FINDING FREE WILL:
CAUSATION IN AN INDETERMINISTIC WORLD

Thomas Hebner

TC 660H

Plan II Honors Program

The University of Texas at Austin

May 13, 2020

_____

Dr. Robert Koons

Department of Philosophy

Supervising Professor

_____

Dr. Galen Strawson

Department of Philosophy

Second Reader

Abstract

Author: Thomas Hebner

Title: Finding Free Will: Causation in an Indeterministic World

Supervising Professors: Drs. Robert Koons and Galen Strawson

Free will is an oft disputed topic in popular culture, religion, and philosophy, yet much of its refutations are built upon theories of determinism and necessitation, whose conclusions are untenable. Any discussion of free will is necessarily a discussion of cause and effect, so any claims made about freedom must first establish what it is to be a free agent. This thesis challenges deterministic causation, showing that even classical Newtonian mechanics gives us indeterminate solutions. In doing this, we show that causation itself is a fundamental truth, built from an ontology of causal powers, the implications of which we explore in detail. From this metaphysical framework we explore a plausible route by which free will may emerge: the theory of Agent Causation, which argues that at the core of every free action is an irreducible causal relation between a person and some appropriate mental or internal event that triggers later elements of the action. Agent Causation will be shown to be a theory of causal powers implied by our ontology, with responsibility and agency thereby emerging.

Acknowledgements

I cannot thank my family and friends enough for their support during this process.

Any words I put here are insufficient to properly express how helpful they were as both critical eyes and emotional encouragers as I (very) slowly made my way through this.

Lastly, special thanks to my Supervisor and Second Reader, Drs Robert Koons and Galen Strawson, for their extraordinarily detailed analysis of my work.

Sometimes the best advice is from those who disagree.

Table of Contents

Introduction

The core of any discussion of free will is agency: can I actually be the cause of my actions? To exist and function day-to-day requires some level of belief about your own control over your actions. You can't *not* believe that you are in control of what you yourself do—you may pay lip service to the idea of external control, determinism, God, or whatever exterior cause you can create, but the very idea of acting, doing, being, and living requires agency. To argue against this idea itself requires some kind of doublethink, for to argue is to act and acting requires an actor. The actor acts, the dancer dances, the orator orates; to verb is to be something that verbs, to cause is to be a causer. This argument can be dismissed as semantic, taking one of a number of possible routes—determinism, primary movers, grand design---all of which reach the same conclusion: the doublethink is necessary and correct, and free will is a delusion. While we may believe we are in control of what we do, this control extends no further than the end of the thought and translates not at all to action. Action, to the denier, is irreconcilable with actual agency despite its derivative usefulness, so we might as well go on pretending the fallacy.

Is such fatalism really all that is left to us? Is the question itself even important? To answer, we must first define what it is we are talking about when we talk about free will. Free will is a power, a power which confers on us both freedom of action and responsibility for that action. Whether we believe in it or not, we feel proud of our good acts and guilty for the wrong when we feel that we are responsible[1] for said event. *Free will*, then, is the *power of agents to be the ultimate*

[1] For this paper, we will not attempt to answer many ethical questions of the consequences of free will. Many of them are obvious, such as the appearance of responsibility mentioned above, but as Galen Strawson points out in *Freedom and Belief*, moral responsibility carries with it a deep sense of true responsibility: we are "truly deserving of praise and blame" for it is "truly up to us" (26). I agree with this and leave further discussion to others.

creators (or originators) and sustainers of their own ends or purposes[2]. To actively will (or freely act) is to be the 'primary mover' of one's own events and purposes. While relatively obscure and functionally unfathomable, the notion of one's ultimate creation of one's motives and purposes is the basis beneath understanding of free will. Imagining a causal chain, we trace the causes or explanations of events back to their sources: the purposes of the acting agent. For the free will faithful, these chains must terminate in the agent's willings[3], which would be the choices, decisions, or efforts of the agent.

For better or worse, we accept moral responsibility for our actions and assume we are to the highest degree in control of our own actions. If we have a choice, we are free, and we always seem to have a choice. But is that a simple delusion? Determinism and strict physical causation appear to imply a complete chain of events throughout space-time; a picture entirely incompatible with free will. If the human mind is a completely physical and deterministic system, under pre-quantum interpretations of matter it was safe to assume the mind was deterministic as well, thereby removing free will and reducing human action to being pre-determined and pre-destinated. There is thereby a prima facie tension between free will and determinism, often referred to as incompatibilism.

At this point, it may be asked why there is even an issue here between determinism (necessitation) and free will. After all, hasn't modern science moved beyond universal determinism? The problem lies not in physics but in all other sciences: developments in biology, neuroscience, psychology, psychiatry, social and behavioral sciences have all moved towards

[2] This relatively uncontroversial definition and some of the discussion that follows is taken from Robert Kane's *The Significance of Free Will*. It is used as our definition here because it broadly fits our purposes.
[3] We will point out later that this is not true in all cases—there are uncaused actions and actions caused by something other than the agent—but extreme agency has been argued, albeit unsuccessfully. If these willings were in turn caused by something else, i.e. the environment, God, or fate, then ultimacy would lie with something other than the agents.

deterministic pictures. Today more than ever, people believe that much more of our behavior is determined by far more causes outside of our control and understanding than we previously believed (Kane, 2002). Genetics and heredity, gut biomechanics, drugs, and unconscious influences on behavior are just a few of the things that modern science claims to dictate our actions and beliefs. The mere existence of indeterminism in the world does not suffice to reject the determinist's challenge to free will[4].

Free will is inherently an issue of causation and how we fit into it. We will argue that the 'never-ending disputation'[5] should be framed in terms of causation, which must be defined as more than a common-sense idea. The emergence of indeterminism, while not a complete rejection, challenges our beliefs about causality writ large: the idea of the causal chain is key, for without a mechanism of cause and effect there is no agency, so how can there be causation in an indeterministic world? Much of the effort at the start will thus be to craft some idea of causation, with some surprising results.

The purpose of this thesis is to build a metaphysical model of the world based on a theory of causal powers. An ontological theory of powers, while controversial, must be the basis for our discussion of free will for powers are fundamental to agency and freedom. Agents have properties, beliefs, desires, and wishes, and it is these properties that do causal work, doing so because they are powerful. Events exist as well. Events that are caused will be produced by powers at work. Events will be the changes produced when powers exercise themselves. These changes can also push other powers together, so when we talk of events in causation, we are mainly talking about

[4] For more on these challenges, see works by Taylor and Dennett, Paul Russel, Richard Double, Benjamin Libet, and Henrik Walter.

[5] This is a reference to the following poem:

> "There is a disputation [that will continue] till mankind are raised from the dead between the Necessitarians and the partisans of Free Will"
> --Jalalu'ddin Rumi, twelfth-century Persian poet

how the empowered properties got together, and it is these partnered properties that do the causing.[6]

These models will be explored in greater depth. While they are discussed, we must keep in mind the ultimate goal: free will. Once established in theory, they will be extended to the physical world as an approximate theory of Agent Causation, with agents classified as "powerful" beings capable of acting, willing, and perceiving change. From this model, which has been established and based entirely on prior theoretical work from Anscombe, Aristotle, Marmodoro, Mumford, Anjum, and Molnar, we will discuss some theories of free will in the presence of indeterminism discussed primarily by Timothy O'Connor. Once each view is established, we will attempt to fit or show the inability to fit the theory with the Powerist theory discussed at the beginning of the paper. The model with best fit--Timothy O'Connor's Agent Causation—will be analyzed further, discussing challenges with the picture and concluding with what we have accomplished.

[6] Agent Causation is fundamental to our ontological picture, and how that fits in to the Powerism picture is discussed in the final two chapters.

Determinism and Necessitation

Determinism

To begin, we should describe what is it is we mean by "determinism" in the most general sense. For our purposes, we will consider determinism to refer to "nomological" or "causal" determinism, or the idea that the facts about our past which we generally consider to be true contain the relevant information when coupled with the necessary laws to determine all the facts of our future. According to van Inwagen (1974), we can consider determinism to be an entailment: determinism is "a relation of entailment that holds between, on the one hand, statements of law and statements of particular fact at a time, and, on the other hand, statements of particular fact at other times."

When we discuss determinism, we consider both finite, closed, irreducible actions and the state of the entire physical world[7]. By the latter we mean any and all relevant information about the universe—a large set, but functionally reducible to our own world for the purposes of our refutation. By 'state' we may understand a particular set of events at a particular time, frozen and locked into place[8] with no facts about it containing information about future actions (i.e. "Maria will turn around in 2 minutes"). Formally, then, we can define determinism as follows:

(a) For every instant of time, there is a proposition that expresses the state of the world at that instant[9]

[7] This and the following discussion on determinism relies heavily on Peter van Inwagen's "The Incompatibility of Free Will and Determinism" from 1974

[8] The notion of "time-slicing" can itself be challenged. Time may be infinitely reducible, which would entail an inability to individuate moments from which we can construct our theories. We will use infinite divisibility later in the section on Norton's Dome

[9] Arguing that we can reduce the state of the world to a proposition is admittedly rather reductive. Gödel Incompleteness implies an inability to know all the facts about any system (including the system of our universe). This itself has been used to argue for free will; for more on this, see Feferman, S. 2011. "Gödel's Incompleteness Theorems, Free Will and Mathematical Thought." In *Free Will and Modern Science*, edited by Richard Swinburne, 102–22. New York: Oxford University Press."

(b) If *A* and *B* are propositions that express the state of the world at two different times *t1* and *t2* such that *A* precedes *B*[10], then the conjunction of *A* with the laws of physics entails *B*

In the definition above, it is clear that 'the laws of physics' are made to do work. If the laws were vague maybes rather than hard code, determinism would be trivially false[11]. What exactly a law of physics is supposed to be is unclear, but for our purposes we do not need to settle this question here. Rather, we shall argue from as uncontroversial of a position as possible, given we are discussing issues of quantum indeterminism and causation, both of which are highly contentious.

Of further interest is the Beta principle, which goes as follows: Let "Np" abbreviate "*p*, and nobody has, or ever had, any choice about whether *p"*, and consider Beta: If Np and N(p ⊃ q), then Nq. Beta is the central rule of interference in van Inwagen's "Consequence Argument" for the incompatibility of freedom and determinism[12]. He proposes the case of the judge "who had only to raise his right hand at a certain time, *T,* to prevent the execution of a sentence of death upon a certain criminal." The judge proceeds to refrain from raising his hand at that time, and that this inaction resulted in the criminal's being put to death. The judge is said to have decided against raising his hand at *T* "only after a period of calm, rational, and relevant deliberation [in which he had] not been subjected to any 'pressure' to decide one way or another about the criminal's

[10] The length of time between A and B is arbitrary. They can occur one after the other directly or be thousands of years apart. According to a simplified determinism, in either case A entails B. This of course ignores probabilistic events, but determinists may be found to argue that even probabilistic events are only probabilistic when all relevant information is unknown.

[11] Again, this is not necessarily true. There are obvious indeterminism/quantum mechanics challenges to determinism, but we can restrict the definition of determinism or cause-effect conjunctions to be derivatively true rather than ontologically necessary. See the section on Norton's Dome for further discussion on the subject.

[12] See "Counterexamples to Principle Beta: A Response to Crisp and Warfield" by Erik Carlson for more discussion

death." His decision is understood to be perfectly rational, free from influence, and under the assumptions of free will entirely his own intellectual product (Van Inwagen, p. 186-191, 1974).

Informally, he argues that if determinism is true as constructed, then there was never any ability for the judge to have raised his hand. In fact the judge had no free will, rational thought, or individual intellectual product whatsoever. His decision was made for him but not by another free agent but by the entire state of the universe leading to this particular event. Van Inwagen's formal argument proceeds in the following manner: by 'P_0' we mean the proposition relating to the state of the world at time 'T_0' which is itself some arbitrary instant of time before the judge's birth (and thereby out of the realm of his influence). 'P' denotes the state of the world at T, and 'L' the conjunction of all laws of physics into a single law[13]. Here is the argument:

1) If determinism is true, then the conjunction of P_0 and L entails P.
2) If the judge had raised his hand at T, then P would be false.
3) If (2) is true, then if the judge could have raised his hand at T, the judge could have rendered P false.
4) If the judge could have rendered P false, and if the conjunction of P_0 and L entails P, then the judge could have rendered the conjunction of P_0 and L false.
5) If the judge could have rendered the conjunction of P_0 and L false, then the judge could have rendered L false.
6) The judge could not have rendered L false.

$\therefore$ 7) If determinism is true, the judge could not have raised his hand at T.[14]

[13] While apparently incompatible on many levels, we may understand the assumption to be a sweeping generalization of the "rules of the game" which we all play by and cannot escape from.
[14] Van Inwagen, *Incompatibility*, 191.

