LIBERTARIAN FREE WILL AND CIRCUMSTANTIAL MORAL LUCK

Daniel Rubio

Rutgers University

1. INTRODUCTION

Discussions of moral responsibility and discussions of free will often go hand in hand. One of the primary motivations for positing free will is that it is what justifies assignments of moral responsibility. This is why moral luck poses such a problem: it seems that, even when granting free will, our moral assessment of people is unduly influenced by factors completely beyond the control of the agent. Thomas Nagel first identified the problem of moral luck with examples like the following: Suppose Bob lived a quiet life in Argentina. However, had Bob been born in Nazi Germany, he would have become a Nazi sympathizer.[1] It is merely a matter of luck that he was not born in Nazi Germany. Likewise, had he been born in Nazi Germany, a parallel case could be constructed: Had Bob been born in Argentina, he would have lived a quiet life in Argentina. When he is born in Argentina, Bob is morally lucky in his circumstances; when he is born in Nazi Germany, he is morally unlucky in his circumstances. Thus, we have a case of circumstantial moral luck.

The interaction between circumstantial moral luck and free will brings about a problem: the problem of counterfactual moral responsibility. After spelling out the problem, I will argue that—while compatibilist conceptions of free will have no answer to the problem of counterfactual moral responsibility—incompatibilists can solve the problem by denying that any of the kinds of counterfactuals that circumstantial moral luck depends on are true. I will then argue that the ability of incompatibilist accounts of free will to avoid the problem of counterfactual moral responsibility gives the defender of free will concerned with moral responsibility good reasons to adopt the incompatibilist position.

Daniel Rubio

2. THE PROBLEM

The problem of moral luck comes because we consider the following principle, often referred to as the control principle, strongly intuitive:

(CP) We are morally responsible only for those things under our control

From this follows another strongly intuitive idea about moral judgment:

(CJ) Two people ought not be given different moral assessments if the only difference between them is due to factors beyond their control.

Nagel's counterexample about Bob seems to defy this principle. It is completely beyond Bob's control where he is born, and consequently it does not seem that the German collaborator should be judged less harshly than the Argentinean non-collaborator. Michael Zimmerman splits the idea of control into restricted[2] and unrestricted control. After dismissing the version of the puzzle relying on restricted control, he outlines the interesting puzzle of moral luck like this:[3]

(1) *P* is responsible for *e*'s occurring only if *P* was in unrestricted control of *e*
(2) No event is such that anyone is ever in unrestricted control of it
(3) No event is such that *P* is morally responsible for it

Although he finds (1) in the argument clearly false, and indeed it seems so,[4] Zimmerman gives the following substitute that seems to capture the same intuition:[5]

(1*) if   (i) *P* made *d* in what he believed to be *s*
           (ii) *P** would have made *d* in what he believed to be *s*, and
           (iii) *P**'s being in a situation that he believed to be *s* was not in *P**'s restricted control

then whatever moral credit or discredit accrues to *P* for making *d* also accrues to *P**

(1*) does not do the same work as (1) in the puzzle, for it leaves open another option: rather than abandon moral responsibility, we assign moral responsibility *counterfactually*. If Bob is such that he would have been a Nazi collaborator had he been born in Nazi Germany, then he is as morally blameworthy as someone who was born in Nazi Germany. The difference is merely one of circumstantial moral luck, and so if we are going to hold CP, and by extension CJ, we must assign Bob the appropriate blame. Zimmerman comes to this same conclusion.[6] But for what is Bob to blame? For being such that he would have been a Nazi collaborator had not luck intervened.[7]

What Zimmerman has identified can be called counterfactual moral responsibility, and it depends crucially on counterfactuals of freedom (CFs). A counterfactual is a CF iff it takes the following form: If *S* were in *C*, *S* would have *A*ed, where *S* is an agent,

C is a counterfactual circumstance, and *A* is a free action. On Zimmerman's account, Bob is morally blameworthy for having the following CF true of him: If Bob were in Nazi Germany, Bob would have become a Nazi collaborator.

Counterfactual moral responsibility is highly undesirable, because if it is true then, if ought implies can, it is not the case that each person ought to live a life free of moral blame. Consider:

(1) Every person ought to live a life free of moral blame (assume for *reductio*)
(2) Ought implies can (premise)
(3) Counterfactual moral responsibility is true (premise)
(4) If counterfactual moral responsibility is true, at least one person is counterfactually responsible for at least one morally blameworthy act (definition)
(5) At least one person is counterfactually responsible for at least one morally blameworthy act (3 & 4)
(6) If at least one person is counterfactually responsible for at least one morally blameworthy act, then it is not the case that every person can live a life free of moral blame (4, definition)
(7) Every person cannot live a life free of moral blame (5 & 6)
(8) If every person cannot live a life free of moral blame, and ought implies can, then it is not the case that every person ought to live a life of moral blame (2 & 7
(9) It is not the case that every person ought to live a life free of moral blame (7 & 8, Contradiction)

The argument is valid, so one of its premises must be false. (2), (3), and (4) are the only options. (4) is extremely plausible—for it to be false, each person ought to be such that in any circumstance, she would avoid doing the morally blameworthy thing. There is strong empirical evidence to the contrary. (2) is also extremely plausible, and if ought does not imply can, then we are saddled with an obligation we cannot fulfill. This is to be avoided. That leaves (3), and (3) is counterfactual moral responsibility.

