THE IMPOSSIBILITY OF SOURCE FREEDOM ON DETERMINISM

By

JACK DEEMING

_____

A Thesis Submitted to The Honors College

In Partial Fulfillment of a Bachelor's Degree
With Honors in

Philosophy

THE UNIVERSITY OF ARIZONA

M A Y 2 0 1 7

Approved by:

_____

Dr. Michael McKenna
Department of Philosophy

Abstract:

Free will can be understood as a particular type of control exercised by an agent over her actions. This type of control requires both the existence of genuine alternatives and a degree of authorship regarding the mental states that motivate action. This paper argues that neither of these criteria can be met if metaphysical determinism obtains. Free choice, however, may still exist just so long as an agent can willfully generate action without constraint or outside coercion. It is this freedom of choice which is required for moral responsibility.

Free will has been a perpetually thorny puzzle for Western philosophy. Is there or isn't there? Can machines have it? How about a God? And what does determinism have to do with any of it? To answer these questions, a steadfast definition of terms seems to be in order. Even here (and perhaps primarily here) exists immense controversy. That admitted, let us characterize a few positions necessary in getting the discussion off the ground. Especially helpful in conceptualizing the various positions in the philosophy of free will is Alfred Mele's gas station analogy. Mele has stated in various forums that definitions of freedom roughly fall into three categories comparable to regular, mid-grade, and premium gasoline.

"Regular", or compatibilist-friendly freedom describes an agent's ability to make real time choices according to her psychological state without being forced by outside influences, without a gun to her head and without a neuroscientist's microchip redirecting her brain. It is mostly agreed upon that these considerations are indeed necessary to freedom. But are they sufficient? Some philosophers prefer to set the bar higher before proclaiming genuine freedom. This (roughly libertarian) view comes in two varieties. Opting for the first of these mid-grade options, some posit that in addition to the absence of coercion, freedom further consists in having the ability to do otherwise, and thus the existence of bona fide alternatives. In the second case of this mid-grade classification, there is the view that the further ingredient necessary for freedom is not the ability to do otherwise as such, but rather that the agent's psychological state constitutes the source of her action in some special way. A sizeable minority of commentators go even further, advocating even more stringent conditions for freedom. This skeptical, "premium" position insists that freedom only obtains in the case that all of the above conditions are met, and also that an immortal soul exists which allows the agent to be the

ultimate cause of her action, a prime mover unto herself. To illustrate these options, please see fig. 1.

Fig. 1

A. Freedom from coercion … "Regular"

i. A + could have done otherwise … "Mid-grade"

ii. A + genuine source-hood … "Mid-grade"

1. A + i + ii + immortal soul, ultimate cause … "Premium"

It is my view that because of the warring definitions of freedom being advocated, the differing voices in the literature on the topic are often talking past each other, or otherwise engaged in linguistic squabbles over whose definition is best. In this paper, I will offer my analysis regarding some of these alternatives, examine which ones may obtain, and offer a conceptualization that I find helpful. In so doing, I will pay special attention to source freedom (ii) because I believe the debate is murkiest at this point, and I will ultimately propose an argument for its impossibility. But first, a distinction between choice and will. I take it that questions involving compatibilist-friendly regular grade freedom are uncontroversial in some respect. Nearly everyone seems to agree that we do in fact make choices in real time about whether to go right or left for instance, or to choose between chocolate or vanilla ice cream. Free choice then, for the current purposes, can be understood as this ability to choose or to avoid outcomes in accordance with mental states, since it comes down to un-coerced decision-making. Choice, in this capacity, denotes how the will is realized after it has already been formed. (And perhaps, this is sufficient for any freedom worth wanting, but that is another question entirely.) In my accounting of free will, the further demands of "middle grade" and

"premium freedom" are where questions of the origin of the will come into view. These are concerned with how the will is acquired or produced. Thus, free will can be understood as referring to whether these mid-grade definitions are satisfied in reality. In the latter half of the 20th century, the philosophical debate surrounding the freedom of the will has centered around controversy regarding "middle grade" conceptions. The literature on this topic diverges into the two forks listed above: as an agent's ability to do otherwise, or an agent's exercising a degree of control over the action she performs. On the former, the consequence argument as laid out by Van Inwagen and others has, assuming its soundness, rendered a binary result. The leeway freedom it references either exists in its full metaphysical glory or not at all. There are no shades of gray, no questions of degree. If successful, the consequence argument unequivocally vanquishes this brand of freedom.

