EuJAP | Vol. 15 | No. 2 | 2019
UDK: 123.1:165.023.2
https://doi.org/10.31820/ejap.15.2.5

THE CONCEPTUAL IMPOSSIBILITY OF FREE WILL ERROR THEORY

Andrew J. Latham
The University of Sydney

Review article – Received: 30/04/2019 Accepted: 13/11/2019

ABSTRACT

This paper argues for a view of free will that I will call the conceptual impossibility of the truth of free will error theory - the conceptual impossibility thesis. I will argue that given the concept of free will we in fact deploy, it is impossible for our free will judgements—judgements regarding whether some action is free or not—to be systematically false. Since we do judge many of our actions to be free, it follows from the conceptual impossibility thesis that many of our actions are in fact free. Hence it follows that free will error theory—the view that no judgement of the form 'action A was performed freely'—is false. I will show taking seriously the conceptual impossibility thesis helps makes good sense of some seemingly inconsistent results in recent experimental philosophy work on determinism and our concept of free will. Further, I will present some reasons why we should expect to find similar results for every other factor we might have thought was important for free will.

Keywords*: Free will, error theory, conceptual impossibility, conditional concept, experimental philosophy*

1. Introduction

Strictly speaking, transcendental arguments are arguments that attempt to show that *X* is a necessary precondition for the possibility of *Y* and hence since actually *Y*, therefore actually *X*. Immanuel Kant (1781/1787) is, of

course, the most famous defender of arguments of this kind. We can find examples of this kind of argument throughout many different domains of philosophy. One recent example involves an objection to certain approaches to quantum gravity in the philosophy of time. These approaches are said to be timeless, since they deny there exists any ordered series of events that are temporally or causally connected to one another. However, a necessary precondition to even entertain these theories is having contentful mental states. But having contentful mental states requires causal connections between at least some of our mental states and states in the world those states are about. So, we are only able to entertain these theories if in fact they are false (Braddon-Mitchell and Miller 2018). Of course, one response to a transcendental argument is to just deny what the proponent takes to be undeniable. For instance, in philosophy of mind, a proponent of eliminative materialism, of the kind defended by the Churchlands (1981; 1986), can just deny that you need to have beliefs (rather than other neuroscientific states) in order to argue that there are no beliefs.[1]

A transcendental argument for free will would proceed by showing that the necessary precondition for the possibility of some way things actually are—for instance, our being agents, or deliberators, or the kinds of things that can ask questions about free will—is there being free will. It then follows that since we are such things, there is free will. Robert Lockie (2018) does just this in his new book *Free Will and Epistemology*: If our having libertarian free will (free will incompatible with determinism) is a necessary precondition for the possibility of our having any justified beliefs, then if we believe that we do not have free will, either this belief must be unjustified, if it's true, or if justified, it must be false. In this paper, I will run a different line of argument to the conclusion that we have free will. Roughly, for now, the idea will be that most of our actions being free is a necessary precondition for understanding our ordinary practices as being non-defective, and as they are not defective, we have free will.

In this paper, I will argue that our concept of free will cannot do the job it is supposed to do, and that concept fail to be satisfied. That's because most of our actions being free is a necessary precondition for understanding our ordinary free will practices as being non-defective. These practices involve drawing certain kinds of distinctions between different kinds of actions that we track with our talk of free and unfree. We distinguish actions performed while being coerced, from those performed while fulfilling our desires, and actions performed in the grips of a mental illness, from those performed after some long effortful deliberation. It's important to note that what I am

[1] Thanks to Kristie Miller for bringing these cases to my attention.

intending to pick out in discussing our free will practices is much wider than our moral responsibility practices. Consider for a moment certain kinds of advertising which push us towards choosing one option over another. While these advertisements might impact our behavior in a predictable manner, they do so in a way which is not *mentally mediated*. That is, the advertising seems to impact behavior via sub-personal level processes which are not consciously available to the deliberator. What is important is that the reason we don't like these kinds of advertising pushes is *not* because we think they undermine our moral responsibility, but because they seem to impact our free will in a manner we don't like. For the purposes of this paper I am going to assume we could not engage in these practices without making these kinds of distinctions, and further, that these practices cannot and should not be revised. The argument for this claim about our practices is a job for another paper. Given that these practices are not defective, then, I argue, we have free will. It is, as it were, *conceptually impossible* for us to deploy the concept of free will that we do, and the world fail to satisfy that concept.

An analogy: one might argue that our concept of ordinary objects such as trees, rocks, and so on, are such that even if it turned out that we are living in a computer simulation, or some demon's brain, it will still turn out that there are trees and rocks. We might discover that their underlying nature is surprising, but not that they don't exist (Chalmers 2005). If our concept of tree was something like: whatever thing it is with which I am causally connected, when I have mental states of *this* kind, then, it would simply turn out that if our world is a computer simulation, trees *are* parts of such simulations. What trees are fundamentally made of turns out to be different than we originally supposed, but that doesn't mean there are no trees.

