

Freedom, Determinism and Gale's Principle

Alexander R. Pruss

Abstract

In simplified form, the argument that I am defending holds that the incompatibility of our freedom with determinism follows from the conjunction of (1) a plausible supervenience claim which says that whether a human agent is free depends only on what happens during the agent's life and (2) a freedom-cancellation principle of Richard Gale which says that an agent is not free if all of her actions are intentionally brought about by another agent. Improved versions of (1) and (2) are also considered.

1. Introduction

I shall show that if determinism holds, Napoleon never performed a free action in his life. The argument generalizes to all of us and makes use of two crucial assumptions. The first is a supervenience claim, on which whether someone has performed a free action supervenes on purely qualitative facts about the person's life history. The second is refined version of a principle of Richard Gale's that should be granted by many compatibilists to the effect that if all of *A*'s actions are intentionally caused by a person *B* making use of non-stochastic means, then none of *A*'s actions are free—*A* is just a puppet of *B*.

I will first give a simpler version of the argument assuming the denial of essentiality of origins claims, and then give a more sophisticated version compatible with essentiality of origins claims.

2. The simpler argument

Let w_0 be the actual world. I first make the following particular supervenience claim:

(S1) Suppose w_1 is a world whose history between 1000 and 1821 AD, both inclusive, matches our world's history. Then Napoleon performs a free action by 1821 in w_1 if and only if he performs a free action by 1821 in w_0 .

Claim (S1) is extremely plausible. Whether Napoleon performed a free action during his life surely only supervenes on how Napoleon developed, how he acted and *perhaps* the effects of his actions. Thus, (S1) would remain plausible even if we strengthened it by replacing the year 1000 by 1769, the year of Napoleon's birth. The more generous lower bound, however, is used here to avoid the objection that being an agent conceptually requires being a part of a certain kind of community, and a community requires a history. Just as perhaps Davidson's (1987) Swampman, produced at random by lightning from the matter of a swamp would not count as an agent at least at first, perhaps something that suddenly arose around 1769 might not count as a community.

Next, observe that both compatibilists and incompatibilists agree that there are "freedom-canceling conditions", conditions under which an action is unfree. An initially very plausible such condition is as follows:

(FC1) If all of A 's actions are intentionally brought about by an agent B , then none of A 's actions is free.

This condition has been defended by Richard Gale. Freedom is incompatible with one's life being completely manipulated by another.

Note that the requirement that *all* of A 's actions are intentionally brought about by another is important. For, to use an example of Gale's (personal communication), you may know that I would not refuse some little favor to a friend, and so by asking me for that favor you would be intentionally bringing it about that I do you the favor. Thus it may not be freedom-canceling for an action if you intentionally bring about that action. In fact, this could even be true if you intentionally bring it about in such a way that I could not do otherwise. For perhaps I am such a friend of yours that I literally could not refuse to do you the favor, and yet nonetheless I do you the favor freely. But the action would arguably not be free if you intentionally brought about the full history of the development of my character that made the favor inevitable.

Gale's principle, however, has at least one potentially problematic consequence. Suppose that I am able to predict with 80% certainty how you will act under any particular circumstances. I arrange

circumstances so as to make it very likely that you act a particular way, and I get lucky—you do act that way. It is not plausible that I have thereby automatically canceled your freedom. Moreover, suppose that I get lucky in this way in regard to *each* of your decisions. I had set up the circumstances so I was 80% sure of each outcome, and I was lucky each time. This is exceedingly improbable if the number of decisions you made is large, but is still possible. And it does not seem that I have taken away your freedom, even though it is true that each of your actions is intentionally brought about by me, assuming—and perhaps we should not—that making an outcome 80% likely and then having the luck of that outcome occurring counts as intentional bringing about.

Another potential counterexample to (FC1) is given by Frankfurt cases (Frankfurt, 1969). If Black is watching you with the neuroscope and will act to ensure you do what he wants, but in fact you do what he wants without his interference, then Black intentionally ensured your action. Whether intentionally ensuring *E* is a case of intentionally bringing about *E* is dubious, but if it is, then this would indeed be a problem for (FC1).

Thus I will weaken Gale's principle some more to make it even more plausible. Say that an agent *B* intentionally *deterministically* brings about an effect *E* provided that the agent intentionally brings about *E* via a deterministic causal chain from *B*'s decision to bring about *E* to the occurrence of *E* (there is an ambiguity here as to whether the agent intends the chain to be deterministic; for definiteness, I will answer this in the affirmative, though I think nothing hangs on this here). Then the following principle is even harder to dispute than (FC1):

(FC2) If all of *A*'s actions are intentionally deterministically brought about by an agent *B*, then none of *A*'s actions is free.

