more in common with the probabilistic event-causal view that Pereboom eschews.²⁷

His theory relies upon the notion that physical states and quantum indeterminacy probabilistically fix the chances that the agent will act in virtue of some reasons for action over others. He assures his readers that it is an “indeterminacy of a nonrandom kind”²⁸, but he fails to introduce any meaningful conceptual distinction between indeterminate brain states and random brain states. Searle is also committed to the causal closure of the physical universe²⁹ which requires every physical event to be wholly explained in terms of prior physical causes,³⁰ which would entail that an agent has no causal powers beyond the causal powers bestowed upon her by prior physical events, rendering her future choices probabilistically fixed by prior physical events.

Searle’s adherence to the causal closure principle precludes his endorsement of any notion of an intrinsically active power possessed by the agent. Searle’s naturalism requires him to argue that the agent’s causal powers (and the exercise thereof) are solely inherited from prior physical events. If the agent’s exercise of causal power is not due to any prior physical event, then it has no prior physical cause and would be a paradigmatic example of the violation

²⁶ John Searle, *Freedom and Neurobiology: Reflections on Free Will, Language, and Political Power*, (New York: Columbia University Press, 2007), 64-65.

²⁷ John Searle, *Mind: A Brief Introduction*, (New York: Oxford University Press, 2004), 232.

²⁸ Ibid.

²⁹ Ibid, 24.

³⁰ Jaegwon Kim, *Physicalism, or Something Near Enough*, (Princeton: Princeton University Press, 2005), 15.

of the causal closure of the universe. The agent-cause inserts causal power into the universe from above, energy which was not transferred from prior physical states or events. Despite these shortcomings, Searle's attempt to reconcile free will and physicalism is informative in that it demonstrates the incompatibility between the causal closure thesis (a principle of physicalism and naturalism) and free will. In the next section, I will consider what the causal closure thesis entails in terms of the causal inheritance possessed by higher-level features and contributed by lower-level phenomena. A distinction will be made between a naturalistic understanding of personal agency and the agent-causal understanding thereof.

A Hierarchical Conception of Reality

Naturalists usually view causation as being bottom-up. In other words, the fundamental physical entities are microphysical entities (e.g. particles). In reality there exists a hierarchy of entities that range from microphysical entities to macrophysical entities (e.g. chairs, humans, buildings, galaxies). The higher levels are causally and ontologically reducible to the lower levels, so that sciences that engage with lower-level phenomena (such as physics or chemistry) may always have their language translated into the scientific languages that refer to higher-level phenomena (such as geology or astronomy).

While the concepts behind these languages are not conceptually identical, they may nevertheless be causally and ontologically reducible to lower-level phenomena (since all of the

causal powers of geological and astral bodies are derived from lower-level entities such as atoms and quarks). In other words, we can entail (or determine) the higher-level entities from the lower-level entities. There is an obvious and undeniable dependency relationship between all levels of physical reality that can be accounted for in an ideally completed theory of physics. Any complete lower-level description of all of the entities will not leave out any causal powers or properties (features) which we find in the higher-levels due to the fact that the higher-levels derive their causal efficacy solely and completely from the lower-level entities.

There is what is known as a supervenience relationship.³¹ Higher-level entities and properties of entities depend upon lower-level entities and properties. Due to this, there is no change in a micro feature without some change in a macro feature (however small it may be). Similarly, there is no change a macro feature without some change in a micro feature. This is because the macro is nothing over and above the micro. The micro features are simply a narrower description of the broad whole (which may be described using the language of sciences that deal with macro-entities and features). The emergence in this view is simply an apparent “epistemological” emergence; our knowledge of the macro-properties is emergent upon the micro-properties (of which we may know nothing).³² In reality, the micro-properties are simply a narrower description of the broader macro-phenomena. There is a tight feed-back

³¹ The supervenience relationship and the naturalistic hierarchical understanding of reality are described in Searle, John. "Why I Am Not a Property Dualist." Accessed November 16, 2013. <http://www.imprint.co.uk/pdf/searle-final.pdf>.

³² The section “The Supervenience Relation and Emergence” in Chapter 2 of this thesis further explains the difference between the naturalistic understanding of emergent properties and O’Connor’s stronger understanding of emergence which he exclusively applies to the phenomena of mental states and free will (though he accepts the naturalistic reductionist model I mention above but he applies it to other naturalistic phenomena).

relationship between all levels of physical reality, though this is not to be taken as a suggestion that micro-entities have independent causal powers from the macro-entities (or vice-versa), but rather that the constituents of macro-entities are explained in terms of micro-entities (this is the bottom-up picture).

Now, contrast this with free will. Free choices are not determined by the lower-level microphysical entities. In fact, instead of there being a bottom-up causal chain (where the lower-level properties contribute towards the causal efficacy and causal manifestation of higher-level properties) there is the notion of top-down causation (or acting from above). The macro-entity (the human agent) forms an intention (in the absence of any bottom-up determining influence) and the executed intention causes the microphysical entities to produce an effect (e.g. the rising of one's arm). While the agent-causal capacity may exist in virtue of the micro-entities, the performance of an action is carried out by the agent and is not a consequence of a state or event in which micro-entities participate.

It is understandable why a naturalist would be uncomfortable with that proposition. For Searle, an advocate of the causal closure thesis, top-down causation (or top-down explanation) is only a reality in virtue of bottom-up causation. He argues:

“Since all of the surface features of the world are entirely caused by and realised in systems of micro-elements, the behavior of micro-elements is sufficient to determine everything that happens. Such a 'bottom up' picture of the world allows for top-down causation (our minds, for

example, can affect our bodies). But top-down causation only works because the top level is already caused by and realized in the bottom levels.”³³

In other words, all macro-level phenomena are explained in terms of micro-level phenomena. This precludes the notion that an agent has an irreducible active power to produce changes in her physical behavior, and that such a power (and its execution) is not explained in terms of deterministic or probabilistic micro-level phenomena. For Searle, every execution of an intention can be explained solely by reference to the events occurring at the micro-level (e.g. the interaction of molecules). The faithful physicalist has no other choice.³⁴

Perhaps not all is lost for the physicalist. Perhaps one may reject the causal closure thesis and still be a faithful physicalist and naturalistic libertarian. In what follows, I will discuss the conditions required for free will, only then to discuss a particularly naturalistic agent-causal libertarian theory of free will which attempts to reconcile freedom and a commitment to philosophical naturalism. In particular, I will discuss Timothy O’Connor’s theory of a naturalistic agent-causal libertarianism. In doing so, I will argue that philosophical naturalism (and more specifically, anthropological physicalism) is conceptually incompatible with what I understand to be robust freedom of the will (namely agent-causal libertarianism). In the next chapter, I will elucidate an agent-causal theory of free will.

³³ John Searle, *Minds, Brains, and Science*, (Cambridge: Harvard University Press, 1984), 94.

³⁴ No pun intended.

CHAPTER 2: DEFINING “FREE WILL”

There are variegated notions and conceptualizations of free agency that are represented by the philosophical literature³⁵, understandings which are commonly understood as being mutually exclusive and irreconcilable. The conceptualization that will be discussed is as follows: An agent exercises her free will if and only if:

- (1) The agent is the ultimate originator of an action (executed intention) at a time ‘t’
- (2) At some time or other prior and up to time ‘t’ (which is the time of the action), the agent had the inherent (intrinsic) capacity (as an intrinsic property) in conjunction with the ability (it is within the agent’s power) to either refrain from the action or to execute the action (also applicable to the formation or refrain from forming an intention to act)
- (3) The agent’s reasons (ends or purpose) are not solely sufficient to causally determine the agent’s choice of action, though the agent acts in virtue of such reasons (though not always consciously aware of such reasons) the content of which are directed toward a goal or teleological end.

