

Mill on Free Will: A Millian Defence of Compatibilism

Sebastian Köhler

MSc Philosophy of the Social Sciences, 2008

This essay will be concerned with John Stuart Mill's solution to the problem of free will. I will proceed in three steps. First, I will state what is exactly meant by “the problem of free will” and outline Mill's solution. Second, I will bring forward a possible objection to Mill's solution and third, I will show why this objection is flawed. Finally, I will briefly sketch a theory of free will that could be developed out of Mill's solution to the problem of free will. I will start my investigation now with the presentation of the problem of free will.

The Problem of Free Will

The problem of free will is the problem whether free will is compatible with determinism. Determinism is the doctrine that says that the occurrence of every event E is sufficiently conditioned by the states of the world prior to the occurrence of E . A person has free will if she controls her actions or we can call her decisions “her own”. One major problem in the discussion of the problem of free will is that it is not clear what it means to say that a person has “control” over her actions or that her decisions are “her own”. This rough definition will nevertheless suffice to show why the compatibility of free will with determinism is problematic: Imagine Albert, who has just bought a bottle of orange juice. If determinism is true, then his doing so was sufficiently conditioned by the states of the world prior to his doing so. It was even sufficiently conditioned by the states prior to his birth!¹ But how

¹ This is the case, because if the occurrence of E at T_n is sufficiently determined by the state of the world at T_{n-1} and the state of the world at T_{n-1} is sufficiently determined by the state of the world at T_{n-2} , then E is also sufficiently determined by the state of the world at T_{n-2} , (by simple logic: if $p \rightarrow q \rightarrow r$, then $p \rightarrow r$). Therefore, if every event E is sufficiently determined by the states of the world prior to the occurrence of E , then E is sufficiently determined by every state of the world prior to the occurrence of E . Therefore, of the event ‘Albert buying orange juice’ is already sufficiently determined before Albert is born.

can we say that Albert has control over his decision or that it was his *own*, in any serious sense, if its occurrence was sufficiently conditioned prior to his birth?

This is the problem of free will: If determinism is true, then all of our decisions are sufficiently conditioned by the states of the world prior to them and even prior to our birth. But if that's true, then it seems – on first sight – rather dubious to say that we could, under any circumstances, have free will. Since we now know what the problem of free will is, we can turn to Mill's solution of it.

Millian Compatibilism

While Mill believed that determinism was true, he was a proponent of compatibilism, which is the position that free will is actually compatible with determinism. Mill's argument against incompatibilism (which is obviously the negation of compatibilism) has two steps. First, he argues that incompatibilists conflate determinism with fatalism and that this is the reason they think that free will is incompatible with determinism. This leads into the second step of his argument, which is an analysis of "free will", that is compatible with the truth of determinism.

So, why do incompatibilists conflate 'determinism' with 'fatalism', why is this conflation problematic, and why is it false? To answer these questions, we first have to look at what is meant by 'fatalism'. Fatalism is the doctrine that it is true for every event E that E will come about – regardless of whether one is trying to prevent E from coming about. It is clear that fatalism is not compatible with free will: If fatalism is true, then it does not matter which choices an individual makes because all events that occur come about regardless of what individuals choose. But if that is true, it can't be said that the individuals have control over their actions because their actions will come about anyway, regardless of what they want or choose. Now we can see why an identification of determinism with fatalism is problematic for the compatibility of free will with determinism. Since fatalism is *clearly* not compatible with free will, determinism – if it means the same thing as fatalism – is incompatible with free will.

It is also easy to see why fatalism can be easily conflated with determinism. Fatalism says that every event will come about and nobody can do anything to prevent that. Determinism says that whatever event will come

about, its coming about will be sufficiently conditioned by the prior stages of the world. So, it seems that if determinism is true and E occurs, nobody could have done anything about that, since the occurrence of E was sufficiently conditioned by the prior stages of the world. But (and this is what Mill highlights) it is not true that determinism and fatalism say the same thing. Fatalism says that it does not matter which actions you take with regard to the occurrence of E, because E will occur anyway. But in determinism the actions people take matter, because if people take actions that do not result in E, then E will not occur. It also matters what people *want* to do, because we can assume that their wants are among the things that determine which actions will come about. This means that if determinism is true, counterfactual statements like “If I had not wanted to do X, X would not have occurred” can be true, while if fatalism is true, they cannot. Therefore, fatalism and determinism do not mean the same thing.