This is the basis from which most arguments for Incompatibilism arise. In short, we can define Incompatibilism to be the thesis that there is no world in which determinism and free will can coexist. Of note in this definition of determinism is the reliance upon the entailment relation between two propositions, the conjunction of P_0 and L, and P. This kind of relationship may appear to imply that causation itself entails necessitation, which is a misconception. The Consequence Argument itself does not mention causation—it is a discussion of determinism, which can also be characterized as simply 'nothing can ever happen otherwise than it does.' What we instead argue is that causation is independent of determinism, meaning that we can reject determinism without paying an ontological price.

In *Causality and Determination: An Inaugural Lecture*, Anscombe formalizes the idea as follows: "Causality is some kind of necessary connexion (sic), or alternatively, being caused is — non-trivially — instancing some exceptionless generalization saying that such an event always follows such antecedents" (1). Thus, if an effect occurs in one case and another effect occurs in an identical version of the case, there must be some further relevant difference. It is a proof by contradiction: if we assume that the judge did raise his hand instead of leaving it down, then given the deterministic laws of necessitation and causation, there must have been some difference in circumstance between the two cases. However, the two cases are identical, and therefore the judge could not have raised his hand[15]. Deterministic causation and free will can thereby not coexist.

[15] Interestingly, Robert Kane's picture of the free mind does not appear to require many of the commitments we will discuss about causation, rather relying on the indeterministic system of a spinning self-network of will. In fact, the circumstances could be exactly the same as established and the judge could make a different decision—we just need a different understanding of the "laws of physics"

The necessitation relation of cause and effect has been argued for centuries. The conception of their constant conjunction is often ascribed to Hume,[16] who argued that the relation between cause and effect "could not be found in the situation, objects, or events called 'causes' or 'effects' but was to be found in the human's mind being determined, by experience of constant conjunction" (Anscombe, p. 4). The constant conjunction view is reinforced by Thomas Hobbes, who argued for a type of simultaneous necessitation in *Elements of Philosophy:*

> "A cause simply, or an entire cause, is the aggregate of all the accidents both of the agents how many soever (sic) they be, and of the patients, put together, which when they are supposed to be present, it cannot be understood but that the effect is produced at the same instant, and if any one of them be wanting, it cannot be understood but that the effect is not produced."[17]

Kant continued this tradition by establishing causation as an *a priori* conception. Instead of the instantaneous manifestation of effect and cause, he argues for an objective time order arising in conformity with classical consequent/precedent conceptions: from the conditions of the precedent event/state necessarily arises the consequent event.

Necessity and Sufficiency

Before proceeding with this discussion on causation, it would prove useful to provide a stricter definition of necessity. While there are a number of difficulties[18] for the so-called "Standard Theory" of necessity and sufficiency, we will only treat such ideas as derivatively true and will

[16] In Strawson's *David Hume: Objects and Power*, he argues against the idea that Hume holds a "straightforward 'regularity' theory of causation, according to which causation is nothing more than the regular succession or constant conjunction." Instead, Strawson claims that Hume believes in what he calls "causal power" or "natural necessity," the claim that all we can ever actually observe is regular succession, which is not to say that all causation actually *is* is the constant conjunction of objects. For more, read his excellent piece.

[17] Interestingly, this simultaneous mutual manifestation is going to be used heavily by the Powerists we discuss later. In their context, powers manifest their effects at the instant of mutual manifestation.

[18] For further discussion, see https://plato.stanford.edu/entries/necessary-sufficient/

thereby attempt to avoid potential problem areas such as ambiguity[19]. The standard theory[20] of necessity relies upon a truth-functional definition of the proposition *if*. In classical logic, the statement 'If *p* then *q*' (p ⊃ q) is only false in the case when *p* is true and *q* is false. If the conditional holds, and *p* is true, then *q* must also be true. In the common understanding, the truth of *q* (the consequent) is necessary for the truth of *p* (the antecedent) while the truth of the antecedent is merely sufficient for the truth of the consequent. In the case of van Inwagen's judge, the birth is necessary for the judge's raising his hand.

In classic 'material' or truth-functional conditionals, there is a reciprocity between necessity and sufficiency. Consider the following conditional: "If you are a human, you have a heart." With weird problem cases and artificial hearts excluded, we can consider the consequent, having a heart, to be necessary for the truth of the antecedent, being human. Conversely, being human is merely a sufficient condition for having a heart; after all, plenty of other animals have hearts too. Necessity can also be identified using an "only if" clause: in the case of the judge, he can raise his hand only if he was born (and determinism is false). There is thus a fundamental link between ideas of necessity and determinism, so for us to escape we must take a different route.

Causation

Let us consider now what it is we mean by causation. As we have mentioned, the common assumption about causation lies in necessitation within the cause-effect relation; that there is a non-trivial generalization to be made about effects being derived from causes. Of important note here is that just because something is said to be a cause of an effect—like a virus being the cause of a

[19] Namely, we operate under the assumption that necessity and sufficiency are not ontological truths, or that modal-ontological dependence is not true. This is a claim that needs further defense. For more on the topic, see works by Rosen, Marcus, or Fine.

[20] This analysis of truth-functional logic primarily follows Humean analysis. There are a number of different ways to do this, but for our purposes this analysis is sufficient and simple.

disease—does not mean that such causes *always necessitate their effect*. The case of the disease is a clear one: you do not always get sick when you are around someone who is also sick for a variety of reasons, one of which being your body is able to resist.

How, then, are we do define causality? At its most basic level, causality consists in the derivativeness of an effect from whatever its causes may be. To say effects arise from causes is not controversial, at least in this understanding of temporal procession. Anscombe argues that we should be able to "derive knowledge of the effect from knowledge of the cause, or *vice versa*, but that does not shew (sic) us the cause as the source of the effect." In other words, she is arguing that causation is *not* to be associated with necessitation in all cases[21]. Put formally:

> "If *A* comes from *B*, this does not imply that every *A-like* thing comes from some *B-like* thing or set-up has an *A-like* thing coming from it; or that given *B*, *A* has to come from it, or that given *A*, there has to be *B* for it to come from. Any of these may be true, but if any is, that will be an additional fact, not comprised in *A*'s coming from *B*. If we take 'coming from' in the sense of travel, this is perfectly evident" (Anscombe 8).

So, then, causation is not to be inextricably linked with necessitation. This gives rise to an important question: is causation *ever* to be associated with necessitation? To answer, we must undertake in something slightly painful: mathematics, which we will use as an alternate but definitive attack upon necessitation and determinism.

[21] We will show later with Norton's Dome that the necessitation relation can be functionally removed from our understanding entirely.

Norton's Dome

Background

In a 2003 article published in *Philosophers' Imprint* titled "Causation as Folk Science," John D. Norton argued against the "common sense" conception of causation. The widely held principle of causality asserts every effect is causally necessitated by a cause and such causal conjunctions rule the world around us, a stance that Norton takes issue with. He posits both a negative and a positive thesis as new models of how we think of causation. In the negative thesis, Norton urges that "the concepts of cause and effect are not the fundamental concepts of our science and that science is not governed by a law or principle of causality" (Norton 1). This causal skepticism is taken in the light of our most mature scientific theories and is motivated by the introduction of possibility into causal models. In the positive thesis, Norton argues that "ordinary scientific theories can conform to a folk science of causation when they are restricted to appropriate, hospitable processes; and the way they do this exploits the generative power of reductive relations, a power usually used to recover older theories from newer ones" (Norton 2).

The "generative power" he describes here is in reference to a common move over the course of scientific progress: recovering older theoretical models of the world and universe that have been found lacking in some area. For example, Norton points out the utility of treating gravity as a force as claimed by Newton despite Einstein's theory of general relativity as well as the treatment of heat as a conserved fluid. This reductive view can be enormously convenient, as many of the problem cases for such reductions are not applicable in everyday situations. So too causation: "causes and causal principle are recovered from science in the same way and have the same status: they are heuristically useful notions...but we should not mistake them for the fundamental principles of nature" (Norton 2). Thus, the two theses, negative or "skeptical" and

positive or "constructive" combine to challenge the idea that causation is a physical fact of the world that we can recover from our sciences.

This type of causal skepticism is not unique to Norton. In 1917 Bertrand Russell raised similar concerns, dispensing with causes in the following way:

> "All philosophers, of every school, imagine that causation is one of those fundamental axioms or postulates of science, yet, oddly enough, in the advanced sciences such as gravitation astronomy, the word 'cause' never occurs...The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving like the monarchy, only because it is erroneously supposed to do no harm." (Russell 132).[22]

What Russell and Norton challenge is the notion of causal fundamentalism, the claim that nature is governed by cause and effect. It is thereby the duty of the individual sciences (physics, chemistry, biology, etc.) to find expressions of causation within their fields. In effect, this is an *a priori* stance on how the world works, similar to the commonly accepted case of energy. A significant section of our mature sciences—in particular, physics, chemistry, and biology—deal with the various manifestations of energy. The interchange of kinetic and potential energy governs mechanistic actions. Field energy, put very simply, is magnetism and electricity, and chemical energy allows us to act. Energy appears as heat, work, light, atomic bonds, and so on, spanning many subjects in a necessary and central manner. For causal fundamentalism to succeed, it must too appear clearly in the various fields.

Causal fundamentalism has sometimes been thought to be equivalent to determinism, a position refuted by Anscombe, Norton, and many others. When the pre-existing conditions are fixed to the proper degree, under the laws of determinism the future becomes itself fixed and

[22] Russell actually changed his mind on this topic later on, consistently making essential use of notions of cause. The quote here is used as an example of widespread skepticism about causation.

unchanging. From this arises the classic Laplace Demon, a "calculating intelligence" which has the ability, given sufficient knowledge, to determine the entirety of the past and future. Laplace claims famously that "we ought then to consider the present state of the universe as the effect of its previous state and the cause of that which is to follow" (Laplace, p. 4, 1814). While the existence/possibility of such an intelligence may be emotionally appealing it falls to a number of challenges, namely Gödel Incompleteness.[23]

This conception of causation toppled quickly with the advent of modern quantum theory. Central to quantum theory is entanglement: two particles, once interacted with, can be infinite distances apart yet both be affected instantaneously by changes to just one of the pair. Consider the case of gravity as well, which has no speed of propagation and can be found to act where it does not appear to be present. We thereby have action at a distance, another apparent problem for determinism. If the universe is causal in nature, "it must rule out *a priori* the possibility of action at a distance" which is in direct contradiction to the scientific consensus on both gravitation and quantum theory.[24]

Quantum theory also proffered mere probabilities of occurrence, not necessities. Even a perfectly complete model of the universe "cannot determine whether some particular Radium-221 atom will decay over the next 30 seconds (its half-life); the best we can say is that there is a chance of ½ of decay" (Norton 5). Decay is uncaused and non-necessitated, but it can be predicted probabilistically. This gives rise to probabilistic causation. However, prediction is not what we are

[23] For more on this, see "Gödel's Incompleteness Theorems, Free Will and Mathematical Thought" by Solomon Feferman

[24] This paragraph doesn't apply to interpretations such as Bohmian Mechanics. For an in-depth discussion of the subject, see the Stanford Encyclopedia of Philosophy entry on Bohmian Mechanics.

up against. Determinism involves necessitation instead of probability; there are interpretations of free will in a probabilistic event space, Robert Kane's theories most notable among them.

The determinist has a response here: theories of quantum mechanics and general relativity do violate our common conceptions of causation and determination, yet they do so only in extreme cases like the two-slit or gravitation. Determinism, they claim, still stands upright in every case applicable to us and our lives. This is not the case. We turn to an ingenious example typically referred to as Norton's Dome, in which a mass at rest in an unchanging surrounding physical environment for an indefinite period of time spontaneously and indeterministically begins to move in an arbitrary direction. This example will be used to show the failure of even simple Newtonian systems of mechanics to utilize any deterministic principle or law of causality[25].

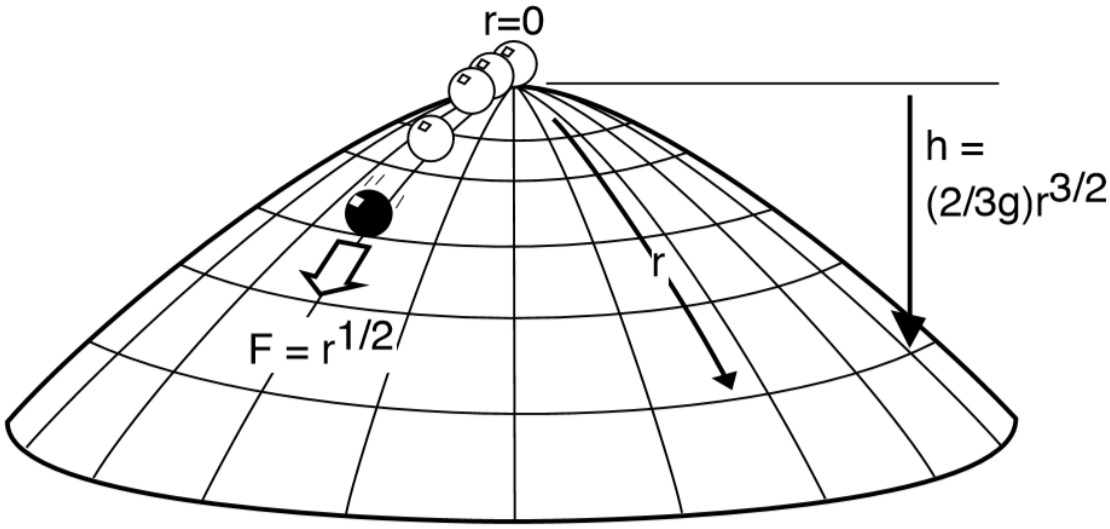

Figure 1a. Mass sliding on a dome

[25] My supervising professor, Dr. Robert Koons, points out that interestingly, quantum mechanics is more nearly deterministic than Newtonian mechanics, in the sense that the Schrödinger dynamics guarantee a unique result in cases like this (basically, the ball simultaneously rolls down in all directions, and stays on top, as well). It's only when a "wave collapse" occurs (through measurement) that we find indeterminism in QM.

