**Indirect Compatibilism**

[*forthcoming*]

Andrew J. Latham

Department of Philosophy and History of Ideas, Aarhus University

**Abstract:**

In this paper I will introduce a new compatibilist account of free action: *indirect conscious control compatibilism*, or just *indirect compatibilism* for short. On this account actions are free either when they are caused by compatibilist-friendly conscious psychological processes, or else by sub-personal level processes influenced in particular ways by compatibilist-friendly conscious psychological processes. This view is motivated by a problem faced by a certain family of compatibilist views, which I call *conscious control views*. These views hold that we act freely when we act in a way that is caused by certain conscious psychological processes. One problem for such views is that current neuroscience suggests that most of our actions are not caused by such processes. Instead, many of the actions we typically suppose are free are caused by sub-personal level processes and hence would count as unfree according to contemporary conscious control views. I argue, contra these views, that many actions caused by these sub-personal level processes are *indirectly* free. Further, most of the actions we ordinarily judge to be free are free in this *indirect* manner.

In this paper I introduce a new compatibilist account of free action: *indirect conscious control compatibilism*, henceforth called *indirect compatibilism*. I take compatibilism to be the view, very broadly, that we act freely just in case our actions are caused by some mechanism *M*, where *M* is compatible with determinism. Then different versions of compatibilism have very different views about what *M* is. The class of compatibilist views with which I am particularly interested in this paper, are those I call *conscious control views*. On such views we act freely when our actions are caused by the right kinds of conscious psychological processes. The right kinds of conscious psychological processes might be desires, reasons, conscious deliberations, standard decision-making processes, and so on, or some combination of these.Indirect compatibilism is motivated by a problem faced by these views. The problem is that we have good reason to think that most of our actions are not caused by such conscious psychological processes. Instead, most of our actions are caused by unconscious sub-personal level processes and hence would count as unfree according to conscious control views. Contrary to these views, I will argue that many actions caused by such sub-personal level processes are *indirectly* free, and that many of the actions we ordinarily judge to be free are free in this indirect manner.[[1]](#footnote-2)

To understand conscious control views and their appeal, let us begin with an example. Katie is deliberating about whether to pursue philosophical research in the United States of America or psychological research in the United Kingdom. Before making her decision, she consciously deliberates. Living in the UK, she weighs up the costs of uprooting and moving away from her friends, family, and partner, against the benefits to her education, job prospects and passion for knowledge and learning. She simulates various possible outcomes for each choice. Eventually, she chooses one option or the other, and that choice causes her to perform one series of actions over the other. For example, if Katie chooses to pursue philosophy in the US, she then performs the series of actions that will cause her to research there. Furthermore, in doing so, she does not perform the series of actions that would have caused her to research in the UK.

If this seems to you to be a paradigm case of free action, then you share my intuition that, in some very broad sense, actions caused by the right kinds of conscious psychological processes are free. What are the right kinds of conscious psychological processes in the example? They could be her desires, or her reasons, or the outcome of her conscious deliberations, and so on.[[2]](#footnote-3) Perhaps they are all or some subset of these psychological processes. Different compatibilist versions of conscious control views will take different stances about which psychological processes are necessary and sufficient for an action to be free. I will not take a stand on this matter in this paper and will simply call whatever those conscious processes are (along with their neural correlates) that directly cause free action: *freedom-conferring mechanisms*.

My aim in this paper is not to argue for compatibilism and against incompatibilism. Nor is it to argue for conscious control compatibilism and against other kinds of compatibilism. I will simply assume that conscious control compatibilism is plausible, as what I want to do is develop a version of the view that better responds to what I see as a major problem for such views. That problem is that we have reason to think, as a matter of fact, that very few (if any) actions *are* caused by freedom-conferring mechanisms, and thus that few (if any) actions are free.

Some readers might be satisfied if only a few of our actions are free; but I am less sanguine. That is because if only a few of our actions are free, then depending on the conceptual relationship between free action and free will, it may turn out that we do not have free will. For as I see it, our concept of *free* *action* is about the aetiology of our actions, and our concept of *free* *will* is about the general capacity of an agent to reliably act freely. Thus, if we do not reliably act freely, then I think it a very real threat that we do not have free will, or our conception of free will will need to be radically revised.

In this paper I will not take a stand regarding how *many* of our actions must be free for us to have free will. But unless the number of actual free actions roughly aligns with the number of actions that we typically judge to be free, there will need to be some serious revision. I tend to think, along with non-philosophers, that this must be a significant proportion of our actions, if it is to be the case that we reliably act freely. Ultimately, this is an empirical point which would be good to examine in experimental philosophy, still I think that it is plausible. Here is the idea. Consider Luca going through her daily morning routine. With little thought, she gets out of bed, has a shower, brushes her teeth, puts her clothes on, has breakfast, goes outside, closes, and locks the door behind her, walks up the road, pats a cat named McSnoots, and so on. At the top of the road, Luca stops, and for the first time that day consciously deliberates over where to get her morning coffee. If you are like me, then it strikes you that we would ordinarily think that Luca acts freely throughout (most of) her morning, and not just while she is getting her coffee. Without evidence to the contrary we default to thinking that we act freely. But, if the only actions that are free are those directly caused by freedom-conferring mechanisms, and if there are very few of these, then we do not act reliably in a free manner. Hence, we do not have free will.

Even if you do not buy this argument, and you think that under these circumstances we would still have free will—just not as much of it as we initially suspected—there is still a good reason to want to resist this conclusion. If a view is one according to which most of the situations in which we take ourselves to be free, are ones in which we are unfree, then surely this a cost to the view. For it fails to categorize our actions into those that are free, and those that are unfree, in anything like the ways that we typically care about, and the ways we track with our ordinary talk of free action, free will, and moral responsibility. Much better, then, if a compatibilist account can vindicate most of these judgements.

My aim in this paper is to show that there is a compatibilist version of a conscious control view that can do just that.

My solution is to show that many actions are *indirectly* caused by freedom-conferring mechanisms, and that it is (modulo defeaters perhaps) necessary and sufficient for free action that the action be either directly, or indirectly, caused by such mechanisms. This is the core of indirect compatibilism, which is the view that an action is free only if it is caused by some freedom-conferring mechanism, or *indirectly* caused by them via some sub-personal level mechanism or mechanisms. Of course, this general idea is not logically connected to compatibilism. We can easily imagine a version of this account where the freedom-conferring mechanisms and sub-personal level mechanisms they influence are of a libertarian kind**—**indeterministic, whilst not being merely chancy**—**but as I develop the proposal in this paper, I will assume that compatibilism is true.[[3]](#footnote-4)

Here, then, is the plan for the rest of the paper. In Section 2, I show why the number of actions judged to be free by standard conscious control views is likely to be far fewer than the number of actions that we typically judge to be free**—***the discrepancy problem***—**and show that it can be solved by holding certain sub-personal level mechanisms cause free actions. In Section 3, I introduce the notion of an action procedure, a special kind of sub-personal level mechanism, and show how action procedures influenced by freedom-conferring mechanisms can be used to solve the discrepancy problem. In Section 4, I describe some important kinds of influence that freedom-conferring mechanisms can have on action procedures which result in them causing free actions. In Section 5, I discuss some of the different ways an action being free might depend on the influence freedom-conferring mechanisms have on actions procedures. Finally, in Section 6, I describe the connection between indirect freedom and moral responsibility, and then conclude.

**2 The Discrepancy Problem**

Let us begin, why should we think that there is a discrepancy between the number of actions classified as free by conscious control views and the number of actions that we typically judge as free? One common (and perhaps best) route to arrive at that conclusion comes via evidence from cognitive neuroscience. There is accumulating, albeit contentious, evidence that freedom-conferring mechanisms cannot be playing the role we suppose they do in causing actions. Consider for the moment the following basic paradigm. Participants watch a time-keeping device and report the time shown when they first feel the urge to perform some simple motor action, such as a wrist flex, or cognitive action, such as basic arithmetic. Shortly after they have performed the simple action, the time-keeping device stops, and the experimental task is reset. It is robustly found that the onset of the neural activity associated with an action occurs well before participants’ first report their urge to perform the action (e.g., Bode, et al. 2011; Bode et al. 2012; Libet, Gleason, Wright, & Pearl, 1983; Schultze-Kraft et al., 2016; Soon, Brass, Heinze, & Haynes, 2008; Soon, He, Bode, & Haynes, 2013). Based on these results some researchers conclude that our freedom-conferring mechanisms cannot be the cause of our actions (e.g., Caruso, 2012; Gazzaniga, 2011; Greene & Cohen, 2004). After all, if freedom-conferring mechanisms were causing actions, then the time participants’ first report their urge to perform an action should occur before (or at least in line with) the onset of supposed neural activity for that action.