I will argue that it cannot be that we deploy the concept of free will that we do, and it turn out that actually we are systematically mistaken about which actions are free, and which actions are unfree. Of course, the idea that there could be such concepts might seem puzzling, so in §2 I will outline and defend the conceptual impossibility thesis. Then in §3 I will show how taking seriously the conceptual impossibility thesis reconciles some apparent inconsistencies in the extant empirical evidence regarding our concept of free will, and determinism. In §4 I will give reasons why we should think that the finding that determinism doesn't matter for our having free will, should generalize to other factors people have thought were important for free will. Finally, in §5 I will conclude.

2. The Conceptual Impossibility Thesis

Before I outline the conceptual impossibility thesis in more detail, some clarifications are in order. The conceptual impossibility thesis is the thesis that given the content of the concept of free will that we, the folk, in fact deploy, it cannot be that the concept is *systematically* misapplied. That is, it cannot be that we are systematically mistaken about which actions are free and which are unfree. Two things are noteworthy here. First, the concept with which I am interested is the *folk* concept of free will. There might be *philosophical* re-conceptions of free will which have quite different content from the folk concept, and I will make no attempt to consider such concepts here. Second, the conceptual impossibility thesis is a thesis about *systematic* error. It is not the thesis that none of our judgements about which actions are free (or not) are false. It is consistent with the conceptual impossibility thesis that some of our judgements about which actions are free (or not) are mistaken.

Why would one accept the conceptual impossibility thesis? Let's call a judgement of the form 'action A is free' a *positive* judgement, and a judgement of the form 'action A is unfree' a *negative* judgement. I will argue that the content of our folk concept is *something* like the following: free will is whatever thing there is in the world which most of our positive judgements track. In this regard, I argue that our concept of free will has a content, which is such that however our world turns out to be, most of our free will judgements (both positive and negative) will be vindicated. The only way this could fail to be is if there were *nothing at all* in common between most of the times we judge that an action is free, and most of the times we judge that an action is unfree, such that we are not tracking anything at all, for there is nothing there to be tracked. But this is clearly not the case: there *are* such similarities. Even free will error theorists don't think that there are no such similarities; they simply think that those similarities are not, in fact, sufficient to vindicate our positive free will judgements.

But why think that the content of our concept is as I suggest?

Consider the kinds of cases that we ordinarily judge positively to be free, and judge negatively to be unfree. Ordinarily, we make positive judgments regarding cases where we are act in accordance with our reasons, in fulfilling our desires, after having mentally simulated numerous courses of actions and their projected outcomes, and so on. Conversely, we make negative judgments regarding cases where we are bound-up, or being coerced and manipulated, or caught in the grips of a psychological or physiological illness, and so on. Of course, neither list is exhaustive of all

the kinds cases that we judge positively and negatively. For the moment, I simply want to roughly flag the kinds of cases that we ordinarily think of as free and unfree.

One way of characterizing the problem of free will is as the worry that there is no metaphysical difference between the cases where we judge actions to be free, and those we judge to be unfree. That would seem to be the case were we to discover that some fact that characterizes those actions we currently class as unfree, turns out to be true of *all* our actions (Dennett 1984; 2013). For instance, if it turned out that all our actions are coerced, or manipulated, or in the grips of psychological or physiological illness, then *prima facie* this would seem to be the discovery that none of our actions are free.

Let us focus, for the moment, on one important metaphysical factor relevant for free will: determinism. Philosophers have traditionally thought that consideration of determinism is important for free will, and so it has received the most empirical attention in experimental philosophy. In §3 I will turn to the empirical data on the relationship between the folk concept of free will and determinism. In §4 I will give some good reasons to think that the lessons of the conceptual impossibility thesis generalize to all other relevant metaphysical facts as well.

For now, suppose we only judge actions to be free if they are not determined. Then indeterminism is necessary for our concept of free will to be satisfied, (as is commonly supposed),[2] and if we discover that determinism is true, then we discover that there is no free will. Notice, though, that if there's no free will, then *all* our actions are akin to being bound-up, or coerced and manipulated, or caught in the grips of a psychological or physiological illness. That, however, seems wrong. Even if there is no deep *metaphysical* difference between the cases, we judge to be free, and those we judge to be unfree, we still want our actions to be like the ones that we ordinarily think of as free. After all, even if, with respect to some particular metaphysical matter of fact, there is no difference between these actions, there still seem to be other relevant differences that we want to track with our talk of free and unfree action. We want to normatively evaluate actions—whether this be moral or prudential evaluation—and to do that we want to distinguish actions that are performed while being coerced and manipulated, from those that are not,

[2] See e.g. Ekstrom (2002), Kane (2005), O'Connor (2000), Pereboom (2001), Pink (2004), Strawson (1986), van Inwagen (1993) to name a few. Contra this some theorists such as Eddy Nahmias (2011) think the folk concept of concept is a compatibilist one and that incompatibilist judgments arise out of people misunderstanding the implications of determinism for free will.

and actions performed while in the grips of a psychological or physiological illness, from those that are not, and so on. Regardless of whether determinism is true, we can be expected to care whether our friend stood on our foot because, having deliberated about it, she decided this is what she wanted to do, and proceeded to do it, or because she was pushed over by the person next to her. *Mutatis mutandis* for all these kinds of cases.