3. CFS AND LIBERTARIAN FREE WILL

The key step in the development of counterfactual moral responsibility deploys CFs. If there are no true CFs, or if the class of true CFs cannot support counterfactual moral responsibility, then there is no counterfactual moral responsibility. Although the idea of true CFs seems no less troublesome than any other true counterfactuals, I will argue that because of the indeterminism involved in the libertarian view of free will, there are few true CFs if the libertarian view of free will is correct.

A CF purports to establish what an agent would have done in a given non-actual situation. For the libertarian, the processes that govern these decisions are indeterministic. Consequently, using Lewis's semantics for counterfactual logic, such a contradiction can be shown. First, take a formal definition of libertarian free will:

(1) $(P \Diamond\!\!\to Q)$ & $(P \Diamond\!\!\to \neg Q)$

P = S were in *C* (*C* includes all circumstances save the ones to be described
 below)
Q = *S A*s

(1) reads "if *S* were in *C*, it might be that *S A*s and if *S* were in *C* it might be that *S* does not *A*." It is analytically true that this entails and is entailed by libertarian free will. Now, take the paradigm CF

(2) $P \Box\!\!\to Q$

(2) reads "If *S* were in *C*, *S* would *A*."
   Two major systems of counterfactual logic have been proposed: that of Lewis, and that of Stalnaker. On Lewis's system, the "nearer than" relation is only transitive and reflexive, while for Stalnaker it is also total and antisymmetric. Consequently, the Lewisian system allows for ties between worlds for being the nearest possible world, while Stalnaker does not. If Lewis right, (2) is logically equivalent to (3):

(3) $\neg(P \Diamond\!\!\to \neg Q)$

Of course, (3) is the negation of the second conjunct of (1), and therefore contradicts the formal definition of libertarian free will. If Stalnaker is right, however, (1) entails:

(4) $(P \Box\!\!\to Q)$ & $(P \Box\!\!\to \neg Q)$

On Stalnaker's system, (4) is a contradiction (on Lewis's system, its conjuncts are merely contraries). So either way, the truth of (1) implies the falsity of (2), and therefore of libertarian CFs.
   There are three kinds of CF that avoid the Might-Would Interdefinability (MWI) Objection. The first two classes are familiar: counterpossibles and semifactuals with necessarily true consequents (for eample: if John were to go to the store, 2 squared would still be four). These include semifactuals where the antecedent entails the consequent (for instance, tautologous semifactual CFs). One other class is less obvious: CFs where the action involved is derivatively free.
   CFs where the action is derivatively free concern those cases that are typically problem cases for libertarians. Addiction is a prime example, although what applies to addiction applies to any situation in which a libertarian must argue for derivative rather than direct freedom. In those cases, (1) of the MWI Objection is not applicable, because its second conjunct is false.
   Consequently, there are some true CFs, but the class of true CFs is quite small. Furthermore, none of those remaining is suitable to establish counterfactual moral responsibility. Counterpossibles never could have obtained, and if the universe is indeterministic in the way libertarian free will supposes it to be, action-counterfactuals will hardly ever be included amongst those where the probability of the consequent,

given the truth of the antecedent, is equal to 1. Finally, in the case of addiction, the freedom (and therefore the responsibility) derives from previous choices, about which there are no true CFs. Without the right sort of true CF, the argument for counterfactual moral responsibility fails. Thus, on the libertarian view, (3) in the *reductio* argument is false, and so the problem of counterfactual moral responsibility dissolves.

4. Objections

4.1 A Probabilistic Counterargument

However, this is not the whole story. Zimmerman thinks that circumstantial moral luck is ineliminable.[8] This seems correct, for, it may be argued, a new problem may be proposed with sentences like *L*:

(L) $P(C|N) > P(C|A)$[9]

C = Performing the act of Nazi collaboration
N = The circumstance of living in Nazi Germany in 1942
A = The circumstance of being born in Argentina after 1945

In this case, while libertarian agents cannot be placed in circumstances where they *would* commit morally blameworthy actions, they can be circumstantially excluded from certain classes of actions. Furthermore, they can be placed in situations where the probability of good actions is increased, thereby (if good actions make for improved character) making blameworthy actions less likely.

There are two problems with this counterargument. The first is that *L* is false (as I shall argue shortly). However, even if it were true (and certainly *L*-like sentences are true of some people) *L* would still be insufficient to establish counterfactual moral blame, and so the circumstantial moral luck that remains will not be able to generate the problem.