To fully appreciate the effectiveness of the consequence argument, one must first notice how it functions in the broader philosophical discussion. Although convincing as a standalone, the consequence argument is best understood as a proof for a premise in a more fundamental incompatibilist argument. The Basic argument for Leeway Incompatibilism (BLI) follows:

1. If an agent acts according to her own free will, then she could have done otherwise. (per i)
2. If determinism is true, then no one can do otherwise than they actually do.
3. Therefore if determinism is true, no one acts according to their own free will.

For quite a stretch of time, the disagreement over free will came to a standstill over this line of reasoning. Incompatibilists endorsed into it; compatibilists denied it. And opposing

considerations for each point of view were offered on a level playing field. Unfortunately, there seemed to be little hope of progress, because the crucial second premise appeared to be up for grabs. This remained the case until the consequence argument burst onto the scene, underlining the incompatibilist assertion of premise two with a powerful argument and thereby breathing new life into an extended stalemate.

The consequence argument rests on the notion that if determinism is true, then our actions are entailed as products of the past and the laws of nature. Since we have no choice about which past events occurred or what the laws of nature happen to be, we likewise have no choice about the consequences of those facts. Although there are more technical formulations of the argument in circulation, the following will be sufficient to for this discussion:

1. No one has power (nor could anyone ever have any power) to alter the facts of the distant past and the laws of nature.
2. No one has power (nor could anyone ever have any power) to alter the fact that the distant past and the laws of nature unavoidably entail one specific future.
3. Therefore no one has power (nor could anyone ever have any power) alter the facts of the future. (I.e. No one has the power to do otherwise.)

This argument appeared to give incompatibilists an upper hand in the dialogue, but the compatibilists were not going to be outdone so easily. In the literature generated around the consequence argument, compatibilists have attempted to mount a counterattack on several fronts including assumptions about the past, the laws of nature, and the transfer principle (The principle that allows one to conclude that *q* is unavoidable for an agent assuming it is

unavoidable for her that *p* obtains and that *p* implies *q*.). Further reiterations of the argument eventually came onto the scene along with even more objections. Whether these rebuttals are at all effective could be a volume unto itself, but the consequence argument (or some version of it) has withstood them all in the eyes of some commentators. Indeed, many incompatibilists today premise their view on the basis that it is in some way sound. What is important to keep in mind for the purposes of the present discussion, an uncontroversial point, is that the ability to do otherwise either obtains or it does not with no in-between state. And it certainly does not, if the consequence argument succeeds.

This binary aspect of the debate may stand in stark contrast to the other construal of freedom on offer. With respect to the "middle-grade" conception of source freedom, the case appears to be quite the opposite. This conception of freedom, although subject to various descriptions throughout the literature, appears always to turn on control, volition, and the like, where the agent is playing some necessary part in the causal chain. In essence, an agent has source freedom if she can exercise control over her actions. To distill this concept further, let us assume a simple working definition of control as something like this: Control obtains just in case an agent's decision is the primary causal determinant of her actions. There is obviously much complexity concealed in the phrase, "primary causal determinant". At the risk of trespassing into the deeply controversial debates surrounding the metaphysics of causality and the philosophy of action, let us assume that "primary causal" just means that the cause is the crucial or most operative factor with respect to producing the effect. Let a determinant be an event that has the capacity such that when it is altered, a later event (the effect in question) is altered in accordance with it. To illustrate, take the example of Smith, who is holding a gun to

the head of Jones, intending to kill him. Acting according to his intention to kill Jones, Smith pulls the trigger resulting in Jones' death. Now a question: what was the primary causal determinant of Jones' death? Although there are many events that were necessary such that Smith's murdering Jones could occur, Smith's decision to act on his intention feels as though it performs special role in modulating the unfolding of events in this causal story. Smith's decision is implicated as the primary causal determinant presumably because it stands in relation to Jones' death in a way that the big bang, Smith's surviving a childhood illness, and even the ignition inside the cartridge of Smith's gun do not. These other events are necessary to the outcome of Smith successfully murdering Jones, no doubt, but Smith's decision may have a distinct quality in that it seems to be crucial in determining the outcome while the other events listed lack the same apparent salience. Predictably, digging into the specifics of this distinct quality is a tricky endeavor, as we shall see. But for now, with this cursory understanding of control in hand, we can ask what function the concept might serve in the framework of the free will debate.