Norton's Dome

The dome pictured on the previous page exists in a downward-directed gravitational field with classic gravitational acceleration *g*.[26] The dome has radial coordinate *r* on its surface and is rotationally symmetric around origin $r = 0$, the peak of the dome. The height of the dome is calculated as a function of the radial coordinate *r*. For our purposes, we will have $h = \left(\frac{2}{3g}\right) r^{\frac{3}{2}}$. The surface of the dome is frictionless so that a point-like unit mass will slide frictionless over the surface under gravity, with gravitational force only able to accelerate the mass along the surface. At any point on the dome, the magnitude of the gravitational force tangential to the surface is $F = \frac{d(gh)}{dr} = r^{\frac{1}{2}}$ and is directed radially outward, resulting in no tangential force at $r = 0$. When on the dome, the mass experiences a net outward downward directed force field of magnitude $r^{\frac{1}{2}}$. When applied to the mass on the surface, Newton's Second Law ($F = ma$) sets the radial acceleration $\frac{d^{(2)}r}{d(t^2)}$ equal to the magnitude of the force field:

(1) $\frac{d^{(2)}r}{d(t^2)} = r^{\frac{1}{2}}$.

Let us place the mass at the apex, $r = 0$. If the mass is initially at rest with no force applied, there is one clear solution to Newton's Second Law for all times *t*:

(2) $r(t) = 0$

This is rather obvious: the mass, initially at rest, with no outside force or change in its circumstances, will remain at rest indefinitely. However, there are alternative solutions. For any radial direction, the following is a solution

[26] Image and construction taken from "Causation as Folk Science" by John D. Norton

(3) $r(t) = \begin{cases} \left(\frac{1}{144}\right)(t-T)^4, & t \geq T \\ 0, & t \leq T \end{cases}$ where $T \geq 0$ is an arbitrarily chosen constant.

We can confirm that (3) is a solution to Newton's Second Law (1):

(4) $\frac{d^{(2)}r}{d(t^2)} = \left(\frac{1}{12}\right)(t-T)^2 = \left[\left(\frac{1}{144}\right)(t-T)^4\right]^{\frac{1}{2}}$ for $t \geq T$ and 0 otherwise so that

$\frac{d^{(2)}r}{d(t^2)} = r^{\frac{1}{2}}$ with $r = \left(\frac{1}{144}\right)(t-T)^4$

This solution amounts to a violation of the statement above: that it will remain at rest indefinitely. The mass will have motion at some indeterminate time, and this motion will be uncaused, occurring at time $t = T$. The physical conditions on the dome are identical to before—symmetric in every direction, so the direction the ball goes is arbitrary.

Note also that there are no probabilities inherent to the problem here, as probabilities do not have much place in Newtonian mechanics. Even if we were to add probabilities for the direction of the ball, it would amount to an equal probability to travel in every direction. However, there is no way for us to add probabilities for the time T that accurately represents the solution (3), as all potential times T are necessarily treated equally. Summed over the infinite number of time intervals (0,1), (2,3), (3,4), ... we would have dx probability assigned to each interval, entailing a probability of zero for each one.

To properly assign probabilities, Norton argues that we would need to graft unnatural additional physical properties onto the system. Per Norton:

> "For example, consider the natural condition that, at any time t, we always have the same probability of no excitation occurring over the next (arbitrarily chosen but fixed) time interval Δt, given that no excitation has occurred by the start of that time interval. This condition uniquely picks out the exponential decay rule $P(t) = \exp\left(-\frac{t}{\tau}\right)$ where *P(t)* is the probability

of no excitation over the time interval (0,t) and τ is some positive time constant. At any time t, the probability of excitation in the ensuing time interval Δt is just $\exp\left(\frac{t+\Delta t}{\tau}\right)/\exp\left(-\frac{t}{\tau}\right) = \exp\left(-\frac{\Delta t}{\tau}\right)$, which is independent of t as required. The problem is that the dynamics of excitation is governed by the magnitude of the time constant τ, which is the mean time to excitation. A small τ means that we likely will have rapid excitation; a large τ means we will not. Nothing in the physical setup of the dome and mass enables us to fix a value for τ. We must fix its value by arbitrary stipulation, thereby inventing the new physical property of rate of decay, which is not inherent in the original physical system." (Norton 10)

Further concerns can then be raised about Newton's First Law, but reformulating the law in the following way is consistent with the solutions found:

For times $t \leq T$, there is no force applied, since the body is at position r=0, the force-free apex; and the mass is unaccelerated.

For times $t > T$, there is a net force applied, since the body is at positions r>0 not at the apex, the only force free point on the dome; and the mass accelerates in accord with F=ma.

What is clearly crucial to this construction is the time T. The solutions we found in (3) entail the following acceleration:

$$(5)\quad a(t) = \begin{cases} \left(\frac{1}{2}\right)(t-T)^2, & t \geq T \\ 0, & t \leq T \end{cases}$$

Thus at time $t = T$ the mass is still at $r = 0$ and has acceleration $a(0) = 0$. Hence at time $t = T$ there is no force and the mass us unaccelerated, but at any subsequent time $t > T$ there is a non-zero force and acceleration.

Herein lies the crux of the issue: how and when does this acceleration/force occur? The natural move is to look for some "first instant" of motion and from there find the cause of the

motion at that instant. The equations above seem to imply that there does exist an instant $t = T$ at which the mass moves. This is a misconception and the cause of confusion at first glance. Rather time $t = T$ is the *last instant at which the mass does not move*. There is no first instant at which the mass moves. Motion occurs only over the interval $t > T$ and this interval has no first instant (Norton 11). That is, for any $\varepsilon > 0$, it is the case that $|t - T| > \varepsilon$ for all real numbers *t, T* with $t \neq T$. There is no possible first instant candidate which cannot be preceded by an earlier candidate. Thus we conclude that there is no first instant of motion and thereby *no first instant at which to seek the initiating cause*. Norton's move here is sound and is made to avoid a vicious infinite regress. As we can allow for indeterministic solutions in a classically deterministic picture, the "trigger" being a non-event at an unspecifiable time is not an issue.

Non-Causal Fundamentalism

We return now to Norton's negative thesis of non-causal fundamentalism, or the idea that science is not at an ontological level strictly cause and effect. To clarify, this is *not* to say that cause and effect are mere delusions, rather that they are not a necessary part of a modern scientific theory. The place of causes in our scientific ontology can be likened to that of superseded theories of gravitation or particle physics: gravitation is not a force but a curvature and particles are both particle and wave simultaneously, yet in many applications we treat them as if the prior case were correct. We continue to exercise the "truths" of the superseded theories in a number of limited cases—so too with ideas of deterministic causation.[27]

[27] Norton further references the material theory of heat (Norton, *...and the caloric*) here to draw comparisons between a defunct theory and a new one. In the eighteenth and early nineteenth century, heat was conceived of as a conserved fluid. The temperature measured the density of the fluid, and the natural tendency of the fluid to flow from high to low density was manifested as a tendency to flow from high to low temperature. The material theory fell apart with the introduction of energy conversion into things like work and conduction. As long as the process contained no conversions, the treatment of heat as a conserved fluid once again becomes applicable.

Newer, conceptually dense and computationally obtuse theories can be found to have the "generative capacity" to recover older, simpler theories. Our general, common sensical daily application of cause and effect may not be indicative of deeper ontological truth but in the limited circumstances of our lives the white lie is immensely useful. To withhold reality from such entities like heat, gravity, and causes because they are not fundamental is to risk an infinite regress: if there is no reality to the things we work with if they are not a necessary part of the ontology of our mature sciences, can one be confident about the reality of anything at all? The regress here is one of meta-powers, or that the powers we work with actually get their reality from a deeper layer we do not see, those powers from a layer beneath them, and so on. One way to stop the regress is by finding a fundamental layer to stop at, but one must have clear reasons for stopping there. For there to be reality under this view, we would need to have confidence that our sciences of today will not be surpassed by future sciences that toss out our assumptions as bosh, a confidence which we cannot and should not have.

We rather take the position that such concepts are persistent in our world but do not lay claim to the same reality as what is part of our fundamental ontology, whatever that may be. Cause and effect are thereby part of a reality *built entirely from a fundamental theory of powers*. The existence of such derivative realities should not be a foreign concept, as much of what we work with—schools, books, work, play—are in some way derivative of the more fundamental picture of life. This more plastic and less precise model of causality will allow us to build Timothy O'Connor's theory of free will through agent causation later on without having to make strong ontological claims as to the nature of reality itself.

Toward Causal Powers
Process

Up to this point, the bulk of the work has been on the defensive, rejecting the claims of the "opposition" and arguing *reductio ad absurdum*. We have come to a stage in which the following theses have been (hopefully) convincingly argued:

1. Free Will is *not* compatible with Determinism
2. Causation is *not* to be associated with Necessitation
3. Deterministic Causation is *not* ontologically fundamental[28]

These have all be negative truths, defining the space around the questions which we are trying to answer; namely:

1. Can Free Will exist in an Indeterministic world?
2. If cause and effect are not necessities, what are they? Do they exist at all?
3. What is the ontological truth resting underneath Causation?

To give some insight into this chapter's initial divergence from the common thread thus far, we should explain the desired outcome. The stated goal is to build a metaphysical model of the world based on a theory of powers. Using Molnar's arguments in *Powers*, we will discuss why, at a basic level, we turn to a theory of powers. This will not involve a deep dive into the minutiae of Powerism, but a breakdown of the relevant features will be included. To an extent, we are forced to assume the ontological truth of Powerism without justification, as such justification has no room

[28] Norton's claims concern the non-fundamentality of deterministic causation. If causation is indeterministic in some way, it is possible to make the argument that it is, in fact, fundamental, but this does not pose as much of a problem to free will as deterministic causation. Indeterministic causation, while not a perfect solution. still leaves the door open for possible agent control

in this paper. This involves a reliance on realism and its necessary components; the ideas of universals and particulars.

We thus claim that powers are the metaphysical bedrock of reality, and the argument shall flow as follows: powers are ontologically dependent on other powers[29]. Powers are monadic properties; not a network of polyadic relations. We then move towards a Dispositional theory of powers, arguing that in the most general sense, causation is generative behavior of objects that is governed by their properties. A disposition or power is something that has possible manifestations though it may nevertheless exist unmanifested. This is where our argument commits us to realism, which we will not challenge.

It is properties that do causal work, and they do so because they are powerful (properties will be defined as bundles of powers). Events exist as well. Events that are caused will be produced by powers at work. Events will be the changes produced when powers exercise themselves. These changes can also push other powers together, so when we talk of events in causation, we are mainly talking about how the empowered properties got together, and it is these partnered properties that do the causing.

From here, we will discuss a Vector Model of causes proposed by Mumford and Anjum, which will lead us into Mutual Manifestation. We will use Marmodoro's partner-power model in conjunction with Mumford/Anjum's idea of mutual manifestation. We will move on to a discussion of Simultaneity of the manifestation of cause/effect, and how we are to avoid challenges of infinite regress. We will follow with an account of dispositional modality and probabilistic models of causation, which will lead us into a discussion of a macro-model of causation based

[29] For this to not lead to an infinite regress, we claim that these powers are ontologically codependent on their partner powers and not upon some unrelated or meta-level of power. For more on ontological dependence, see https://plato.stanford.edu/entries/dependence-ontological/#SomProForModExiAna

around powerful Agents and their dispositions. This model will be what we conclude the discussion of Powerism with, and it will be the model against with we compare O'Connor's theory.

Powerism

Let us begin by defining what it is we mean by Powerism, and why we have turned to it in the first place. At a basic level, any theory of Powerism takes the attribution of causal powers, occasionally referred to as 'dispositions', to be ontologically fundamental. Things or objects in our world have a number of facts about them that describe how they act. Take a wine glass: it has a number of dispositions fundamental to its composition, such as the disposition to shatter when struck, to allow light to pass through, to change temperature according to its contents. These dispositions do not lead inevitably to what they describe, but rather imply a tendency towards an event when in the presence of some partner: the hammer to shatter, the light to shine, the cold liquid to chill.

