1 The Free Will Debate

When we order from a restaurant menu or plan the weekend events, we typically assume that we could have chosen differently. When we are angry with wrongdoers for their immoral actions, we often suppose that they could have refrained from acting as they did. More generally, it's commonplace to hold that human beings have free will. However, there are reasons for believing that none of us has free will, reasons that stem from various sources. One traditional source is theological; everything that happens, including human action, is causally determined by the divine will.[1] Another is the nontheological, naturalistic view that the past states of the universe together with the laws of nature causally determine a unique future. Given the prospect that some version of causal determinism is true and that consequently all of our actions are causally determined by factors beyond our control and beyond our causal reach, is it reasonable to believe that we have free will?

One concern to keep in mind in answering this question is that the term "free will" as it is used in various debates has a number of distinct senses, and the answer may depend on which sense is intended. A first sense is illustrated by the examples just cited; to have free will is to have alternative possibilities for choice and action. To streamline this discussion, we might think of choices or decisions as a type of action, a mental type, by contrast with actions such as raising one's hand, which include mental and bodily aspects. With that in mind, we can define this sense of free will as follows:

> **free will AP** (for "alternative possibilities"): free will is an agent's ability, at a given time, either to act or to refrain: that is, if an agent acts with free will, then she instead could have refrained at that time from acting as she did.

Another sense of "free will" singles out freedom from causal determination:

> **free will ND** (for "not determined"): free will is an agent's ability to act without being determined so to act by causes beyond her control.

The focus of **free will ND** is different from **free will AP**, and as we will see in later sections, it may be possible to set up examples in which an agent is not causally determined by factors beyond her control and yet could not have refrained from acting as she did, and also to devise cases in which an agent could have refrained from acting as she did but nonetheless is causally determined to act by factors beyond her control.

A further sense of "free will" focuses on a feature of human action that distinguishes us from certain nonhuman animals. We have free will because we

[1] For a discussion of theological determinism, see Vicens and Kittle (2019).

can act on the basis of reasons, using our rational or intellectual capacities, while certain animals do not have free will because they lack this ability:

> **free will RR** (for "reasons-responsive"): free will is an agent's ability to act rationally, that is, to act in a way that is responsive to reasons.

The characterization of free will that René Descartes provides in his *Meditations on First Philosophy* arguably features all three of these senses, even though he advertises it as simple:

> . . . the will (*voluntas*), or free choice (*arbitrii libertas*) . . . simply consists in this: that we are able to do or not do (that is, to affirm or deny, to pursue or avoid); or better, simply in this: that we are carried in such a way toward what the intellect proposes for affirmation or denial or for pursuit or avoidance, that we feel ourselves determined to it by no external force. (Descartes 1639–40/1996, 4th Meditation, AT 7, 57; Ragland 2016, 8)

A further sense of "free will" is not represented in this passage from Descartes but dominates in the current free will debate, for reasons we shall explore. Intuitively, we must act with free will in order to be praiseworthy or blameworthy for an action, that is, to be morally responsible for it. A fourth definition accordingly links free will to moral responsibility:

> **free will MR** (for "moral responsibility"): free will is an agent's ability to exercise the control in acting required to be morally responsible for an action.

We will now examine the debate about free will to ascertain which of these senses should be our focus in this discussion.

In the history of the free will debate, causal determinism of some sort has been taken to be the main threat to our having free will:

> **causal determinism**: every event has causal antecedents that render it inevitable.

The parties to the free will debate are traditionally grouped into camps with reference to whether causal determinism and free will are compatible:

> **compatibilism**: our having free will is compatible with causal determinism, with all of our actions being causally determined by factors beyond our control.
> **incompatibilism**: our having free will is not compatible with causal determinism, with all of our actions being causally determined by factors beyond our control.

Incompatibilists, in turn, divide into those who hold that determinism is false and that we have free will – the **libertarians** – and those who hold that determinism is true and that we lack free will – the **hard determinists**. Libertarians endorse:

causal indeterminism: not every event has causal antecedents that render it inevitable.

In the definitions of compatibilism and incompatibilism, we can insert any of the definitions of free will we've canvassed. Here is one principle for settling on a definition in such contexts: select the definition that allows us to distinguish contrasting positions that divide parties in the philosophical debate. If we follow this advice, we can set aside **free will RR** immediately, since virtually everyone agrees that in general human beings can be responsive to reasons for action. This is not to say that **free will RR** is an unimportant sense of free will, but only that our having free will in this sense isn't controversial. It is controversial, as we shall see, whether **free will RR** is the crucial necessary condition for moral responsibility, as many contemporary compatibilists contend.

