<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Harvard Review of Philosophy*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Svedberg, M., Tännsjö, T. (2017)

Consequentialism and Free Will: The Conditional Analysis Resuscitated.

Harvard Review of Philosophy, 24: 23-41

<https://doi.org/10.5840/harvardreview2017558>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-329279>

(Preprint version)

Consequentialism and Free Will: The Conditional Analysis Resuscitated

Maria Svedberg and Torbjörn Tännsjö
Stockholm University

Abstract: Many moral theories incorporate the idea that when an action is wrong, it is wrong because there was something else that the agent could and should have done instead. Most notable among these are consequentialist theories. Relatively little attention has been given to the question of how to understand the meaning of ‘could have’ in this specific context. However, without an answer to this question, consequentialist theories fail to yield determinate verdicts about the deontic status of actions in real scenarios. It is here argued that a conditional analysis provides the required answer and gives us the most plausible version of consequentialism. Such a conditional analysis has been universally rejected as an analysis of the general meaning of ‘could have,’ but we show that in the specific context of specifying the meaning of ‘could have’ in a criterion of right and wrong action, all the standard objections to it fail.

Key words: consequentialism, free will, conditional analysis, could have acted otherwise, regress argument, coma argument, phobia, manipulation, coercion, unintentional action

I. Introduction

Many moral theories incorporate the idea that when an action is wrong, it is wrong because there was something else that the agent could and should have done instead. The argument we present here applies to any of these theories. However, for simplicity, we primarily focus on the perhaps most notable one, consequentialism:

(Consequentialism) An action A, performed by an agent in a concrete situation S, is *wrong* if and only if there was another action B that the agent could have performed instead of A in S such that, if the agent had done B, and not A, in S, the consequences would have been better.

An action A is *right* if and only if it is not wrong.

Consequentialists differ about what constitutes a good consequence and a bad consequence, but we need not go into that controversy here. Regardless of this disagreement, they all need to be able to specify what it means for an agent, who performs a certain action, to be able to do something different from what she does. The general meaning of ‘could have’ (or ‘was able to’) has been widely discussed in the literature on free will.¹ However, relatively little attention has been given to the question of how to understand the expression in this specific context. But without an answer to this question, the consequentialist criterion of right and wrong action fails to yield determinate verdicts about the deontic status of actions when we apply it to real scenarios. We argue that the following analysis roughly provides the required answer:

(The simple conditional analysis) An agent who did A in S could have done B instead of A in S if and only if, had the agent decided to do B in S, she would have done B, and not A, in S.

The simple conditional analysis was often advocated by compatibilists up to the 1960s. Early renditions were put forward by Thomas Hobbes and David Hume.² During the twentieth century its proponents included G. E. Moore and A. J. Ayer.³ However, due to various counter-examples proposed in the second half of the twentieth century, the simple conditional analysis is nowadays generally dismissed.⁴ We here show that, even if it fails as a general analysis of the meaning of ‘could have,’ a slightly modified version of it is exactly what we are looking for in a criterion of right and wrong action.

We first argue that a conditional analysis helps us find the most plausible interpretation of consequentialism. We then show, after having distinguished our view from a dispositional view, that all the standard objections to it turn out to be irrelevant when we restrict the analysis to the consequentialist criterion of right and wrong action. Objections such as the regress argument, the coma argument, the phobia argument, the argument from unintentional actions and the manipulation argument can all be dealt with, *inter alia*, by distinguishing between acting wrongly on the one hand, and being an apt target for blame and punishment on the other hand.

II. Why Accept the Conditional Analysis?

Two considerations count in favour of understanding ‘could have’ in normative criteria of rightness and wrongness along the lines of the simple conditional analysis.

¹ We will use the expressions ‘could have’ and ‘was able to’ interchangeably.

² Hobbes 1962, Hume 1955.

³ Moore 1912, Ayer 1954.

⁴ See, e.g., Campbell 1951, Broad 1952, Chisholm 1964: 20–25, Lehrer 1968, and van Inwagen 1983: 114–119.

