British Journal of Undergraduate Philosophy, 2(1), April 2007

Limitations of semi-compatibilism: a defence of the principle of non-responsibility

Gregor Ulm
King's College, University of London
gregor.ulm@gmail.com

Introduction

A proponent of semi-compatibilism argues that moral responsibility, free will and causal determinism are compatible. Formal difficulties notwithstanding, we should scrutinize the practical applicability of this if we intend to judge the merits of it as a philosophical position. Therefore, I will expound Harry Frankfurt's attack on the principle of alternative possibilities and its utilization by John Martin Fischer and Mark Ravizza, the most eminent semi-compatibilists. Afterwards, I will raise an important objection to their reasoning. This will lay the groundwork for my defence of Peter van Inwagen's famous argument for the principle of the transfer of non-responsibility, which involves a special emphasis on the difficulties underlying the implications of causal overdetermination.

Frankfurt's attack on the principle of alternative possibilities

Where a compatibilist argues that free will and causal determinism are not mutually exclusive, which is by itself anything but self-evident, a semi-compatibilist takes a rather daring additional step. She accepts Harry Frankfurt's argument against the principle of alternative possibilities (PAP) (1969, 1971), she takes moral responsibility to be compatible with determinism, and she furthermore assumes that determinism may exclude free will (e.g. Clarke: 2003: 10).

Frankfurt aimed to undermine PAP, a principle once claimed to be an a priori truth, at least by some philosophers (Frankfurt, 1969: 829). It states that an agent can only be held morally responsible for an action if

she could have done otherwise. Harry Frankfurt famously showed that this line of reasoning could be challenged, so that an agent *can* be held responsible for an action even if he could not have done otherwise. Let us consider a typical example (cf. Frankfurt, 1969: 829; Fischer and Ravizza, 1998: 29).[1] David holds a loaded slingshot in his hand, intending to kill Goliath, and does so. However, unknown to him, an evil force exerts control over his brain. This mysterious force is able to monitor David's mental state and to manipulate his behaviour if need be. Would it have been the case that David was to abstain from his original plan to murder Goliath, the evil force would have intervened, thus leading to the very same and well-known result – David would have fired his slingshot and killed Goliath.

However, as it turns out, such an intervention is unnecessary because David does not deviate from his original plan – he takes aim, shoots and kills his enemy. Even though David voluntarily killed Goliath in the actual case it is obvious that he could not have done otherwise as he was bereaved of any alternative possibility – he was bound to act in accordance with his original plan because it just was not possible for him to resist the influence of the counterfactual intervener. As will be seen, the plausibility of semi-compatibilism hinges on the implications of this controversial thought experiment.[2] The question we have to deal with is whether such a case would indeed, as Frankfurt claims, justify the statement that an agent, in our example David, can be held responsible for his actions even though he could not have done otherwise.

At first sight, an analysis of the situation seems to be remarkably unspectacular. We are able to discern two outcomes, a factual and a counterfactual one:

(1) In accordance with his original plan, David kills Goliath.

---

1 In the original example, Black wants Jones 'to perform a certain action' (Frankfurt, 1969: 829). As the adaptation of Fischer and Ravizza is filled with rather cumbersome technical details, I opted for a less ostentatious example.

2 Frankfurt's thought experiment is controversial to this day, see e.g. Widerker and McKenna (2003).

(2) David is struck by a sudden pang of conscience. However, at the very same moment he is about to loosen his firm grip on his slingshot, the evil force intervenes and David proceeds to fire his slingshot, leading to Goliath's premature death.

As it seems, David has no real choice. In every case, Goliath is moribund as we take it as given that his opponent will not miss. In (1) it is indubitable that David assumes the role of an agent. On the other hand, in (2) the case is quite different. For an observer the scene of the action appears to be indistinguishable from (1), but since the act of killing is due to the intervention of the evil force it seems appropriate to charge the intervener with responsibility for the murder and not David as the intervener represents the *conditio sine qua non* without which David would not have fired his slingshot. Consequently, David assumes the role of a mere assistant to the evil force.[3]

A possible objection

As a solution to the problem, I propose to split the role of the agent in order to better handle the intervener's manipulation. Therefore, in the actual case (1) we are confronted with $David_1$, the murderer of Goliath, whereas in the counterfactual case (2) $David_2$ acts as the assistant of the evil genius. Thus, I suggest that we should restate the two cases in a slightly different way:

(3) In accordance with his original intention, $David_1$ kills Goliath.