Whether we have **free will ND**, the ability to act without being causally determined by factors beyond our control, is an issue of great interest, since, among other things, it raises the question of whether the universe is fundamentally deterministic or indeterministic. With naturalistic determinism as the point of reference, the universe is indeterministic just in case there is more than one possible future given the laws of nature and some state of the universe at a time. Regardless of how this crucial issue resolves, **free will ND** cannot serve as the key sense in the debate between compatibilists and incompatibilists, since compatibilists distinctively hold that free will is compatible with causal determinism, and they typically also maintain that free will is compatible with indeterminism, at least if the indeterminism doesn't interfere with control in action. For notions that are at issue among libertarians, compatibilists, and hard determinists, this leaves **free will AP**, the ability to act and to refrain, and **free will MR**, the control in acting required for moral responsibility. Discussions of each of these notions are complex in illuminating and instructive respects. Incompatibilists about **free will AP** hold that if causal determinism were true, we would lack the ability to act and to refrain. Compatibilists about **free will AP** contend that even if an action is causally determined by factors beyond the agent's control, she may still, at the time of the action, have been able to refrain from performing it. As we'll see, the history of the debate features several compatibilist strategies for securing this outcome.

Regarding **free will MR**, virtually everyone holds that causal determinism is compatible with our having the control in action required for moral responsibility in *some* sense, for example, in a sense focused solely on forward-looking objectives such as the moral reform of a wrongdoer and reconciliation in relationships (Pereboom 2014, 2017a, 2021; to be discussed). But participants in the debate disagree about whether causal determinism is compatible with our

having the control in action required for moral responsibility in a sense involving a resolutely backward-looking notion of desert, specifically *basic desert*:

> For an agent to be *morally responsible for an action in the basic desert sense* is for the action to be attributable to her in such a way that if she was sensitive to its being morally wrong, she would deserve to be blamed or punished in a way she would experience as painful or harmful, and if she was sensitive to its being morally exemplary, she would deserve to be praised or rewarded in a way that she would experience as pleasurable or beneficial. The desert at issue is basic in the sense that the agent, to be morally responsible, would deserve such blame or punishment, praise or reward, just by virtue of having performed the action with sensitivity to its moral status, and not, for example, by virtue of consequentialist or contractualist considerations.
> (Pereboom 2001, xx, 2014, 2, 2021, 11–12; cf., Feinberg 1970; McKenna 2019)

We can add that the imposition of basically deserved pain or harm, pleasure or benefit, is conceived as noninstrumentally good since such imposition is not envisioned as good only insofar as it brings about a further good. Rather, it is conceived as good in itself (McKenna 2019).

Basic desert contrasts with nonbasic desert, which invokes further goods, such as good consequences, to justify desert claims. John Rawls (1955) presents a two-tiered theory in which lawyers, judges, and juries appeal only to backward-looking reasons for punishment, while the practice itself is justified on forward-looking, consequentialist grounds. In a similar vein, Daniel Dennett (1984, 2003; Dennett and Caruso 2021) and Manuel Vargas (2013, 2015) endorse positions on which justifications for blame and punishment in our practice of holding people morally responsible appeal to what people deserve, while that desert is not basic since at a higher level that practice is justified by anticipated good consequences, such as moral reform and advancement. On the accounts developed by Dennett and Vargas, our practice of holding people morally responsible in a desert sense should be retained because doing so yields the best overall consequences relative to alternative practices (cf., Doris 2015). Others, such as James Lenman (2006) and Ben Vilhauer (2009b), ground nonbasic desert in social contractualist considerations.

For many, the intuition that wrongdoers deserve to be punished concerns basic desert specifically, in contrast with its nonbasic relative. That this is so might be shown by a type of thought experiment derived from Immanuel Kant (1797/2017), in which there is no instrumental good to which punishing a wrongdoer would contribute. Imagine that a person on an isolated island viciously murders everyone else on the island and that he is not capable of moral reform due to ingrained hatred and rage. Thus, there are no good consequences that the punishment might aim to realize, and there is no longer

a society on the island whose rules might be determined by contract. Many nonetheless have the intuition that this murderer deserves to be punished severely. The desert would be basic since the specifics of the example eliminate nonbasic desert.