There have been two broad categories of criticisms of these findings: *methodological criticisms*, which point out issues, either with, the participants, methods, and analyses, and *conceptual* *criticisms*, which point out issues with the kinds of actions being investigated in these experiments and used to generate conclusions about the nature of free action and free will. In this paper I will not engage in detail with methodological criticisms of these kinds of studies, even though there are many good ones in the literature that are certainly worth developing. For instance, you might wonder whether people can reliably and accurately report the timing of their awareness to perform an action (e.g., Banks & Isham, 2009; Lau, Rogers & Passignham, 2007). You might also wonder whether the earliest recorded neural activity is related at all to the action that participants will go on to perform (e.g., Miller, Shepherdson, & Trevena, 2011; Schurger, Sitt, & Dehaene, 2012).[[4]](#footnote-5) For argumentative purposes I will set aside such critiques. That is, I will assume people are able to make accurate and reliable introspective reports and that the various psychophysiological components being reported are representative of some form of preparatory neural activity for a participant’s action. While critiques of this kind might be right, it is far from clear that they are, and so it is worth seeing how far we can get assuming that they are not.

The most common conceptual criticism of this body of evidence is that the actions being investigated in these experiments are not free or are a very rare, or special, kind of action (e.g., Tallis, 2011; Mylopoulos & Lau, 2014). This means that investigating them does not reveal anything interesting to us about the nature of ordinary free action. It is not hard to motivate this kind criticism, especially once you see the kinds of task instructions that Libet and colleagues (and other researchers for that matter) present to participants. For instance:

“[…] let the urge to act appear on its own at any time without any preplanning or concentration on when to act” (1983; pg., 625).

Many people find such an instruction utterly mysterious, and it is unclear how participants would go about implementing such an instruction. Furthermore, if this kind of instruction could be followed accurately, then it would seem to rule out the relevance of these experiments for determining whether freedom-conferring mechanisms cause actions. After all, the experimental instruction I just quoted appears to be an instruction to not use freedom-conferring mechanisms at all in the experiment. At best what this body of results would show is that some of our actions are not free because they are not caused by our freedom-conferring mechanisms; but that is hardly an interesting result as no one ever thought that all our actions were free.

Now if this is what I thought was occurring in these experimental tasks, then I might conclude that there is already enough reason to think these results are no major problem for conscious control views. After all, if the actions which are being investigated are those that we always considered to be not free, or are some very rare, or special kind of action, then it is possible that most of our actions are still caused by freedom-conferring mechanisms. But I think quite the opposite is true. I think that these results reveal to us the nature of a perfectly mundane and common kind of action, one which we commonly think of as being free.

But why think that? Here is one way to motivate the idea. Consider the previously described researchers in cognitive neuroscience investigating free action. I find it very striking that these researchers ever thought that the simple, repetitive actions in these experiments were free. Yet they must have, since otherwise, why would they think that their results speak to the issue of free action? I do not think that the judgement that these simple, repetitive actions are free is wrong, quite the opposite. I think that this judgment is right and reveals something interesting about our concept of free action. In these experiments people were asked to repeatedly perform a very simple action for hundreds of trials. Once people had freely decided to participate, most people would have automated their performance so that their actions were not caused by freedom-conferring mechanisms but instead were being caused by a sub-personal level process. Insofar as we think that we correctly judged that the actions performed during these experiments are free, then *prima facie* it seems as though some sub-personal level mechanisms cause free actions.

Of course, this thought generalizes beyond the experimental context. For instance, consider the case of commuting to work on an ordinary day. While conscious control is necessary to get the act of getting to and from work up and running, over time it can take place with less-and-less conscious influence, even when it involves complex actions such as driving a car (though we are unlikely to admit to that). This does not mean there is no thought occurring while we are commuting. Rather as we require less conscious thought to commute, we free ourselves up to be able to deliberate over other matters. For instance, what our goals are for the day and what out plans are to achieve them. Or perhaps more likely, what Netflix series we hope to binge once we return home. The fact that we are free to think on other matters while commuting also explains why we are sometimes surprised by the fact that we have arrived at work or back home. That is because once the act of getting between home and work becomes automated it can be accomplished without the need of conscious oversight.

Or consider again the earlier case of Katie choosing between philosophical research in the US and psychological research in the UK. Katie chose philosophical research and as a result performs the set of actions that will cause her to research in the US. The choice of language in the original case was intentional as what I am particularly interested in, in this paper, is the status of the actions that bring about Katie’s choice to perform philosophical research in the US. Once Katie has freely chosen philosophical research, she might automate at least some of her actions to bring about her choice to research in the US. That is, at least some of her future actions might be caused by a sub-personal level mechanism whose function it is to bring about her choice to research in the US. For example, it is easy to imagine Katie packing up her belongings and cleaning her residence without much thought, and instead thinking about where in the US she might like to live. Further, once Katie has settled on places she would like to rent in the US, we can easily imagine her submitting the completed rental applications without much thought, and instead thinking over the details of her future research topic, and so on. The thought then is that even if only a few of our actions are directly caused by freedom-conferring mechanisms, most of our actions might still be free if they are caused by a certain kind of sub-personal level mechanism: those which have been influenced in some way by freedom-conferring mechanisms.

If this thought is right, then it is easy to see why, at least broadly speaking, we can avoid the discrepancy problem. For the discrepancy problem arises because we find that many of our actions are caused by sub-personal mechanisms and not by freedom-conferring mechanisms. And so, if we think that the only mechanisms that confer freedom are these freedom-conferring mechanisms, then we must conclude that most of our actions are not free. On the view I am advancing here, however, there is more than one way for an action to be free: it can be directly free, by being caused by a freedom-conferring mechanism, or it can be indirectly free, by being caused by a particular sort of sub-personal mechanism. But just what sort of sub-personal mechanism causes free actions? I turn to that matter in the next section.

1. **Controlled Action Procedures**

So far, I have said that one major problem for conscious control views is that experimental evidence from cognitive neuroscience seems to show that freedom-conferring mechanisms do not cause actions. Many theorists think this conclusion (that there are no free actions) is wrong because the actions being examined in Libet-style experiments are a very rare or special kind of action. While I agree that the conclusion is wrong, contrary to others, it is because I think the actions performed during these experiments are both free and representative of most of our actions. More specifically, I think they are representative of actions caused by a special kind of sub-personal level mechanism, which I will call an *action procedure*. In this section I will describe what action procedures are and begin to explain why they are important for free action.

An action procedure is a sub-personal level mechanism that causes actions, or influences some other action procedure, both in a *contained* and *reliable* fashion. So, to start with a simple case, we can imagine a simple fist-bumping action procedure, which causes a fist-bumping action. The fist-bumping action procedure is both contained and reliable, in the sense that it will only produce actions of a fist-bumping kind (contained) and whenever it is activated will rarely fail to produce fist-bumping actions (reliable). So, to be contained a sub-personal level mechanism must only cause a certain kind of action or influence other sub-personal level mechanisms in a certain kind of way. To be reliable, a sub-personal level mechanism when activated must reliably cause actions or influence other sub-personal level mechanisms. Sub-personal level mechanisms that are *both* contained and reliable are action procedures.[[5]](#footnote-6)

But what about the ways action procedures might influence other action procedures? Matters here are more complex as there are numerous ways this might occur. However, I think that my account becomes much more concrete once I outline some of the possibilities.