(4) $David_2$ is struck by a sudden pang of conscience. However, at the very same moment he is about to loosen his firm grip on his slingshot, the evil force

[3] Peter van Inwagen provides a similar objection, albeit in a different context when he writes that the agent then is degraded to the status of a mere instrument (1983: 129-34).

> intervenes and $David_2$ proceeds to fire his slingshot, leading to Goliath's premature death.

It should be obvious that $David_1$ could not have done otherwise. Considering that (3) does not allow for an alternative possibility nor for free will, as there is only one path open to him, we should therefore refrain from holding him morally responsible for his action. This is due to the fact that we take the wind out of Harry Frankfurt's objection as the manipulation that takes place in (4) is of no relevance for $David_1$'s action in (3). Hence, the purported counterfactual case is irrelevant for our judgement of the action of $David_1$ as we actually have to deal with two separate agents. Therefore, we have elegantly sidestepped any possible objections a proponent of Frankfurt's point of view could raise.[4] In what follows this result will enable us to successfully defend Peter van Inwagen's principle of the transfer of non-responsibility (TRANSFER NR) against the objections put forward by John Martin Fischer and Mark Ravizza.[5]

Transfer of non-responsibility

Peter van Inwagen (1983) opposes Frankfurt as he contends that we cannot justly hold an agent responsible for something he has done if he

[4] This is also the reason why I consider myself justified in not having to dwell on Fischer and Ravizza's concept of 'guidance control' (1998: 30). They understand this concept to allow for attributing moral responsibility to an agent even if causal determinism obtains. This is based on their interpretation of Frankfurt. As my criticism is more fundamental, I am justified in leaving this point largely uncommented. Nonetheless, it might be worthwhile to point out that Fischer and Ravizza deduce moral responsibility from Frankfurt-type examples for the sole reason that the agent is understood to possess the aforementioned feature of 'guidance control' (1998: 34-5). By splitting the role of the agent I avoid their objection. Of course, one might object to my way of reasoning. On the other hand, I contend that my argument is more convincing than Fischer and Ravizza's as they merely postulate the existence of 'guidance control' without giving sufficient reasons for accepting that position. The difficulties of such a manoeuvre should soon become evident. As will be seen, we can add as many counterfactual interveners as we want to without finding a way to avoid my fundamental objections.

[5] I borrow the term 'TRANSFER NR' from Fischer and Ravizza (1998).

lacked control over the causes that led to this specific outcome.[6] His argument is as follows: If *p* obtains, and no one is even partly morally responsible for *p*; and if *p* obtains, then *q* obtains, and no one is even partly morally responsible for the fact that if *p* obtains, then *q* obtains; then *q* obtains, and no one is even partly morally responsible for *q* (1983: 182-8; Fischer and Ravizza, 1998: 152).[7] This abstract syllogism is then illustrated by means of a rather graphic example: no one is responsible for the fact that John is bitten by a cobra on his thirtieth birthday; and no one is responsible for the fact that if John is bitten by a cobra on his thirtieth birthday, then he dies on his thirtieth birthday; then no one is responsible for the fact that John dies on his thirtieth birthday (Fischer and Ravizza, 1998: 154; van Inwagen, 1983: 187).

Fischer and Ravizza now try to refute the plausibility of the principle TRANSFER NR, which we just exposed, by modifying the scenario. They opt for complicating matters by adding a sharpshooter to the scene (1998: 166).[8] The most important detail is that she has achieved such a high level of marksmanship that she is able to judge when to fire the rifle so that the bullet will lead to John's death at the very same moment as the snake's venom. Technical intricacies notwithstanding, they claim that this modification refutes Peter van Inwagen's original example as he is only able to back the principle TRANSFER NR because it allows for just one path leading to John's death. The added sharpshooter is supposed to open up a second path leading to the same outcome.[9]

---

[6] This stance seems to be plausible enough. An opposing point of view, namely that lack of free will does not bereave us of responsibility, is brought forward by Derk Pereboom (2003).

[7] The original example is to be found and explained in van Iwagen (1983: 182-8). I opted for the reformulation by Fischer and Ravizza as I will scrutinize one of their examples later on.