Well for one, this approach seems a more promising conception for the compatibilist precisely because it lacks the all or nothing wager of its leeway counterpart. Control can presumably occur in degrees and often seems to in many cases. For instance, common sense reports that I have no control over the fact that my eyes are blue, that I have some control over whether I score well on an exam, and that I have an even greater degree of control over whether I will pick up the cup in front of me with my left hand or my right. To rephrase, some versions of control feel more robust than others. If this intuition is taken as veridical, then apparently the incompatibilist is at a disadvantage, because irrespective of whether

determinism is true, our everyday experience screams that we possess some modicum of self-control. In other words, our sense of agency, be it real or illusory, lines up with the compatibilist on this point. And once the admission is made that any directly effective causal relationships of any degree exist linking agents and actions, the natural compatibilist tendency will be to ask of those relationships, "Isn't that just what it means to be free?".

At this point, the incompatibilist seems to be left in the unenviable position of identifying an arbitrary amount of control that supposedly constitutes legitimate freedom and then arguing that on determinism, said standard is unattainable. It is unclear how they would recover the high ground in the discussion thereafter. But what if the sense of control employed above could be shown to be illusory? What if our intuitions are lying to us on this point (as they are often wont to do) and that the control necessary for source freedom is actually like the ability to do otherwise, either existing or not existing without any gradations? What if certain metaphysical propositions could level the playing field in this way? Now that would certainly alter the stakes, but is this a possibility that obtains?

Before we move toward an answer, we will need to get clear on the notion of agency as well. For our purposes, an agent is anyone who acts, although there a plenty of nuanced views of agency to be found in the literature. (We will flesh out this conception later with regard to agential action.) Now, in questioning control as it relates to agency, no one is arguing that individuals don't make real time decisions, that they lack free choice. Take it as given that they do. Rather, the point of contention concerns the reasons why a person exercises their causally effective powers in the first place and whether that's sufficient for control or ultimately, freedom with respect to those decisions. In other words, if an agent's causally effective powers

can only express themselves in a single outcome predetermined by the past and the laws of nature, one might do well to question the effectiveness of those powers.

To press on this notion of supposed control more poignantly, a popular manipulation example is instructive: a malevolently ambitious neuroscientist implants a chip in a patient's brain to induce a mental state including desires, beliefs, intentions, etc. (*DBI*) such that the patient would choose to perform a given action (*A*). For the sake of argument let us suppose that our scientist perfectly replicates the causes of action that a compatibilist would label as free. When the patient inevitably performs (A), one might inquire whether the action was at all controlled by the patient in such a way that it could reasonably be called free. I submit that the answer is no, even supposing she had no knowledge of the chip in her brain and sincerely insisted that she generated the relevant *DBI*. The reality, however, would be that she lacks control precisely because she had no part in producing *DBI.* Nor, the incompatibilist will argue, do any of us have any part in producing our *DBI* in a deterministic universe. Like the example, finite beings beholden to the natural order don't elect their mental states - so how can they control the direct results of those mental states? To rephrase, if our actions are analogous in all relevant respects to this brief manipulation parable, whence cometh the control requisite to source freedom?