Proposition (FC2) embodies the intuition that if an agent ever acts freely, then the forensic buck stops with the agent in the case of at least some of the agent's actions, where the forensic buck stops with an agent provided that there is no earlier *agent* fully responsible for the action. This version of Gale's principle is not subject to the luck objection, since the lucky manipulator is not initiating deterministic chains of causes leading to *A*'s action, nor does the Frankfurt objection apply, since Black is not initiating any chains of causes leading to *A*'s action.

A different worry, however, can still arise in respect of (FC2) if one is impressed with the following kind of Frankfurt example based on determination. Black initiates a causal chain independent of the agent's causal powers that actually culminates in the agent doing the action that Black desires, while simultaneously the agent *also* acts in a way that causally overdetermines the action. It is far from clear whether in this case the agent counts as acting freely. After all, Black has *caused* the action, though so has the agent. This example resembles and is ultimately inspired by an example of Mele and Robb (1998). In Mele and Robb's case, however, there is no actual overdetermination. Instead, the independent causal process (in the above, started by Black) *would* have terminated in the action were the agent not to do it on his own, but in the actual world, that process is pre-empted by the agent doing the action on his own. It is, thus, much more plausible in Mele and Robb's case that the agent acts freely than in my case above where Black actually causes the action, though does so in an overdetermined way. However, Mele and Robb's case is not a counterexample to (FC2), since the action of the agent is not deterministically caused by another in their case.

If we think that we would not be free if all of our actions were overdetermined by another agent, then we can stick with (FC2). Otherwise, we need to weaken (FC2). But to do that is easy and natural: we simply require that *B* brings about the actions of *A* without overdetermination. Moreover, to make the principle even more plausible, we can add that *B* brings about all the character traits of *A*:

(FC3) If all of *A*'s actions and character features are intentionally deterministically brought about by an agent *B* without *B*'s activity being the initiation of only one or more members of a larger set of overdetermining sufficient causes of the actions and character features, then none of *A*'s actions is free.

Observe that all three principles are compatible with the denial of the Principle of Alternate Possibilities (PAP), where PAP states that if *A* acts freely, then *A* could have done otherwise. In fact, principle (F3) appears to be compatible with all the standard Frankfurt-style counterexamples.^[1]

Suppose now then that one of the freedom-cancellation conditions (FC1), (FC2) and (FC3) holds, and assume the supervenience condition (S1) holds. It follows that if determinism holds, Napoleon never acted freely. For imagine a world w_1 with the same laws as those of our world starting in 1000 AD, and which is just like ours between 1000 and 1821 AD, but where at some point before 1000 AD there was a

super-powerful demon, Fred, who intentionally acted in order to bring about the complete state of the universe in 1000 AD as a means to that state's deterministically resulting in Napoleon exhibiting precisely such-and-such actions and character features, where these specified actions and character features happen to exactly match those of our world w_0 , and where that complete state of the universe in 1000 AD in w_1 also happens to exactly match that of w_0 . Fred sets up things before 1000 AD, then ceases to exist still before 1000 AD, and then in 1000 AD the w_1 has the same state that w_0 has on the same date. By determinism, then, w_1 develops just as w_0 afterwards.

Thus, w_1 is exactly like w_0 from 1000 AD onward. By (S1), it follows that Napoleon performs a free action in w_1 if and only if he performs a free action in w_0 . But in w_1 , all of Napoleon's actions and character features are intentionally deterministically brought about by Fred. Moreover, this is not a case of overdetermination. Fred is not initiating an independent causal chain that overdetermines the outcome, with Napoleon also overdetermining the same outcome. Rather, Fred is causing the conditions that, given the deterministic physics, cause Napoleon to act as he does and to acquire the character features he acquires. This is a case of transitivity rather than of overdetermination. Fred is causing Napoleon to cause various outcomes.

Hence by any one of (FC1), (FC2) or (FC3), it follows that Napoleon is not free in w_1 . Thus by (S1) he is not free in the actual world w_0 , if determinism actually holds. Q.E.D.

This argument more generally applies to all of us. It would not, however, apply to an agent like God who had no causal origins.

Observe that one might think that if all of A 's actions and character features are intentionally deterministically brought about by B , then A is not even an agent.^[2] This claim, however, is compatible with each of the principles (FC1)-(FC3). For if A is not even an agent under such conditions, then A does not perform any actions, and hence it is trivially true that none of A 's actions is free. Some may even think that we could extend the above argument to conclude that if determinism holds, then Napoleon was not an agent, but whether this is so depends on whether an agent needs to act or just have a potentiality for acting.