One way an action procedure might influence other action procedures is by selectively filtering the information we have available, and in doing so makes some action procedures more, or less, likely as options than other action procedures. In some cases, an action procedure might selectively filter *all* the information about a particular action procedure causing it to not appear as an option to the agent. For instance, imagine Luca again, who now walking home from work now finds herself feeling very thirsty. She stops and begins looking around for somewhere to get drink of water. There is no shortage of water sources around Luca. Prior to her walking home there was a heavy downpour of rain, and so there are plenty of puddles of water on the ground. On the street Luca is walking along each home has an external hose tap in the front garden. The home Luca is walking by is wide open with the kitchen in plain sight, and so on. Still, Luca concludes to herself that there is nowhere she can get a drink of water and so continues walking home.

One reason Luca might have thought there are no options available to her apart from continuing to head towards home was because she is *sincerely* unaware of the puddles on the ground, the external hose taps, and the stranger’s open kitchen. An action procedure filtered out all these pieces of information and that meant Luca had no good reason to think that the option to get a drink of water was afforded to her. And, if Luca had no good reason to think that the option of getting a drink of water was afforded to her, then she had no good reason to hang around.

Another way an action procedure might influence other action procedures is by influencing which possible courses of action we can deliberate over that are set in motion through action procedures. Crucial, here, is that not all information, or options we are consciously aware of, are ones which enter the causal process of deliberation. The action procedures that we deliberate over represent the pool of action procedures that we can freely choose from when deciding what to do. If an action procedure is not part of the deliberation pool, then it is not an option for choice, even if we can entertain the idea, and there are no constraints on what we can choose within the deliberation. So, Luca might be aware of the puddles of water on the ground, the external hose taps, and the stranger’s open kitchen, and she might judge that if she were to make use of any of these options, she would be able to satisfy her thirst. But, if an action procedure has filtered out the action procedures that would set in motion these courses of actions, then the option of getting a drink of water (around here) is not there for Luca to choose, even though she can entertain the idea of drinking from a puddle, an external hose tap, and the stranger’s kitchen.

Alternatively, an action procedure might influence other action procedures by influencing which possible courses of action, which are set in motion through action procedures, that we can choose. So, Luca might be aware of the option of drinking water from the puddle on the ground, the external hose taps, and the stranger’s open kitchen, and that each of these options would be able to satisfy her thirst. In addition, the action procedures that would set in motion these courses of action are all in her deliberation pool. However, an action procedure might cause her to be unable to choose from among any of these options. The key difference between this case, having an effect at the point of choice, and the previous case, having an effect at the point of deliberation, is that in the former there is an apparent option to choose, whereas in the latter there is not even the apparent option. The influence of the action procedure is at a later stage: at the point of choice, rather than at the point of choosing which options to deliberate over.

Finally, an action procedure might influence other action procedures by influencing which actions action procedures can cause. So, Luca might be aware of the option of drinking from the puddle, the external hose taps, and the stranger’s open kitchen, and that these options would satisfy her desire to drink water. Luca might also find each of these courses of action and their corresponding actions procedures are in her deliberation pool, and so she is able to choose from among them. But, after Luca chooses one of these courses of actions, she still finds that she is still unable to drink any of the water. For instance, Luca might choose to drink water from the puddle on the ground, and is able kneel by the puddle, bring her head towards the water, and so on. However, the action procedure in this case actively filters the actions which are being caused by the action procedure so that no matter what, Luca cannot drink the water in puddle on the ground.

Now, if *all* our action procedures are independent of our freedom-conferring mechanisms, then I think that those actions are unfree. However, I think that many action procedures are influenced by freedom-conferring mechanisms, and what is more, I think that action procedures that have been influenced by freedom-conferring mechanisms cause free actions, albeit *indirectly*. Put roughly for now, this is the essence of indirect compatibilism: an agent acts freely either when their actions are directly caused by freedom-conferring mechanisms, or when they are indirectly caused by freedom-conferring mechanisms via some action procedure.

I will call an action procedure that has been influenced by a freedom-conferring mechanism a *controlled action procedure*. Perhaps the simplest kind of influence a freedom-conferring mechanism can have on an action procedure is its *creation*. For example, an agent might create a fist-bumping action procedure that causes them to fist-bump whenever someone fist-bumps at them. Or an agent might create an action procedure that causes them to be unable to kill other people by influencing other actions procedures. Important for my purposes now, an agent might create an action procedure to automate their actions while performing a Libet-style experiment. This controlled action procedure would reliably cause them to flex their wrist in a timely manner sometime after the start of an experimental trial has been cued. The reason why the actions caused during the experiment are free is because the action procedure which causes them was caused by a freedom-conferring mechanism or mechanisms of the agent.[[6]](#footnote-7)

Exactly what sorts of connections between a freedom-conferring mechanism and an action procedure count as *influence* will be discussed in detail in the next section. For now, I will stick with creation as a paradigm instance in which a freedom-conferring mechanism influences an action procedure.

To illustrate further, consider Alfred Mele’s (2014) description of his experience and approach to performing a Libet-style experiment:

At first, I waited for conscious urges to flex to pop up in me so I’d have something to report when it was time to make the *consciousness* report. I waited until I was pretty sure the urges weren’t just going to pop up on their own. I wondered what to do, and I hit on a strategy. I decided that I would just say “now” to myself silently, flex my wrist in response to the silently uttered cue, and then, a little later, try to report where the hand was on the Libet clock when I said “now”. (pg. 14)

While Mele’s experience is of course his own and different people will report different experiences, there are two important points to take away from this report: points that I think generalize beyond any one individuals’ approach to the experimental task. First, actions simply do not appear out of nowhere. That is, it is hard to imagine how participants could bring about wrist flexes (or any other action, including cognitive ones) without any thought for the requirements of the experimental paradigm they were participating in. Second, actions that might seem to appear out of nowhere are in fact caused by some controlled action procedure of the agent.

With action procedures and controlled action procedures now explained we can return to Mele’s performance of a Libet-style experiment. First, as Mele describes, simply waiting for urges to flex one’s wrist seems like a non-starter to successfully complete a Libet-style experiment. For one, there is no guarantee that the required urge would ever arise over the course of the experiment—if at all—if there were no controlled action procedure in place to cause the required actions. This issue is compounded by the fact that these experiments rely on there being numerous successfully completed trials. At the very least, successful experimental completion requires some controlled action procedure to cause some wrist flex at some time during each experimental trial. As far as I am aware, there has never been a reported case of some poor participant spending an experiment simply sitting and waiting, with no success, for the urge to flex their wrist to arise on its own.

Of course, I do not think that creation (or other kinds of influencing for that matter) requires that people make the conscious decision to create a specific controlled action procedure, or bring it about that one controlled action procedure, as opposed to another, activates, and causes actions. Instead, I think that participants’ conscious decision to participate in the experiment is realized by the creation of an action procedure which automates and causes their actions for the duration of the experiment. There will be some relatively rare cases where someone does create a very specific action procedure via conscious control. For example, expert musicians might create an action procedure to cause a very precise and specific set of actions via conscious control. I vividly recall performing finger dexterity exercises on the piano which began by consciously controlling individual finger movements in a precise manner repeatedly until such time as these precise finger movements became entirely automated. Expert athletes very likely do something similar in their training.

If this general thought is correct, then the reason Libet-style experiments fail to show there are no free actions (and free will for that matter) should be clear. It is not because these researchers are mistakenly targeting some special or rare kind of action, as some theorists have thought. Rather, the opposite is true. It is because these experiments do not target the special and relatively rarer kind of action: the act of a freedom-conferring mechanism creating (or influencing) some action procedure. For, as I have suggested, it is the fact that freedom-conferring mechanisms are involved, in an appropriate manner, in the influencing of action procedures which is responsible for many of our free actions, and hence for vindicating the thought that we have free will.[[7]](#footnote-8)

1. **Influence and Freedom**

As I have said, my solution to the discrepancy puzzle takes the form of an account of free action: indirectly free action. According to indirect compatibilism, we are not only free when we perform an action which was directly caused by freedom-conferring mechanisms, we are also free when we perform an action caused by an action procedure which has been influenced by a freedom-conferring mechanism or mechanisms. So far, I have suggested that we are free when we perform an action caused by an action procedure that we have created. But creation is not the only kind of freedom-conferring mechanism whose influence on an action procedure might make the actions caused by one free. In this section I will describe some other kinds of freedom-conferring mechanisms, distinct from creation, whose influence seems important on the freedom of actions. This will not be an exhaustive list of all the kinds of freedom-conferring mechanisms that are relevant to indirectly free action, but it will cover what I think are the most important and interesting ones. Then in the next section, I will describe some different ways these freedom-conferring mechanisms could be related to our concept of free action, including how they might be traded-off against one another.