These properties of the glass and their partner powers are what do our causal work, doing so because they themselves are powerful. Powers have possible manifestations though they may nevertheless exist unmanifested. Under our definition, these properties do not exist without being tied down to some *thing*. Properties are then properties of objects or things: length, volume, mass all have to be *of some thing*. So, when we say that the hammer broke the glass, we make claims about a cause-effect event and the objects involved therein. The hammer broke the glass and the glass was broken by the hammer because they have some causally powerful property that relates to this event. This rather general definition is a type of substance causation, in which "it is a something about the substance that does its causal work" (Mumford and Anjum, 1994).

What, then, is an "event"? The commonsense construction of causal events is used by everybody all the time: pulling on the faucet causes the water to flow, swinging the hammer breaks the glass, and smoking causes cancer. Such construction is causal, but why is it so? We assume

some base level of background knowledge when we make these generalizations; we know it is not the act of smoking that causes cancer and swinging a noodle at a glass (probably) will not break it. It is rather the properties of the objects involved that do the causal work; properties such as mass, velocity, fragility, and so on[30] are the truthmakers of our causal construction. So, then, events do exist, and we can explain them as the causal productions[31] of powers at work. Event causation is merely the generative behavior of objects governed by their properties. In other words, events in our conversation about causation are the situations in which partnered properties come together—how the sugar met the tea or how the mass met the fragile glass—but the events themselves are not what does the causing. It is rather the properties.

There are both particular and general causal truths. Billiards balls rolling around a table are clearly particular causal events due to the chaotic nature of the game's construction. General truths are a bit more obscure: smoking causing cancer is vaguer than the cue ball hitting the 8 at the right angle to go into a pocket, but the causal truth is still apparent. General causal claims mean that some cause C disposes towards some effect E only, which is not to say that every particular c will produce an effect e. There may be some singular causal truth that Tim's smoking caused Tim's cancer and there can be a general truth about smoking causing cancer, but this one case does not commit us irreversibly to the thesis of smoking causing cancer in all circumstances. There is a necessary context-dependence in all causal events, which appears to apply some kind of

[30] The "smoking causes cancer" example, a common one used to illustrate long-term causation, can be confusing. As far as we understand it, it is not the individual inhalation through a cigarette that causes cancer but the cumulative detrimental impact on a person's lungs from years of smoking that then causes the cancer. It is impossible to point to one specific event or inhalation and call it the sole cause of the cancer. Keep in mind that even this point was disputed for decades, albeit by those with a vested interest in such a connection's falsehood, but the example serves to illustrate causal ambiguity to some extent.

[31] This definition holds only for events that are caused. To explain uncaused events, see Norton's Dome and discussion of indeterminism.

metaphysical priority to particular truths.[32] In other words, the particular causal connections of some situation do not supervene on the causal powers of the agents. That is, fixing the causal powers does not fix whether there are causal connections between powers and particular events.[33]

Properties

In this talk of properties, all we have done to explain what they are is provide a few simple examples, neglecting to define what a property or a disposition actually is. Properties are clusters of causal powers[34]. The ontology is as simple as possible: powers are the base from which causation arises. Powers are thus the metaphysical bedrock of reality. All powers also have partner-powers, each of which serves as a necessary condition for the existence and manifestation of the other[35] (Marmodoro, p. 58, 2017). Each power can thereby be identified jointly by the type of activity its manifestation consists of and the possible types of partner-powers compatible with it, allowing for reciprocity of causation in a variety of causal scenarios.

Consider heat, a property fires have and cold fingers (relatively) don't. Heat as a property then has two partners: the power to warm and the power to be warmed. When a fire causes my fingers to become hotter, the fire's power to heat is activated, manifesting by my fingers heating up.

[32] There is really a token-type distinction here between particular and general truths. A token has to be of a type and each type exists in its tokens, and the degree to which one has priority over the other is unclear. Given the issues of context dependence we have raised, we will lean towards priority for particulars.

[33] Countervailing powers explain much of this. The agent with the power to move their hands cannot do so when handcuffed, the ball cannot roll away when it is held, and so on. The manifestation of partnered powers requires specific circumstances to occur, particularly the presence of the partner powers. When Norton's ball rolls down the dome, its powers are manifesting simultaneously at some indeterminate first instance.

[34] This definition and our construction of events as moments of mutual manifestation is derived from a number of sources, of which the most prominent is *Causes and Powers* by Mumford and Anjum.

[35] Of course, there can absolutely be more than just one partner, which allows for possible worlds arguments. We won't get into these in this paper.

However, the fire's heating of my fingers can take place only if my finger's power to be heated is activated mutually and manifests itself by getting my fingers hotter.

Some properties are a bit trickier to define. Take a property like sphericity: if properties bestow their bearers with causal powers since they are themselves clusters of causal powers[36] what powers make up "being a sphere?" To explain, we must argue that sphericity is in fact a covert disposition[37], such as the disposition to roll in a straight line down an inclined plane or to reflect light in a particular pattern or manner when illuminated. A soap bubble sticks to an inclined plane, a ball weighted unevenly rolls erratically, and balls of clay deform and smush, but these are just spherical things that did not manifest this particular disposition. They instead manifested some other more powerful property. In the case of the soap bubble, while it is disposed to roll in a straight line due to its sphericity it fails to do so due to its countervailing power of stickiness, which is clearly more powerful due to the way in which it manifests. The weighted ball is disposed to roll in a straight line forwards but is overpowered by its simultaneous power to roll left towards the weighted section, and so on with the clay. These powers still exist even though they are unmanifested: they are simply too weak to reveal themselves in most cases.

A causal power is then the potentiality to bring about some change or the activity of bringing about that change (Marmodoro, 2017). Nothing inert or impotent is needed within the power's nature that anchors the power to reality[38]. Rather, there is nothing more to a power that its powerfulness: what the power can and cannot do. A power can potentially be its actuality, by which we mean that the potentiality is for the power's *activity* when it is being exercised. The

[36] These clusters of causal powers are made up of a number of distinct single powers. If they have substantially different powers than is apparent from their component powers, the issue of emergence arises, which we will not discuss here.
[37] Sydney Shoemaker has a different theory, identifying such properties with conditional dispositions: i.e. the disposition to roll a certain way if the object were to become rigid and persistent. This is another possibility which makes properties functions of powers
[38] For a critique of this position, see John Hawthorne's paper "Causal Structuralism"

powerfulness of a power is then its capacity to actively engage in an activity such as a doing or a suffering. This is *not* equivalent to saying that the powerfulness of a power is its readiness to get activated or to instantaneously 'jump' into another power—the activated power is still the power under our view.

This idea of partnered powers and mutual manifestation may entice us into thinking that, at a fundamental level, reality is a web of relations. This is not the case, as powers are not polyadic relation properties. Furthermore, there is no polyadic relation connecting a power in potentiality to its manifestation since *the manifestation of a power is numerically the same power in a different state*. Nor is the essential nature of a power *P* understood to be a different power *Q* to which *P* in potentiality is related; *P is one with* its essential nature (see Marmodoro 2009; 2013).

The essence of a power tells us what type of entity the power is; namely, if it is the sort that can do such and such type of activity to this or that type of thing. To be clear, there is no polyadic relation connection a power to its power partners. Powers are ontologically dependent on their partners, but ontological dependence is not a polyadic relation[39]. Powers are instead monadic properties of a special kind in that they have intentionality or 'directedness' towards something. Aristotle discusses the idea in the *Categories*:

> "We call *relatives* all such things as are said to be just what they are, *of* or *than* other things, or in some other way *in relation to* something else. For example, what is larger is called what is *than* something else (it is called larger than something); and what is double is called what is *of* something else (it is called double of something); similarly with all such other cases" (Aristotle, *Categories* 6a36-b3)

[39] Strawson argues that this idea is true only when powers are 'extrinsically' specified, , i.e. specified with respect to something else. He says that the basic 'power being' of a thing is just its intrinsic nature considered wholly independently of anything else. Thus: x is ABCDEFGH (description of its intrinsic nature). This is why x does Q to y (has the power to do Q to y) and does R to z (has the power to do R to z), and so on and so on. Alternatively one can say that it is its x being B in particular that is the reason why it has the power to do A to y and so on.

There is clearly then some ontological relation between these properties, and we can call such a relation an interdependence.

> "If there is no master, there is no slave either…When there is a slave there is a master; and similarly with the others [*sc.* other relatives]…Also, each carries the other to destruction; for if there is not a double there is not a half, and if there is not a half there is not a double. So too with other such cases" (Aristotle, *Categories* 7b6-22)

These ontological interdependencies of causal powers are thereby not polyadic relations holding between them. They are rather the mutual *conditions* that enable their respective existence and activity. A power P is manifesting when such and such a power P' (where $P=/=P'$) is satisfying such and such conditions, for example that it exists, it is appropriately located, nothing impedes it, and so on. In the case of the fire warming my fingers from before, we can consider the fire's heating power to be P and my fingers power to be heated P'.

Causation, too, is not a web of relations. Causation often involves a change. When it does so, there is an event, and events as we have constructed them are to be understood essentially as changes rather than as property exemplifications. A particular exemplifying a property at a time would be better understood as a state of affairs. Causation is then not a simple relation between two events but rather a state of affairs coupled with powerful properties that make an event occur.

We are then in a position where we have powerful partnered powers with mutual ontological codependence but no polyadic web of relations. A cause is understood to be something that disposes to some degree towards an effect, so causation is events happening as they do because the agents therein have the powers that they have. In the fire example, the fire should be understood as having the disposition to warm a human body among other things. This is true in both the general and particular case: all fire has the disposition to warm things colder than it, and this particular fire disposes to the warming of this particular body that is in its vicinity. Every event

can be interfered with to a variety of degrees by countervailing powers. To explain this, we turn towards a vector model of causes.

Vector Model

We first need some method of explaining what change looks like. What we called a state of affairs before can be reduced to a one-dimensional "quality space" such as the one proposed by Lombard. When one thing causes another to warm, it is causing a movement in the quality space for "heat." The quality space, pictured below in Figure 2.1, is the background against which events can occur in our vector model. We decide here to employ vectors as a simple yet illustrative model of causation as they have both direction (intentionality/directedness) and intensity (powerfulness), thereby likening them to powers rather cleanly.

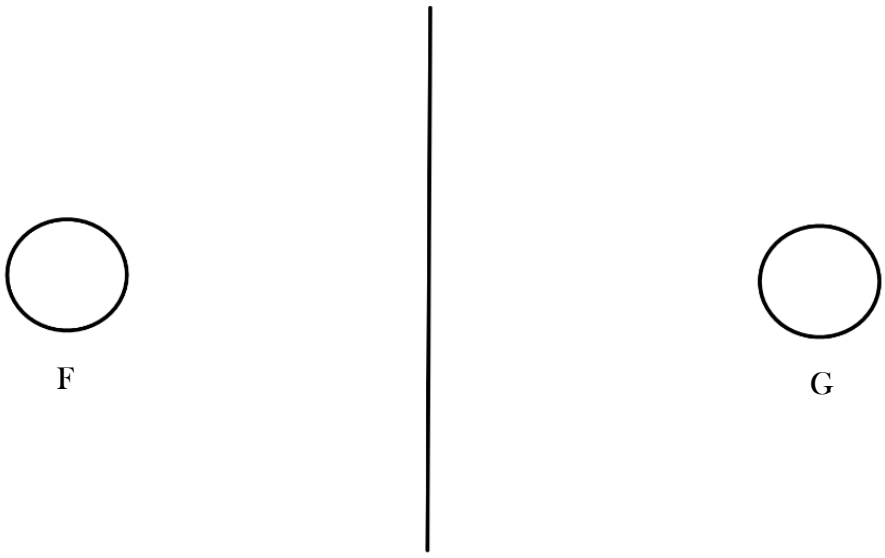

Figure 2.1: A one-dimensional quality space

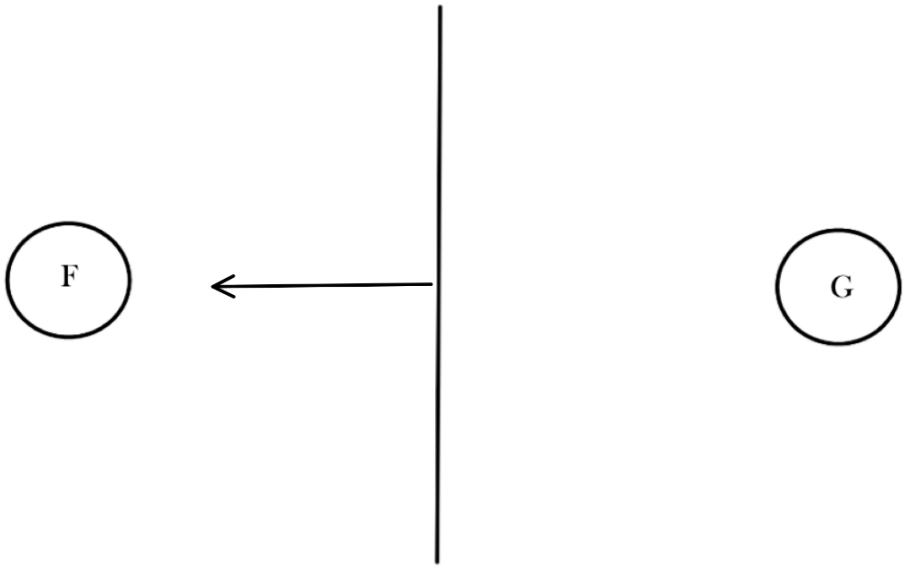

Figure 2.2: A disposition towards F

Think of F and G as the two sides of the heating coin, with F being "hot" and G being "cold." In Figure 2.2 above, we see a single vector plotted on the same one-dimensional quality space. To be precise, the figure illustrates a certain moment at a certain starting point on the F-G quality space where there is a single disposition towards F in operation (exercising its power). From this simple picture, a large list of questions arises. First and foremost, this picture is nothing like the real world. In every real situation there will be a number of relevant powers disposing in wildly different directions. Some, like the picture above, may dispose towards heating or cooling, but others may be completely orthogonal and dispose towards a color change or a gain in weight. We can thus have an object undergoing a number of simultaneous changes. However, these vector diagrams are meant to represent only one specific causal situation relating solely to the powers that are causally relevant to one dimension of the property, impacting only one subject.