These criteria comprise what has been traditionally understood to be an agent-causal account that presupposes the requirement of an agential source to ground responsibility (in

³⁵Libertarian and Compatibilist models of free agency as discussed in *The Oxford Handbook of Free Will*, ed. Robert Kane (New York: Oxford University Press, 2011).

other words, these criteria are necessary (and jointly sufficient) conditions for the agent-causal libertarianism I will be discussing. (1) Is the sourcehood condition,³⁶ (2) is the categorical ability condition, and (3) refers to the agential condition. In what follows, I shall consider each condition jointly and defend them against alternatives.

Necessary Conditions for Freedom

The sourcehood condition will be argued to be the foundation for human control in action from which the categorical ability condition follows and consequently, the agent-causal condition is to be understood in light of both. Traditionally, it has been argued that the ability to have done otherwise was the foundation and ground for agential responsibility. “Could have done otherwise” refers specifically to the idea that given two otherwise isomorphic possible worlds with the same description of a past sequence of events and the same set of laws of nature, the agent may either choose to exert her free will in acting or refrain from exerting her ability to act. The past events of both worlds are isomorphic and yet the agent makes radically different choices which are not determined (or probabilistically fixed) by the past.

If they were either determined or probabilistically fixed by past events, the relevant choices would be contingent upon past events. The understanding that the agent’s choices are

³⁶ Kevin Timpe uses this phrase to describe the agent’s being a source of her actions. Timpe advances a different argument than I do for this condition. For his argument, see Kevin Timpe, *Free Will: Sourcehood and its Alternatives*, (New York: Continuum, 2008), 103-118.

not in any way contingent upon past events is known as the categorical ability of an agent to have done otherwise (as opposed to the mere conditional or hypothetical ability, had the past been different). The Principle of Alternative Possibilities has been argued to have been the very grounds by which an agent is in control of her action.

The philosopher Harry Frankfurt illustrates his rejection of the Principle of Alternative Possibilities (shortened to PAP) as foundational of responsible agency by showing how an agent can act freely and yet be precluded from doing other than what she in fact had done. In the counterexamples he offers, the agent freely makes a choice in the actual world, but had she chosen not to have freely perform an action, she would have been coerced into making it. The idea here is that a hypothetical instance of coercion which acts as a fail-safe in case the agent does not freely choose in her own does not affect the fact that in the actual scenario she freely chose to make that same decision she would have been coerced into making had she chosen not to freely engage in that act.

Frankfurt recognizes the hypothetical scenario to involve compulsion and therefore precludes free choice because the agent is coerced either by physical threat or by an action-guiding device inserted inside of her brain which would coerce her to make the choice. She is asked to freely decide to perform an action (e.g. punch someone in the face) and if she decides to do so then she acted freely. If she had decided not to do so, she would have been coerced

into doing so. This is how Frankfurt eliminates PAP as a requirement for free agency, since the agent no longer has any alternative choices according to the argument.³⁷

There are several problems with his proposed counterexamples. Frankfurt assumes (but does not demonstrate) that a rejection of PAP will allow determinism to become compatible with moral responsibility and free agency.³⁸ PAP was a generally recognized principle; both determinists and non-determinists accepted it. I briefly noted earlier the categorical ability of being able to have done otherwise. This is to be contrasted with the conditional (or hypothetical) analysis of the ability to have done otherwise offered by deterministic accounts of free will.³⁹ In the conditional analysis, the agent could only have otherwise had past events been different in some relevant way such that the difference would have caused the agent to have done otherwise.

In other words, the agent would have done otherwise had some prior condition obtained which would have caused the agent to have done otherwise. For example, had I desired to have punched a stranger in the face then I would have done so. Note that the action is conditional upon the prior desire to do so but the desire may be determined by antecedent events, and the desire itself (along with other determinants) may have caused the action. This is known as a compatibilist account of free will, and Frankfurt seems to advocate this causal

³⁷ Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy*, 66, no. 23 (1969): 834-835.

³⁸ *Ibid*, 838-839.

³⁹ For a comprehensive account of both abilities and their relationship to the compatibilist-incompatibilist debate, see Kadri Vihvelin, "Free Will Demystified: A Dispositional Account," *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*, ed. Paul Russell, Oisín Deery (New York: Oxford University Press, 2013), 179-184.

understanding of human action, although he argues that the conditional analysis is not even necessary in order for a person to have free will in a deterministic context.

This issue becomes clear when Frankfurt argues that if one eliminates any alternative possibilities (as the counterexamples attempt to do) then determinism would seem to be compatible with free will. In the examples he gives, Jones would be coerced in a hypothetical scenario if she chooses not to do what Black wants⁴⁰. What Frankfurt fails to realize is that according to determinism, any action Jones performs (whether being coerced by Black or not) is still a form of compulsion (or necessitation). Her strongest desires cause her actions, while her desires are caused by antecedent events. Therefore, in the actual scenario in which Jones chooses what she desires (and is undisturbed by Black) she is still acting under compulsion (or necessitation) in that she is being acted upon by antecedent events and she is passively participating in an event that is beyond her direct control. The sources of the agent's actions do not truly originate from within the agent, but are instead traced to events ultimately outside of her control. Prior events necessitate her action, and so it seems to be fair to count her action as being compelled to occur. External necessitation and compulsion seem to be synonymous in these instances.

Therefore, eliminating the requirement of alternative possibilities would not seem relevant to formulating conditions necessary for moral responsibility. Even if we allow for alternative possibilities, the alternative possibilities envisioned by Frankfurt are understood as being conditional. Alternative possibilities within a deterministic universe are not categorical,

⁴⁰ Ibid, 834-837.

but rather a full description of Jones' beliefs, desires and disposition alongside a full description of the Universe prior and up to the instant of her decision will suffice to determine exactly what choice she will make.

A categorical understanding of PAP precludes a full description of the universe and the constitution of the agent up to the moment of decision from entailing or determining any choice. This implies that there is a causal gap between the moment prior to and the moment of the choice of the agent.⁴¹ The event prior to the choice and the event after the choice are not causally connected; the agent is causally connected to the event of the choice and the event caused by the choice. This causal gap is due to the fact that the causal efficacy of the agent was not determined by prior events but rather the agent contributes *sui generis* causal power since the agent acts as a first cause (as opposed to acting as part of an event).

Categorical PAP seems relevant to free choice, but hypothetical (i.e. conditional) PAP seems irrelevant to free choice because even when Jones performs an actual "free" choice he seems to make that choice by physical necessity. Frankfurt rightly eschews PAP but the PAP he rejects is the conditional understanding of it. Such an understanding of PAP is irrelevant to moral responsibility (due to the consequence of being externally necessitated to act). He has not yet formulated an argument against the categorical understanding of PAP. Here Frankfurt might protest that a categorical understanding of PAP has not yet been established as being relevant to free choice. If the conditional analysis fails, why should the categorical analysis fare

⁴¹ In my view, the causal gap closes at the moment of the action, for the agent creates a new causal connection with an event at that instant.

any better? My criticism of determinism did not make mention of the categorical analysis of PAP, so how can it be relevant to moral responsibility?