So there seems to be a concept of free will that tracks superficial differences between the cases we judge to be free, and the cases we judge to be unfree. For ease of explication I will call this a *social kind concept*.

One might, however, object. Consider for the moment a potentially analogous case involving water and ice. According to the story I have provided so far there are *two* different social concepts. The *social* concept of water, which is sensitive to the stuff that fills the oceans, flows through the rivers, falls from the sky whenever it rains, and so on, and the *social* concept of ice which is sensitive to the stuff found in glaciers, around the poles of the Earth (for now), falls from the sky as hail, and so on. Yet while perhaps once we thought that water and ice were different kinds of things, as a result of scientific investigations we have discovered that there is no deep metaphysical difference between water and ice: they are both H_2O. So, we now believe there is only one *natural kind concept*, which both water and ice fall under.

Surely, we should expect the same thing to occur in the case of free will: discovering some deep metaphysical fact that characterizes actions we currently judge to be unfree, to be shared with actions we judge to be free, gives us warrant to conclude that both sets of actions are of the same *metaphysical* kind, and that both are unfree. For example, if determinism is in fact true, and we judge that such a metaphysical fact makes actions unfree, then we should judge that none of our actions are free.[3]

Thus, there seems to be a concept of free will that tracks some deep metaphysical feature of our actions. I will refer to the concept of free will that is relevantly similar to a natural kind concept, a *metaphysical kind* concept.

The idea that free will might be a natural kind has been expressed in the free will literature before (Heller 1996; Deery 2019). Such a view is a natural extension of the paradigm-case view advanced by Antony Flew

[3] Thanks to David Braddon-Mitchell for the Ice and Water case.

(1955), who suggested that the meaning of 'free will' is fixed by the paradigm cases.[4] However, if it's a conceptual constraint that something falls under the concept of free will only if that thing forms a natural kind, then the meaning of 'free will' is fixed by whatever natural kind is uniformly in common between all (and only) the paradigm cases.[5]

One consequence of thinking of free will as a natural kind is that it admits a family of views which vary according to what you think is in common between all the paradigm cases. For instance, on the one hand, free will might form a metaphysical kind and so carve nature at its joints. This seems to be the case when we think that free will is whatever allows our actions to be *indeterministic*, whilst not being *merely chancy*. On the other hand, free will might form a psychological, functional or social kind. While these latter kinds do not carve nature at its joints, they nevertheless carve nature up in a useful fashion. Perhaps free will is a psychological capacity or suite of psychological capacities, or perhaps free will is just the practices themselves of judging certain actions to be free and unfree. Finally, and most permissively, free will might just be whatever is a member of the set of paradigm-cases. On this view free will could be anything at all.

It is my view that we should treat this family of natural kind views as a kind of prioritized hierarchy.[6] By that I mean that if the metaphysical kind is there and in common between the paradigm cases, then that's what free will is and necessarily so. Else, if the psychological kind is there and in common between the paradigm cases, then that's what free will is, and necessarily so, and so on. Then perhaps, finally, if there is no natural kind in common between the paradigm cases, then free will just is the paradigm cases. While I think that there is something in common between the paradigm cases I am not taking a stand in this paper on exactly what that is. Further, I am not advocating that it is possible for anything at all to count as free will which would seem to be the case if there is nothing at all in common between the paradigm cases, aside from being a member of the set of paradigm cases. While I think that it's open for someone to think that, it is not my view.

For the ease of ongoing discussion I will restrict myself to just the social and metaphysical kinds. Given these two apparent concepts of free will,

[4] Thanks to an anonymous referee for making me aware of this existing and growing literature.

[5] Though for arguments against the paradigm-case view and free will as a natural kind, see van Inwagen (1983) and Daw and Alter (2001).

[6] I will have much more to say about this kind of prioritized hierarchy when I come to discuss the idea of the folk concept of free will being a conditional concept with respect to determinism in §3.

there are two conceptual impossibility theses: one *weak* and one *strong*. The *weak conceptual impossibility thesis* is that the social concept of free will and the metaphysical concept of free will are both important. If some underlying metaphysical feature is missing (i.e. determinism is true) then on the metaphysical concept of free will, error theory will be true. However, according to the weak conceptual impossibility thesis the social concept of free will is also important, and on the social concept there will be free will regardless. The conceptual impossibility thesis is true of the social concept. The *strong conceptual impossibility thesis* is that while both the social concept and metaphysical concept exist, it's only the social concept that matters, so the conceptual impossibility thesis is true of the concept that matters. Let me elaborate on both these theses.

2.1. The Weak Conceptual Impossibility Thesis

There are two apparent concepts of free will: a metaphysical concept which is open to the possibility that there is no free will (analogous to the discovery that since ice is just H_2O, in some deep sense there is no ice) and a social concept according to which as long as there *are* differences between paradigm cases we judge to be free and paradigm cases we judge to be unfree, this guarantees there is free will. On the weak conceptual impossibility thesis, both concepts are needed, and the social concept is guaranteed to be satisfied.