We must also explain what it is we mean by a "moment" in these diagrams. The vectors in a quality space represent only the operating dispositions without actually showing any real change. If change were to occur, we would have to represent it with an entirely new vector diagram. That being said, a moment should not be thought of by default as an infinitely small time slice or indivisible instant. A world of powers implies a world of "active, dynamic particulars" (see Harre and Madden 1973) which yields some problems for any attempts to time slice (see Mumford 2009). Issues of dynamism and flux are also relevant. We are thus to think of a moment as allowing for some extension through time, as powers can take time to do their work[40]. Moments, therefore, are temporally open events.

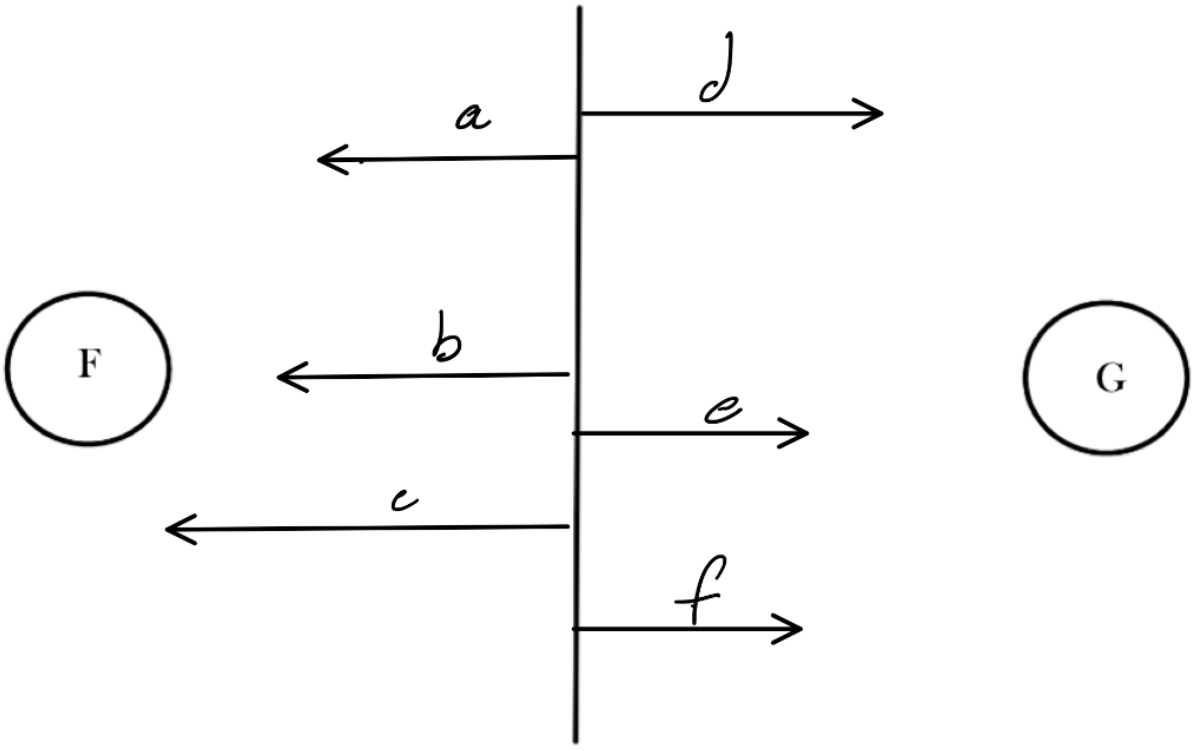

Figure 2.3: Multiple powers (countervailing)

[40] A clear example of temporally extended powers is the fireplace in a cold room. During the winter months, especially in the older, poorly insulated homes, by the time you get home from work the temperature in your home will be far below comfortable, so you build and light a fire. While rewarding and pretty to look at, this fire doesn't suddenly solve the issue of the cold—its power to heat the air around it takes time, and the heat transfer process in the air molecules does not occur instantaneously. The "moment" to describe the room being heated could be indefinite—after all, the room is poorly insulated and there is a simultaneous countervailing cold power from the outside world.

Figure 2.2 is not an entirely accurate image of the world as we know it. A more realistic model is given in Figure 2.3, where multiple powers dispose the object in either direction. In order to determine the resultant vector, simple vector addition will suffice; remember that the length of the vectors corresponds to the "powerfulness" or magnitude of the power, so that when summed there will be some cumulative effect resulting from the input of all relevant dispositions.

The model then shows what it is to be *a* cause[41], namely, to be one of the operating powers represented as a vector that disposes towards an effect. Whether, how, and to what extent the effect occurs will be determined polygenically by vector addition[42]. As you can see above, countervailing powers can affect the timing, chance, or extent of an effect. Are they then considered part of the cause of the effect, even though they dispose in the opposite direction? We should here say that countervailing powers were not necessarily a cause of the effect without ruling out the possibility that those same powers could be a cause of the effect happening in a particular way.

Effects are thereby shown to be the product of many powers acting together[43], and it is possible that the sum of these powers may have a zero resultant vector. This is a non-issue in the vector model and in real life—a tug of war is essentially the countervailing powers principle. The same power can produce different overall effects depending on which other powers combine with it. Powers can thus have different partners for the production of different mutual manifestations. This can clearly be shown in the vector model, which allows a power to be depicted working with

[41] The vector-causal model comes entirely from Mumford and Anjum's *Getting Causes from Powers*

[42] A distinction between causes and background conditions cannot have any real ontological strength because the effect is not triggered until they are all present, which can in any case be a momentary matter. At that moment, it matters little from the point of view of the effect that one of the factors that contributed to it had been around for ten hours while another had only been around for a second.

[43] This is not always the case—there are possible single powers in operation in specific scenarios, such as radioactive decay. These are not incompatible with this vector model: it's just a one-vector quality space in a (long) temporally extended moment!

other powers on different quality spaces. Similarly, we can imagine the vector model being extended to a multi-dimensional universe, with causes coming from any and all angles without issue.[44]

Compatibility

Can this vector model be squared with the mutual manifestation conception of causation we raised before? Let us recall that these vectors, like in real life, indication dispositions towards an action and are not all realized when countervailed. Two orthogonal force vectors acting simultaneously on an object do not pull the object first one way and then the other—they act together, pulling the object on the bisector of their respective angle. So too with causal powers. To explain, let us break down a seemingly simple example: tossing a ball in the air.

As we have established, causation can be thought of as the mutual manifestation of powers composing some broader event. When I toss a ball into the air, I am manifesting my power to throw and the ball its power to be moved in a direction by a sufficient force. As the ball rises, the air around it is resisting and slowing the balls movement, while the ball's low friction and aerodynamic properties allow it to continue to move upwards. The mass of the Earth is also manifesting itself in the slowing of the ball coupled with the ball's necessary relativistic tendencies. At the peak of the throw, the ball has a velocity of zero, but a number of powers are continuing their manifestation—gravity to pull it down, air to hold it up, friction to hold it in place, and whatever spin it has to distort the air around it, all of which can be imagined in respect to their partners. As the ball falls, a similar process occurs and it returns to my hands, where powers and dispositions continue to drive action ad infinitum.

[44] For a discussion of issues of overdetermination, see Mumford and Anjum's *Getting Causes from Powers*

It may appear that Mumford and Anjum's vector addition picture commits one to determinism—once the vectors are in place, the effect seems to be directly determined—but this is not the case. Vectors are not all compositionally the same. While some have consistent physical power, others can be probabilistic, manifesting only some of the time, or even indeterministic, manifesting all by themselves without any interference to produce changes. Vector addition performed with these must not be straightforward summation but rather have some other mechanism by which these powers are exercised and can either contribute to the cause which would have occurred without their existence or turn towards some entirely different event. Rather than committing us to determinism, the vector model provides us with some structure by which we can understand the manifestation of powers.

So, then, we have some idea of causation, but we have a big question to answer: time. Specifically, when do these powers manifest? Do they occur simultaneously, one before the other, or in some unknowable way? Perhaps most importantly, which has temporal priority, the cause or the effect? There is a risk of regress here, about when specifically there is a change from cause to effect. To answer, we turn to the same argument used in *Norton's Dome*:

> The natural move is to look for some "first instant" of motion and from there find the cause of the motion at that instant...[this] is a misconception and the cause of confusion at first glance. Rather time $t = T$ is the *last instant at which the mass does not move*. There is no first instant at which the mass moves. Motion occurs only over the interval $t > T$ and this interval has no first instant. That is, for any $\varepsilon > 0$, it is the case that $|t - T| > \varepsilon$ for all real numbers *t, T* with $t \neq T$. In other words, there is no possible first instant candidate which cannot be preceded by an earlier candidate. Thus we conclude that there is no first instant of motion and thereby *no first instant at which to seek the initiating cause*.

To extend this issue to the manifestation of powers, there is then no first instant of manifestation. We can consider then the cause and the effect to be temporally extended events which overlap in a number of cases. The effect as we think of it is in fact the event of mutual manifestation—the hammer hitting the glass, the ball falling off the shelf, the keyboard sending signals to the screen. There is no restriction on the time it takes for the effect to manifest, for much of what we perceive as effects are highly delayed results. Effects can occur instantaneously, such as in cases of quantum entanglement, or over the course of years in macro-events such as weather patterns and economic turns. That there is no gap in time between powers being suitably partnered and acting is not a contradiction because causation involves "continuous processes that are extended through time and involve many changes of properties" (Mumford and Anjum, 2014).

To reiterate, we do *not* think of deterministic causation as a fundamental building block of reality, but rather of powers as the pieces of the puzzle and causation as the image thereby derived, meaning that causation in some way is, in fact, fundamental. Where these powers exist and what this causation looks like is a question to be answered in the next chapter, but we have thus far constructed some reasonable model of how to think of powers, causation, and indeterministic events. Where does Free Will fit into all of this?

Agent Causation

Background

Any discussion of Agent Causation and what it entails is a discussion of the metaphysics of free will, which we have yet to do. Agent Causation theorists argue that causally powerful agents are necessary features of any theory of freely chosen action, even though there are a number of possible intentional activities without free action all together[45]. Agent Causationists are also mostly Incompatibilists, the thesis that there is no world in which determinism and free will can coexist. They, like us, argue that the absence of determinism is not enough for freedom to occur, so they posit some basic requirements for models of free, responsible agency. AC Theorists must give a clear account of what they mean by 'agent control': the manner in which a particular piece of behavior is connected to, controlled by, or is 'an outflowing of' the agent. They must also allow for the possibility of alternative, viable courses of action that are genuinely open to the agent.

There is some degree of intuitive pull towards these theories. The agent has a number of internal states with "objective tendencies of some determinate measure to cause certain outcomes" (O'Connor, *Oxford*, 340)[46] This then provides an opening into what could be agent control, in which the agent can freely select from a plurality of real alternatives, but once again the mere existence of an opening does not lead inevitably through said door. AC Theorists provide that causal impetus: the agent.

