

Philsoc Student Essay Prize, Trinity term, 2020 – 3rd Equal Prize

How might Free Will be compatible with Determinism?

By David Molyneux

Introduction

Two observations.

Firstly, some human beings (for at least some of the time) seem to possess *free will*. We seem to have the ability to control our choices and actions and make decisions that appear to be ‘*up to us*’. Closely linked to this putative ability to make our own decisions is the concept of *moral responsibility*. Because our decisions seem to be freely chosen, we can be appropriately praised or blamed for these decisions.

But secondly, we also appear to be part of a deterministic universe. We (including that part of us involved in making choices) are subject to the laws of physics and chemistry, and these in turn are determined by what has gone before. If *determinism* is true, then this would suggest that our choices are not ‘*up to us*’.

Worryingly, these two plausible observations seem to be in tension. If determinism is true, then we do not have free will. Conversely, if we possess free will, then at least some of our decisions are not determined. The claim that the existence of free will and the truth of determinism are not compatible is called *incompatibilism*. By contrast, *compatibilism* is the view that an agent can still possess free will even though his choices are determined.

In this essay I will discuss four arguments that have been suggested in favour of compatibilism. I will argue that these arguments are not completely convincing.

Some definitions

I need to make precise the vague definitions I have offered so far.

a) Determinism

Determinism (more precisely, *causal* determination) is the claim that, given a specified way that things are at time *t*, the way things go thereafter is fixed as a matter of natural law. (Hoefer, 2003) If we are aware of the position and properties of all the objects in the universe, then we could determine the events of the future. Importantly, this principle does not just apply to inanimate material (for example atoms, planets), but to living organisms, including human beings. The truth of determinism can be questioned (for example by considering quantum indeterminism) (Hoefer 2003), but in line with the essay question, I’m going to assume a deterministic universe.

b) Free will

I introduced free will (vaguely) as the idea that an agent’s decisions are ‘*up to him*’. However, there are two widely accepted versions of this principle in the literature.

Firstly, free will is the ***ability to do otherwise***. This can be formalised in the so-called *Principle of alternative possibilities* (PAP) (Griffith 2014).

PAP states:

A person acts freely only if she could have done otherwise.

Secondly, free will is being the ***ultimate source of one’s actions and choices***. The use of the word ‘ultimate’ indicates that the action/choice *originates* with the agent.

Both these specifications capture the idea of free will. An agent who acts freely is both the source of his action and has the ability to do otherwise.

c) Moral responsibility

Moral responsibility plausibly depends on free will. There are various accounts of moral responsibility, but perhaps the most influential is the Strawsonian account. Moral responsibility (Strawson 1974) is the proper object of *reactive attitudes* - such as praise and blame. If someone

lacks free will, then they do not have the power to control their choices, and thus it is difficult to see how they could be blamed/praised for their decisions. Determinism therefore threatens both free will and moral responsibility.

The challenge to compatibilism - Van Inwagen's consequence argument

In the introduction to this essay, I hinted at an incompatibility between free will and determinism. Van Inwagen (1983) has expressed this formally in what has become known as the *Consequence argument*. Here's my version of the Consequence argument, building on the discussion of free will in the previous section:

- a) No-one has power over the facts of the past and the laws of nature
- b) The events of the past and the laws of nature entail every fact of the future
- c) Therefore no one has power over the facts of the future - so no one can do other than they actually do, and no one is the ultimate source of their actions

The compatibilist response

Given the consequence argument, how can compatibilism respond? There are actually many widely differing compatibilist responses, so I can only outline the territory.

First strategy - classical compatibilism.

The ability to do otherwise (and perhaps also to be the ultimate source of one's action) is simply choosing or acting without *external* constraints. In other words, doing what one wants to do. This is historically an early strategy - Thomas Hobbes was a compatibilist and saw free will as:

"no stop in doing what he has the will desire or inclination to do (Hobbes 1967).

But this seems an unpromising strategy. Imagine a mentally ill patient in the grip of a psychosis. The patient does what he wants to and is not constrained externally. But his behaviour seems a paradigm of *unfree* action. Something has gone wrong.