But what do I mean when I say both concepts are needed? Well the fact that water and ice are both H_2O plays an important explanatory role in our best scientific theories; such as why ice and water exhibit the same chemical properties. So, there is an important sense in which there is not both water and ice, there is just H_2O. Perhaps philosophers, too, will conclude that there's no metaphysical difference between those cases that we ordinarily judge to be free, and those we judge to be unfree. However, aside from generating an apparent problem for free will, I am not sure what purpose we have for taxonomising our actions according to their deep metaphysical nature. For instance, what is gained by classifying our ordinary actions by the lights of determinism? I will return to this point shortly when I describe the strong impossibility thesis. I leave it open, here, that there could be good reasons for classifying our actions according to their metaphysical nature (i.e. determinism), and thus to in some sense collapse the distinction between free and unfree actions on the metaphysical concept.

Even if we do so, however, there is clearly some relevant distinction between the actions we judge to be free, and those we judge to be unfree. To see this, return to the case of water and ice. Suppose we agree that there

is no metaphysical difference between water and ice, and hence that in some good sense we can collapse the distinction between them. Nevertheless, there is a clear sense in which despite this, there *is* both water and ice despite there being no metaphysical difference between them. That's because we care about the role the superficial differences between water and ice plays in ordinary matters. If I am thirsty and ask for a glass of water at a restaurant, I would be amused to receive a glass filled with ice.

Similarly, even if there is no deep metaphysical difference between actions we judge to be free, and to be unfree, we still care deeply about whether actions fall into one, or instead the other, category. We care whether or not we act for our reasons, in order to fulfil our desires, or after some process of deliberation as opposed to being bound-up, coerced and manipulated, or caught in the grips of a psychological or physiological illness. What this social concept of free will tracks then, is *whatever it is* which vindicates this difference.

The weak conceptual impossibility thesis holds that the distinction between free and unfree actions is like the distinction between water and ice. Just as there are two ways of thinking about water and ice, there are two ways of thinking about free and unfree action. On the metaphysical concept, we group the cases according to their metaphysical nature, and so decide that there are no free actions if determinism is true. This is analogous to the sense in which there is not water and ice, there is only H_2O. On the social concept we group the cases according to some, perhaps more superficial, difference between them, a difference that we care about for our ordinary purposes. This is analogous to the sense in which we ordinarily treat water and ice as distinct despite there being no metaphysical difference between them. That's because what we are often just as, if not more, interested in, is the role such distinctions play in ordinary matters, and not their deep metaphysical nature. So, while error theory is true of our metaphysical concept of free will, the conceptual impossibility thesis is true of our social concept of free will.

2.2. The Strong Conceptual Impossibility Thesis

What of the strong conceptual impossibility thesis? According to that thesis, while there are two concepts of free will, only the social concept *matters* for any important purposes. In the water and ice case, the metaphysical concept on which despite superficial differences, both water and ice are H_2O, plays an important role in our best scientific explanations in the chemical sciences. That's why the metaphysical concept matters. But there seems to me to be nothing analogous in the case of free will that

justifies taking seriously the idea that just because something about every free action turns out to be like the unfree actions, that that feature is crucial for freedom. The strong conceptual impossibility thesis says that the free and unfree distinction is not like the distinction between water and ice because while we certainly care about the superficial differences between those actions we judge to be free and unfree, there's nothing analogous to the chemical sciences which justifies taxonomising our ordinary actions according to deep metaphysical similarities. Error theory might be true on the metaphysical concept of free will, but no one ever cared about that concept because it doesn't matter for any of the purposes for which we deploy that concept. So, on the only concept that matters, the social concept, the conceptual impossibility thesis is true.

In the next section I will show how the conceptual impossibility thesis has important consequences for the interpretation of extant empirical work on our folk concept of free will and its relationship to the thesis of determinism. Then later, I will give some reasons to think that all factors that we might have thought mattered for free will (such as determinism) *don't.*

3. Experimental Philosophy, Determinism and the Folk Concept of Free Will

One metaphysical factor that many people have supposed matters for free will is determinism. The thesis of determinism holds that the entirety of particular facts about the past, in conjunction with the laws of nature, entails every truth about the future. Is our concept of free will compatible with determinism being true? Compatibilists answer affirmatively. According to them, if determinism is true then provided agents have some preferred set of abilities, which vary according to the version of compatibilism at issue, then free actions are those produced by those abilities. For ease of explication I will refer to whatever the abilities are that when exercised in the production of an action makes that action free according to compatibilism: *compatibilist powers*. Conversely, incompatibilists take it to be a necessary condition for our having free will that indeterminism is true. *Libertarians* are incompatibilists who think there is free will. Call whatever the abilities are that when exercised in the production of an action makes that action free according to libertarianism: *libertarian powers*.

If the conceptual impossibility thesis is true, then the folk concept of free will must be compatible with determinism. But, while it's often been assumed that the folk concept of free will is an incompatibilist one, there

is excellent evidence from experimental philosophy that the folk concept is a compatibilist concept (e.g., Nahmias et al. 2005; 2006) and also that it is an incompatibilist concept (e.g., Nichols and Knobe 2007).