Let us start simply with the example introduced in the previous section, of a freedom-conferring mechanism *creating* an action procedure. Now consider again a Libet-style experiment in which a participant repeatedly performs a simple action (e.g., wrist-flex, left-or-right button press, simple addition, or subtraction, and so on) in a timely manner, following some task cue, for the duration of the experiment. It is plausible to think that the simple action in question is being caused by a controlled action procedure, because the action procedure was one that was *created* by the participant. Upon being told the instructions for the Libet-style experiment, the participant created, and implemented, an action procedure which unconsciously causes the actions required by the task. That action procedure makes it so that when a task cue is presented, the necessary actions are unconsciously caused in a timely manner. It seems plausible to think that the actions caused by this action procedure in this case are indirectly free because the action procedure that caused them was itself caused by a directly free action.

As well as creating action procedures, another important influence agents might have over their actions procedures is the capacity to *destroy* them, should they freely want to do so. Destroying an action procedure reflects an agent’s capacity to remove an action procedure they possess.[[8]](#footnote-9) Such a capacity might be deployed whenever an action procedure is no longer required or has become detrimental to the agent because of changes in the environment. For example, it is easy to imagine ourselves creating an action procedure that automatically causes a simple action that is reliably rewarded whenever it is cued in the environment. But imagine the environment suddenly changes so that the unconsciously caused action, which was reliably rewarded, is now reliably punished. It is easy to imagine being motivated to forget such an action procedure as soon as possible. But, if I leave the punishing action procedure alone when there is nothing preventing me from destroying (or updating it, something that I will describe shortly) it, then contingent on their being the capacity to destroy action procedures (or turning it off, something that I will describe shortly too), it seems plausible to think I am freely continuing to get punished.[[9]](#footnote-10) So, the capacity to destroy action procedures seems relevant for actions being indirectly free as well.

You might think that being able to destroy and recreate action procedures at will means you will have no action procedures at all. Indeed, action procedures do collapse into conscious control in cases where creation and destruction are *instant*, and *costless*, but in reality creation and destruction are never instant or costless, even if they sometimes get close. The reason we adopt action procedures, is precisely because they reliably convey these benefits over our conscious control. Due to being sub-personal in nature, action procedures can reliably do things much more quickly and effortlessly than what can be done under our conscious control.

To illustrate this point it will be helpful to consider a case. In the previous section, I described the finger exercises that I would use to improve my piano performance. The purpose of this training was not so that I improved the conscious control of my finger movements (though this might be a by-product of such training), but so that those movements no longer required conscious control. The process of creating that action procedure was not instant, but effortful and timely, and the benefit was that I could then consciously control other elements of my performance such as prosody. But now imagine that I notice that I am making some reliable mistakes with my fingering while performing. I might engage some destruction process that sub-personally removes the action procedure over time and start over, but this does not seem like a great thing to do in this context. After all, many features of the original action procedure will be just fine and if we were to start over from scratch there is no guarantee that those good making features of the original action procedure would reappear. Instead, rather than destroying an action procedure we might instead freely *update* the function of an action procedure.

Rather than destroying an action procedure at will then, we might *update* an action procedure so greatly that it is effectively a removal and replacement. To illustrate, consider once again a simple action procedure created to perform a Libet-style experiment. The controlled action procedure causes a simple action, or performs a simple cognitive operation, some small time after the presentation of a task cue. Instead of intentionally removing the action procedure, we might remove the action procedure by altering it in ways that prevent it from performing. For example, we might remove the cue component so the action procedure no longer causes an action, or we might remove the response component, so the action procedure no longer has an action to cause. Or we might set a long time-delay between the cue and action production, such that there would be no reasonable expectation an action would be caused even when the action procedure is implemented.

But, just as an agent might have the capacity to update an action procedure to prevent its performance, so too might they have the capacity to update an action procedure to improve or change its performance. To illustrate this, once again imagine a someone who has an action procedure capable of automating performance in a Libet-style experiment. Now imagine that they are brought back to the laboratory to perform a slightly modified version of the task. While they are still required to perform some action, sometime after a cue is presented, their body is now hooked up to an accelerometer. Each time they perform an action, feedback is presented about the speed of their action. The feedback indicates whether the action was performed at the correct speed, and when it indicates the action was performed at the incorrect speed, whether they should speed-up or slow-down their action. Instead of creating a new action procedure, the participant in this example might simply update the original action procedure so the caused action comes to be performed at the required speed. Of course, the process of updating an action procedure might go through numerous iterations as the agent adjusts their action speed, and then looks to the feedback to see whether their update was successful. Still, once the correct speed has been identified, the updated action procedure will be able to automate performance in much the same manner as the original action procedure. It seems plausible to think that the action caused by the updated portion of the action procedure is indirectly free if the updates were caused by a directly free action.

Finally, an agent might already possess an action procedure which can automate and control their actions in the manner they want. In this case an agent might freely *turn on* an action procedure they already possess. To illustrate, consider a participant who already has an action procedure capable of automating actions during a Libet-style experiment. Such an agent does not have to create another action procedure; instead, they can simply turn on the action procedure they already have, to automate and control their task-relevant actions. Once again, this is not equivalent to direct control. As noted previously, due to being sub-personal in nature, action procedures can do things much more quickly and effortlessly than what can be done under our conscious control. And, while I am not able to consciously control my return service to a 180 kph serve in tennis, I am able to choose the kind of return service I will make by turning on the appropriate action procedure[[10]](#footnote-11). It seems plausible to think that the actions caused by the controlled action procedure are indirectly free because the action procedure that caused them was turned on by a directly free act.

Conversely, sometimes we can find ourselves performing the wrong kinds of actions, or simply wanting to change the kinds of actions we perform and so want to *turn off* the current action procedure. For example, consider the earlier example of an action procedure that automatically causes a simple action that is consistently rewarded. However, we can imagine a niche environment where that action procedure turns out to be maladaptive and the ordinarily rewarded actions are now consistently punished. If I leave the now punishing action procedure alone when there is nothing preventing be from turning it off, then contingent on their being the capacity to turn off action procedures, it seems plausible to think I am freely continuing to get punished. Thus, the capacity to turn action procedures on and off seems relevant for actions being indirectly free.

1. **Indirect Freedom**

So far, I have outlined a few of the important ways an agent might influence action procedures. These are important because they seem relevant to whether an action procedure is a controlled action procedure and so causes free actions, albeit *indirectly*. For instance, an agent might have the capacity to create an action procedure or destroy an action procedure. They might have the capacity to monitor their environment and turn an action procedure on or off. Finally, they might have the capacity to update an action procedure they possess to improve upon or modify its function. Of course, there might be other kinds of freedom-conferring properties that I have not described here. That is fine; nothing I say here hinges on having successfully described *all* the important ways an agent might influence their action procedures. Still, I think I have described some of the most notable and obvious ways this might happen. For ease of ongoing discussion, I will collectively refer to the capacity to create, destroy, monitor, and update action procedures as: *freedom-conferring properties*. Up to this point I have only suggested that these freedom-conferring properties seem relevant for indirect freedom. In this section I will begin the process of outlining how indirect freedom depends upon these properties.