Second strategy - the conditional account.

The next move a compatibilist could make is to interpret the ability to do otherwise in conditional or counterfactual terms. A free agent *could have chosen otherwise* if he happened to have different wants or desires (McKenna 2004). But this seems problematic too. If determinism is true, then not only will the choices of the agent be determined, but his wants and desires too. So, the claim that an agent could have done otherwise isn't immune to the deterministic challenge. Here's another problem. Imagine an agent with a deep-seated phobia of dirt. She is unable to enter a dirty kitchen in order to (say) prepare a meal. The conditional analysis would have it that she has free will and could have (freely) chosen to enter the kitchen. But this is the very thing that she cannot do. A paradigmatically unfree action is thus classified by the conditional analysis as a free action.

Third strategy - undermine the PAP using Frankfurt-style cases.

One compatibilist approach is to deny the principle of alternative possibilities (PAP). This strategy was introduced by Harry Frankfurt (Frankfurt 1988). Frankfurt-style cases all have a similar form. Imagine a controller 'C', who is able to monitor and control the brain of an agent 'A'. 'A' has a plan to shoot and kill a victim 'V'. However, should 'A' change his mind and not go through with the assassination, 'C' is on hand to activate the controlling mechanism and compel 'A' to perform the killing. As it happens, 'A' does go ahead with the killing, with no intervention from 'C'.

It seems that 'A' both had free will and would be considered morally responsible for the killing - *yet he could not do differently* (because of 'C'). Frankfurt was able to argue that free will and moral responsibility do not require the ability to do otherwise, and this allows him to undermine a key component of the incompatibilist challenge. In response, it is worth noting that the Frankfurt-style arguments are controversial, and the conclusions that Frankfurt reaches have been widely criticised. However, even if Frankfurt-style arguments are successful, they only undermine the PAP, and not

'ultimate source' arguments. The agent in the Frankfurt cases may have had no alternative but to kill the victim, but he certainly was the *source* of the decision to proceed with the assassination.

Fourth strategy - what really matters for free will.

More recently there have been compatibilist arguments that have re-described the relationship between free will and determinism. What these arguments share is the thought that the existence of free will is not threatened by the fact that our actions and decisions are deterministic. What really matters for free will (say these compatibilists) is that the decisions are in some way *ours*. There is insufficient space to mention all these attempts, but I would like to mention just two approaches.

Frankfurt's 'mesh' theory (1988) claims that our first order desires are freely chosen when they can mesh with (be endorsed by) our second order desires. This challenges the worry that we lack freedom because of faulty conative and cognitive processes.

By contrast, **Fischer and Ravizza's reasons-responsiveness theory** (1988) argues that an agent acting freely is sensitive to, and will modify behaviour in response to, appropriate reasons. Both of these approaches are controversial.

One persuasive criticism of both approaches is based on manipulation arguments - it seems that both desire-meshing and reason-responsiveness can be induced by external manipulation - which would imply that these two properties are not a reliable surrogate for free will.

Conclusion

I have considered several attempts to make free will compatible with determinism. Do they succeed? Ultimately, the success of the compatibilist project can be assessed by asking if the accounts of free will developed by compatibilists are convincing. It seems that no compatibilist account of free will captures the richness of free will and moral responsibility. Determinism may be compatible with *something*; it is not compatible with free will.

Bibliography

Frankfurt HG (1988). *Freedom of the Will and the Concept of a Person*. Springer

Griffith M (2013). *Free will: the basics*. London. Routledge

Hobbes T (1967). *Leviathan*. London. Penguin Classics

Hoefer C (2003). Causal determinism. *Stanford Encyclopedia of Philosophy*

McKenna M (2004). Compatibilism. *Stanford Encyclopedia of Philosophy*

Fischer JM and Ravizza M (1988). *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge. Cambridge University Press

Strawson PF (1974). *Freedom and resentment and other essays*. London. Egmont Books (UK)

Van Inwagen P (1983). *An Essay on Free Will*. Oxford. Clarendon Press.