One might object that no world with the same laws as those of our world could contain such a demon as we need in w_1 . That this is so, however, is not clear, for even deterministic physical laws could be *ceteris paribus* and trumped by the action of a non-physical demon, and the laws of our world might well

be *ceteris paribus*. Moreover, we do not need to assume that the laws of w_1 match those of w_0 from 1000 AD on. We do not need them to match prior to 1000 AD.

3. Essentiality of origins worries

Nonetheless, my argument is problematic if Kripkean essentiality of origins claims hold. For if it is an essential property of Napoleon that he be the son of Carlo Buonaparte and Maria Ramolino, and the essential property of each of these parents that they have the parents they did, and so on, then world w_1 where the state of the universe in 1000 AD was produced by the demon is not guaranteed to be one that contains Napoleon, since it might not contain some of Napoleon's ancestors prior to 1000 AD.

There are two ways out of this difficulty. The first is to suppose that w_1 is a world that contained the same individuals just prior to 1000 AD as w_0 does, but where the demon rearranged these individuals and their properties in a way that made the state at 1000 AD match the state of the actual world w_0 then. The argument then seems to go through.

Unfortunately, one may then have a similar worry about the identity of particular events in Napoleon's life. For if the identity of an event depends on its complete causal history, then none of the events that happen in Napoleon's life in w_0 happen in w_1 , since the causal histories are different prior to 1000 AD. Napoleon fights a battle at Waterloo in w_1 , and that battle is just like *the* Battle of Waterloo, but is numerically distinct from it.

A better way out of the difficulty is to strengthen (S1). Observe that, intuitively, whether someone acts freely should, in some appropriate sense, not supervene on the numerical identity of the agent or of the events in the agent's life, but only on the *purely qualitative* properties of the agent and of the events in the agent's life. It would be an odd kind of favoritism that said that Napoleon was free, but someone just like Napoleon who lived a life just like his was not. Here, a property is purely qualitative if it does not make *de re* reference to any concrete individuals, and does not involve any haecceities (if there are such things). Thus, *being Napoleon* is not purely qualitative, nor is *being shorter than Caesar*, while *being an emperor named "Napoleon"* and *being the tallest adult human ever* are purely qualitative.

Moreover, only the purely qualitative properties of the agent and the agent's environment that specify on what goes on *during the agent's lifetime* should matter for determinations of freedom, or at most those purely qualitative properties that specify what goes on during the agent's lifetime together with a

history sufficient to ensure that the agent can be in a community plus maybe a few moments after death to allow the start-up of causal chains started by the agent in the last moment of the agent's life.

To make this intuition somewhat more precise, say that two worlds w and w^* *qualitatively match* in a space-time region R provided that what exists and goes on in R at w is qualitatively just like what exists and goes on in R at w^* . In particular, any object O existing at a space-time point x in R at w corresponds to an object O^* existing at x at w^* , with O and O^* having exactly the same purely qualitative R -properties, where an R -property is a property that makes no reference to what exists or goes on outside of R . (Thus, if R is Spain in the 19th century, then *being taller than any person in France* is not an R -property.)

Let N be the property of being a short emperor who was born in 1769 and died in 1821. In our world, only Napoleon has N . In light of the fact that freedom should not depend on the numerical identities of objects or events, we can replace (S1) with:

(S2) Suppose w_1 is a world that qualitatively matches our world, w_0 , everywhere between 1000 and 1821 AD, both inclusive. Then Napoleon performs a free action by 1821 in w_0 if and only if the unique person with N performs a free action by 1821 in w_1 .

Assuming determinism, we now construct a world w_1 with the same laws as those of our world from 1000 AD on, and whose state is qualitatively the same as that of our world in 1000 AD, but where that state is produced by a demon with the express intention of ensuring that the unique person with N should have such-and-such a character and do such-and-such actions. By any of (FC1)-(FC3), that person with N is not free in w_1 . Hence by (S2), Napoleon is not free in w_0 . And thus, if determinism holds, none of us are free. ^[3]

References

- Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association*, 60, 441-458.
- Fischer, J. M. (1999). Recent work on moral responsibility. *Ethics*, 110, 93-139.
- Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66, 829-839.
- Gale, R. M. (1990). Freedom and the free will defense. *Social Theory and Practice*, 16, 397-423.
- Mele, A. & Robb, D. (1998). Rescuing Frankfurt-style cases. *Philosophical Review*, 107, 97-112.

^[1] See, for instance, the still-excellent survey by John M. Fischer (1999).

^[2] I am grateful to Wayne Davis for this observation.

^[3] I would like to thank Wayne Davis, Richard Gale and Jonathan Kvanvig for related discussions.