In response, I would argue that the principle of alternative possibilities (PAP) is a necessary but not a sufficient condition for moral responsibility. Traditionally it has been understood that to be morally responsible is nothing more than to simply be able to have chosen otherwise. I have already shown why this is an inadequate understanding of moral responsibility. Even determinists have an understanding of PAP which precludes genuine free choice. Moral agency cannot be understood simply in terms of PAP. There is another notion which is more fundamental, namely the requirement of origination (sourcehood) or of being an ultimate source for your action (as was mentioned earlier).

An agent can have the ability to do otherwise and yet fail to be the ultimate source of her choices. She can be the immediate source of her actions in that her intentions are formed by her reasons and deliberation. However, the act of deliberating and the beliefs and desires which form the intention are caused to do so by events that are traceable to events ultimately outside of the agent's control. In this respect, a deterministic theory of action fails to ground the control sufficient for moral agency. She is simply not in control of her choices. Here is where Frankfurt's possible objection would be placed: Why assume that a categorical understanding of PAP (one where past events do not determine whether or not a specific choice will be made and thus the agent is the difference-maker) is metaphysically significant as opposed to a conditional understanding of PAP?

PAP is relevant to moral agency but it is only in virtue of the agent's being the ultimate intrinsic source for one's actions that PAP becomes metaphysically relevant. PAP is merely a necessary byproduct of the agent's actions being traceable to the agent alone. The agent's control over forming intentions (not the agent's features forming intentions beyond the agent's control) is what grounds moral responsibility and what entails PAP as a byproduct. It is a byproduct of the agent's control only if the agent is not necessitated to cause the intention (action) and therefore by implication could have done otherwise. Note that PAP is not fundamental to our understanding of moral responsibility. The origination or source requirement is fundamental, but PAP is necessary in that the categorical analysis of PAP is entailed from the origination requirement. Free agency requires the agent to be the ultimate source of her action.

Moreover, free agency requires that nothing about the agent's constitution should determine or necessitate the choices she makes. This means that it is wholly up to the agent whether to act or not. Since it is not a part of the agent's constitution whether a specific choice is made, she can also refrain from making a choice. She does not act by a necessity of her own nature (choice x is not necessarily entailed by her own nature). For the agent to be free she must be the origin of her action and she must not act by necessity. This feature grounds a sufficient amount of control (as opposed to nomological or metaphysical necessitation).

This sourcehood entails a categorical understanding of the principle of alternative possibilities. Consequently, as the sole originator of her action, the agent's reasons for action

cannot be the sole cause of the agent's subsequent act of will.⁴² This feature would preclude any theories of free will which require the event of a free action to be completely causally determined by antecedent events which are not within the control of the agent (and do not ultimately originate from within the agent). I will now turn to the agent-causal notion of explanation in human action and O'Connor's strategy for dismantling possible objections to the absence of contrastive explanation in agent-causal human action. After that discussion, we should have an adequate understanding of agent-causal freedom (as O'Connor defends it) and objections to his view will be considered.

Choice and Contrastive Explanation

There is nothing about the reason acted upon that compels the person to act for that reason. If an agent is deliberating between two competing and equally impressive reasons, the only explanation to be given as to why the agent chose reason A over reason B is that the agent was simply more impressed with A rather than B. In O'Connor's view, there is no further explanation for that contrastive fact.⁴³ If there were an explanation as for why the agent was more impressed by A rather than B, then we would be able to determine beforehand what mechanism would make reason A seem more impressive to the agent than B (and this would

⁴² Randolph Clarke has developed an agent-causal model wherein reasons act as co-causes alongside the agent. I do not pursue this view here, but it will suffice to assert that reasons cannot be the sole cause of a free action. Clarke's integrated agent-causal account is defended in Randolph Clarke, "Toward A Credible Agent-Causal Account of Free Will," *Noûs*, 27, no. 2 (1993): 191-203.

⁴³ Timothy O'Connor, *Persons and Causes: The Metaphysics of Free Will*, (New York: Oxford University Press, 2000), 76.

become explained in terms of events that involve the agent as a constituent rather than being a purely agential cause, which is what O'Connor is attempting to avoid). To give an example, Laura has the opportunity to either (A) work for the family business, or (B) attend graduate school at a prestigious university. Laura initially finds both of these reasons to be equally compelling, but ends up choosing A as opposed to B.

One cannot give an explanation for the fact that Laura found the reason(s) supporting A more impressive than the reason(s) supporting B. She just found them more compelling and that is all. If one could explain why she found the reasons for A more compelling, then not only would one have an explanation for why she had chosen A, but one could also explain why she could not have chosen B. Given this explanation, she could not have chosen B but must have chosen A. This explanation would presumably lie outside of the agent herself (the reasons would likely have sufficient causal efficacy in producing the intention to act) and would have to determine how and why she chooses A over B.

No longer would the agent be responsible for the forming of the intention to act, but it would be the event of her possessing certain reasons which would explain the forming of the intention. The laws of nature would ensure that whenever an event of this type occurs (the same possession of the same reasons by the agent in identical circumstances), the same intention would be formed. This is a classic example of causation in terms of events (involving an agent) as opposed to causation triggered ultimately by the agent. In the former, the agent has no independent say in the matter (since events are essentially connected by laws of

nature). Only an agent-causal picture would allow for the agent to be in full control of her choices, although we would have to sacrifice a traditional understanding of explanation in order to give the agent full control in the forming of her intentions.

The same principle noted here applies to the timing of certain decisions. Why did Laura choose to act upon her reasons in March instead of in April? Reasons certainly played a significant role in this act, but more significantly, there were reasons involved which she found compelling (reasons regarding the timing of said decision). One might argue that her acting for these reasons is non-explanatory, but it can also be argued that her acting for these reasons is self-explanatory. It is self-explanatory in the sense that she chose to work for her family business in April because she found the reasons for it more compelling, in conjunction with the fact that her possession of agent-causal freedom allows her to act for reasons without there being a contrastive explanation for why she did not act otherwise. The explanation for why there is no contrastive explanation available is that free actions by definition cannot have contrastive definitions.⁴⁴ This does not mean, however, that free actions have no explanation simpliciter. The explanation for why Laura chose to work in the family business is that she found the relevant reasons associated with that choice ultimately persuasive.

An advocate for event-causal free will might argue that agent-causal libertarianism would refute the Principle of Sufficient Reason if it is the case that Laura's choice to act upon the reasons for A rather than the reasons for B is not explainable in principle other than the

⁴⁴ Timothy O'Connor, *Persons and Causes*, (New York: Oxford University Press, 2000), 91-95.

simple restatement of the fact in question (viz. she was simply impressed by the reasons for A over the reasons for B, which begs the issue). The Principle of Sufficient Reason is the thesis that every fact has an explanation for its obtaining. Leibniz (the most well-known formulator of this thesis) asserted: "...we hold that there can be no fact real or existing, no statement true, unless there be a sufficient reason why it should be so and not otherwise..."⁴⁵

One must follow the logic of this assertion. Not only is the explanans (the entity or proposition doing the explaining) going to explain why the explanandum (the entity in need of an explanation) is true, but it must also explain why it is not such that it must be false. In Laura's case, all she can demonstrate is that she chose reason for A because she was impressed by it (she can explain why the explanandum came to be true) but she cannot explain why it was not the case that she chose to carry out the reason for B instead (or in other words, why it cannot be the case that it is false that she did not choose and carry out A). Thus she must explain why it is not the case (or why it is false) that A was not chosen. All she can do, however, is simply restate the fact that she did chose A which does not answer the challenge to offer a contrastive explanation for why one state of affairs obtained as opposed to another.