How should we make sense of this apparent inconsistency? Roskies and Nichols (2008; though see also Björnsson 2014; Latham 2019) noticed a difference in the experimental materials used. While Nahmias and colleagues situated some of their determinism vignettes in the *actual* world, Nichols and colleagues situated them in *hypothetical* worlds. In order to confirm their suspicion that participants' free will judgements to deterministic vignettes differed as a result of where they were being evaluated, participants were evenly split between considering deterministic vignettes in the actual world or in some other hypothetical world. Consistent with the authors' hypotheses, where the deterministic scenario was situated significantly impacted participants' free will judgements. Participants' free will judgements were significantly higher when the deterministic vignette being evaluated was in our own world relative to when the deterministic vignette being evaluated was in some hypothetical world.

3.1. Determinism and a Conditional Concept of Free Will

Roskies and Nichols (following Braddon-Mitchell 2003; though see also Latham 2019) argued that these results suggest that the folk concept of free will takes a conditional form with respect to determinism. So:

> If the *actual world* is indeterministic, and agents have libertarian powers, then these libertarian powers are what free will is and must be.
> Else, if the actual world is deterministic, and agents have their preferred compatibilist powers, then compatibilist powers are what free will is.

To make things even clearer, this conditional analysis of free will can be organized into a simple two-dimensional diagram (see Figure 1).

		Possible World	
		I	**D**
Actual World	**I**	**T**	**F**
	D	**T**	**T**

'Some agents have free will'

Figure 1. Two-dimensional diagram showing the conditional analysis of free will with respect to determinism, given the sentence 'some agents have free will'.

Here is how to read the two-dimensional table: along the top we see two classes of worlds, indeterministic worlds (I) and deterministic worlds (D). Let's suppose for ease of explication that all indeterministic worlds contain agents with libertarian powers and all deterministic worlds contain agents with compatibilist powers (this assumption can easily be removed with a much more complex diagram). These are 'worlds considered as counterfactual' relative to each other. Down the left-hand side, we see the same two classes of worlds, but here they are not thought of as counterfactual alternatives to each other, where one is actual and the other is an alternative. Instead, they are alternatives about how the actual world itself, for all we know *a priori*, might be.

What we are doing when we read this table, is considering our judgments about whether or not some agents have free will, relative to different contexts (ways things might be, for all we know, only one of which is actual), from the perspective of different indices (ways the actual world might turn out to be). Suppose, then, that the actual world turns out to be indeterministic. From the index of an indeterministic actual world, if we look at counterfactual worlds that are also indeterministic then we will judge that it is true that some agents have free will. This is reflected in the T value in the world at the top left cell of our table. That world is being evaluated from the perspective of an actual indeterministic world (specified on the left of the table). The top right cell contains an F. There, we evaluate what to say about the truth-value of 'some agents have free will' at a deterministic world, from the perspective of an indeterministic

actual world. In that case, since we judge that those deterministic worlds do not contain agents with free will, that sentence comes out as false.

On the other hand, suppose that the actual world turns out to be deterministic. Now consider our judgements about 'some agents have free will' at a deterministic counterfactual world (the cell on the bottom right). Since compatibilist powers are sufficient for free will, we will judge that the sentence is true in that counterfactual world. Furthermore, since having either compatibilist or libertarian powers is sufficient for having free will conditional on the actual world being deterministic, it follows that we will judge that in any worlds with those powers, regardless of whether they are deterministic or not, agents have free will. Hence 'some agents have free will' will be true when evaluated in counterfactual indeterministic worlds, conditional on the actual world being deterministic. This is reflected in the bottom left cell of the table.

Let's tie this back to the empirical results. When a vignette is taken to describe the actual world, we should expect that if people deploy a conditional concept, they will judge that agents are free in the deterministic world considered as actual, and will judge that agents are unfree in the counterfactual deterministic world. People are inclined to judge that people in the counterfactual deterministic world are unfree, because people in fact believe that the actual world is indeterministic and so think, unless told otherwise, that indeterminism is a necessary condition for free will.[7] So, far so good; but this evidence is only consistent with the folk having a conditional concept of free will with respect to determinism. The reason these results do not show that people in fact possess a conditional concept of free will is because we do not have data and responses to all the conditions necessary to determine whether or not there is a conditional concept.

Recently, Latham (2019) tested more directly whether or not the folk concept of free will is a conditional one with respect to determinism. They noted that the conditional account makes two key predictions regarding people's free will judgments to various conditions, which they called the *weak* and *strong signal for conditionality*. The weak signal for conditionality is what Roskies and Nichols (2008) identified might be present in their data. Given that people tend to believe the actual world is

[7] As a descriptive matter of fact, the overwhelming majority of ordinary people think that the actual world is indeterministic. For example, Nichols and Knobe (2007) found over 90% of participants chose the vignette describing an indeterministic universe, not a deterministic universe, as being most like the actual world. Similarly, Latham (2019) found 81.6% of participants selected the indeterministic universe as being most like the actual universe.

indeterministic, if they possess a conditional concept and are asked to evaluate the actual deterministic world, they should be expected to respond that there is free will in such a world. That's because according to the conditional concept, indeterministic and libertarian powers are only necessary for free will if they obtain actually. The strong signal for conditionality was more novel. For the minority of people who believe the actual world is deterministic, if they possess a conditional concept and are asked to evaluate a counterfactual deterministic world from the perspective of an actual indeterministic world, they should be expected to respond that there is no free will in that world. That's because according to the conditional concept, indeterminism and libertarian powers are necessary for free will if the actual world is indeterministic.