The resulting version of **free will MR** does distinguish the parties in the debate. Given this selection, here are the characterizations of the three traditional positions:

> **hard determinism**: because causal determinism is true, we cannot have the sort of free will required for moral responsibility in the basic desert sense.
> **compatibilism**: even if causal determinism is true, we can have the sort of free will required for moral responsibility in the basic desert sense, and we do in fact have it.[2]
> **libertarianism**: because causal determinism is false, we can have the sort of free will required for moral responsibility in the basic desert sense, and we do in fact have it.

Features of our moral responsibility practice that don't invoke basic desert have not been a focus of contention in the free will debate. For example, someone might be held morally responsible because his tendencies to act wrongly are apt to be modified or eliminated partly by blaming, and his dispositions to act rightly might be strengthened by praising (Schlick 1939; Nowell-Smith 1948; Smart 1961). For a variation on this idea, someone might be held morally responsible by asking her questions such as: "Why did you decide to do that? Do you think it was the right thing to do?" She may then come to appreciate the moral reasons for changing her dispositions, attitudes, and behavior, resulting in moral reform and reconciliation in relationships. Such an exchange can be viewed as engaging a forward-looking *answerability* sense of moral responsibility (Scanlon 1998, 2009; Shoemaker 2011, 2015; Smith 2012), by contrast with a backward-looking *accountability* sense, conceived as licensing angry responses and the imposition of harmful sanctions (Watson 1996; Shoemaker 2011, 2015). Incompatibilists would not regard the control required for answerability as conflicting with causal determinism, and it is open to hard determinists to endorse this notion of moral responsibility. The accountability sense, conceived specifically as invoking basically deserved pain or harm, does divide the parties in the debate.

It's important to note that contemporary skeptics about free will are not typically hard determinists, exactly, since, for reasons we'll explore in the

[2] There is a broader sense of compatibilism on which one might be a compatibilist and deny that we have free will of the sort specified (Strawson 1986, 6). For example, one might believe that this sort of free will is compatible with determinism, but since our actions are never produced by conscious willing, we lack such free will.

next section, they don't think that we're in a position to ascertain that the universe is causally deterministic. Instead, most free will skeptics maintain that whether or not the universe is causally deterministic, our having the controversial sort of **free will MR** is either highly unlikely (Pereboom 2001, 2014; Levy 2011; Caruso 2021) or impossible (Strawson 1986, 1994). By contrast with the hard determinists, these are the free will skeptics we'll consider, specifically in Section 7.

2 Libertarianism

Libertarians contend that determinism is false and that at least some of our actions are not causally determined by factors beyond our control. They accordingly maintain that we have **free will ND**. Furthermore, they hold that our having **free will ND** makes room for our having **free will AP**, for the ability to act and to refrain. For many libertarians, our having **free will ND** and **free will AP** are required for the contested sort of **free will MR**, the control in action required for moral responsibility in the basic desert sense. Most libertarians would agree that **free will RR**, reasons-responsiveness, is also required and thus that the contested **free will MR** entails the other three senses of free will. In some contexts, it's assumed that free will just is the conjunction of these four senses. For example, if you're a member of the Free Will Baptist Church,[3] this is the position you're apt to defend. As we shall see, many contemporary compatibilists maintain that of these senses, only **free will RR** is entailed by the controversial **free will MR**, and thus, their requirements for this crucial notion are less stringent.

2.1 Libertarianism and Indeterminism

One route to **free will ND** involves claiming that we are at least partly nonphysical, as mind–body dualists maintain, and that this allows for the requisite indeterminacy. We'll soon discuss an objection to such dualistic libertarianism. Can undetermined freely willed decisions be reconciled with our bodies being governed by deterministic natural laws? How might a free decision to raise one's hand harmonize with the rising of the hand being governed by deterministic laws?