In order for this contrastive condition to be fulfilled, the reason acting as the explanans must be sufficient for the action such that its existence is sufficient for the action (which consequently makes the action inseparable from the reason). Excising the action from its occurrence would eliminate the reason from existence. Similarly, affirming the existence of the

⁴⁵ Gottfried Leibniz, "The Monadology," *Modern Philosophy: An Anthology of Primary Sources*, ed. Roger Ariew, Eric Watkins (Indianapolis: Hackett, 2009), 278.

reason and the strength thereof would necessarily entail the action which follows. An agent-causal libertarian would argue against the necessary entailment relationship between a reason and the agent's acting upon it, but this is the situation the libertarian finds herself in.

If one wants to explain why this state of affairs is not such that choice A is not acted upon, one must appeal to the reason for A and that reason must act as a sufficient and necessitating condition which brought that state of affairs about. If the reason for A exists in the state and strength to which it does exist, it must have brought about the action associated with it. Agent-causal libertarianism does not sit well with this view, as a reason can exist in any state and still not compel an action to occur.

In this case, one may have to settle for a weaker version of the Principle of Sufficient Reason, since there are certain facts (such as contrastive facts) which cannot be explained. However, it may be that contrastive facts regarding free actions are self-explanatory. In other words, due to the nature of agent-causal freedom, certain facts do not require any further explanation. Paradoxically, the explanation for why there is no explanation is that no explanation is needed. Just as there are necessary self-explanatory facts (analytic truths such as: 'All bachelors are unmarried men') there are also contingent self-explanatory truths (truths regarding acts of freedom such as: 'Laura acted on reasons for A as opposed to reasons for B'). O'Connor has embraced a similar weak form of the PSR.⁴⁶

⁴⁶ Ibid.

O'Connor has argued that reasons (belief-desire pairs in his view)⁴⁷ are not sufficient to form an intention, though they are a necessary condition for an intention (and so weakly-explains why the agent acts). There is the remaining issue of internal influences upon our choices. Certain habits make it more likely that we prefer to act upon certain reasons over others. For O'Connor, reasons "increase the probability" of an agent's forming a certain intention. Reasons shape the propensity or likelihood of future intentions, but O'Connor makes it clear that reasons are not probabilistic causes of future intentions. Only the agent causes intentions, though her propensity to do so is structured by the reasons she possesses. Therefore, reasons structure the tendency or propensity of an agent to solely cause her intentions, though the reasons themselves are causally inefficacious.⁴⁸ Now that we have an agent-causal theory of libertarian free will from which to work with, in the next section we shall analyze the metaphysics of reduction and supervenience in order to adjudicate O'Connor's view vis-à-vis naturalism and physicalism.

The Supervenience Relation and Emergence

Having this model of free will in mind, one must ask whether or not this model is compatible with a broadly naturalistic worldview. Timothy O'Connor has proposed a view that would allow this form of free will (generally referred to as an 'agent-causal libertarian')

⁴⁷ Ibid, 72.

⁴⁸ Ibid, 95-101.

conception of free will) to be reconciled with our best empirical theories.⁴⁹ While O'Connor holds to a generally naturalistic view of the human person, he manages to allow for the irreducibility of consciousness (he is not a reductive materialist when describing mental states).⁵⁰ He is, however, a substance materialist regarding human persons. The human person is not an immaterial soul or a mystical entelechy, but rather a completely physical substance with full-fledged physical features which nevertheless exhibits irreducibly non-physical states of consciousness which supervene over the physical states and processes of the person (namely, brain states and processes). The notion of supervenience must be unpacked further.⁵¹

The supervenience thesis asserts that there is a necessary covariance between two properties such that if property A is supervenient on property B, there can be no changes in property B without there being changes in property A. Supervenience qua supervenience is a very broad notion that has many different definitions. Timothy O'Connor's model relies on a notion of mental supervenience which asserts that supervenient emergent mental properties are causally preserved by a subvenient base property (more on this later).

The pertinent point is that supervenience qua supervenience applies to more than one ontological relationship of covariance between properties. For instance, one may discover that A supervenes on B because A is identical to B (e.g. water is H₂O). Or perhaps A supervenes upon B because B (or a group of B) either constitutes part of or all of A. Or B non-causally explains A

⁴⁹ Timothy O'Connor, *Persons and Causes*, (New York: Oxford University Press, 2000), 108-110.

⁵⁰ Ibid, pp. 115-120.

⁵¹ The definition that is to be given is found in David Lewis, "New work for a theory of universals," *Australasian Journal of Philosophy*, 61, no. 4 (1983): 358-359.

(e.g. Murder is wrong because it is the killing of an innocent person). In the murder case, there is an explanatory co-variance (co-variance of A and B using an “in virtue of” relation).

In O’Connor’s case, the supervenience relationship most pertinent to his model of consciousness (and even free agency) is one that relates a causally sustaining subvenient base property to an irreducible emergent property (this description will be further elucidated).⁵² All of these aforementioned kinds of co-variance relationships between properties are subsumed under the umbrella of supervenience and can be described, as David Lewis himself described it, as a denial of independent variation and an affirmation of dependent variation. O’Connor argues that consciousness is an irreducibly simple kind of feature (he notes that it has “nonstructurality”)⁵³ that causally emerges from the complex causal interaction of the parts, processes and features of the physical organism.

These micro-physical parts have properties that are paired with other properties and mutually interact and manifest property-complexes, but when this pattern of structured substantial parts and properties reach a certain threshold of biological complexity, the whole organism exhibits a new simple feature that it did not initially possess.⁵⁴ This feature is not describable using scientific predicates or the objective languages of biology, physics and chemistry. The feature here is subjectivity or consciousness. The predicates of physics cannot begin to describe the feature or processes of subjective phenomenal states, according to

⁵² The description is to be found in Timothy O’Connor, *Persons and Causes: The Metaphysics of Free Will*, (New York: Oxford University Press, 2000), 117-118.

⁵³ *Ibid*, 116.

⁵⁴ *Ibid*, 110-115.

O'Connor. This particular claim will not be pursued here, but this assertion of emergence is relevant for O'Connor claim regarding the presence of free will within human organisms.

An emergent property, for O'Connor, is a nonstructural (simple) feature of an entity that cannot be reduced to (or be explained in terms of, or be identified with) a set or collection of properties belonging to the entity's parts or its whole.⁵⁵ In order to illustrate this by means of its opposite, consider a structural property (or feature) of an entity (or substance). Water is a paradigmatic example in this respect. Water has the property of being identical to H₂O. Is "identical to H₂O" a structural property or a nonstructural property? One can successfully identify H₂O with a set of properties belonging to its parts (or properties which are true of its whole), namely one oxygen atom (its part) having the property of being bonded with two distinct hydrogen atoms which have the similar property of being bonded to one another. The property of being "identical to H₂O" is nothing over and above (and therefore nothing more than) the possession of the aforementioned set of properties. Timothy O'Connor and Hong Yu Wong have endorsed a succinct definition for their understanding of a structural property, which is as follows:

"A property, S, is structural if and only if proper parts of particulars having S have properties not identical with S and jointly stand in relation R, and this state of affairs is the particular's having S."⁵⁶

⁵⁵ Timothy O'Connor, *Persons and Causes*, (New York: Oxford University Press, 2000), 111-112.