The crucial point raised by manipulation examples then, is that in order to have control, an agent must be to some extent the self-generative source of their DBI and more broadly their entire psychological structure, the mental state that frames these dispositions. But is this possible? Assuming determinism, agents cannot be self-generative in this way. They are merely parts to a whole, that is the universe, which by definition can only unfold exactly as it does

according to the past and the laws of nature. Along the same line of reasoning, would this not entail that the effective powers we interpret as control are not self-generated and cannot ever be for finite creatures? Furthermore, could this proposition be evidenced conclusively by cold, hard, logic? Well the idea is hardly new. Such a proof would follow from the success of an argument that turns on the concept of ultimacy, versions of which have been popularized by Galen Strawson and offered by many others before him. Here is my reiteration, the Argument Against Control (AAC):

1. We only act as we act in any given instance according to our specific psychological state (including specific DBI) at that time.
2. In order to have control over our actions, we must have control over the psychological states that cause our actions.
3. We have no control over our psychological states at any given time.
4. Therefore, we have no control over the way we act.

This argument makes use of three concepts, which need to be clarified at the outset. The first concept to note in this argument is agential action, which we shall define as goal directed execution of intent that is at least potentially morally relevant. This is essentially a refinement of our working definition of an agent as one who acts, by linking that concept and identifying a special relationship to the sort of behaviors that matter in the context of the free will debate. Second, is an agent's psychological state. I take this to include all aspects of the agent's current mental functioning encompassing but not limited to DBI, including thoughts, memories,

predispositions, neurocognitive capacities, and the like – really anything that plays a role in generating intentional action. Finally, recall from above that control, as referenced here, can be interpreted as a directly effective, causal relationship beginning with the agent.

Once the terms are agreed upon, the discourse can begin to take shape. The argument seems solid at a glance, and it is logically valid. But are the premises true? Critics have produced several objections against them that perhaps successfully attempt to preserve the notion of source freedom. At the very least these criticisms, if they have the intended effect, caution strongly against the declaration of its demise. Notably among these philosophers, Randolph Clarke calls into question the argument's effectiveness in his paper, "On the Impossibility of Moral Responsibility". Therein he addresses the most popular version of the argument, Strawson's formulation that seeks to disprove moral responsibility. Strawson uses the same logic outlined above, replacing the concept of control that I emphasize with moral responsibility. (I chose to focus on control in this paper, because it is necessary to the source freedom essential to many compatibilist accounts of action unlike moral responsibility which is *arguably* separable from this freedom.) Clarke points out an implicit assumption lurking in the background of the first premise: that we act as we act solely due to our psychological state at that time. To him, it's an open question, one that must be answered before we can accept Strawson's argument or the *AAC*. One might pursue this by pointing to numerous instances of individuals acting apparently in discord with their character, their nature, and thus presumably their psychology. Does this not reflect a degree of control? Or is it merely a psychological state manifesting itself in an unexpected yet in principle predictable fashion? I will argue conditionally for the latter.

First, let me address whether it is possible to have an action as I have defined it which is not the direct result of a psychological state. No, in short. Now it is clear that not all physical movement requires intent, but that's hardly the point. The volitional actions relevant to the free will debate are aimed at the execution of certain ends based on the DBI of the agent. Any action therefore necessarily incorporates intent, which is a psychological state. Moreover, this definition stipulates a causal connection beginning with intention and ending in action. Please notice that it adds nothing toward the case for control even if the relationship is incidental or probabilistic. In that case, the degree of control may be ratcheted down, but it's the existence of *any* degree of control in the first place which is on offer. An unruly theory of mental causation is strictly irrelevant to this point. Action must be the result of exercising some particular agential capacity for any conversation of libertarian or compatibilist freedom to even get off the ground. In other words, that actions are the result of mental states is necessary to any conception of free will. Any affirmative view in the source freedom debate therefore require free actions to be the necessarily connected to mental states. There is no freedom to be had if action and mentality only line up incidentally on occasion, though this would not bother a free will skeptic in the least. A more striking question may be: Do mental states directly, causally generate the actions they produce or are they merely reliable predictors?