[8] One might assume that it would have been more convincing if Fischer and Ravizza had tackled the validity of van Inwagen's original syllogism, and not just one of his examples, because even if we assume that in a nearby possible world nobody dares to contradict their claims, then one could still hold against them that their findings do not adequately reflect ordinary cases like the one represented by the original and unmodified principle TRANSFER NR, and thus they only managed to refute a very special case (note: a special case which they themselves introduced!).

[9] Again, the sharpshooter is understood to be responsible for his action as she exerted 'guidance control'. Please see footnote 5 for my objection to this concept.

On the contrary, I contend that even though we now have to deal with a case of causal overdetermination. This is due to the simultaneous effects of the snake's venom and the bullet. So the main problem remains unresolved. Instead of fully detailing my objection one more time, I refer the reader to my splitting of the agents introduced above (if we are tempted, we could even try to introduce a counterfactual intervener ensuring that the sharpshooter kills John). The result is that the sharpshooter cannot be responsible for his action if causal determinism obtains. The additional problem of causal overdetermination, however, remains to be dealt with.

Problems of causal overdetermination

In general, and opposing Fischer and Ravizza (1998: 165), I think it is quite implausible to assume that we can ever have two different paths leading to the same outcome. As it seems, their whole argument is based on an inadequate interpretation of Peter van Inwagen's example. Obviously, there is nothing wrong with letting the predicate variable *q* stand for 'John dies on his thirtieth birthday', but once we alter the scenario, we should watch our steps carefully because once we decide to add poisonous venom and a sharpshooter to the scene we will need to distinguish the cases. The statement that John dies in both cases is certainly true, but proper reflection should lead to a reformulation like the following:

(5) John dies on his thirtieth birthday because of the venom of a snake.

(6) John dies on his thirtieth birthday as a consequence of the simultaneous effects of poisonous venom and a bullet entering his head.

When looked upon superficially, both cases might appear to be identical. However, we are faced with a problem. Fischer and Ravizza take each of their cases, (5) as well as (6), to be, at the same time, instantiations of *q*. This approach is inconsistent, however, because as we have seen the events leading to John's death are quite different. In

order to refute TRANSFER NR both cases would have to be shown to be identical, which is nothing short of impossible – it is clearly false to assume that (5) equals *q* and then assume that (6) equals *q* as well. As both cases are by no means equal, due to the fact that John's death is the result of different causes, Fischer and Ravizza do not live up to their claim to produce a valid counterexample which functions against the principle TRANSFER NR. Each of the cases, (5) or (6), can only in isolation be accepted as a valid instantiation of *q*.

Finally, I would like to remark that we can very well accept a certain amount of imprecision in Peter van Inwagen's original example, as there is just one way for John to die. Nonetheless, one might still raise objections against my arguments. One could assume that the instantiation of *q* is exactly the same in two cases, while allowing for the possibility that the events leading to *q* differ. I would concede that such a scenario is quite possible – just imagine that the doctor who performed the autopsy on John's dead body in the two possible worlds comes to the same conclusion. In both cases he scribbles down 'cause of death: snake venom' on the report. But an omniscient external observer would have experienced two different incidents: in the first one, John's niece opens the snake's cage, and in the second his nephew did this. As a result, we are still able to distinguish two cases of *q*, namely 'John died from a snake's venom (as a direct consequence of a negligence of his niece)' and 'John died from a snake's venom (as a direct consequence of a negligence of his nephew)'. Thus, I contend that my objections to Fischer and Ravizza are quite solid.

While I do not pretend that my reasoning is able to refute semi-compatibilism altogether, I nonetheless argue that my objection poses a serious threat for any supporter of this position. It is my understanding that so-called Franfurt-type examples should be irrelevant for the debate regarding free will and determinism. Furthermore, I have raised an objection against cases of causal overdetermination. If I am right on these two counts, then one would indeed be justified in reducing semi-compatibilism to compatibilism of free will and causal determinism.

Bibliography

Clarke, Randolph, 2003, *Libertarian Accounts of Free Will*, Oxford: OUP

Fischer, John Martin, and Ravizza, Mark, 1998, *Responsibility and Control: a Theory of Moral Responsibility*, Cambridge: CUP

Frankfurt, Harry G., 1969, 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66, pp. 829-39
1971, 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy*, 68, pp. 5-20

Pereboom, Derk, 2003, *Living without Free Will*, Cambridge: CUP

van Inwagen, Peter, 1983, *An Essay on Free Will*, Oxford: OUP

Widerker, David, and McKenna, Michael (eds.), 2003, *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, Aldershot: Ashgate