[45] C.A. Campbell, 1967; John Thorp 1980; Allan Donagan 1987; Randolph Clarke 1993, 1996; and Timothy O'Connor 1993, 1995, and 2000 are among those who have specifically taken this view

[46] This and much of the discussion to come is drawn from two sources: Timothy O'Connor's "Libertarian Views: Dualist and Agent-Causal Theories" in *The Oxford Handbook of Free Will,* p. 337-355, and *Persons and Causes: The Metaphysics of Free Will.*

Simple and Causal Indeterminism

Before we dive deeper into Agent Causation, we should first discuss a failed preliminary attempt at explaining free will in a causally indeterministic[47] world, for it is the precursor to what is to come. Specifically, I am talking about Carl Ginet's Simple Indeterminism, the thesis that agent control is noncausal in nature and that there is a simple mental action lacking internal causal structure at the core of every causally complex action. He argues that these acts instead have some "actish phenomenal quality" that separates them intrinsically from unbidden thoughts or images. In other words, these conscious mental actions have a feeling of direct production or control.

This action, as conceived of by Ginet, is some volition that causally generates a person's voluntary exertion of her body directed to the immediate present. In the case of raising one's finger, this volition becomes a "fluid mental activity over time" (Ginet, 1966). Whereas our actions may be uncaused, at the bottom they are not inexplicable. Reasons and reason codes explain our actions by entering into the contents of whatever the guiding volitions may be. Ginet discusses action in the following manner:

> "To determine an event is to act in such a way that one's action makes it the case that the event occurs. Let us grant that if the event is not one's own action, then this requires that the event be causally necessitated by one's own action and thus that it not be an undetermined event. But if the event is one's own action, then one's determining it requires only that one perform it; and one's performing it, which is just the action's occurring, is compatible with that event's being undetermined, not causally necessitated by antecedents" (Ginet, p. 127, 1966).

[47] A discussion of causal indeterminism as O'Connor conceives of it is to follow, but we shall be able to square it with the models we have established in previous chapters.

Such an account is unsatisfactory. The notion of my *determining* an event to occur does not have independent content and is not equivalent to my *bringing it about* as there is no necessary causal content. Determining an event in this context is necessitating it as part of a causal chain outside of agent control, while to bring an event about is to actively control the event's outcome. Is an event an action merely because of the phenomenal character of the mental event as its origin? The phenomenal quality of such events should be entirely irrelevant.

So, this account, with the addition of an account of the lack of external control, involves causation only in a negative manner. It requires the absence of certain kinds of controlling factors or what we have called thus far 'countervailing powers.' Positively, it consists merely in the occurrence of an event having an "actish phenomenal quality," which is implausible in the same way that broader Phenomenalism is implausible about the ontological structure of the world. We are committed in our theories to Realism: seemings are not sufficient for realities.

The agent then must be able to control her actions by way of her prior reasons and thought processes, which *may* cause but not necessitate whatever her actions may be. O'Connor calls this thesis "causal indeterminism" and sets out to refute it. Deterministic or probabilistic causal theories can successfully underwrite a weak kind of agent control, yet not any kind of free will, for the theories fail to allow for any form of real alternative choices or possibilities for the agent to take. Carl Ginet's simple indeterminism yields such possibilities, but there is no apparent agent control. If we instead consider causal indeterminism, we may have an avenue of attack. The causal relation of the deterministic picture fails to link some belief-desire complex to the subsequent action; it instead links some set of prior mental states and an action with no active agent control.

It is entirely possible that this connection between mental states and action is purely probabilistic. For the causal indeterminist, free will would still be possible only if there was some

level of control over the probabilities of mental states leading to action. If rather there is some set of probabilities about competing internal desires that dictate which action the agent takes, it is unclear where the agent control comes in to play in the causal indeterminist picture. If we can conceive of a causal relation that constitutes some form of agent control, there is no clear reason why this relation must be necessitating. Our beliefs and desires vary, but many of them are enduring and constitute central features to our own self-conceptions. When we act in alignment with these complexes, or when our actions are causally produced by these central features, they are products of "who we are" at the moment of the effect's occurrence. This does not require that the actions be a deterministic outcome of prior mental states, only a causal outcome. The causal indeterminist holds that in many circumstances, persons have distinct desire-belief complexes or reasons that point towards different courses of action and that the performance of any of these would coherently graft onto precisely the same prior circumstance. Thus, each reason is a potential cause of the corresponding behavior, and whichever action is undertaken will have been caused by its match[48].

Notice that this picture does not solve the picture of free agency. While causal indeterminism allows for genuine variety of possibilities, it does no more for agent control than claim that such actions are merely 'outflowings of' the agent, not freely chosen by the agent. Rather, it falls into a statistical or quasi-statistical tendency that governs the general pattern of behavior. Probabilistic tendency is an insufficient mechanism for free control. While agents still are the cause of events in a probabilistic picture, it is unclear how they actually exert control freely instead of being mere conveyors of probability under causal indeterminism.

[48] This does not contain any explanation for uncaused events

Before going further into O'Connor's theory of agent causation, we discuss Storrs McCall's picture of causally indeterministic freedom of choice. Without getting too deep into the fundamental physics of his theory, his conception of the universe is useful and provides us with some further requirements for a sufficient picture of free will. According to McCall, reality is a multi-variated space-time structure that is branching along the time axis, each branch of which is a four-dimensional manifold that constitutes a physically possible future. The present is the first branch point, with the past below it (the universe's 'trunk') and the future the set of branches beyond it. Objective probabilities are a function of branch proportionality, and the flow of time is constituted by branch attrition—the falling away of all possibilities relative to a given moment save one.

He then argues that the prima facie problem for the causally indeterministic account is as follows: absent determinism or at least a high probability of a particular outcome, one cannot speak of an event as being caused—it is instead random. Thus, an understanding of responsible free agency requires, in addition to significant indeterminism in the process leading up to a choice, a mechanism of nonrandom selection that permits an intentional explanation of the choice. Any such explanation is irreducible to explanation in terms of the causal probability of the outcome, for as we have discussed previously such reductions lead to mere outflowings with no apparent real agency or possibility for control.

O'Connor's Agent Causation

The basic tenet of Agency Theory is the following: at the core of every free action is an ontologically irreducible causal relation between a person and some appropriate mental or internal

event that triggers later elements of the action[49]. This claim is subject to a number of challenges, a few of which we shall discuss. The first is the idea of an "Infinite Regress of Choice" in which each potentially causing event, specifically internal mental events, is itself caused by some level of meta-cause, with this meta-cause itself being caused by a higher level of cause, and so on ad infinitum. In 1788, Thomas Reid put the challenge in the following way:

> "'Liberty,' they say, 'consists only in a power to act as we will; and it is impossible to conceive in any being a greater liberty than this...To say that we have power to will such an action, is to say, that we may will it, if we will. This supposes the will to be determined by a prior will; and, for the same reason, that will must be determined by a will prior to it, and so on in an infinite series of wills, which is absurd" (Reid 501)

According to O'Connor, this is a mischaracterization of the theory: we do not need to have performed a prior act of will in order to have determined the action-initiating volition. We exert active power (which we conceive of through its effects/manifestations) in so determining it; we determine the will directly, and the exertion of the active power is not itself a type of volition. This is a reasonable response. We can conceive of volitions as a type of mental event akin to what we have called intentions: causally initiating behaviors of a certain type which often are expressed immediately in some way. There is no apparent need for a prior will here, for the two are merely simultaneous.

However, even if the exertion of active power is not to be understood as a prior willing, isn't it a prior event of some sort or other? If so, this collapses back into simple indeterminism. To avoid this, we have to clarify what exactly we mean by the "exertion of an active power." An exertion of active power is to be understood as not an event at all. It is rather the instantiation of a causal relation between agent and volition, which is *exactly* how we have previously categorized the

[49] Part of the job is reconciling two prima facie different conceptions of causation: event and agent causation

mutual manifestation of powers. Looking ahead, it is entirely consistent with the Powerist picture we have provided to conceive of these volitions and wills as themselves pieces of the mutual manifestation of partnered powers, which occurring at a specific time constitutes agents acting in events. Are these instantiations not causally complex events? The willing as a component of these instantiations can be compared to temporally extended events we have discussed before, such as the heating of a room. Volitions and their exertions are analogous to shorter events composing longer events as parts.

To discuss where agent causation comes in, we must first note that a particular which freely and directly brings about an effect has to be an agent that can represent possible courses of action to himself and have desires and beliefs about those alternatives (Reid 1788). O'Connor thus argues that agent causes bring about immediately *executive* states of intention to act in various ways:

> "This direct causing by agents of states of intention goes like this: parallel to event causes, the distinctive capacities of agent causes ('active powers') are grounded in a property or set of properties.[50] So any agent having the relevant internal properties will *have it directly within his power to* [sic] cause any of a range of states of intention delimited by internal and external circumstances. However, these properties function differently in the associated causal process. Instead of being associated with 'functions from circumstance to effects,' they (in conjunction with appropriate circumstances) make possible the agent's producing an effect. These choice-enabling properties ground a different type of causal power or capacity—one that in suitable circumstances is freely exercised by the agent himself." (O'Connor p. 72, 2000)

This picture carries with it a number of ontological commitments. First, universals and particulars are taken as basic ontological categories. Concerning particulars, also known here as agents, we

[50] At a glance, this appears to be a problem: our previous construction of powers and properties has properties defined as mere clusters of causal powers, while this claims the reverse. This is entirely reconcilable. Our picture still holds, and what O'Connor here calls "agent causes" and "active powers" are either simply properties or clusters of properties. Either remain compatible with our ontology.

require that they include things that endure through time, wholly existing at each moment of an extended temporal interval (not perdure). Agents are not required to be a kind of substance radically diverse from physical substance, although the powers of agents must not be reducible to the powers of their micro-physical constituents. Properties are understood to be universals[51] that have essentially their dispositional tendencies—tendencies as interpreted by the causal powers account of causation, which is compatible with our previous picture. Lastly, universals are immanent constituents of the physical world (Armstrong, 1997) as against the universals-as-transcendent-forms Platonic picture.[52]

The agency theorist must still explain how reasons come in to play. Many philosophers, specifically under the light of compatibilism, argue that reasons can explain an action only to the extent that they directly produce it.[53] The agency theorist disagrees with this conclusion, instead believing that reasons can explain an action without leading directly to the action's production. This allows O'Connor and his contemporaries to provide a schematic sufficiency[54] of free action in terms of an antecedent desire:

---

[51] Without contradiction here, we can refer to properties as corresponding to natural classes or resemblance classes of particulars

[52] These commitments can clearly be controversial and argued over ad nauseum. It is not the goal of this work to argue them any further, and they appear to hold up at least to common sense. Aristotle discusses many of the issues here in all of his works.

[53] This claim about the general philosophic understanding comes from O'Connor, and while generalizations about what philosophers agree upon are rarely correct this one serves its purpose as a strawman effectively.

[54] In a footnote of his own, O'Connor briefly discusses his reasoning for non-necessitation, which is based upon the potential for divine action in creation: "I do not provide necessary conditions on the concept of acting for a reason because a belief-desire causal theory is an alternative way this concept (though not that of freely acting {at a reason) could be realized. How about for the more specific concept of acting for a reason in the agent-causal way? I'm again hesitant to say that the conditions I provide are necessary, although I think they come close. The hesitancy concerns the rather extraordinary case of the possibility of divine action in creation. One way of thinking about this—common in medieval philosophy—has it that there is no change in God as a consequence of His creating the world. (Were He to have created a different world, or none at all, His intrinsic state would have been exactly the same) I think sense can be made of this idea by conceiving God's intention to create not as a purely intrinsic state—one that would vary, depending on which world He created—but as a causally relational state between Himself (whose intrinsic state is properly characterized as a state of willing Himself) and the resulting creation." (O'Connor 85n1)

"The agent acted then in order to satisfy his antecedent desire that θ if:

1. prior to this action, the agent had a desire that θ and believed that by so acting, he would satisfy (or contribute to satisfying) that desire;
2. the agent's action was initiated (in part) by his own self-determining causal activity, the event component of which is the-coming-to-be-of-an-action-triggering-intention-to-so-act-here-and-now-to-satisfy-θ;
3. concurrent with this action, he continued to desire that θ and intended of this action that it satisfy (or contribute to satisfying) that desire; and
4. the concurrent intention was a direct causal consequence (intuitively, a continuation) of the action-triggering intention brought about by the agent, and it causally sustained the completion of the action" (O'Connor 86)

These reasons can themselves be indeterminate on things like timing, environment, and other various factors without contradiction. Similarly, an agent can have reasons for an action which he performs yet these reasons not be the direct cause, for reasons can plausibly cause actions without determining them. It also allows for the limiting case: total conscious ignorance of one's intention in acting. In such circumstances, the response should be plain: the action is not free, and there is no direct agent control over such action. The existence of unfree action is not a contradiction to free action at large, but rather just an example of an agent's lack of complete and total control. Similarly, mis-remembered or mistaken perceptions of one's intentions occur frequently in real life—actions with unintended consequences are often realized to be the direct result of an internal mistake after the fact, and the agency theorist has some awareness of his slip-up.