To develop this objection, consider Van Inwagen's instructive case, which Clarke includes in his discussion. In it we are faced with a thief who intends to rob the poor box of a church. But when he lifts the lid he is confronted with the memory of his mother, who on her deathbed beseeched him, making him promise to live an honest life. Upon the recollection of this promise, he closes the lid and goes along his way. So what are we to make of the thief's

decision not to rob the poor box? What's specifically under scrutiny here is the following chain of events: First, the thief experiences the psychological state of desiring to rob the poor box. Then he experienced another psychological state of remembering his promise, which led to another psychological state, that of not wanting to rob the poor box. It was this mental state that was translated into action. Clarke is basically asking the Strawsonian, "How do you know that there is not a decision tree embedded in this parable that allows that the thief's psychological state may lead to the decision to rob the poor box *or* that it may not?" This question can best understood when broken down into two questions, that present two possible locations for this decision tree. On the one hand, what factors modulated the thief's mental state allowing for the transition from intention to memory to a separate intention? Maybe there is a break in the causal chain at the intersection between mental events for the agent to inhabit. On the other hand, what factors determined which psychological state would be translated into action? Unless the agent can be barred completely from having an effect anywhere in this chain of events in either case, then source freedom stands a fighting chance. The mere possibility must be negated for the success of Strawson's argument against moral responsibility and the *AAC;* Clarke's rejected alternative seems to leave space for libertarian or free will unless there is no such alternative on offer.

Re-enter our old friend, the consequence argument. On indeterminism, Clarke's objection cannot be disproven. The randomness, or the probabilistic nature of the universe as the case may be, allows for at least the possibility that the agent plays a crucial role. Perhaps with the exception of ruling out the existence of personal agents, there could in principle be an intervention in the causal chain at the transition between mental states or the decision

regarding which mental states are translated into action. By contrast, determinism allows for no such interruptions. It locks in that chain of events, because by definition, there is only one way for the universe to unfold and it unavoidably must unfold in that way. If the consequence argument works and determinism turns out to be true, this conclusively rules out the alternative that Clarke offers since it rests on the existence of a decision tree.

So can any other model of source freedom succeed in a deterministic universe where Clarke's objection fails? Reasons-responsiveness theories such as the one offered by Fischer and Ravizza attempt to do exactly that, thereby restoring a version of source freedom. They set the standard somewhat lower than "middle grade" when defining control, namely, that it obtains when an agent is capable of evaluating and rationally responding to reasons for and against the course of action in question. Reasons-responsive theories are attractive accounts of moral agency because they emphasize what is most important about freedom to the average person, the weighing of options and the production of a real-time decision based on those internal calculations. If this definition sounds familiar, it should. It is essentially one and the same as free choice, as defined above. This intuition could be verified empirically, and perhaps experimental philosophy will soon perform its due diligence on this point and acquire the actual sociological data, discovering what matters to the majority of people in an account of freedom. But even if it does not, it is hard to imagine that the typical conception of freedom outside the academy includes universal metaphysical propositions or elaborate explanations of mental causation. No; a person is considered free in society if they were capable of appreciating, rationally appraising, and eventually acting according to reasons. In other words, free choice is sufficient for moral responsibility. No free will is required on this view. Fair enough, a critic may

retort; this may be a solid appraisal of what societies look for when assigning praise or blame, but does it constitute genuine control?

No. At least not at all with respect to the notion of the agent being the primary causal determinant of her actions. That is why it might be useful to consider the controversial distinction on offer. Assuming that moral responsibility and "middle grade" freedom are separable for the moment. This would mean that moral responsibility is at bottom a societal convention, a claim about how reliably a person will act in a certain way and whether they were constrained by extenuating circumstances such as a violent coercion or a brain tumor. It is about behavioral proclivities, the likelihood of recidivism, and the practical need to categorize our social interactions to the benefit of the greater human interest. Moral responsibility on this view, can function regardless of the exact nature of causality and is unaffected by determinism. Freedom then, by contrast, is a metaphysical claim involving "middle grade" concerns that are at least influenced by the truth or falsehood of determinism, if not unraveled by them. Perhaps freedom, properly understood, is identified in a set of facts about how the universe operates when it coalesces into the miraculous little centers of agential unity we call human beings. Free will posits a difference in the how the world is, a difference that makes a difference. In order to be free in the truest sense, we would need a universe in which agential outcomes are genuinely up for grabs. These outcomes would need to be affected by their primary causal determinants such as to take a special turn in how they unfold, a turn at the locus of the individual will. To summarize, moral responsibility is a convenience - an important one, a vital one – a convenience we rightly base our lives upon. The gravity of this fact is not diminished in the slightest even if the distinction is ultimately nominal. Free will is a deeper metaphysical claim

about the nature of reality, causality, and human volition. My contention is that it is at least possible to separate the two, that there are possible worlds where agents are not free but are still morally responsible and vice versa.