L claims that the probability of Nazi collaboration given being born in Argentina after 1945 is less than the probability of Nazi collaboration given living in Germany in 1942. Since being born in Argentina after 1945 entails being unable to collaborate with the Nazis, $P(C|A)=0$. Thus, for *L* to be true, $P(C|N)$ must always be greater than 0. Intuitively, this seems true; the libertarian is committed to the principle of alternate possibilities, so for any libertarian agent with the opportunity to collaborate with the Nazis, possibly that agent chooses Nazi collaboration. Consequently, all that is required for the truth of *L* is what I will call the possibility-probability inference (PPI):

PPI: $\Diamond(E \text{ \& } K) \rightarrow P(E|K) > 0$[10]

E = Any event
K = All relevant background information (history of the world to the time of the event, complete description of the world at the time of the event, laws of nature, etc)

In essence, PPI represents the assumption that every broadly logically possible event on a given set of background information has a probability greater than 0.

While PPI seems plausible, it fails. The probability calculus works on standard analysis; thus, while for any even probability distribution over a finite field, the probabilities sum to one, this fails for a countably infinite field. Anything with a chance of $1/\infty$ is undefined (or, as a convention, assigned a probability of 0). Thus, if there are contingent events on a given set of background information with probabilities of $1/\infty$, then it is not true that every contingent event on a given set of background information has a non-zero probability, and PPI fails. Standard examples of such a chance include the chances that an infinitely fine dart hit any given point on a dartboard and that a dart hit a mathematical line across a dartboard. So there are chances of $1/\infty$, and consequently PPI fails.

If PPI fails, then it is possible for a contingent event to have probability 0, (or undefined) and consequently L is falsified, for it may be that some libertarian agent has sufficiently good character that even were she placed in Nazi Germany in 1942, the probability of her collaborating with the Nazis is 0, and therefore not greater than the probability of her collaborating with Nazis when she is placed in circumstance that entail her not collaborating with Nazis.

Nevertheless, suppose L were true: would that still establish counterfactual moral blame? It seems not. On the libertarian view, no matter what the objective probability of an act, unless it is merely a derivatively free action, it is indeterministic whether or not the agent will commit any given possible act. If that is true, then the possibility of doing otherwise will always eliminate the true "would counterfactual," and with the true "would counterfactual" eliminated, so is counterfactual moral responsibility.

4.2 A Final Objection

Norvin Richards argues that we assess agents based on what they do, not on what they would have done.[11] We do this because the only way to know what an agent would do in a given situation is to see what an agent in fact did when that situation obtained. Two agents may well be *deserving* of the same assessment, but luck makes that desert more or less clear.[12] This is why we do not assign blame counterfactually. While this may go some way toward explaining our moral practices, it is does not solve the problem of counterfactual moral responsibility. The objection is merely epistemic. We do not wish to be only epistemically morally blameless, but actually morally blameless. Consequently, while our epistemic position may explain why we assign blame differently based on luck, it cannot be used as a basis for a lucky difference in deserved blame.

5. Conclusion

The puzzle about moral luck arises because we find the control principle to be a highly intuitive part of our received view of moral responsibility. The same can be said of the dictum "ought implies can." If the goal of morality is to live a life free of moral blame and we allow the right sort of true CFs, then the problem of counterfactual moral responsibility arises. Adopting the libertarian view of free will[13] allows a way out of

the problem, for if there are no true CFs of the right sort, then there is no problem of counterfactual moral responsibility. This does not completely eliminate moral luck. There remains resultant moral luck, and a probabilistic version of circumstantial moral luck, but neither of these can produce the problem.

If what I have argued is correct, then the problem of counterfactual moral responsibility provides one strong *prima facie* reason to accept the libertarian view of free will: the ability to hold three major elements of our received view of moral responsibility in conjunction. The compatibilist and the determinist seem to be committed to true CFs, so they must deny either the control principle, "ought implies can," or the idea that every person ought to live a life free of moral blame. While this option remains open, the account of moral responsibility that denies one of these three is less plausible than the account that can retain them, and that provides one strong *prima facie* reason to accept the view of agency that can hold them all.

<div align="center">NOTES</div>

1. Thomas Nagel, "Moral Luck," *Mortal Questions* (Cambridge: Cambridge University Press, 1979), 26.
2. Restricted control = df X has restricted control over e iff X can bring about e or prevent e being brought about. Unrestricted control = df X has unrestricted control over e iff X has restricted control over e and all events on which e is contingent. See: Michael Zimmerman, "Luck and Moral Responsibility," *Moral Luck*, ed. Daniel Statman (Albany: SUNY Press, 1993), 219.
3. Zimmerman 220.
4. If e is an action of Ps, e is contingent on P being born, and P has no control over this.
5. Zimmerman 225.
6. Zimmerman 226.
7. Zimmerman 230.
8. Zimmerman 228.
9. The probability of Nazi collaboration given living in Nazi Germany in 1942 is greater than the probability of Nazi collaboration given being born in Argentina after 1945.
10. Read: If it is possible that E and K, then the probability of E conditional on K is greater than 0.
11. Norvin Richards, "Moral Luck and Desert," *Mind* 95 (1986):378.
12. Richards 199.
13. And Lewis's semantics for counterfactuals.