This conclusion allows Mill to take the second step in his argument. We know that determinism is compatible with counterfactual statements like “If I had not wanted to do X, X would not have occurred”. But if such counterfactual statements are true under determinism and it is true under determinism that what a person wants or doesn’t want determines which actions she does, then determinism seems to be compatible with free will: If it is true that X, *ceteris paribus*, only occurs when the person wants it to occur and otherwise doesn’t, then it seems that the action can be said to be under the control of the person and thus, the decision for the action is the persons own. Finally, the action occurs because the agent *wants* it to occur. In what other sense could a decision be an agent's own? Mill thus answers the question of whether free will is compatible with determinism with the following analysis of “free will”: A person has free will, if and only if her actions would not have obtained, if she had not wanted them to obtain. This analysis is plausible because on first sight it seems uncontroversial to assume that an action is under the control of agent if it fulfills the suggested criteria as well as to assume that in this case a decision can be called the agent's own.

The Requirement of ‘Ultimate Control’

The question is now whether this analysis *really* gives the necessary and sufficient conditions for what we mean by ascribing ‘free will’ to a person. One objection to this analysis, which Mill himself discusses, is readily apparent:

It may be the case that counterfactual statements like “If I had not wanted to do X, X would have not occurred” can be true if determinism is true. But it is not sufficient to say that if such a statement is true of a person then the person has free will because if determinism is true the facts about what a person wants do not *ultimately* depend on *her* choices, but on factors external to her. Since, if determinism is true the occurrence of every event E is already sufficiently conditioned by the states of the world prior to E, the fact that a person wants to do E at a time T, is also sufficiently determined by the stages of the world prior to T. The same fact holds here that holds for actions: Whether or not the person wants to do X is sufficiently conditioned even before their birth! But that seems to imply that they have no real control over their actions and that their choices are not their own because although her actions depend on the things they want, they have no *ultimate influence* on the things she wants.

We can see why this may be problematic by looking at the following example: Imagine Charles, a supernatural being with mind controlling abilities who can influence what people want. One day Charles meets Daniel, a clown. Because Charles is afraid of clowns and adores monks he manipulates Daniel’s will so that Daniel wants to become a monk. Daniel then becomes a monk. Now, in this situation it is true that Daniel would not have become a monk if he had not wanted to become a monk. And yet it seems natural so say that Daniel’s will is not free, because he has no ultimate control over what he wants. The same holds under determinism: It might be true that I now want to become a philosopher, but if determinism is true, then my so wanting was sufficiently conditioned by the states before my birth and I therefore have no ultimate control over this want. I have no choice over whether I want to X or not because, although these volitions might result of, and be influenced by, other volitions of mine, the ultimate reasons for X lie outside myself. This is because my wanting X or deciding against wanting X at moment T is sufficiently conditioned by stages of the world prior to T, even stages that occurred when I was not even born. And it makes no sense to say that I have an influence on events that occurred before I was born. This objection seems extremely powerful and greatly strengthens the incompatibilist position. How might a compatibilist reply?

The Compatibilist Response

I think that he should reply by rejecting the notion of free will that the incompatibilist uses. The incompatibilist demands that our actions and wants are *ultimately* under our control. ‘Ultimately under our control’ means that there is no other influence in the decision except for the person herself. To have free will in this sense, we must be capable of building our own character, of determining our own wants and actions without relying on any factor that is not determined by our selves. But how can we possibly choose our character or wants without relying on something that is not determined by our own selves? Assume that you make an ultimately first decision of whether you want one particular character trait or another. On which grounds are you making this decision? If you have reasons for your decisions, then there must be a fact of the matter that explains why these reasons are reasons *for you*. This might be either a fact about you, or a fact about the reasons. If it is a fact about you, since it is (as we stipulated) the first decision you make about yourself, then this fact of the matter must be determined by a source external to you. If it is a fact about the reasons, this fact must also be determined by a source external to you, since it is still your first decision and you couldn’t have decided (by stipulation) before this decision which reasons are reasons for you. Therefore, if you decide on the grounds of reasons, then your first decision couldn’t be ultimately yours in the sense that the incompatibilist uses the word, because your decision depends on facts that are not determined by you. Therefore, these decisions cannot be free in the incompatibilist sense.