Assuming the truth of this highly contentious differentiation, we can view with greater clarity the crux of the issue regarding compatibilist accounts of freedom such as reasons-responsiveness theories. These accounts are excellent for assessing moral responsibility as defined above, but fail to address any difference that makes a difference in the causal chain, making them insufficient for free will. In essence, reasons-responsive theories save the control necessary to a robust source freedom by watering down the concept until it's unrecognizable, something quite apart from "middle grade" considerations. Such theories do this by removing the idea of the agent playing any genuine role in determining the chain of causality. Naturally, compatibilists will contest the phrase "genuine role" in the preceding, but these complaints can only register if the notion of determination can be applied equally to reasons-responsive theories of freedom and their incompatibilist counterparts alike.

So where are we with respect to the dual notions of freedom we looked at from the start? Leeway freedom fails if any version of the consequence argument succeeds and must be evaluated on its own terms. Source freedom is a more complex case but is simplified in the light of a careful analysis of control. Control construed as primary casual determination requires options, and the primacy of the agent's decision point in generating which option obtains. Crucially, those options are only available in an undetermined universe. Source freedom requires the agent to be a special factor, a determining factor in directing the unfolding of events. On indeterminism, we can look at the case of Smith shooting Jones and see that the

causal uncertainty in the unfolding of events was laid to rest in Smith's decision. The authentic uncertainty of events up until the moment Smith decides to act on his intention provides an account of the distinctly crucial quality described when we first outlined this example. On determinism, there never was any uncertainty, metaphysically speaking, to implicate Smith's decision as controlling the outcome in any way. Indeed, Smith's decision is merely another link in the chain of events in such a universe, another hash mark in the timeline undistinguishable from any other. As noted earlier, source freedom appears to occur in degrees, and these degrees are only made possible and actual in light of the open future offered by indeterminism.

In light of this, the AAC comes back to life, if only in within the context of a deterministic universe, stipulating the impossibility of control by denying source-hood. In particular, it would deny the degrees of instantiation that we would require in Smith's decision to kill Jones in order to call it a true exercise in control and thereby free will. But as we have seen, and in contrast to Strawson's original formulation, the AAC must take into account a rigidly determined metaphysics. With that in view, consider the following:

5. If determinism is true, we only act as we act in any given instance according to our specific psychological state (including specific DBI) at that time.
6. In order to have control over our actions, we must have control over the psychological states that cause our actions, meaning we must control either the transition as one psychological state is replaced by the next, or the process by which some psychological states are translated into our actions but not others. (Or both.)

7. But assuming determinism, we control neither the transition nor the translation of psychological states.
8. Therefore, on determinism we have no control over our psychological state at any given time.
9. Therefore, on determinism we have no control over the way we act.
10. Therefore, on determinism we have no source freedom.

If successful, this argument pulls together the threads of this paper to make the case for source incompatibilism. And there are certainly many controversial steps contained in these premises. But assuming some basic definitions of control and freedom, as well as the accounting of mentation and action as laid out above, then the reasoning has a chance to go through, possibly rendering source freedom and determinism mutually exclusive conditions.

Bibliography

Clarke, Randolph. “On the Impossibility of Moral Responsibility”. *Midwest Studies in Philosophy*. 2005.

Fischer, John Martin. Ravizza. Mark. “An Essay on Moral Responsibility”. Cambridge University Press. 1998.

Kane, Robert. *The Significance of Free Will*. Oxford University Press. 1996.

McKenna, Michael. “Ultimacy and Sweet Jane”. *Essays on Free Will and Moral Responsibility*. 2008.

Mele, Alfred. *Free Will: Why Science Hasn’t Disproved Free Will*. Oxford University Press. 2014.

Strawson, Galen. “The Impossibility of Moral Responsibility”. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. 1994

Van Inwagen, Peter. “An Essay on Free Will”. Oxford University Press. 1983