Earlier, I roughly characterized a controlled action procedure as an action procedure that has been influenced by a freedom-conferring mechanism. We can now begin to be more precise. An action procedure is a controlled action procedure if it is created, and could be destroyed, triggered, monitored, and updated, by freedom-conferring mechanisms or another controlled action procedure. That is, it is sufficient (but not necessary) to be a controlled action procedure when an action procedure possesses all the freedom-conferring properties. Equally, though an action procedure might be a controlled action procedure if it were *not* created by a freedom-conferring mechanism, but, nevertheless, could be destroyed, monitored, and updated by a freedom-conferring mechanism. That is, it seems plausible for an action procedure to be a controlled action procedure provided it possesses some degree of freedom-conferring properties. Thus, the concept of a controlled action procedure is a kind of cluster concept, which means there are plausibly, numerous ways action procedures get to be controlled action procedures and a wide taxonomy of controlled action procedures to explore. Unfortunately, I do not have the space in this paper to explore every species of controlled action procedure. As a result, to simplify matters going forward let me first broadly cleave controlled action procedures into two varieties which we might call *paradigmatic controlled action procedures* and *non-paradigmatic controlled action procedures*. Something is a **paradigmatic controlled action procedure** iff it is a controlled action procedure that has all the freedom-conferring properties, whereas, something is a **non-paradigmatic controlled action procedure** iff it is a controlled action procedure that does not possess at least one of the freedom-conferring properties.

In this paper I do not want to enter an extended discussion of paradigmatic controlled action procedures (henceforth P-CAPs) as there should be *no* disagreement (at least) amongst (would be) indirect compatibilists that actions brought about by P-CAPs are indirectly free. Just as there should be *no* disagreement among indirect compatibilists that actions brought about by actions procedures independent of freedom-conferring mechanisms are not free. The reason for lack of disagreement about actions caused by P-CAPs is because they seem to possess *all* the features that we might think of as being important for free action. First, actions are caused by a sub-personal level mechanism that the agent was the source of (in the compatibilist sense).[[11]](#footnote-12) That is, the actions are being caused by a sub-personal level mechanism freely created or updated by the agent to perform the way that it is. Second, the actions are caused by a sub-personal level mechanism that the agent has on-going control of (in the compatibilist sense). Namely, the actions are being caused by a sub-personal level mechanism that the agent can monitor and freely turn on and off (or even destroy) if they wanted. Similarly, I do not think that anyone will be tempted by the thought that actions caused and influenced by sub-personal level mechanisms that an agent did not create and has no on-going control over are free.

One reason to look more closely at NP-CAPs is because while I think that we have clear and fixed intuitions about scenarios involving P-CAPs and actions procedures which have no freedom-conferring properties, matters become much less clear when we consider scenarios where there is only a subset of freedom-conferring properties. I have mixed thoughts about what to say about cases where our actions are caused, or influenced, by a sub-personal mechanism which we created, but have no on-going control over. Or, cases where our actions are caused, or influenced, by a sub-personal level mechanism which we have on-going control over but had no role at all in creating. I do not think I am alone in this thinking and that most indirect compatibilists will be divided on such cases. Thus, NP-CAPs will be a useful foil to highlight the different ways we might want to spell out the relationship between indirect freedom and these freedom-conferring properties.

Another reason to look more closely at NP-CAPs is because I think that cases where we have the capacity to create, destroy, monitor, and update action procedures are relatively rare, especially if we think that each of these capacities can be had in degrees as well. NP-CAPs then are likely to be representative of most of our controlled action procedures. And, if what I said earlier in this paper about our ordinary concept of free action being one that classifies most of our actions being free is true, then most NP-CAPs must cause free action. That is, our concept of freedom is not one that demands that each of our sub-personal mechanisms hold *all* the relevant freedom-conferring properties to be free. What is up for grabs is what counts as enough, and whether any of these freedom-conferring properties should be thought of as necessary for our freedom.

Perhaps the simplest account that we might have is that we act indirectly free if our action was caused or influenced by an action procedure that has any *one* of the freedom-conferring properties. Of course, it might possess more than just one. Such an account of indirect freedom does not distinguish between different kinds of freedom-conferring properties and says we are indirectly free just in case the action procedure in question has *any* freedom-conferring properties at all.

The problem with such a simple account is that it strikes most of us as being obviously wrong (though see Latham 2019 for an extended defense of something like such an account of free action). For instance, you might think that the freedom-conferring property of creation is *sufficient* for indirect freedom. Provided that the action procedure is one that you created, then despite lacking any other freedom-conferring properties the actions it causes, and influences, are indirectly free. But it seems that most people are going to think that some kind on-going control over the function of an action procedure is *necessary* for free action. Consider the case I described earlier of an action procedure that was reliably rewarded and then due to some change in the environment came to be reliably punished. It seems to me that in such cases we do not think that you continue act freely because the action procedure was one that you freely created. Instead, it is the very fact that you appear to lack any on-going control over your action procedure that instead makes us think that you are not acting freely at all.

Perhaps though it is *sufficient* for indirect freedom that you have (some kind of) on-going control over the action procedure. For instance, you might think that the freedom-conferring property of monitoring is *sufficient* for indirect freedom. Provided that the action procedure is one that you can monitor and turn on and off, then despite lacking any other freedom-conferring properties the actions it causes, and influences, are indirectly free. But this too will be contentious as most people are also going to think that being the source of the action procedure is *necessary* for free action as well. For instance, if you learned that the reason you possessed some action procedure is because some alien or nefarious neurosurgeon put it then in the dead of night, bypassing your ordinary psychological processes, this counts against the thought that you are free.

Does this mean that *any* of these freedom-conferring properties are necessary? I do not think so. Some people might think that it is necessary for freedom that the agent be the source of the action procedure. Similarly, some people might think that it is necessary for freedom that the action procedure be one that the agent has ongoing control over the action procedure. Other’s might be indifferent to the different kinds of influence the agent might have for their action procedures and just think that all is required is that the agent has *enough* influence. I will not take a stand on these issues here, though I am partial to thinking that what is necessary and sufficient for indirect freedom is an action procedure just having *enough* freedom-conferring properties. For now, I just wish to highlight that it is open to disagreement whether some individual freedom-conferring property is necessary or sufficient for indirect freedom and that indirect compatibilism can accommodate a plethora of different views on this point.

Still, it seems that the simplest account of indirect freedom will not do. It is not enough for indirect freedom that an action procedure simply has *a* freedom-conferring property. Previously, I noted that people will agree that actions caused by P-CAPs will be free and that actions caused by actions procedures independent of freedom-conferring properties are not. Where we become unsure are cases where only some subset of freedom-conferring properties is present. How might we capture this thought? Here are two suggestions: (a) an action is indirectly free if it is caused by an action procedure that has *enough* freedom-conferring properties or, (b) freedom of action comes in degrees, and there is some association between freedom of action and the degree of freedom-conferring properties.

Consider (a) for the moment. We might say that some action is free if the NP-CAP that causes it has enough freedom-conferring properties above certain vague threshold. Why a vague threshold? Because it is not clear to me that there is some straightforward number of freedom-conferring properties that would successfully cleave the NP-CAPs which cause free actions from those that do not. In addition, the idea that there is a sharp cut off between those NP-CAPs which cause free actions and those that do not seems wrong. That is, I do not think we think that actions caused by an NP-CAP with freedom-conferring properties just below a threshold value do not strike us as being just as unfree as ones caused by action procedures with no freedom-conferring properties at all. On this view, actions caused by NP-CAPs above the vague threshold are free, actions caused by NP-CAPs below the vague threshold would be no free, and actions caused by NP-CAPs that fall within the penumbra of the vague threshold would be neither determinately free not free. Further, some actions in the penumbra will be near its center and others will be more towards the edges, meaning some actions in the penumbra will be closer to being free and others closer to being unfree. This means we can also capture some sense of indirect freedom being graded without needing to posit that indirect freedom itself comes in degrees.

Now, consider (b). We might say that some action is free to a non-zero degree if the action procedure that caused *A* has *some* degree of freedom-conferring properties. Of course, reflecting on the apparent failure of the simple account we started out with reveals that this cannot have be a straightforward association. But that need not be an issue. For instance, we might think that some action is free to a non-zero degree if the NP-CAP which caused it has freedom-conferring properties above a vague threshold. Actions caused by NP-CAPs that possess freedom-conferring properties below such a vague threshold are free to a minimal degree, instead they are just straightforwardly unfree.