⁵⁶ Timothy O'Connor, and Hong Yu Wong, "The Metaphysics of Emergence," *Noûs*, 39, no. 4 (2005): 663.

By contrast, a simple (and therefore nonstructural) emergent property is one which is not constituted by a more basic set of properties and thus cannot be ontologically reduced to that set. It cannot be classified under an understanding of supervenience which relates structural supervenient properties with its more basic (and constitutive) subvenient base. Rather, the only way to understand a genuinely sui generis emergent property is to view it as being causally maintained (as opposed to being constituted) by its subvenient base. The subvenient base literally causes it to exist. O'Connor and Wong say as much when they opine: "Emergent properties are nonstructural properties of composite individuals. We further presume that they arise from and are sustained by underlying microstructures."⁵⁷

This explains O'Connor's view on conscious states of the brain. The sensation of pain seems to be utterly basic and irreducible to physical states, and O'Connor categorizes such sensations (or qualia) as emergent properties belonging to the brain. Note that the sensation is still a state of the brain, but rather than being reducible to a collection of physical properties of the brain, it is a genuinely unique feature of the brain that is caused by subvenient physical states (of the brain) rather than being constituted by these relevant states. For O'Connor, the subjective sensation of pain is not in any way identical to a mereologically structured set of objective states of neuron-firings in the brain. However, they are caused by them. In the next section, we will consider the distinctions between a weak understanding of emergence and a strong understanding of emergence.

⁵⁷ Ibid, 664.

Structural Properties and Emergence

In some of the philosophical literature on the topic of emergence, structural properties have been said to “emerge” but only in the sense that they have a different level of description than its more basic constitutive properties. In other words, they are semantically (or conceptually) emergent properties. This amounts to the claim that structural properties of a substance are described (and understood) differently than the individual properties (or even proper parts of the substance) individually considered. They are emergent only by way of its description. A new description entails a new concept which is expressed by said description. Simply because we have a new concept does not mean that we have a sui generis basic property that is caused to exist by its subvenient base.

We certainly learn new concepts from “emergent” properties that are structural in nature, but this knowledge simply amounts to knowledge of the interrelatedness of its spatial and geometric properties and the powerful interactions of its parts. We have a different concept of water as we know it than we do of H₂O as a molecular structure. Simply because we have these conceptual distinctions does not mean that water’s properties are caused by H₂O. One does not cause the other precisely because one is nothing but the other. They are one and the same. This weaker form of emergence is identical to the understanding of supervenience I briefly summarized earlier in this thesis; structural properties are ontologically identical (though not epistemologically indiscernible) to a set of simpler properties, even though one may

describe this identity relationship as one of co-variance such that the supervenient base is dependent upon the subvenient base.

In the literature, this form of “emergence” has been called “weak emergence”⁵⁸ and “epistemological emergence” in contrast to “strong emergence” and “ontological emergence”⁵⁹ which characterizes O’Connor’s description of emergent consciousness. Correspondingly, in the former model of emergence, there are no causal or sustaining hierarchies of being or reality (there are no properties or proper parts of entities causing higher-level properties to exist). There are only levels of description from the inside-out or from the bottom-up. A collection of atoms does not cause a chair to exist; the chair is composed and is nothing more than the collection of atoms described by different linguistic predicates (expressing slightly different concepts). In weak emergence, there are higher and lower-level descriptions of the same phenomena which entail one another. John Searle borrows an example from Roger Sperry to illustrate the naturalistic understanding of emergence:

“Consider a wheel rolling down hill. The wheel is entirely made of molecules. The behavior of the molecules causes the higher-level, or system feature of solidity. Notice that the solidity affects the behavior of the individual molecules. The trajectory of each molecule is affected by the behavior of the entire solid wheel. But of course there is nothing there but molecules. The wheel consists entirely of molecules.

⁵⁸ See Chalmers, David. consc.net, "Strong and Weak Emergence." Accessed March 22, 2014. <http://consc.net/papers/emergence.pdf>.

⁵⁹ Michael Silberstein, and John McGeever, "The Search for Ontological Emergence," *The Philosophical Quarterly*, 49, no. 195 (1999): 182-200.

So when we say the solidity functions causally in the behavior of the wheel and in the behavior of the individual molecules that compose the wheel, we are not saying that the solidity is something in addition to the molecules; rather, it is just the condition that the molecules are in. But the feature of solidity is nonetheless a real feature, and it has real causal effects.”⁶⁰

Searle’s language of “causation” can be misleading since solidity is not caused to exist by the behavior of the molecules, but rather the behavior of the molecules explains the phenomenon of solidity (which is nothing but the behavior of the molecules). Similarly, his assertion that solidity has an effect upon the individual molecules may lead one to believe that solidity has autonomous causal powers distinct from the behavior of the molecules, but this would be a misunderstanding of weak emergence. Instead, it is more appropriate to understand the solidity of the wheel as having inherited causal powers (powers inherited solely from the behavior of the molecules) since solidity is simply a higher-level (or broader) description of what happens at the lower-level (solidity is a description of the structure and behavior of the molecules, but solidity qua solidity leaves out molecule interaction in its definition). Solidity does not exert any unique causal powers not already found in the lower (or micro) levels. The naturalistic understanding of supervenience and property emergence is radically different from O’Connor’s view of both. Searle concurs:

“Since all of the surface features of the world are entirely caused by and realised in systems of micro-elements, the behavior of micro-elements is sufficient to determine everything that happens. Such a 'bottom up' picture of the world allows for top-down causation (our minds, for

⁶⁰ John Searle, *Freedom and Neurobiology: Reflections on Free Will, Language, and Political Power*, (New York: Columbia University Press, 2007), 48-49.

example, can affect our bodies). But top-down causation only works because the top level is already caused by and realized in the bottom levels.”⁶¹

O’Connor’s discussed model of emergence, on the other hand, relies upon a bottom-up depiction of reality in which the subvenient neurons (for example) undergo a process of firings causing a higher-level feature (pain) to come into existence. Here we have the notion of causation (or, at the very least, determination) that brings pain into existence. O’Connor and Wong state:

“If their appearance in certain systems is to be explained at all, they must be explained in terms of a causal, not purely formal, relationship to underlying, immediately preceding structures.”⁶²

Unlike Searle’s model, O’Connor allows for the macro-property of pain, for instance, to cause certain changes within the micro-properties of the person without the need for pain itself to be realized in the bottom-levels. Pain (as a property) comes into existence from nothing and so neither it nor its parts (as it has no property-parts) can be realized in lower-level features of the brain. In the next chapter, I will bring O’Connor’s theory of agent-causalism into view and examine the metaphysical implications it has for naturalism and physicalism.

⁶¹ John Searle, *Minds, Brains, and Science*, (Cambridge: Harvard University Press, 1984), 94.

⁶² See Timothy O’Connor, and Hong Yu Wong, "The Metaphysics of Emergence," *Noûs*, 39, no. 4 (2005): 664.