First of all, this understanding focuses on the right thing; it focuses on the way people can, *through their decisions*, control their *actions*. In this it differs from similar conditional analyses cast, for example, in terms of 'wants.' The issue that we have set out to settle here is a semantic one: how should we understand the expression 'could have' in order to get the most plausible version of the consequentialist criterion of rightness and wrongness? In formulating the criterion we have made the obvious choice to restrict moral evaluation to *actions*. We have not included entities such as character traits and wants, because we want a criterion of rightness and wrongness based on how we affect the world, and how we, in turn, can be influenced. We affect the world through our overt actions, and indirectly through our decisions, which are mental actions that typically precede in time overt actions. Certainly, such things as wants, character traits, and so forth, are of importance too, but the only 'channel' to them for us agents is through our actions. If we want to change our wants and character traits we have to do so through decisions and actions. To focus exclusively on actions and decisions here makes sense of a common practice. We constantly evaluate other people morally and tell them when they did something wrong with the intent to affect them to improve morally over time. Our criterion makes sense of this because it attributes 'wrongness' to the entities that are most easily affected, our actions, and because the set of alternatives is delineated in terms of decisions, which are actions and thus directly open to influence too.

When we consider with which alternative actions we should compare a concrete action in order to determine whether it is wrong, we are not interested in what the agent would have done if she had had different wants, a different character, different skills and so forth. We are rather interested in what would have happened if she had had the same wants and so forth, but had *decided* differently. In this context, a possible world where the agent makes a different decision is typically much closer to the actual world than a world in which she has different wants. The determination of the correct criteria for closeness between worlds is, we submit, a semantic rather than a metaphysical question. As David Lewis points out: "Counterfactuals are infected with vagueness, as everyone agrees. Different ways of (partly) resolving the vagueness are appropriate in different contexts."⁵ In the present context we have resolved this vagueness in the way that yields the most plausible version of consequentialism.

The second consideration counting in favour of the simple conditional analysis is that it allows, without relying on any controversial metaphysical assumptions, that agents in a world like ours sometimes act wrongly. It would count heavily against a putative consequentialist criterion if it turned out that it yields the result that all our actions are right, since there are no alternatives to them (or, as some may prefer to say, that it doesn't for this reason even apply to our actions). It is generally assumed

⁵ Lewis 1979: 457.

that 'ought' implies 'can.' If an agent cannot abstain from performing an act, then it is not true that she ought to have abstained. And wrongness can be defined in terms of 'ought': an action is wrong if and only if the agent ought to have abstained from performing it. Thus, if the agent could not have abstained, the action is not wrong. So, if no one could ever have acted otherwise, then no one ever acts wrongly. In any criterion of right and wrong action referring to alternatives, it therefore makes sense to stipulate a meaning of 'could have' that allows that we, at least sometimes, have alternatives.

This is achieved on the simple conditional analysis. Regardless of whether our world is deterministic or not, there seem to be many situations where it is true that if the agent had decided to act differently, then she would have done so. In the closest possible world(s) in which she decides to act differently, she acts differently. And the possible world(s) where the agent decides to act differently don't have to be physically possible, i.e., they don't have to be compatible with the actual history up to the decision, together with the laws of nature operating in the actual world. If determinism is true, then a possible world where the agent decides to act differently is not, from the point of view of the actual world, physically possible. Such a world has a past which is different from the actual past, or laws of nature that are different from the actual laws, or both. Most philosophers hold that agents do not have access to such a world.⁶ But the simple conditional analysis does not require that the agent has access (whatever that may mean) to a world where she decides to act differently, and consequently does so. The only thing that is required is that in the closest world(s) where she decides to act otherwise, she acts otherwise. We think that this requirement is met in many situations. There seem to be many cases where it is true that if the agent had decided to perform an action B, which she in fact did not perform, then the agent would have performed B. Remember, too, that this is a semantic question, and not a metaphysical one. And as far as we can see, the assumption that this requirement is sometimes fulfilled is compatible with physical determinism, as well as physical indeterminism, or at least the kind of physical quantum indeterminism some philosophers claim to be true of our universe.