Latham (2019) found that people who believe the actual world is indeterministic respond that there is free will in an indeterministic actual world and a counterfactual indeterministic world from the perspective of a deterministic actual world. Further, they respond that there is no free will in a counterfactual deterministic world. Interestingly though, people who believe the actual world is indeterministic are unsure whether or not there is free will in the deterministic actual world (the weak signal for conditionality). People who believe the actual world is deterministic respond that there is free will in the deterministic actual world, the indeterministic actual world, and counterfactual indeterministic actual world from the perspective of a deterministic actual world. Again, interestingly, people who believe the actual world is deterministic are unsure whether or not there is free will in the counterfactual deterministic world, from the perspective of an indeterministic world (the strong signal for conditionality).

While people don't straightforwardly respond in a manner predicted by the conditional concept, they do respond in a manner that supports the idea that we possess a conditional concept with respect to determinism. That's because I don't think it is mere coincidence that people who believe the actual world is indeterministic are unsure how to respond to an actual deterministic world. Nor do I think it's a coincidence that people who believe the actual world is deterministic are unsure how to respond to a counterfactual deterministic world from the perspective of an actual indeterministic world. Both these conditions are correctly identified as being important with respect to people's concept of free will once it has been identified that our concept of free will might be a conditional concept with respect to determinism.

Why are people unsure how to respond in conditions associated with the weak and strong signal for conditionality? Let's start with the weak signal

for conditionality. Imagine someone believes the actual world is indeterministic and is then asked to evaluate whether there is free will in the actual deterministic world. It's extremely unlikely that people change their beliefs about the actual world in order to perform such evaluations. Instead, what people most likely do is simulate how they would respond if they counterfactually believed the actual world is deterministic. Importantly, this cognitive process does not mask the effects of what people actually believe, which is what explains why people are unsure about how to respond. If someone has a conditional concept and believes the actual world is indeterministic, then they should also think that indeterminism and libertarian powers are necessary for free will. So according to their actual belief there is no free will in the deterministic actual world. But if they succeed in simulating what they would think if they counterfactually believed the actual world is deterministic, then they should also think compatibilist powers are sufficient for free will. So according to their simulated counterfactual belief there is free will in the actual deterministic world. Thus, there is a response conflict between their responses generated in accordance with their actual belief, and their simulated counterfactual belief.

This also explains why we observe that people who believe the actual world is deterministic are unsure how to respond in the condition associated with the strong signal for conditionality. Imagine now someone who believes the actual world is deterministic and is asked to evaluate whether there is free will in a counterfactual deterministic world from the perspective of an indeterministic actual world. If that person has a conditional concept with respect to determinism, then they should also think that compatibilist powers are sufficient for free will. So according to their actual belief there is free will in the counterfactual deterministic world. But if they succeed in simulating what they would think if they counterfactually believed the actual world is indeterministic, then they should no longer think that compatibilist powers are sufficient for free will. Instead they should think that indeterminism and libertarian powers are necessary for free will. So according to their simulated counterfactual belief there is no free will in the counterfactual deterministic world. As a result, there is a conflict between free will responses that are generated in accordance with someone's actual and simulated counterfactual beliefs.

3.2. Determinism and the Conceptual Impossibility Thesis

If the folk concept of free will is a conditional concept with respect to determinism, then the conceptual impossibility thesis too, at least with respect to determinism, is correct. That's because no matter how things turn out actually to be—with respect to the world being deterministic or

not—if we possess that concept, we will judge that we possess free will. Once you hold fixed the compatibilist powers and libertarian powers in all these worlds, all worlds considered as ways things might actually be contain agents with free will. So even if determinism is actually true, and we only possess compatibilist powers, we will judge that we are free. On the other hand, if indeterminism is actually true, and we possess libertarian powers, we will judge that we are free, and that indeterminism and libertarian powers are necessary for free will.

This means there is something we could discover, if the conditional story is correct, which would make us think that indeterminism and libertarian powers are necessary. But that doesn't mean that the conditional concept of free will is inconsistent with the conceptual impossibility thesis, because there is nothing we could discover about how things are *actually* that would make us judge that actually there's no free will. Remember, we're holding fixed here that there are actually either compatibilist or libertarian powers. So, my claim is just that nothing we could discover about determinism would lead us to judge that we are unfree. As I suggested earlier, I think the conceptual impossibility thesis generalizes beyond determinism, but I have no empirical data that can support that contention here. Still in the next section (§4) I will give some good reasons why I think we should expect this.

So, with regard to the world being deterministic or not, free will is compatible with anything that we could discover about how things actually are. But if that's right then how did we become convinced that the folk concept of free will is an incompatibilist one? The conditional analysis offers up a ready explanation. If people think that the actual world is indeterministic and contains agents with libertarian powers, then they will judge not only that we are free, but also that deterministic possible worlds containing only agents with compatibilist powers lack free will (see footnote 7). So to the extent people are confident that actually, the world is indeterministic and there are libertarian powers, they should be expected to deny that compatibilist powers are sufficient for free will. From the perspective of a world where indeterminism is true, some, but not all, counterfactual worlds will contain agents with free will.