CHAPTER 3: NATURALIZING FREE WILL

O'Connor not only holds to a bottom-up causal picture, but he also defends a top-down causal picture in which pain causes a person to behave in certain ways. In other words, the feature of pain exerts a downward causal impact upon the neurons of the brain which result in the person developing a new plan of action.⁶³ Now it seems appropriate to introduce O'Connor's conception of free agency. O'Connor views free will as a dispositional capacity⁶⁴ possessed by rational human persons, the same persons which possess rational faculties sufficient for rational deliberation. O'Connor accepts an agent-causal account of libertarian free will and eschews an event-causal account of free will. In his account, the agent has a set of beliefs and desires which represent possible courses of action (for O'Connor, a belief-desire pairing constitutes a reason for action).⁶⁵

When an agent is satisfied with a proposed belief-desire pair and its represented goal for which to act, the agent executes or brings about the desired action via an intention. Usually, this takes the form of an "executive intention" which is an intention to immediately carry out an intended act.⁶⁶ Therefore, the intention causes a bodily movement, though your intending is not an event caused by prior events (though it causes a subsequent event, namely the bodily

⁶³ Timothy O'Connor, "Agent Causation," *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. Timothy O'Connor (New York: Oxford University Press, 1995), 179-180.

⁶⁴ *Ibid.*, 181.

⁶⁵ Timothy O'Connor, *Persons and Causes: The Metaphysics of Free Will*, (New York: Oxford University Press, 2000), 72.

⁶⁶ *Ibid.*

movement). This is how the agent-causal feature distinctively characterizes O'Connor's approach over that of the compatibilist which seeks to reconcile event-causation with free will.⁶⁷ The person possesses a sui generis nonstructural property which is the intrinsically active power to form intentions to act. O'Connor argues that the causal influence of the emergent property is not in any way inherited nor is the emergent property reducible to the subvening properties:

“...the occurrence of an emergent property is a function of certain joint causal potentialities of underlying base properties. Consequently, the continuing instantiation of the emergent property is completely dependent on some set of properties or disjunctive range of properties in the object's microstructure. Yet it [the emergent property] exerts a causal influence on the micro-level pattern of events that is not reducible to the immediate causal potentialities of subvening properties.”⁶⁸

This notion of top-down causation will be an issue for O'Connor's reconciliatory effort in the following section.

⁶⁷ O'Connor rehearses the dialectics of the both the compatibilist v. incompatibilist and the event-cause v. agent-cause debate in Timothy O'Connor, *Persons and Causes*, (New York: Oxford University Press, 2000), 67-73.

⁶⁸ Timothy O'Connor, *Persons and Causes: The Metaphysics of Free Will*, (New York: Oxford University Press, 2000), 111-112.

Causal Unity and Non-Reductionism

Having now described Timothy O'Connor's model of libertarian free agency, some worries will be raised for his reconciliatory view on naturalism and agent-causal libertarian free will. O'Connor describes two theses in his work 'Persons & Causes'. The first thesis is the "Causal Unity of Nature" thesis, while the second is the "Micro-Macro Constitution" thesis. As mentioned previously, O'Connor is not a reductionist regarding the nature either of consciousness or of free agency. He eschews the "Constitution" thesis which asserts that consciousness is nothing but a set of micro-properties (a structural property).⁶⁹ In this view, a mental state like pain is nothing but the firing of neurons and c-fibers (which each are composed of atomic components which have their own set of properties and relational features). O'Connor argues that consciousness is a simple emergent property as opposed to a structural property that is reduced to (and identified with) a set of micro-properties, and therefore rejects constitution in favor of the Causal Unity thesis, which asserts that there is a connected fabric of nature which causally connects micro-physical phenomena to micro-physical phenomena. The Constitution thesis entails the Causal Unity thesis, but the Causal Unity thesis can be defended without recourse to the Constitution thesis.⁷⁰ O'Connor argues that his position is consistent with naturalism because he accepts the causal unity of nature (all non-basic properties in the universe ultimately supervene upon physical properties).

⁶⁹ Ibid, 109-110.

⁷⁰ Ibid.

O'Connor uses the Causal Unity thesis to allow for the emergence of the simple property of consciousness from mereologically complex micro-properties. Having defended the emergence of consciousness as a non-physical and nonstructural (i.e. simple) property in his work, O'Connor adopts a similar strategy in maintaining that the agent-causal capacity (faculty) of the physical person is a strongly emergent property. The relevant property supervenes upon the microstructure of the agent while not being reducible to the microstructure and structural properties (e.g. spatial configurations and interactions between the micro-parts) that relate each of the micro-entities that constitute the agent. In this sense, the agent-causal capacity one possesses is fundamentally irreducible. It is a *sui generis* property that comes into existence via the structural properties of the whole organism. In other words, it is caused to exist.

O'Connor has acknowledged the causal coming to be of the emergent property but notes that this is not an instance of "something coming from nothing"⁷¹ but an instance of the determination of the emergent by the more fundamental features of the entity in question. However, this acknowledgement of his fails to address the fact that this emergence is an instance of creation *ex nihilo* (creation out of nothing), though of a simple property rather than of a substance. By the term "simple property", I refer to a monadic property (or quality) belonging to a substance (a simple property like solidity), as opposed to a merely relational feature (e.g. an arrangement) which represents a configuration that parts of a whole stand in relation to each other. Emergentists such as O'Connor posit the coming to be of monadic (simple) properties by the causal interaction of micro-entities arranged in a certain structure

⁷¹ Ibid, 123.

(such as in the human organism), with the micro-entities possessing certain monadic properties and causally interacting with one another (relational properties/features). Both kinds of properties are the properties of the person which give rise (causally, not merely explanatorily) to the nonstructural (simple) monadic property of the agent-causal capacity.

This seems to be a classic example of creation out of nothing. The causal powers of the agent-causal capacity are not inherited by the combined causal powers of the parts (such as in weak emergence). The sui generis emergent property has its own unique powers not reducible to those of its subvenient sustaining base. This means that the agent qua agent has causal powers that cannot be explained as the mere aggregation of the causal powers of the micro-entities that compose the agent, and thus these micro-entities cannot predict or determine the causal activity of the emergent freely-acting agent (keep in mind O'Connor's rejection of the "Micro-Macro Constitution thesis"). The whole organism has agent-causal powers that are not determined by the laws of nature nor by the parts of the organism that are determined by the laws of nature. They are, however, causally sustained by the laws of nature together with the micro-constituents of the organism. The human organism is, after all, a physical substance on O'Connor's view (and agents are substances).

Not only are these agent-causal powers independent of (and not determined by) the laws of nature, but they seem to be able to violate the natural laws which regulate the micro-parts (and the macro-entity). Downward causation consists in a macro-organism's (as a free agent) exertion of power to influence and change the behavior of the micro-parts. An instance

of this would be the raising of one's hand. The agent has reasons for and against the raising of her hand, but she is impressed by the reasons for doing it. The agent self-determines herself to form the immediate intention to execute that action, and once it occurs, the agent-causal powers exert an influence upon certain portions of the brain which influence the central nervous system and so forth, leading to the raising of the hand. This resultant event of arm-raising can also be described using the micro-language of physics which describes the language at a lower-level of description (the language of physics or chemistry) rather than a biological one (e.g. the intention causing molecules to interact in certain ways). The point here is that the agent can, by her free act, divert the natural course of events. In the next section, several major objections will be presented against O'Connor's model and some rebuttals will be considered.