Another possibility is that the laws of nature are indeterministic. Before the twentieth century, physicists typically conceived of the fundamental laws of nature as deterministic. In the early nineteenth century, mathematician and

[3] https://nafwb.org/site/

physicist Pierre-Simon de Laplace provided a famous characterization of the universe that accords with this conception:

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes. (Laplace 1814/1951, 4)

However, beginning in the 1920s, quantum mechanics was developed, which has revived the prospect of fundamentally indeterministic physics. On one account, quantum mechanics is an instrumentalist and not a realist theory; while it is a remarkably successful predictive tool, its role does not include informing us about the nature of microphysical reality (e.g., Healey 2012, 2017). Realist accounts, which, by contrast, affirm that quantum mechanics does indeed inform us about the nature of microphysical reality, are another option. For our purposes, the pertinent issue in quantum mechanics is how to get from premeasurement states in which particles have indeterminate superposition – that is, multiple distinct states at the same time, for example, a superposition corresponding to a particle being spin-up and deflected upward and spin-down and deflected downward – to well-defined outcomes subject to measurement. The wave function equation that describes this process is probabilistic, and the question at hand is whether there is a realist metaphysical interpretation of this equation that is fundamentally probabilistic and thus metaphysically indeterministic.

Albert Einstein famously said of this metaphysics, "God does not play dice," and that thought inspired the realist and determinist interpretation of David Bohm (1952), according to which there are hidden factors or variables that render the universe deterministic despite the probabilistic wave function equation. In *Bohmian mechanics*, particles are fundamental in the sense that they do not reduce to waves, and the wave function together with initial particle positions fixes the specific particle position at all times. However, on another realist interpretation, developed by Ghirardi, Rimini, and Weber (1986), the universe is fundamentally indeterministic. The *Ghirardi–Rimini–Weber (GRW) spontaneous-collapse theory* adds a probabilistic equation to the standard quantum dynamics developed by Erwin Schrödinger, with the result that every particle, which on this view is reducible to waves, has a small probability per unit time of undergoing a "hit," in which its state jumps to a state that is relatively localized, that is, a "collapse." The prior state of the system

determines a probability distribution for the location of the particle, but knowledge of the prior state cannot facilitate the prediction of specific locations with certainty because those locations are not causally determined in virtue of the prior states. Alternatively, according to the *many-worlds theory* developed by Hugh Everett (1957), there are no hidden variables, and there is no supplemental equation that probabilistically predicts locations. Consider a state involving an electron in superposition corresponding to its being spin-up and deflected upward and spin-down and deflected downward. Everett's view is that this is both a state in which the electron is deflected upward and one in which it is deflected downward. But these states are causally isolated from each other, and thus, they are distinct and might be thought of as occurring in different worlds. This view is deterministic, but everything that happens appears indeterministic from the point of view of an observer restricted to a specific world. The relevant upshot of this discussion is that the metaphysics of quantum mechanics isn't settled, and it is epistemically open that the universe is indeterministic, but also that it's deterministic.[4]

Even if the indeterministic GRW spontaneous-collapse theory turns out to be true, this wouldn't all by itself settle whether we have **free will ND** of a significant sort. Free decisions would require the indeterminism to be suitably located, plausibly at the level of the neural constitution of decisions, while quantum indeterminacy would, in the first instance, be located at a microlevel more fundamental than the neural. It's a serious possibility that quantum micro-indeterminacies, on the supposition that they exist, are ordered with enough redundancy so that at the neural level, indeterminacy all but vanishes. For this sort of reason, the engineering of roads and bridges doesn't need to countenance fundamental quantum indeterminacy. The result would be neural-level determinism or near-determinism. If all the actions we perform turned out to be 99.9 percent probable on their causal antecedents, we would technically have **free will ND** but not to a degree that would make a practical difference relative to the contrasting deterministic picture.

For alternative possibilities to be significantly probable, there would have to be mechanisms that facilitate the "percolating up" of significant microlevel indeterminacies to the neural level, on the analogy of a Geiger counter that senses microlevel events and registers them at the level of the moving of a macrolevel indicator. This issue was recognized and addressed by physicist Roger Penrose (1989, 1994) and anesthesiologist Stuart Hameroff (1998), who suggested that free will and consciousness arise through the enhancement of

[4] This discussion is indebted to Peter Lewis's (2016) *Quantum Ontology*. See this book for a clear and thorough account of the metaphysics of quantum mechanics.

quantum effects within subcellular structures internal to neurons known as microtubules. This hypothesis has been energetically debated but remains speculative (for a recent overview, see Atmanspacher 2020).