Event Causation and Wayward Causal Chains

The question we are left to answer now is thus whether an agent's originating activity is itself an event with causes, which would reduce any theory of agent causation to event causation. Event causation comes from the Neo-Humean model of causation, in which laws, dispositions, and

powers are *not* fundamental truths about the world. Rather, categorical truths about the distributions of qualities and other general features about space-time are fundamental. Causal connections are thereby derived in this picture. Events are understood here as the landscape of the environment at a particular time. One event immediately causes another event through their respective space-time relation, the internal qualities of the two events, and the laws of nature. Since this theory is blatantly incompatible with the models of Powerism established in previous chapters, we must ensure that our causal theories do not collapse into this type of causation.[55]

Event causation's reliance on Neo-Humean theory makes it an insufficient mechanism for explaining free will. Richard Taylor highlights the irreducibility of free action to event causation, arguing that actions are to be analyzed as behavior with certain specific types of mental causes such as volitions or belief-desire states, claims that any event-causal view inescapably faces two fatal difficulties: (1) a behavior of the agent's being caused by a mental event is consistent with that mental event itself being caused by a manipulative agent and (2) if the reductive account is given as an analysis of intentional action, a further difficulty arises: wayward causal chains, according to which reasons *can* cause the expected outcome but in such an unusual way that the output is clearly not the original intent of the agent.

Taylor's approach involves supplementing to agent theory in such a way that all actions involve an irreducible causal relation between the agents and the events internal to his action; problem (1) above then couldn't be described as an action the agent performs, as the agent had nothing to do with the event's occurrence (Taylor p. 94, 1966). O'Connor provides an alternative analysis of wayward causal chains:

[55] In the closing section, we discuss the distinction between agent causation and event causation.

> "Consider a familiar sequence of events. My finger presses the doorbell button, the doorbell rings, and your cat jumps in fright. We may sensibly say that my finger's pressing the button causes the causal sequence*, the ringing of the bell's causing the cat to jump* . But what we mean here is simply that it caused the sequence indirectly, by causing the first element of the sequence, the bell's ringing. We may also sensibly say that the electrician's wiring of the doorbell system was a cause of the sequence, *the depressing of the button's causing the bell to ring*...we may term this latter episode one of 'structural' causation, which consists in establishing a causal pathway — here, the wiring and power supply — between two objects or systems that is subsequently triggered by some appropriate event. Here we mean only that the 'structuring' cause provided a context in which some causal factor exerted its characteristic effect. It is not to say that the establishment of an electrical pathway in any way brought about or enabled the button's depression's exerting its characteristic influence on its immediate environment, only that it will determine one important wider effect of that influence." (O'Connor p. 53, 2000)

Both of the scenarios discussed above discuss causes of causings, yet neither seem to support the idea of a cause of an agent's causing her own intention. In the first understanding, there is no contradiction or connection to agent causation—the agent is acting freely as part of a temporally extended causal chain. In the second understanding, it is clearer that free action is involved, yet there is still no need to invoke some other layer of mental state to explain the wayward causal chain. There are countless external factors influencing the causal activity of any basic cause, so saying that an agent causes his own intention should be rejected outright—there is not an event-causal tie between reason and action here.

Analysis

We thus argue that agent causation is an ontological primitive. Within the terms of an agency theory of action, one cannot reduce the notion of agent causality to other, more basic notions. There are a number of conclusions to draw. Firstly, it is important not to misconstrue talk of *an*

agent's exertion of active power. This term, in O'Connor's view, does correspond to a special type of event, but this event is not to be thought of as prior to and constitutively separate from the intention that is the agent's immediate effect. Second, reflection on the nature of an agent's originating activity suggests that there cannot be a cause that produces it. That something can bring about the causing of this event is apparently absurd. The agent's causing of an event and the event itself are not separate existences, so nothing could directly cause the first without directly causing the second. There is no principled way of stopping at one specific level of metacauses and positing an infinite series of metacauses is logically vicious and prima facie false. We also argue that exerting active power is intrinsically a direct exercise of control over one's own behavior.

This does not lead us perfectly into free will. O'Connor acknowledges this, and presents a few further questions to be answered: how is the agency theorist thinking of causation more generally such that we can understand the idea of agent causation as a kind of causation distinct from event causation? How do factors structure the range of an agent's active power or influence the exercise of such power, consistent with the claim that such exercises are not the sort of events that may be produced by any such factors? Does the strong conception of freedom of action within the agency theory, together with uncontroversial observations about the nature of human deliberation and action, suggest that freedom of action is far less frequent in human beings than ordinarily supposed? Lastly, how can we give a consistent philosophical account of the limited presence of active power in a world that is fundamentally event-causal in character, and can any such account be reconciled with the emerging scientific conception of nature?

The common-sense view of ourselves as fundamentally causal agents—not "unmoved movers" but rather "not wholly moved movers"—is apparently internally consistent and theoretically comprehensible through the lens of event causation. The primitive element of

causation conceptually is the producing or bringing about of an effect. As we have shown previously, causation is thereby not the constant conjunction of cause and effect nor a form of counterfactual dependence. A causal relation can be obtained between an agent and some event internal to herself, which does not imply that the type of event effected on one specific occasion will or would always be produced under relevantly similar circumstances—in other words, there is no necessitation relation in causation. *All causation is agent causation*—substances jointly exercising causal powers. Free action is intentional agent causation, but this is simply a species of the agent/substance causation we see in the inorganic world.

Under the classic causal powers picture we consider powerful particulars, which when placed into the appropriate circumstances can manifest their causal powers into observable effects. The powers of these objects are based on their underlying nature—their physical form, genetic composition, chemical constitution, and dynamical structure. Circumstances or states of affairs prompt the exercise of a power in one of two ways: by stimulating a latent mechanism into action, or by removing inhibitors to the activity of a mechanism in a state of readiness to act. This construction parallels that of our mutual manifestation picture as well as the vector models posited in the previous chapter; the vectors are the causal powers and the event space is the state of affairs in one dimension.

From this schematic follows free action, yet interestingly nowhere therein is consciousness found. These rules could be plausibly fulfilled by some unaware automata—is that all we are? Without going into deep neurobiological concepts, taking agency theory seriously requires at least some discussion of the issues. Principal among these is precisely upon which underlying properties or physical attributes agent-causal capacity lies. What features, functionally or otherwise, constitute some system being a free agent, and what would it take for a free agent to no longer be free? There

is enormous difficulty in this issue, and it is ultimately an empirical matter not able to be solved philosophically, but the agency theorist casually conjectures that "a function of biological consciousness, in its specifically human manifestations, is to subserve the very agent-causal capacity" we've been discussing (O'Connor, p. 122, 2000).

We are avoiding here any reference to issues of consciousness, scientific pictures of the world, or emergence as a concept. While we do use emergence to some degree, we do not have the capacity here to argue out its finer details. Similarly, biological issues of consciousness are not here addressed, for I am far from qualified for discussing issues of the mind. We rather provide in this paper some plausible theory of the way the world is structured and how humans as free agents fit into it. There are of course issues to be argued further, and this is not the final word on free will and agency. Consider this rather a (hopefully) reasonable step in the right direction.

Compatibility, Objections, and Conclusion

Compatibility

We then have a picture of agent causation as an explainer for free action and will, with issues of control, sufficiency and determination, reasoning, and causation explained to some degree. How does this theory square with our previously established ontology? In particulars, which here we call agents, we find the powers we argue to be fundamental, with their ability to cause thereby derived. The existence of powerful particulars slots perfectly on top of our picture, a few dimensions above but plausible constructed from our basic vector models. The mutual manifestation of powers doesn't create contradictions either—there are a huge number of partnered powers manifesting at every given moment or event, so when zoomed out we can think of the agent and her respective powers pairing with the powers of the environment surrounding them to produce the effects desired. To avoid issues of overdetermination, we have both powers and causation as fundamental to our ontology.

What about freedom? This too fits. Consider what the powers are: they are all powers *to* something. They have direction and intention and inclination towards some object. Their collective impact on a particular or an agent constitutes intentionality, specifically in the presence of conscious thought. The impact is understood to be the agent's reasoning for actions, and the indeterminate nature of these powers and causation at large, as discussed previously, allows for the agent to have agency. In providing the schema for control over actions, O'Connor and the agency theorists provide the emergence of free action into the world. Note that this is free action is *not* an addition to our ontology but rather a derivative truth, functionally useful and conceptually sound enough for us to claim freedom at an agent/human level of existence.

Event vs. Agent Causation

However, just because agent causation is coherent does not mean it is correct. Much of the work of this thesis has been to build a theory of causation and powers from which free will emerges. At the moment, all we have done is build some parallel theories which are interesting but potentially entirely irrelevant to one another, so to connect the two we must answer questions and challenges about the central feature of both: causation and how it is put to work.

In the previous section, we discussed a problem O'Connor needs to address: how is the agency theorist thinking of causation such that we can understand the idea of agent causation as a kind of causation distinct from event causation? He holds that agent causation *alone* is essentially intentional and purposive. The locution of freedom is then in the agent's causing an intention for a reason. Event causation then has no intentionality therein, but event causation too involves the exercise of some primitive capacities in an indeterministic manner. Both hold that whatever happens is made to happen for some cause, but the cause, be it agent or event, operating indeterministically implies only that it has a positive tendency in total circumstances toward more than one type of outcome (O'Connor 2002). Consider Norton's Dome: the ball is then an agent with two possible outcomes, rolling or staying, with indeterministic tendencies towards both.

So the division between event and agent causation lies in intentionality. Can we reduce event causation further? To do so, we discuss what is actively *causing* when we discuss events. We talk about the 'total cause' of some event, and what makes up this total cause in our ontology is powerful particulars. Event causation is then merely the generative behavior of objects governed by their properties. Events are the situations in which partnered properties come together but the events themselves are not what does the causing. It is rather the properties of the agents therein.

Event causation is then second class to agent causation, and any discussion of it is essentially a discussion of a collection of empowered agents acting in various manners.

The theory of powers thus answers the question of event vs. agent causation. Event causation is a derivation of the actions of various powerful agents collectively contributing to an event, agents which can be free or unfree without contradiction. Agents in these circumstances are powerful, which allows us to call event causation a mere functional truth, useful when discussing macroscopic situations but not a sufficient explainer of action. The powers of events themselves are then structural, governed by the agents therein, and the effects caused by events are the result only of agents. We see empowered agents partnering with other empowered agents to mutually manifest some effect, reducing events to empowered agents. The distinction between event and agent causation is in fact not a distinction at all, but a connection.

Causal Structuralism

We have claimed that properties drive causal power, but how does this occur? What in the nature of properties allows them to confer causal power? One view is Hawthorne's "causal structuralism," which argues that for each fundamental property, there is a causal profile that constitutes the individual essence of a property.[56] Of note here is that this view does *not* hold that there is something to a property (Hawthorne calls it a 'quiddity') that is over and above its causal profile. The two key theses of causal structuralism are thus: "(i) That for any given natural property, there is some causal profile such that having that profile is sufficient for being that property and (ii) that for any given natural property, there is some causal profile such that having that profile is necessary for being that property" (Hawthorne 362).

[56] For an in depth discussion of the matter, see "Causal Structuralism" in *Philosophical Perspectives*, 2001

The strongest case for thinking that "the causal profile of a property exhausts its nature" is methodological: don't invoke what you don't need. We don't need quidditative extras in order to understand our world. Take for example the negative charge of an electron—this is not something 'over and above' the causal role it plays. All our definitions and knowledge about negative charge is essentially knowledge of its causal power. There is no additional thing or quiddity that is in fact "the thing that plays the charge role." There is no need to invoke anything above and beyond what we already have and need: causal structure to fundamental properties.