Creatio Ex Nihilo and Emergence

The main objection I will lay against O'Connor's naturalistic model of free will concerns the emergence of free will from physical complexity *ex nihilo* (out of/from nothing). O'Connor acknowledges that his view commits him to the conclusion that a genuinely unique disposition (or capacity) is created that is itself not constituted by the physical properties which cause it to exist. This strongly emergent view would seem to commit him to understanding this agent-causal property as being created out of no pre-existing properties. It would amount to a unique form of creation out of nothing.

It is paradigmatically non-naturalistic for the agent-causal property to emerge ex nihilo as opposed to simply being a new way in which micro-particulars are spatially and geometrically arranged (which thereby stand in different relations to one another and possess the properties of standing in such relations). The latter view understands capacities and dispositions in terms of the ways in which certain molecules are arranged, which entail certain behaviors between the molecules (which are described in terms of capacities and dispositions). The former view goes even further and claims that the ways in which certain molecules are arranged can cause a new property to emerge that is not described as a structured or complex way (property/feature) in which the molecules exist (namely their behavior and arrangement).

The molecule-collection themselves do have interesting properties, however. They have the disposition to cause the organism that they constitute to possess free will. This latent capacity is manifested when these molecules are arranged and interact in a very complex fashion, according to O'Connor. The disposition to cause the agent-causal capacity to exist within an agent is not identical to the agent-causal disposition itself. This disposition to cause properties (namely the agent-causal capacity) to exist from nothing is uncharacteristic of other physical properties, and should cause any naturalist to pause when categorizing this dispositional property as naturalistic.

To press the issue further, if properties can emerge ex nihilo after a certain threshold of complexity is reached, there seems to be a number of unwelcome consequences to this allowance. For instance, one could allow for the possibility that a subvenient base's properties

necessitate a supervening property that is the disposition to create objects ex nihilo (as opposed to the disposition to freely enact intentions, as in O'Connor's account). A naturalist would find it difficult to allow for such qualities (or dispositions) to be admitted into her naturalistic ontology. How would a naturalist allow for structural properties which necessitate (via creation ex nihilo) a further genuine and sui generis property which allows a substance to create other substances ex nihilo and yet deny material objects the ability to create other substances ex nihilo?

There seems to be no principled metaphysical distinction between a substance that can create something (e.g. another substance or a proper part of itself) from nothing and a substance that acts freely in virtue of the creation of a sui generis nonstructural property from nothing. If one is paradigmatically unnatural (namely matter creating something from nothing), then the other must also be admitted into a supernaturalistic ontology. The same would be the case if a substance had the intrinsic property of extinguishing from existence certain other substances (the ability to annihilate certain substances, or even properties). Such a quality is undeniably unnatural and unaccounted for by any naturalistic epistemology, but such a property is not radically disparate from the dispositional property to create the agent-causal capacity ex nihilo, or the property to act in the absence of sufficiently necessitating event causes.

There seems to be no significant metaphysical difference between the disposition to create an agent-causal capacity from nothing which subsequently allows for intention-

formation in the absence of events and the disposition (the disposition itself also emerging ex nihilo) to create matter ex nihilo. Neither disposition is to be found among the physical predicates of either physics or chemistry. Both dispositions seem to betray a supernaturalistic ontology. In what follows, I will present another objection which I argue to render O'Connor's reconciliation between naturalism/physicalism and agent-causal libertarianism impossible.

Naturalistic Libertarianism vis-à-vis Panprotopsychism

The structure of Frank Jackson's Knowledge Argument is helpful in understanding the direction of my next argument. According to Jackson's argument, no knowledge proffered by the physical sciences can yield phenomenal knowledge acquired by direct acquaintance.⁷² To summarize Jackson's original argument, we are to imagine a scientist named Mary who has studied all facts relating to the human eye and color in a black-and-white room (she also has black-and-white glasses, for the sake of argument). She learns all of this theoretical information from within a room, but she eventually leaves the room and takes off her glasses, upon seeing a red flower she learns something new about the wavelength we know as red (but she did not recognize as having that quality with which we are acquainted). The argument is intended to show that phenomenal facts about red are not reducible to theoretical facts about the wavelength or the spectrum of light.

⁷² Frank Jackson, "What Mary Didn't Know," *The Journal of Philosophy*, 83, no. 5 (1986): 291-295.

Philosopher Yujin Nagasawa formulates Jackson's argument as follows:

(1) Mary (before her release) knows everything physical there is to know about other people.

(2) Mary (before her release) does not know everything there is to know about other people (because she learns something about them on her release).

Therefore,

(3) There are true propositions about other people (and herself) that escape the physicalist story.⁷³

I, however, am not arguing for such a strong and controversial thesis. Rather, let us modify Jackson's argument slightly such that it is not an argument for phenomenal knowledge as non-physical but (let us stipulate) an argument against knowledge of proto-consciousness, for example. Several philosophers have given parity-versions of Jackson's argument in order to directly refute panprotopsychism in relation to physical substances.⁷⁴ Panprotopsychism is the

⁷³ Yujin Nagasawa, *God and Phenomenal Consciousness: A Novel Approach to Knowledge Arguments*, (New York: Cambridge, 2008), 4.

⁷⁴ An excellent assessment of David Chalmers' panprotopsychistic developed position and its relation to the Knowledge Argument is found in Yujin Nagasawa, *God and Phenomenal Consciousness: A Novel Approach to Knowledge Arguments*, (New York: Cambridge, 2008), 109-114.

metaphysical thesis in Philosophy of Mind that all physical substances are imbued with proto-consciousness (which itself is not consciousness) such that in certain substances with a particular structural arrangement of proto-consciousness of a particular threshold will have full-blown consciousness of the kind we know and love. One particular objection to this argument is the 'combination problem', namely, how do the properties of proto-consciousness combine to yield consciousness?⁷⁵ On the one hand, the threshold to which these properties must be structurally arranged seems utterly contingent. Moreover, the properties themselves seem utterly mysterious and empirically unobservable (either directly or by implication). The thesis of panprotopsychism states, however, that the relationship between proto-consciousness and physical entities is one of strong metaphysical necessity (ranging across all possible worlds in which these physical entities exist), and proto-consciousness is not reducible to physical states of the entity.

According to Jackson's argument, Mary is confined in a black-and-white room and is not able to experience certain phenomenal states (e.g. seeing green, red, or blue). She is tasked with learning everything physical there is to know about other people. This information includes information about the visual mechanisms by which humans experience what they call green, red, blue and other colors. We may conclude that she comes to know everything physical there is to know about other persons. However, she is released from her room and experiences colors other than black and white. Did she learn anything new? The Knowledge Argument asserts that she did, namely the phenomenal experience of colors other than the

⁷⁵ See Chalmers, David. *consc.net*, "The Combination Problem for Panpsychism" Accessed March 22, 2014.

ones she had previously experienced in the room. This knowledge is not only knowledge of her experience, but knowledge of other persons and of how they experience the world. The argument concludes that physical knowledge of other persons is not phenomenal knowledge of other persons.

The Knowledge Argument may be illuminating in this regard: No amount of physical knowledge of physical states can reveal any knowledge regarding states of consciousness. Now in order to find out if one can know about proto-phenomenal or proto-conscious states by learning physical facts, one can retain the structure of the argument and simply substitute the study of physical facts with the study of proto-phenomenal or proto-consciousness facts.