2.2 Event-Causal Libertarianism

The science and metaphysics of libertarianism face certain challenges, and this has motivated three different varieties of this view. First, according to *event-causal libertarianism*, defended prominently by Robert Kane (1996), Laura Ekstrom (2000, 2019), and Mark Balaguer (2010), actions are caused solely by events, conceived as things or substances having properties at times, such as *Annie's wanting at noon today to give Ben his medicine*, and indeterminacy in the production of actions by agent-involving events is a core requirement for moral responsibility.[5] This view contrasts with a position according to which substances, such as cars, asteroids, and people, can be causes and not just by virtue of having a role in events. Those who claim that only events can be causes recognize that we may sometimes speak as if substances are causes, but argue that once we clarify such speech, we will see that the event view is right. For example, suppose a car drives through some rainwater on the street and splashes you. You may say: "The car made me wet!" But speaking precisely, it's not the car, exactly, that made you wet but an event, the car's driving through the rainwater on the street at 725 5th Avenue at 2 PM that caused this effect. By contrast, in *agent-causal libertarianism*, defended in recent decades by Randolph Clarke (1993, 1996, cf., 2003), Timothy O'Connor (1995, 2000, 2008), and Meghan Griffith (2010), freely willed actions are, by contrast, accounted for by agents who, as substances, cause them without being causally determined to do so. According to the agent-causal position, it is essential that the causation involved in freely willed action is not causation among events involving the agent but is rather an instance of the agent as a substance causing an effect.[6] On *noncausal libertarianism*, basic free actions, such as free choices, aren't caused at all. Instead, the agent is their *subject* and not their cause.

One of the most influential objections to event-causal libertarianism is that if actions are undetermined in the way that this position requires for freely willed action, agents cannot have sufficient control in acting to secure moral responsibility. The ancient Epicureans maintained that the universe ultimately consists

[5] In the history of philosophy, event-causal libertarianism can arguably be found in Lucretius (50 BCE/1998) and in Niccolò Machiavelli (DeCaro 2021).

[6] Agent-causal libertarianism can arguably be found in Immanuel Kant (1781/1787/1987) and Thomas Reid (1788/1983) and, besides those just mentioned, is explicitly proposed by Richard Chisholm (1964, 1976), Richard Taylor (1966, 1974), Meghan Griffith (2010), and Christopher Franklin (2018).

of atoms and the void and that if universal causal determinism were true, the atoms would all be falling downward (Lucretius 50 BCE/1998). But to explain the interaction of atoms and also free will, they posited random swerves in the otherwise downward paths of atoms. A traditional question for the Epicurean view is: Do agents plausibly control how and when an atom swerves?

This concern is often framed as *the luck objection* (e.g., Haji 2000; Latham 2004; Mele 2006; cf., van Inwagen 1983).[7] A famous instance of this objection is found in David Hume's *Treatise of Human Nature* (1739/1978, 411–12, cf., 1748/2007, §8), where he argues that if an action is not determined by factors involving the agent, it will not have sufficient connection with the agent for her to be morally responsible for it. This concern might be developed in different ways, but here is my own favorite. For an agent to be morally responsible for an action in the basic desert sense, she must have a certain kind of control in producing that action. But in an event-causal libertarian picture, preceding agent-involving events don't causally determine the action. Suppose that a decision is made in a deliberative context in which the agent's moral motivations favor deciding to A, her self-interested motivations favor her deciding to not-A, and the strengths of these motivations are roughly in equipoise. A and not-A are the only options she is considering. The potentially causally relevant events accordingly render the occurrence of each of these decisions roughly equiprobable. But then the potentially causally relevant events do not settle which decision occurs, that is, whether it is the decision to A or the decision to not-A. And furthermore, because in event-causal libertarianism only events are causally relevant, *nothing*, and in particular, *nothing that's agent-involving*, settles which decision occurs. Given the complete causal role of these preceding events, which decision it is *remains open*. Thus, it can't be the agent or anything about the agent that settles which decision occurs, and the agent therefore lacks the control required for moral responsibility for the one that does. Since the agent "disappears" at the crucial point in the production of the action – when whether it occurs is to be settled – we can call this the *disappearing agent argument* (Pereboom 2004, 2014, 2017b; cf., O'Connor 2008).[8]

Another objection to event-causal libertarianism in the "luck" family, the "rollback objection," originates with Peter van Inwagen (1983). Here is the version formulated by Balaguer (2010, 92ff.). Suppose Ralph is deciding between two options, moving to New York and staying in Mayberry, and his reasons and motivations for each are equally strong. Balaguer calls this a *torn*

[7] For responses to the luck objection, see Robert Kane (2007), Mark Balaguer (2014), and Robert Hartman (2016).

[8] Objections to this argument are set out by Al Mele (2017) and Randolph Clarke (2019); for replies to these objections, see Pereboom (2017b).