One helpful referee asks whether the notion of degrees of freedom and a penumbra suggests that there is something like Freedom with a capital F and freedom with a small f with moral responsibility being proportionate between them. For instance, imagine that you are late for your partner’s work reception because you got off at your regular subway stop when you shouldn’t have. In Case 1 you got off at your regular stop because you had indirectly-freely chosen to get off at this stop every time that you left work. In contrast, in Case 2 you got off at your regular stop because you had directly-freely chosen to get off so that you would be late to the reception. Intuitively it seems that while you act freely in both cases you are not free in the *same way*.

I think that the referee is correct that we want to be able to distinguish actions which are maximally free from those that are free to a non-zero degree (greater than zero but less than one), but I do not think that requires that there be two kinds of freedom. Instead, we could accommodate this thought with a second vague threshold between NP-CAPs which cause maximally free actions and NP-CAPs which cause actions which are free to a non-zero degree (greater than zero but less than one). For example, consider an NP-CAP with freedom-conferring properties to degree .999. We wouldn’t think that an action caused by this NP-CAP would be free to degree .999; instead, we would just think that the action is maximally free. It is worthwhile to note that on this kind of view any additional degree of freedom-conferring properties above this higher vague threshold contributes nothing more to the freeness of the action.

What determines these thresholds? This is another matter which indirect compatibilists will disagree about. Some theorists will think that what determines these thresholds is an objective fact regarding the nature of free action, but I think that is very unlikely. Instead, I think that what determines these thresholds is likely to be subjective and determined by our values and interests in free actions. To illustrate, consider a red-coloured tile. If we were to reduce the colour saturation of the tile, then at some point we might judge that the tile has changed colour to pink. But there is no objective fact regarding the nature of colour that determines when the tile changes from being red to pink; instead, it seems that our judgment about the colour of the tile is bound up with our colour interests. For example, if I play a game where the objective is to choose all the red objects, then even objects with very low red colour saturation might get to still count as being red. However, if I am home decorating and looking for red objects to decorate my home with, then only objects which have a degree of red colour saturation above a vague threshold might get to count as being red.[[12]](#footnote-13)

I think that something similar is likely true in the case of our free actions. In most contexts, the standards associated with free action might be low and so most of our actions get to count as being free or perhaps even maximally free, provided there is some non-trivial degree of freedom-conferring properties. However, in some contexts, such as when the stakes are high, the standards for free action might be significantly higher and rule out most of our actions being free, except for those with high degrees of freedom-conferring properties (cf. Lewis 1996 on knowledge). One example of a context where the stakes are raised is a moral context which helps to explain our divergent intuitions in Case 1 and Case 2 described above. While under ordinary circumstances the freedom-conferring properties associated with your NP-CAP in Case 1 might be enough to cause actions that count as maximally free, the same degree of freedom-conferring properties might only be enough to count as being free-to-a-degree when determining how blameworthy you are for being late to the work reception. In contrast the degree of freedom-conferring properties in Case 2 might be enough to cause actions that count as being maximally free in both an ordinary and moral context. The effect of this context shift on our free will judgments then predicts that we should judge that you are less blameworthy in Case 1 than Case 2. That’s because while we continue to judge that you are maximally free in Case 2 and so maximally blameworthy for being late to the work reception, we no longer judge that you are maximally free in Case 1. As a result, while we may still blame you for being late in Case 1 the degree of blame will be less than in Case 2.

What the indirect compatibilist will think here will depend on prior beliefs about what they take to be most important to free will. Of course, this is not exhaustive of all the ways indirect freedom might be related to freedom-conferring properties. But I think they might be two of the most plausible and promising accounts to develop in the future. I perhaps think that freedom comes in degrees, but which account you think is best will depend on your various intuitions about free actions, and for my purposes, I do not need to take a stand on which is preferable.

1. **Controlled Action Procedures and Moral Responsibility**

So far in this paper I have presented my account of free action: indirect compatibilism. According to indirect compatibilism we act freely either when our actions are caused by freedom-conferring mechanisms, or controlled action procedures that possesses *enough* freedom-conferring properties. The number of freedom-conferring properties an action procedure reflects the amount of influence that the agent has over it. I think that this account looks good because it is able vindicate the thought that most of our actions are free in a manner that also makes good sense of the current brain sciences research being performed on free action. If these are also good making features of the theory for you reader, then so much the better for indirect compatibilism.

Still some readers might not be entirely convinced, perhaps because they want to see how indirect compatibilism fares in other contexts. One reason we are so concerned about arriving at the correct account of free action and free will is because of its supposed tight conceptual connection with moral responsibility. By most theorists lights it is a *necessary* condition for our being morally responsibility that we have free will and act freely. Even better then for indirect compatibilism if it can help get clear and make sense of problematic moral responsibility cases.

The work of giving a complete account of moral responsibility in terms of indirect compatibilism is work for a different paper. Still, it would be useful to describe how the resources afforded by indirect compatibilism might be used to provide clarity for one kind of case: *problematic moral responsibility tracing*. Tracing is a feature of many accounts of moral responsibility. Even if someone is not strictly morally responsible for some action at the time of action, they nevertheless might be morally responsible overall because their moral responsibility can be traced back to an action for which they are morally responsible. The paradigm case of tracing in action is the case of causing an accident while drunk driving. Drunk agents are unlikely to possess the requisite control of their actions at the time of causing the accident to be morally responsible. Of course, this fact does not prevent us from holding them accountable. We think that they are morally responsible overall because they *did* possess the requisite control of their actions when they decided to get drunk. Moral responsibility for the accident can be traced back to moral responsibility for the drinking.

However, we do not just hold people morally responsible in the drunk driving case because the accident can be traced back to an action that they did have control over. But also, because people (are expected to) *foresee* that drinking negatively impacts driving ability and so accidents are a likely consequence. Unfortunately, in the case of most ordinary cases we do *not* foresee all the consequences that our present actions might have for our future actions. I will describe such a case shortly, for now here is the rough problem. In attempting to figure out whether some agent is morally responsible overall we might successfully trace back to an action for which someone is morally responsible, but which the agent could not have been expected to foresee would have the downstream consequences that it did.[[13]](#footnote-14) If such cases, we are seemingly faced with a difficult decision. Either we accept that the agents in such cases are not morally responsible overall and give up many of our ordinary moral responsibility judgments with it. Or, we need to revise how we think about foreseeability and tracing in the context of moral responsibility. The resources provided by indirect compatibilism though offer up another option.

To see that, consider for the moment the following case: Imagine an agent Bodi who has an action procedure that causes him to be unable to kill other people. I will discuss how such an action procedure might come about shortly. Bodi is a community police officer and one evening while he is patrolling, he notices that the traffic barrier on one of the bends has been broken. He pulls his patrol car over, holsters his gun and begins to investigate. Beyond the broken traffic barrier and down a bank he finds a rolled truck with a local driver still trapped inside. The driver is conscious and does not appear to be badly injured. While Bodi tries to free the trapped driver he quickly realizes that it will be impossible without assistance. Bodi begins conversing with the truck driver to keep him calm while they wait for assistance, but as they converse, a small fire catches at the back of the truck and begins spreading rapidly. The truck driver sees the approaching fire and, knowing he is about to die, turns to Bodi and begs him to not let him suffer needlessly by burning to death. Bodi breaks down, for he too knows the truck driver is going to die and does not want him to suffer by burning to death, but there is nothing he can do to prevent the truck driver from dying in this manner. Still, he realizes that if he could kill the truck driver then the driver would not have to suffer needlessly. In that moment Bodi wishes that he were able to kill other people.

Bodi wants to perform the action of killing to prevent a truck driver from burning to death. Unfortunately, Bodi possessed an action procedure that prevented him from killing people. Such an action procedure might do this by making Bodi *sincerely* unaware of the gun holstered on his side. Or it might prevent the killing action procedure from entering his deliberation pool. It might prevent Bodi from choosing the killing action procedure in his deliberation pool. Or it might filter the actions that are caused by the killing action procedure so that no matter what Bodi cannot kill the truck driver. Imagine now, that it was Bodi that freely created the controlled action procedure that prevents him from killing the truck driver. So, killing the truck driver was not an option that Bodi could freely choose to perform. This is not to say that Bodi could not freely choose to perform any number of other options, just that he is unable to perform the action that by his own lights would be best. In cases where an agent’s actions are the product of multiple controlled action procedures how should we decide whether someone is morally praiseworthy or blameworthy for their action?