Reasons

Returning to free agents, the question may be asked: what exactly does the agent cause? Under O'Connor's construction, the agent causes an immediately executive, or action-triggering, intentional state. This intentional state is the agent's "choice" as well as the agent's basic action, which typically constitutes the initial segment of more extended causal processes that result from such choices. However, these explanations in terms of reasons are distinct from explanations in terms of causes. There is a relative strength (or 'powerfulness') to reasons, leading to a noncausal link between actions and the reasons explaining them. The agent recognizing a reason for her action seems to induce or elevate some kind of objective probability for the agent to cause some action. Agent causation is then probabilistically structured by the tendency-conferring states of having reasons to act and also more enduring character traits and long-standing general intentions. The agent is then the sole causal factor directly generating intentions to act but doing so is shaped causally by her total motivational state. The mere fact that she had a reason that gave her a tendency to act does not explain the action.[57]

[57] This and the following discussion of reasons comes from Timothy O'Connor's article in *The Oxford Handbook of Free Will*, which is titled "Libertarian Views: Dualist and Agent-Causal Theories"

This reasons explanation is subject to Davidson's challenge to noncausal reasons: among cases where the agent has more than one reason for performing an action, it is plausible to suppose that in some of them only one reason actually prompted the action, while in others a plurality of factors did so. In what does this difference consist?[58] O'Connor answers the challenge in the following way:

> "I contend that a satisfactory answer to Davidson's challenge requires...the agent causationist to suppose that agents cause executive states of *intention* of a particular sort. The content of these intentions is not merely that I perform an action of type ø, but that I perform an action of type ø *in order to satisfy desire D* (or prior intention I). If intentions have this rich sort of content, then the difference between acting to satisfy desire D1 and acting to satisfy D2 and acting to satisfy both, (sic) will be a function of the content of the intention that I cause to occur. When Davidson asks what accounts for my acting on reason R1 and not R2, given that I was aware of both at the time of acting, the answer will be that we must look to the content of the intention I cause; this will have the form, that I do A for in order to (sic) satisfy reason _____. In a given case, the blank will be filled by either or both of R1 and R2. In actively deciding which action I will undertake, I am *inter alia* deciding which reason I am aiming to satisfy." (O'Connor 2002)

Our Powerism ontology opens up another avenue of attack. We can think of reasons, or the having of reasons, as powers. My having a reason to eat lunch could be identified with my power to eat lunch intentionally. Then we might not need to posit an intervening mental state (the volition) at all. In addition, if I have several reasons, each for incompatible outcomes, then free choice consists in exercising one of these rational powers, to the exclusion of the others. The tie between the reason and the action is not an event-causal tie. Rather, it is a tie between a causal power and its manifestation (and this eliminates the wayward causal chain problem).[59]

---

[58] This challenge comes from Donald Davidson's *Essays on Actions and Events*.
[59] This idea comes from my supervisor, Dr. Robert Koons

What about agents being mere conveyors of probabilities? Again, Powerism helps us here. We can understand the 'probabilities' of our reason states leading to some outcome as the powerfulness of these reasons. Strong reasons for action are the 'high probability' reasons, and weaker reasons for action are the 'low probability' reasons. The powerfulness of these can be thought of as the magnitude of the vectors from our vector addition model, with room for reasons to themselves be indeterministic. Since these reasons do not necessitate action, the free agent is able to choose freely from them—while these reasons have power, they do not determine the outflowing of the agent, and the agent is no longer just a conveyor of probabilities.

Responsibility

To conclude this discussion, we turn to Strawson's (1984: ch. 2, 1994) objection to indeterministic theories of free action in general. He argues that they unwittingly entail an infinite regress at every locus of indeterministic choice of course of action. The way one acts in this view is explained by "how one is, mentally speaking" (M), so for the agent to be responsible for how they act the agent must also be responsible for M. To be responsible for M, the agent must have *chosen to be M deliberately in accordance with reason R1.* For this choice to incur responsibility, the agent must also choose to be moved by R1, requiring a further reason R2, and so on *ad infinitum*. In other words, "free choice requires an impossible regress of choices to be the way one is in making choices" (O'Connor 2002).

How are we to respond to this? To have any sort of responsibility, the free will theorist must have some idea of how the agent is in rational control (and is aware of this control) over the choices she makes when there are no conditions that remove freedom from the equation. For agent causationists who argue free control over action resides solely in the causal efficacy of the agent's reasons, it makes sense to worry about how those reasons came about in the first place. On

the agent-causal account established, the agent causes an 'action initiating intention' *to A for reason R1*, which is then explained by the agent being aware of reason R1 while deliberating and completing the action. The agent is simply in this state, among others, and then deliberates accordingly. The range of possibilities for actions the agent can take is circumscribed by the sum total of the reason-states as well as some idea of the scope of responsibilities that would be incurred as a direct result of the action. But this choice was *not* fully causally determined by nor some probabilistic outcome of those states. Instead, the choice was directly determined by the agent, a choice which can be *explained* by the reason states or M but not a full *result* of those states. The choice was made for certain reasons, but the agent is not constrained by these reasons, and thus there is no need for the agent to have chosen these reasons in the first place.

This hints at an important concern. At birth, we have a number of deep behavioral and attitudinal dispositions, and as we develop we are placed in environments which shape us without any rational choice on our end at how we got there. These dispositions constitute many of our long-term reasons for action, even if nature and nurture do not completely causally determine these ideas. They absolutely influence the range of choices available to the agent: the choices faced by Timmy in a free situation can be substantially different from the choices faced by Todd in the same situation. These contingencies then circumscribe the options available to Timmy upon reflection, which leads into us holding Timmy accountable for his actions. If these factors, unchosen by Timmy, play a large role in his mature choices even in an agent-causal frame, is autonomy enough?

Of course, the agent-causationist can concede that responsibility for character development and the choices that thereby ensue comes in stages and holding Timmy responsible for a choice that is totally out of his wheelhouse or scope of understanding is ridiculous. Similarly, passing on

choices within his range but given limited opportunity to consider can have some degree of reduced responsibility. Immoral actions to one agent may be the only options available to another. Was there any point in the immoral agent's life where more moral paths were open? If those paths were taken, would they have led her to the same position today? Our understanding about these matters is rough at best, and in an indeterministic, non-necessitating world of action, we deem moral judgement appropriate.

The agent causationist must allow for the existence of agents whose basic decision-making capacities are just the same as others yet they lack the moral sensitivity to common human decency through no fault of their own. The AC theorist can also accept that any idea of perfect responsibility for one's choices, character, and constitution is "not just contingently lacking in humans but is impossible" (O'Connor, 2002). Perfect responsibility would require rational detachment and indifference from the outset or complete openness to all possible courses of action. Calling such an idea coherent is doubtful.

There is a final question to be asked: has this done anything for the debate on free will? While this is far from the definitive word on the matter, what we have done here is establish a viable ontological framework from which free will *may* emerge. Powerism has expanded upon agent causation such that we have denied event causation and all of its ontological conclusions, as well as answering issues of wayward causal chains and probabilistic causation. Have we found free will? I am inclined to believe that these theories about causation, agency, and powers have serious merit, so while there is still much discussion to be had on our ontology, I believe we have provided an avenue by which we can indeed find free will. If these are the paths we are walking, we will arrive in a free world, responsible for all the consequences therein.

Annotated Bibliography

1. Anscombe, G. E. M. *Causality and Determination: An Inaugural Lecture*. Cambridge University Press, 1971.

This lecture is used as an introduction to ideas of Causality and what it means for something to be determined. The general assumptions of Hume and contemporary physicalists is Causality is some kind of necessary connection, or alternatively, that being caused is – non-trivially – instancing some exceptionless generalization saying that such an event always follows such antecedents. Anscombe argues that such a definition overstates the "necessitation" of cause to effect, and that causation is not to be identified with necessity whatsoever.

2. Marmodoro, Anna. "Aristotelian Powers at Work." *Causal Powers*, 2017, pp. 57–76.

Marmodoro provides a different definition of Causation in response to the Humean picture. She argues that causation is the exercise of causal powers. What is distinctive about the ontology underpinning this account of causation is that it takes the manifestation of a power to be the activated state of the very same power, and not the occurrence of a new power, allowing for differences in directionality of causation as well as reciprocity.

3. Anjum, Rani Lill, and Stephen Mumford. "Mutual Manifestation and Martin's Two Triangles." *Causal Powers*, 2017, pp. 77–89.

This article discusses Martin's Two Triangles model of causation, which suggests there is no causation over and above the right mereological composition of the component parts or

powers. Anjum and Mumford argue that this is implausible, as there are more complex rules of composition that can involve interaction with and alteration of those powers. Causation can thus involve the production of genuine novelty, which mereology alone does not deliver.

4. Molnar, George, et al. *Powers: A Study in Metaphysics*. Oxford Univ. Press, 2010.

Molnar's incomplete final work covers why we turn to Powers to explain the world, specifically towards a Dispositional Theory of Powers. Molnar agues a distinction between the polygenic nature of effects and monadic/pleiotropic manifestations, as manifestations are isomorphic with powers. He also discusses the idea that laws of nature supervene on (are necessitated by) the simple powers of the objects of the world, with behavior determined holistically by the totality of the relevant laws and contingent initial conditions.

5. Mumford, Stephen, and Rani Lill Anjum. *Getting Causes from Powers*. Oxford University Press, 2014.

From the ontological commitment to Powerism, Mumford and Anjum construct a model of causation, arguing that it is properties that do causal work, doing so because they are powerful. Properties are thereby just clusters of causal powers, bestowing their bearers with causal powers. From here they construct a dispositional theory of causal powers. A disposition or power is something that has possible manifestations though it may nevertheless exist unmanifested, thereby leading us to dispositional modality.

Additional Sources:

6. Aristotle. 1984. *The Complete Works of Aristotle: The Revised Oxford Translation.* Edited by Jonathan Barnes. Princeton, NJ: Princeton University Press
7. Armstrong, D.M. *A World of States of Affairs.* Cambridge: Cambridge University Press, 1997, pp. 80
8. Carlson, Erik. "Counterexamples to Principle Beta: A Response to Crisp and Warfield." Philosophy and Phenomenological Research 66, no. 3 (2003): 730-37. Accessed April 19, 2020. www.jstor.org/stable/20140570.
9. Clarke, Randolph and Capes, Justin, "Incompatibilist (Nondeterministic) Theories of Free Will", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/incompatibilism-theories/>.
10. Davidson, Donald. 1980. *Essays on Action and Events*. Oxford: Clarendon Press.
11. Enç, Berent. "Causal Theories of Intentional Behavior and Wayward Causal Chains." *Behavior and Philosophy* 32, no. 1 (2004): 149-66. Accessed April 20, 2020. www.jstor.org/stable/27759475.
12. Ginet, C. (1966). "Might We Have No Choice?" In Lehrer (1966).
13. Hobbes, Thomas. *Elements of Philosophy the First Section, Concerning Body*. ProQuest, Eebo Editions, 2011.
14. Hume, D. (1748) *Enquiry Concerning Human Understanding*, Oxford: Clarendon Press, 1975, §VIII. (95)
15. Hume, David. 1978. *A Treatise on Human Nature.* Edited by L.A. Selby-Bigge. Oxford: Clarendon Press

16. Hume, David. *A Treatise of Human Nature*. Wildside Books, 2007.

17. Kane, Robert. 1996. The Significance of Free Will. New York: Oxford University Press.

18. Kane, Robert. 1999. "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism." The Journal of Philosophy 96: 217–240.

19. Kane, Robert. *The Significance of Free Will*, Oxford University Press USA - OSO, 1998. ProQuest Ebook Central, https://ebookcentral-proquest-com.ezproxy.lib.utexas.edu/lib/utxa/detail.action?docID=3052339.

20. Koons, Robert C. *Metaphysics (Fundamentals of Philosophy),* Wiley. Kindle Edition.

21. Laplace, Pierre Simon, *A Philosophical Essay on Probabilities*, translated into English from the original French 6th ed. by Truscott, F.W. and Emory, F.L., Dover Publications (New York, 1951) p.4

22. Lombard, Lawrence Brian. *Events: A Metaphysical Study*. Routledge, 2019.

23. Martin, C.B. 2008. *The Mind in Nature*. Oxford: Oxford University Press

24. Norton, John D. November 2003. "Causation as Folk Science," Volume 3, No.4 of *Philosopher's Imprint. www.philosophersimprint.org/003004/*

25. O'Connor, Timothy and Franklin, Christopher, "Free Will", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/sum2019/entries/freewill/>.

26. Penrose, Roger. "The Emperor's New Mind." *RSA Journal*, vol. 139, July 1991, pp. 506–514. *Jstor*, doi: Royal Society for the Encouragement of Arts, Manufactures and Commerce.

27. Robert Kane (2016) *On the role of indeterminism in libertarian free will*, Philosophical Explorations, 19:1, 2-16, DOI: 10.1080/13869795.2016.1085594

28. Russell, Bertrand. 1917. "On the Notion of Cause," Ch. IX in *Mysticism and Logic and Other Essays.* London: Unwin, 1917, 1963

29. Schrödinger, Erwin. *Nature and the Greeks, and, Science and Humanism*. Cambridge University Press, 2014.

30. Strawson, Galen. "Naturalism.org." *Luck Swallows Everything | Naturalism.org*, 26 June 1998, https://naturalism.org/philosophy/free-will/luck-swallows-everything .

31. Strawson, Galen. Free will, 1998, doi:10.4324/9780415249126-V014-1. Routledge Encyclopedia of Philosophy, Taylor and Francis, https://www.rep.routledge.com/articles/thematic/free-will/v-1.

32. Taylor, Richard. *Action and Purpose.* Englewood Cliffs, N.J: Prentice-Hall, 1966.

33. Wilson, Jessica. "The Causal Argument Against Component Forces." *Dialectica*, vol. 63, no. 4, 2009, pp. 525–554., doi:10.1111/j.1746-8361.2009.01216.x.

Biography

Thomas Hebner was born on August 5, 1998 in Princeton, New Jersey. He enrolled in the Plan II Honors program at the University of Texas at Austin in 2016, adding additional Mathematics and Philosophy degrees the following year. He is the oldest of six kids, one of which is concurrently enrolled at the University of Texas on the Track and Field team. While in college, Thomas was a member of the club lacrosse team as well as various intramural teams, none of which resulted in any success or championships. He plans to graduate from UT in December of 2020, attending law school in the fall of 2021.