It seems difficult to image what proto-conscious properties even are, let alone whether they can lead to knowledge of phenomenal properties. It gets even worse for the panprotopsychist, for even if the subject of study were physical facts (as in the original thought-experiment) a study of physical facts would never yield knowledge of proto-phenomenal facts. Therefore, it looks as if proto-consciousness is not physical and faces the same epistemological and metaphysical dilemma as regular states of phenomenal consciousness. No amount of physical information will generate a deductive apparatus by which we can infer any proto-conscious or consequently conscious mental states that are formed as a result of a structural interaction between proto-conscious properties.

O'Connor's naturalistic libertarianism is analogous to panprotopsychism in that O'Connor posits physical properties with supernatural-properties which are such that if

combined with other physical properties, they create a new property (viz. a capacity) which emerges ex nihilo and is not inferable from the subvenient base (the physical substances along with its properties). But neither are the supernatural-properties (which cause the agent-causal capacity to exist) inferable from their physical base. The subvenient-supervenient relationship in both instances seem utterly contingent (as O'Connor himself acknowledges) as is the subvenient-supervenient relationship found in panprotopsychism. In panprotopsychism, the proto-conscious properties arrange and interact in such a fashion that a novel property (viz. the capacity of consciousness) emerges ex nihilo, similar to O'Connor's agent-causal capacity.

Moreover, the actualization of the agent-causal capacity results in the agent's possession of yet another simple (nonstructural) property, the state of acting freely for x (which is not ontologically reducible to nor decomposable into any other property). The state of acting freely is analogous to the state of being conscious for the panprotopsychist. Both states (i.e. both mental states, the former being a freely-formed intentional state) are strongly emergent states of the biological organism, and both are non-identical to the dispositional capacities of the biological organism. So, in reality, the naturalistic libertarian is dealing with two main instances of strong emergence: the emergence of the mere capacity to act freely and the very exertion of a free choice via the forming of an intention. The former has been discussed at length earlier in this chapter and the latter is being introduced now only to compare it with the claims of panprotopsychism. If panprotopsychism is to be rejected tout court by the naturalist for its blatant supernaturalism then it is only appropriate for her to reject O'Connor's libertarianism. In summary, both O'Connor's naturalistic libertarianism and panprotopsychism

seem to posit exotic matter in some form or other, which are a vestige of supernaturalism. I will conclude with some comments on the nature of supernaturalism and whether my arguments are sufficient to refute O'Connor's naturalistic libertarianism.

Free Will and Supernaturalism (Conclusion)

I have already proposed several objections to O'Connor's reconciliatory naturalism, namely the violation of the causal closure (in my interaction with Searle's naturalistic libertarianism in Chapter 1), the creation of the agent-causal capacity ex nihilo, and the parity case between a 'physical' object with the disposition to create other entities ex nihilo (or annihilate other entities) and a substance which can produce the agent-causal capacity ex nihilo, and the parity case (using Jackson's argument) between panprotopsychism and naturalistic libertarianism. Timothy O'Connor (and other reconciliatory naturalists) might rebut my objections by pointing out my failure to demonstrate that agent-causal free will requires any immaterial faculty or disposition. In other words, the disposition is not intrinsically immaterial in the same fashion that, say, a phenomenal or intentional mental property (in O'Connor's view) would be characteristically mental and inaccessible to physics. Indeed, if free will exists, it would be publicly-accessible for all to witness (since it is merely a disposition of a physical substance).

However, I would argue that immaterialism is not the only kind of supernaturalistic position that one may adopt. Immaterialism may be a sufficient condition for supernaturalism, but it is not a necessary condition. One may be a materialist and still reject the metaphysical presuppositions which underlie physicalism and naturalism (e.g. causal closure, bottom-up hierarchical view of reality, impossibility of top-down causation, the naturalistic epistemic stance). Materialism is co-extensive with neither physicalism nor naturalism; materialism is a necessary but not a sufficient condition for either and materialism is a sufficient condition for supernaturalism but not a necessary one. It may seem odd to be a materialist and not a naturalist, but naturalism has certain implications which materialism does not.

For example, a naturalistic universe could not possibly be populated by physical substances (e.g. zombies) possessing the disposition to produce non-physical mental states (like pain) and happen to never manifest those dispositions (the dispositions are never actualized). One may wonder how that would render such a possible world non-naturalistic. It has nothing to do with the presence of immaterial properties, since it has already been stipulated that these dispositions to produce mental states are never actualized. However, the question remains as to whether these dispositions are purely physical or semi or quasi-mental properties. These zombies are not themselves conscious, but if they drank an elixir, perhaps their mental dispositions would be actualized. These very dispositions are not physical properties but rather quasi-mental properties (despite not being immaterial).

This argument suffices to show that immateriality is not a necessary condition for a rejection of naturalism. Similarly, if some powerful being (for example, Zeus) was composed of

some exotic matter which disposed him to create matter from absolutely nothing, this would contradict the relevant theses of naturalism. It is sufficient to show that these relevant properties are not found within the predicates of our most advanced physical theories in science (or perhaps of any future science in principle) in order to conclude that free will is not compatible with naturalism. Although O'Connor's reconciliatory attempt asks for too much, we may be forced to rest content with a materialistic position that eschews causal closure, a solely bottom-up hierarchical view of causal reality, and the naturalistic epistemic stance.

BIBLIOGRAPHY

Clarke, Randolph. *Libertarian Accounts of Free Will*. New York: Oxford University Press, 2003.

Fischer, John Martin. *Four Views on Free Will*. Edited by John Martin Fisher, Robert Kane, Derk Pereboom, Manuel Vargas. Oxford: Blackwell Publishing, 2007.

Ginet, Carl. *On Action*. New York: Cambridge University Press, 1990.

Gottfried, Leibniz. *The Monadology. Modern Philosophy: An Anthology of Primary Sources*. Edited by Roger Ariew, Eric Watkins. Indianapolis: Hackett, 2009.

Howell, Robert. *Consciousness and the Limits of Objectivity: The Case for Subjective Physicalism*. Oxford: Oxford University Press, 2013.

Kane, Robert. *The Significance of Free Will*. New York: Oxford University Press, 1998.

Kim, Jaegwon. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press, 2007.

----- . *Supervenience and Mind*. New York: Cambridge University Press, 1993.

Mele, Alfred. *Free Will and Luck*. New York: Oxford University Press, 2006.

Nagasawa, Yujin. *God and Phenomenal Consciousness: A Novel Approach to Knowledge Arguments*. New York: Cambridge University Press, 2008.

O'Connor, Timothy. *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press, 1995.

----- . *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press, 2002.

Pereboom, Derk. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press, 2014.

----- . *Living without Free Will*. New York: Cambridge University Press, 2001.

Searle, John. *Freedom and Neurobiology: Reflections on Free Will, Language, and Political Power*. New York: Columbia University Press, 2007.

----- . *Mind: A Brief Introduction*. New York: Oxford University Press, 2004.

----- . *Minds, Brains, and Science*. Cambridge: Harvard University Press, 1984.

Swinburne, Richard. *Mind, Brain, and Free Will*. Oxford: Oxford University Press, 2013.

The Oxford Handbook of Free Will. Edited by Robert Kane. New York: Oxford University Press, 2011.

The Philosophy of Free Will: Essential Readings from the Contemporary Debates. Edited by Paul Russell, Oisín Deery. New York: Oxford University Press, 2013.

Timpe, Kevin. *Free Will: Sourcehood and its Alternatives*. New York: Continuum, 2008.