Of course, in most of the free will literature the distinction between judging of the actual world that it is deterministic and that indeterminism is a necessary condition for free will, and judging of the actual world that it is indeterministic, and that indeterminism is a necessary condition for free

will, is not made.[8] What actual philosophers of free will, embedded and entrenched in their philosophical views, would judge when this distinction is drawn is not something about which we have empirical data. Nevertheless, I think this distinction makes a difference to the judgments of ordinary agents.

4. Conditionality and the Conceptual Impossibility Thesis

In the previous section I provided evidence that the folk concept of free will is conditional with respect to determinism. This results in our judgments about whether or not *we* typically possess free will being insensitive to whether determinism is actually true. Instead, the truth or otherwise of determinism only affects our counterfactual judgments about whether agents in other worlds have free will. One way to think of the conceptual impossibility thesis is as a generalization of this.

So far, I have talked about whether determinism is true or false *simpliciter*. But it's important to also consider potential defeaters of free will (of which determinism is just one) in another way: the local way. How might we react if we were to learn that it is *sometimes*, *somewhere* true. While in fact in the case of determinism it is plausible that it's either globally true, or else false, when we generalize from determinism to other factors people might have thought important for free will, this may not be so.

Our judgments about whether we typically have free will are *insensitive* to various apparent defeaters to our free will being true *in general*. Imagine for the moment your favorite free will defeater *X*. If there is no global *X*, then having *X* rules out counterfactual populations from being free (and perhaps niche local populations as well). But if in fact *X* is generally actually true, then it doesn't affect our judgments about counterfactual populations. The presence or absence of *X* does *not* affect out judgements about whether actually *we* are free at all.

Let's work through a couple of examples. Imagine how the account I am offering might deal with another important challenge to free will. If what some brain scientists think is correct, then conscious psychological states do not perform the role we suppose they do for our actions (e.g., Libet et

[8] To the best of my knowledge Peter Van Inwagen (1983) is the only theorist who appears to identify this distinction and thinks that the actual world is indeterministic and that this indeterminism is necessary for free will. In the very last paragraph of his book *An Essay on Free Will*, he writes "…it is conceivable that science will one day present us with compelling reasons for believing in determinism. Then, and only then, I think should we become compatibilists." (p. 223)

al. 1983; Soon et al. 2008). Instead, conscious psychological states, and the actions that we suppose they cause, are both caused by an unconscious common cause. Let's call worlds where all actions are like these brain scientists think: *Libet worlds*. On the account I have been describing, if the actual world is one where conscious processes are causally involved in typical decisions, then we might think that is necessary for free will. But if actually they are not—if our world is a Libet world—then we will say that so long as the typical neural common cause of the action and its accompanying conscious state is in place, then the resultant action is free.

We can also imagine an even more extreme case (even by the lights of the free will literature). Imagine everyone's actions everywhere are being controlled by an alien species called Dromes. These Dromes have total control over both our conscious and unconscious psychological states, and thus our actions as well. For ease of explication, let's call worlds where all actions are controlled by Dromes: *Drome worlds*. On the account I have been describing, if the actual world is a Drome world, then we would still have free will, since free will is just whatever we are tracking that that distinguishes the cases we ordinarily judge to be free and the cases we ordinarily judge to be unfree. But if the actual world is not a Drome world, as is commonly supposed, then only those actions that are not the result of Drome control will be free, and necessarily so.[9]

Of course, we can be almost certain that free will error theory would be true of our metaphysical concept of free will if the actual world is either a Libet or Drome World. Still, despite there being no deep metaphysical difference between the cases we judge to be free and unfree, I think that we can be expected to want our actions to be like the ones that we ordinarily think of as free. Even if, with respect to some particular metaphysical matter of fact, there is no difference between these actions, there are relevant differences that our social concept of free will tracks with our talk of free and unfree action. It also seems that we can be expected to normatively evaluate actions, and to do that we need to distinguish actions that are performed while being, (what we might have ordinarily of thought of as), coerced and manipulated, from those that are not, and actions performed while being in the grips of, (what we might have ordinarily of thought was), a psychological or physiological illness, from those that are not, and so on. Regardless of whether all our actions are the result of unconscious processes, Dromes, *mutatis mutandis* for all these kinds of

[9] You might think that it's consistent with us making the discovery that we have no free will that our free will practices would persist, albeit as a useful fiction. However, on the view that I am advancing here, if the free will practices are what is in common between paradigm cases, then free will just is realism about the practices, and so we do have free will.

cases, we still care that our actions be like the ones that we would ordinarily think of as being free. If that's right, then the conceptual impossibility thesis is true of our social concept of free will.