While we might want to count Bodi as being overallmorally responsible for having done the right/wrong thing overall, prima facie it appears as though Bodi is in fact morally responsible for two different actions. First, Bodi is morally responsible with respect to the action of creating the action procedure that is causally constrains his future courses of action. Second, Bodi is morally responsible with respect to which of his remaining courses of action he freely chooses to perform. Thus, it really seems as though Bodi can count as being morally responsible for having done the right/wrong thing with respect to each of these things. The reason this complicates matters is because it is possible for our rightness and wrongness judgments to be incongruent across them both. For example, someone can do the right thing with respect to their action of creating one action procedure yet do the wrong thing with respect to another course of action, set in motion by other action procedures, that they freely deploy.

Following Braddon-Mitchell (2020) I think that whether someone counts as being morally responsible for having done the right thing or wrong thing overall depends on which of their actions we are focusing on and seeking to morally influence. I will not spend much time on this issue as it is does not speak directly to the central issue of this paper. However, at the very least it is of interest to sketch each possible combination very briefly and to see whether we are likely to think that someone is morally responsible for having done the right thing or wrong thing overall. The purpose of doing this is that by going through a few concrete examples, I hope to show how our ordinary moral judgments might be reconciled with indirect compatibilism. Further, that indirect compatibilism correctly predicts and explains those cases where our intuitions are less clear. Actions which are the product or multiple controlled actions procedures that can be apt targets for incongruent judgements. For ease of explication, I will stick with the case of Bodi.

Consider first the case of the doing right thing causally constraining your future options and choosing the right remaining course of action. This case is simply the original Bodi case presented earlier in this section. Bodi (arguably) does the right thing when he performs the action of creating the non-deliberative cause that removes the option of killing. While Bodi cannot shoot the truck driver he freely chooses to perform the best available remaining option which was to stay with the truck driver and comfort him the best that he can. Given the fact Bodi does the right thing with respect to both actions in this case we should have no issue judging that Bodi is morally responsible for having done the right thing overall.

Now consider a case of doing the wrong thing causally constraining your future options and choosing the wrong remaining course of action. Imagine, now, Bodi\* who is a cunning serial killer and is deciding how he should kill his next victim. Bodi\* freely created a repugnant (and laughably over simplified) action procedure that restricts his courses of action to either killing his victim slowly and painfully or killing his victim quickly and painlessly. The victim pleads with Bodi\* to let them go but this is not an available course of action because of the action procedure. As it turns out, Bodi\* freely chooses to perform the wrong available course of action and kills his unfortunate victim slowly and painfully. It should be uncontentious that Bodi\* does the wrong thing when he performs the action of creating the repugnant actions procedure. Similarly, it should be uncontentious that Bodi\* does the wrong thing when he freely chooses to perform the wrong available course of action, killing his victim slowly and painfully. Given the fact Bodi\* does the wrong thing with respect to both actions in this case we should have no issue judging that Bodi\* is morally responsible for having done the wrong thing overall.

What about doing the right thing causally constraining your future options and choosing the wrong remaining course of action? Let us go back to Bodi. Bodi (arguably) does the right thing when he creates the action procedure that removes the option of killing. But now imagine that instead of choosing to comfort the truck driver as best he can, Bodi runs away from the scene leaving the truck driver to die alone. In this case it seems as though Bodi chooses the wrong remaining course of action. While Bodi does the right thing with respect to the creation of the action procedure, he does the wrong thing with respect to which of the remaining courses of action he freely chooses to perform. I think in this case many people (including myself) focus on the fact that Bodi chooses the wrong remaining course of action. As a result, we judge that Bodi is morally responsible for having done the wrong thing overall.

Finally, the wrong thing causally constraining your future options and choosing the right remaining course of action? Again, Bodi\* does the wrong thing when he creates his repugnant action procedure. However, now imagine when the victim pleads with Bodi\* to let them go, and Bodi\* freely chooses to do the right thing by his lights and kills them quickly and painlessly. Remember that the only other course of action available to Bodi\* is to kill his victim slowly and painfully. While Bodi\* does the wrong thing with respect to creating the repugnant action procedure, he seemingly does the right thing with respect to which remaining course of action he freely chooses to perform. I think in this case many people (including myself) focus on the fact that Bodi\* does the wrong thing when he creates the repugnant action procedure. The reason Bodi\*’s act of murdering someone quickly and painlessly turns out to be right simpliciter is because his repugnant action procedure causally constrained his future choice options to just those with murderous outcomes. Thus, we judge Bodi\* is morally responsible for having done the wrong thing overall.

1. **Conclusion**

In this paper I have introduced a new compatibilist account of free action: indirect compatibilism. On this view actions are free either when they caused by freedom-conferring mechanism, or else by sub-personal level processes influenced in various ways by freedom-conferring mechanisms. Freedom-conferring mechanisms are those conscious psychological processes that cause free actions. This view is motivated by a challenge posed by the brain sciences to free actions, and by extension, to free will. The challenge is that evidence from the brain sciences shows that our freedom-conferring mechanisms cause no (or very few) actions, and so none (or very few) of our actions are free. I argued that we can accommodate many of our free action judgments by acknowledging that most cases we judge to be free are not free because they are caused directly by freedom-conferring mechanisms, but because they are caused indirectly via action procedures which those freedom-conferring mechanisms influenced in some way. Indirect compatibilism also helps predict and explain the cases where our moral responsibility intuitions are less fixed and clear. If I am right that actions can be indirectly free, then most of the cases we judge to be free are vindicated, and if I am right about indirect compatibilism, then people have free will.

**References**

Anderson, M. C., and Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*. 410: 366–369.

Angus, D. J., Latham, A. J., Harmon-Jones, E., Deliano, M., Balleine, B., and Braddon-Mitchell, D. (2017). Electrocortical Components of Anticipation and Consumption in a Monetary Incentive Delay Task. *Psychophysiology*, 54(11): 1686-1705.

Banks, W. P., and Isham, E. A. (2009). We infer rather than perceive the moment we decided to act. *Psychological Science*. 20: 17–21.

Bode, S., He, A. H., Soon, C. S., Trampel, R., Turner, R., and Haynes, J. D. (2011). Tracking the unconscious generation of free decisions using ultra-high field fMRI. *PLoS One*. 6: e21612.

Bode, S., Sewell, D. K., Lilburn, S., Forte, J. D., Smith, P. L., and Stahl, J. (2012). Predicting perceptual decision biases from early brain activity. *Journal of Neuroscience*. 32: 12488–12498.

Braddon-Mitchell, D. (manuscript). Freedom and binding consequentialism.

Caruso, G. D. (2012) *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Lexington Books.

Fischer, J. M., and Ravizza, M. (1998). *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Fischer, J. M., and Tognazzini, N. A. (2009). The Truth about Tracing. *Noûs*. 43: 531-556.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*. 68: 5–20.

Gazzaniga, M. S. (2011). *Who’s in Charge? Free Will and the Science of the Brain*. New York: Ecco.

Greene, J., and Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society B*. 359: 1775–1785.

Hartman, R. J. (2020) Indirectly Free Actions, Libertarianism, and Resultant Moral Luck. *Erkenntnis*. 85: 1417–1436.

Hobbes, T. (1654/1999). *Of Liberty and Necessity*, in Hobbes and Bramhall on Liberty and Necessity, ed. Vere Chappell, Cambridge: Cambridge University Press, 15–42.

Hume, D. (1740/1978). *A Treatise of Human Nature*, P.H. Nidditch (ed.), Oxford: Clarendon Press.

Latham, A. J., (2019). The Conceptual Impossibility of Free Will Error Theory. *European Journal of Analytic Philosophy* Special Issue, Free Will and Epistemology: Transcendental Arguments for Freedom, Bernáth, L., O’Connor, T., & Szigeti, A. (eds.), 15(2): 99-120.

Lau, H. C., Rogers, R. D., and Passingham, R. E. (2007). Manipulating the Experienced Onset of Intention after Action Execution. *Journal of Cognitive Neuroscience*. 19: 81–90.

Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy*. 74(4): 549-567.

Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (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: 623–642.

McKenna, M. (2008). Putting the Lie on the Control Condition for Moral Responsibility. *Philosophical Studies*. 139: 29–37.

Mele, A. R. (2014). *Free: Why Science Hasn’t Disproved Free Will*. New York: Oxford University Press.

Miller, J., Shepherdson, P., and Trevena, J. (2011). Effects of clock monitoring on electroencephalographic activity: is unconscious movement initiation an artifact of the clock? *Psychological Science*. 22: 103–109.

Mylopoulos, M. I., and Lau, H. (2014). Naturalizing Free Will. In A. R. Mele (ed.), *Surrounding Free Will: Philosophy, Psychology, Neuroscience*. Oxford University Press.

Nahmias, E. (2002) When consciousness matters: A critical review of Daniel Wegner’s *The Illusion of Conscious Will*. *Philosophical Psychology*. 15: 527–541.

Roskies, A. L. (2010). How does neuroscience affect our conception of volition? *Annual Review of Neuroscience*. 33: 109–130.

Ryckman, N. A., Addis, D. R., Latham, A. J., and Lambert, A. J. (2018). Forget About the Future: Effects of Thought Suppression on Memory for Imaginary Emotional Episodes. *Cognition and Emotion,* 32(1): 200-206.

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., Blankertz, B., and Haynes, J. D. (2016). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences of the United States of America*. 113: 1080–1085.

Schurger, A., Sitt, J. D., and Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences of the United States of America*. 109: 2904–2913.

Smith, K. (2011). Neuroscience vs philosophy: Taking aim at free will. *Nature*. 477: 23–25.

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

Soon, C. S., He, A. H., Bode, S., and Haynes, J. D. (2013). Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences of the United States of America*. 110: 6217–6222.

Tallis, R. (2011). *Aping mankind, neuromania, Darwinist and the misrepresentation of humanity*. Acumen: Durham.

Vargas, M. (2005). *The Trouble With Tracing*. Midwest Studies in Philosophy, 29: 269–291.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

West, R. (2011). The temporal dynamics of prospective memory: A review of the ERP and prospective memory literature. *Neuropsychologia*. 49(8): 2233–2245.

West, R., and Moore, K. (2002). Adjustments of Cognitive Control in Younger and Older Adults. *Cortex*. 41: 570–581.

1. Of course, such accounts are (by most) best thought of as accounts of what is *necessary* for free action and free will and not what is *sufficient*. It is also necessary that there be no objectionable defeaters of an alarming kind. For instance, we might discover that all the inputs into our ordinary psychology are due to the machinations of an evil demon. I think that many people would think that such a discovery would also be the discovery that we do not act freely or have free will (though for arguments contra see Latham, 2019). I will not take a stance on this matter in this paper. [↑](#footnote-ref-2)
2. See, for example, Hobbes (1654/1999), Hume (1739/1978), Frankfurt (1971), Wallace (1994), Fischer and Ravizza (1998) for some compatibilist accounts of the sufficient conditions for free will and moral responsibility. [↑](#footnote-ref-3)
3. Hartman (2020) has recently developed a possible libertarian account of indirect freedom. An agent is indirectly free iff she performs a libertarian free action in the past such that (i) she believes that performing libertarian free action could cause her to perform similar kinds of actions in the future, (ii) her disposition to now perform those actions counterfactually depends on her libertarian free action in the past, and (iii) there is adequate similarity between her libertarian free action in the past and the actions she is disposed to perform. I agree with Hartman at the highest level of abstraction: there are directly and indirectly free actions, and indirectly free actions are dependent on the directly free ones. However, I disagree with Hartman on what the correct characterization of directly and indirectly free actions is, and what the relevant dependence relation is between them. [↑](#footnote-ref-4)
4. For some more general critiques see also Roskies (2010), Smith (2011), and Mele (2014). [↑](#footnote-ref-5)
5. It is worthwhile to note that I think that action procedures are sub-personal level mechanisms that are defined functionally as whatever those things are which cause actions in a contained and reliable fashion, and when they have been influenced by freedom-conferring mechanisms cause actions which are indirectly free according to indirect compatibilism. While action procedures must have some kind of neural implementation in us, whether it is a unified and systematic implementation such that there could be an independent, neural level theory of action procedures in cognitive sciences is a question on which I remain neutral. Of course, in very different kinds of entities, such procedures will be implemented by very different processes, possibly not even neural ones. I also remain neutral on whether the psychological and sub-personal are reallevels in the world or simply explanatory stances that we can take towards agents. [↑](#footnote-ref-6)
6. See also the experimental literature involving prospective memory tasks (e.g., West, 2011; West & Moore, 2002). In prospective memory tasks people consciously encode some intention in response to some task instruction to be performed at later times. Interestingly, Angus et al. (2017) found that the electrocortical component associated with the conscious encoding of intentions is also present when people are consciously monitoring and updating their performance in response to environmental feedback. [↑](#footnote-ref-7)
7. Some readers might be concerned by another skeptical worry suggested by Libet-style experiments. The worry is that even if freedom-conferring mechanisms do influence action procedures they are not freedom conferring because they originate from other non-conscious neural processes. But if determinism is true, then this is just what we should expect to find, and so it should not be surprising or concerning to the compatibilist (at least). What matters to compatibilists is whether our freedom-conferring mechanisms do the things we suppose they do as part of the causal chain involving our actions, and this is consistent with them being caused by non-conscious neural processes (see Nahmias, 2002). [↑](#footnote-ref-8)
8. A referee asks whether we even have the capacity to destroy an action procedure. For instance, it is difficult to imagine someone voluntarily destroying the action procedures that underlie their expert performance (i.e., LeBron James destroying his basketball talent). Further, even if we have the capacity to destroy an action procedure, we don’t appear to use it. Instead, we ‘destroy’ action procedures by simply not using them. I think it is an open question whether we have the capacity to destroy an action procedure, and that it is plausible that we do not. However, there is suggestive, and accumulating, evidence that we have the capacity to intentionally forget learned word-pair associations and perhaps more impressively, certain autobiographical memories (e.g., Anderson and Green, 2001; Ryckman, Addis, Latham, and Lambert, 2018). While the capacity to intentionally forget memories is undoubtedly controversial, much of this controversy is over the capacity to forget particularly traumatic or highly valanced memories. It is perhaps less controversial to think some prosaic process might exist that allows for the targeted removal of certain memories and certain action procedures as well. However, even if we have the capacity to destroy an action procedure, I think that the referee is right that in most cases we just ‘destroy’ actions procedures over time by not using them. [↑](#footnote-ref-9)
9. A helpful referee points out that there seems to be a stark difference between case in which I indirectly-freely perform the punished action and directly-freely perform the punished action. I agree with the referee and will return to discuss this difference at the end of Section 5. For now, I think that performing the punished action indirectly-freely will be associated with a lower degree of freedom-conferring properties. And, whether the degree of freedom-conferring properties is enough to count as being maximally free or free to a non-zero degree (greater than zero but less than one), depends on context. In ordinary contexts performing an action indirectly-freely might be enough to count as be maximally free. However, when the stakes are higher, such as when consider matters of responsibility, it might only be enough to count as being free to a non-zero degree (greater than zero but less than one) and so we count as being less responsible as well. [↑](#footnote-ref-10)
10. Note that in this case the easier it is to intervene in an action procedure the better. This is not always true: sometimes the point of action procedures is to commit ones future self to certain courses of action. Such a case will be considered in Section 6. [↑](#footnote-ref-11)
11. Libertarian readers can read this section as describing the *necessary* conditions for indirect freedom and read in their preferred account of libertarian powers to turn it into a libertarian account of indirect freedom. [↑](#footnote-ref-12)
12. Thanks to David Braddon-Mitchell for suggesting to me the colour saturation case. [↑](#footnote-ref-13)
13. For example, see Vargas (2005) and McKenna (2008) who have described several cases like this and have argued that they cause problems for tracing moral responsibility. Though see also Fischer and Tognazzini (2009) for a defense of tracing against these kinds of cases. [↑](#footnote-ref-14)