If on the other hand you don’t have reasons for your decision, your decision is totally arbitrary. We could repeatedly rewind the course of history to the point where you made the decision and there is a high probability – since you don’t have any reasons for your decision – that it turns out that you choose different character traits each time. But does it really make sense to say that your decision was your ‘own’, that you had control over your decision, when you decide like this? It wouldn’t even be possible to tell why you did decide the way you decided because there is no fact of the matter that could answer this question (you can’t just say that you wanted to do it, because, as was stipulated you have no wants). This leads to a difficulty: it doesn’t even seem to make sense to say that somebody makes a decision and has no reason for this decision. Can a person make a decision, without

wanting something? I don't think that is possible. It doesn't even make sense to speak of a person who has no wants, no values and no character.

I think these considerations show that any account of free will that requires us to have ultimate control over our actions is not coherent, since ultimate control in this sense is impossible and rules out free will by definition. But first it seems to me that our idea of freedom of the will is not an incoherent idea *per definitionem*. That's why I think we have reason to reject the condition of ultimate control. Second, even if the incompatibilists insist that this condition is part of the way people use the concept of free will, I think we have good reason to reject it and bring forward a revisionary notion of free will that is nearer to what we mean by free will in ordinary language, without falling prey to the objection that it is an incoherent idea.²

A Sketch of a Compatibilist Account of Free Will

I will now sketch a short version of just such a conception of free will that is based on Mill's account. We can take Mill's central statement regarding free will to be the following:

[I]f we examine closely, we shall find that this feeling, of our being able to modify our own character if we wish, is itself the feeling of moral freedom which we are conscious of. *A person feels morally free who feels that his habits or his temptations are not his masters, but he theirs...* (Mill, 1925, pp. 3-4, my emphasis)

I think this statement can be developed into the following approach: Even though it might be true that we don't have ultimate control over the things we want and we don't want, it still makes sense to talk of having free will, if it is either the case that our wants, actions, values and our identity as a person are consistent or if when a inconsistency comes about, the person will want to resolve the inconsistency and if she wants to resolve it, will be able to do so. A person will be unfree if and only if she recognises an inconsistency in her behaviour and wants, but is not able to resolve this inconsistency (either

² Personally I think that a contextual analysis of free will is the most appropriate to what we mean by free will in the sense that we speak of free will mostly in those contexts (maybe "linguistic contexts" or "language games" would be the more appropriate notions) which Mill, for example, highlights (where we can reasonable say that "if I had wanted otherwise, I would have done so") and that in these contexts it makes perfect sense to say so, but if we apply the notion to other contexts (like the question of ultimate control), it becomes nonsense. But such concerns are beyond the scope of this paper.

because she doesn't want to³ or because she wants to, but is for some other reason not able to resolve it). Such a person will feel alienated from her desires, actions or values, because they occur without her willing them to occur. This is why we can say that she doesn't feel like her actions are in any way under her control or her own. On the other side, a persons will can be called 'free' in this sense, if and only if she does the things she wants to do and her wants and her values don't conflict. In the case of a conflict, a free person will be able to resolve the conflict in a way she can identify with; that is, she will be able to resolve the conflict in a way she wants it to be resolved. Such a person will be able to identify with her behavior and will not feel that any of her behavior comes about although she doesn't want it to come about. I think in such a case it clearly makes sense to say that her behavior is under her control and to call it her own. Of course, this kind of account will not convince an incompatibilist who holds out for *ultimate* control over one's wants and actions (although the incompatibilist will first have to show that this idea is intelligible), but I think that it is a good and plausible starting point⁴ for a compatibilist account of free will.

References

Mill, J. S. (1925) *A System of Logic*, Eighth Edition. London: Longmans Green.

3 This I take to be an exceptional phenomenon. An inconsistency between ones desires and behaviour (or between different desires) seems to me such a serious and painful thing that it seems unlikely that a person who suffers from this kind of inconsistency doesn't develop the want to resolve it.

4 This is only a preliminary account of a new kind of Millian free will and some problems remain. It would, for example, have to solve the question of why we would not call Daniel's decision to become a monk 'free', although there is no inconsistency between Daniel's wants and his identity that he can not resolve.