5. Conclusion: The Conceptual Impossibility of Free Will Error Theory

In this paper I have argued for the conceptual impossibility of free will error theory - the conceptual impossibility thesis. There are two apparent concepts of free will: a metaphysical concept that tracks some metaphysical feature of our actions, and a social concept that tracks relevant differences between actions we ordinarily judge to be free and unfree. The weak conceptual impossibility thesis is that while free will error theory might be true on the metaphysical concept, there will be free will regardless, on the social concept. That's because our social concept of free will cannot do the job it's supposed to, and that concept fail to be satisfied. So, the conceptual impossibility thesis is true of that concept. The strong conceptual impossibility thesis is that while both concepts exist, only the social concept matters, and so the conceptual impossibility thesis is true of the only concept of free will we care about.

The conceptual impossibility thesis not only makes good sense of our practices—that we continue to hold people responsible for some actions, and not others, regardless of whether we think that our world is deterministic, and regardless of whether we think that certain neuroscientific findings hold—and it helps us make sense of some inconsistent findings in the experimental philosophy literature examining our concept of free will. This, jointly, gives us some reason to think that the conceptual impossibility thesis is correct, and that there is no way the actual world could be such that we judge that we do not have free will: on at least *one* of our concepts.

Acknowledgements

I am grateful to David Braddon-Mitchell, Kristie Miller, Michael Duncan, James Norton, two anonymous referees, and the metaphysics group at the University of Sydney for their useful feedback on the earlier versions of this manuscript. Thanks to the Ngāi Tai Ki Tāmaki Tribal Trust for their support.

REFERENCES

Björnsson, G. 2014. Incompatibilism and "Bypassed" Agency. In *Surrounding Free Will*, ed. A. R. Mele. Oxford: Oxford University Press.

Braddon-Mitchell, D. 2003. Qualia and analytical conditionals. *The Journal of Philosophy*. 100: 111–135.

Braddon-Mitchell, D., and K. Miller. 2019. Quantum gravity, timelessness, and the contents of thought. *Philosophical Studies*, 176: 1807–1829

Chalmers, D. J. 2005. The Matrix as Metaphysics. In *Philosophers Explore the Matrix*, ed. C. Grau. Oxford: Oxford University Press.

Churchland, P. M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78: 67–90.

Churchland, P. S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. MIT Press.

Daw, R., and T. Alter. 2001. Free acts and robot cats. *Philosophical Studies*, 102: 345–357.

Deery, O. 2019. Free actions as a natural kind, *Synthese*. DOI: 10.1007/s11229-018-02068-7

Dennett, D. C. 1984. *Elbow Room: The Varieties of Free Will worth Wanting*. MIT Press.

Dennett, D. C. 2013. Please Don't Feed the Bugbears. In *The Philosophy of Free Will: Essential Readings From the Contemporary Debates,* eds. P. Russell and O. Deery. Oxford: Oxford University Press.

Ekstrom, L. 2002. Libertarianism and Frankfurt-style cases. In *The Oxford Handbook of Free Will, 2nd edition*, ed. R. Kane. Oxford: Oxford University Press.

Flew, A. 1955. Divine Omnipotence and Human Freedom. In *New Essays in Philosophical Theology*, eds. A. Flew and A. McIntyre. London: SCM Press.

Heller, M. 1996. The mad scientist meets the robot cats: Compatibilism, kinds, and counterexamples. *Philosophy and Phenomenological Research*, 56: 333–337.

Kane, R. 2005. *A Contemporary Introduction to Free Will*. New York: Oxford University Press.

Kant, I. 1781/1787. *Critique of pure reason.* P. Guyer and A. Wood (eds. and trans.), Cambridge: Cambridge University Press, 1997.

Latham, A. J. 2019. *Indirect Compatibilism*. Dissertation. http://hdl.handle.net/2123/20440

Libet, B., C. A., Gleason, E. W. Wright, and D. K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity

(readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106: 623–42.

Lockie, R. 2018. *Free Will and Epistemology: A Defence of the Transcendental Argument for Freedom*. London: Bloomsbury Academic.

Nahmias, E. 2011. Intuitions about Free Will, Determinism, and Bypassing. In *The Oxford Handbook of Free Will, 2nd edition*, ed. R. Kane. Oxford: Oxford University Press.

Nahmias, E., S. Morris, T. Nadelhoffer, and J. Turner. 2005. Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, 18: 561–584.

Nahmias, E., S. Morris, T. Nadelhoffer, and J. Turner. 2006. Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73: 28–53.

Nichols, S. B., and J. Knobe. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*. 41: 663-685.

O'Connor, T. 2000. *Persons and Causes: The Metaphysics of Free Will*. Oxford: Oxford University Press.

Pereboom, D. 2001. *Living Without Free Will*. Cambridge University Press.

Pink, T. 2004. *Free Will: A Very Short Introduction*. Oxford: Oxford University Press.

Roskies, A. L., and S. B. Nichols. 2008. Bring moral responsibility down to earth. *Journal of Philosophy*, 105: 371–388.

Soon, C. S., M. Brass, H. J. Heinze, and J. D. Haynes. 2008. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 11: 543–545.

Strawson, G. 1986. *Freedom and Belief*. Oxford: Clarendon Press.

van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.

van Inwagen, P. 1993. *Metaphysics*. Boulder, Co.: Westview Press.