III. The Categorical Notion

As mentioned above, the general meaning of 'can' and 'could have' has been widely debated with regards to free will and moral responsibility over the last 50 years or so. Many philosophers have suggested that it should be understood in a categorical sense,

⁶ See, e.g., van Inwagen (1975 and 1983) and Fischer (1994). Their claim is challenged by Lewis (1981), Perry (2004, 2008) and to some extent by Saunders (1968). For an evaluation of these different positions, see Svedberg 2014.

with no 'ifs.' However, it seems to be very hard to spell out this categorical sense. Peter van Inwagen once remarked that how to analyse 'can' "is one of the most difficult questions in philosophy."⁷ What appears to be even more difficult is to establish whether or not we have the ability expressed by this categorical 'can.'⁸ Arguments have been presented that such ability is incompatible with determinism.⁸ Other arguments are supposed to show that it is incompatible with indeterminism.⁹ The soundness of these arguments is still under debate, with no foreseeable solution in the near future. Thus, for all we know, it might very well be true that no agent could ever have acted otherwise in a categorical sense. It might even be the case that having this ability is impossible, because the very idea is incoherent.

So, if we understand 'could have' in a categorical sense in a criterion of right and wrong action, then that criterion may turn out to entail that no one ever acts wrongly simply because nobody could ever have acted otherwise. As indicated in the previous section, such an implication strikes us as highly counter-intuitive. The task we have set out to achieve here is to formulate a criterion to capture the common intuition that people sometimes act wrongly because they do not perform the action they could, and should, have performed.

Here someone may retort that there are fully viable ideas of agents having the ability to act otherwise in a categorical sense both in a deterministic universe (e.g., David Lewis's local miracle compatibilism) and in a universe where determinism is false (e.g., agent causation).¹⁰ But these metaphysical ideas are controversial, and our point is that even if it were to turn out that no one ever could have acted otherwise in a categorical sense, the intuition that agents sometimes act wrongly still remains. This indicates that a categorical analysis of 'could have' makes poor sense of the consequentialist criterion of rightness and wrongness. On the simple conditional analysis, on the other hand, we do not have to presume any of these controversial metaphysical ideas to ensure that agents sometimes could have acted otherwise.

IV. More on Metaphysics

This is not to say that it is completely uncontroversial to claim that the simple conditional analysis implies that we sometimes could have acted otherwise. If necessitarianism of the strong variety sometimes attributed to Spinoza is true, i.e., if there are no possibilities, then it makes poor sense to query about what an agent would have done, had she decided differently. In that case, the simple conditional

⁷ Van Inwagen 1975: 188–189.

⁸ See, e.g., Ginet (1966, 1990), Wiggins (1973), van Inwagen (1975, 1983), and Lamb (1977).

⁹ See, e.g., van Inwagen 2000.

¹⁰ Lewis 1981. For ideas of agent causation, see, e.g., Chisholm (1966, 1971, 1976) and O'Connor (1995, 1996, 2002).

analysis seems to imply that no agent could ever have acted otherwise. Or maybe it rather implies in each particular case that the agent both could and could not have acted otherwise. We say this, since someone may be prepared to argue that, if there are no possibilities, then all claims about what would have happened, had an agent decided differently, are true since the antecedent of all such claims is (vacuously) false. Thus, it is true that had the agent decided to act differently, she would have done so, and it is also true that had the agent decided to act differently, she would *not* have done so. In any case, such a strong metaphysical necessitarian view doesn't strike us as very plausible, and few have actually defended it. It has even been questioned if this was indeed Spinoza's view.¹¹ So, it is a virtue of the simple conditional analysis that, with the sole exception of necessitarianism, it is metaphysically neutral.

V. The Conditional Analysis vs. a Dispositional Analysis

We are rarely met with conditional analyses in contemporary discussions. We do meet with a somewhat similar view, however, in the idea that our freedom to act otherwise from how we do should be seen as a combination of a certain ability (a disposition) we have and an opportunity to exercise this ability.¹² This dispositional view invites the question of what constitutes the ability (disposition) in question. In what kind of mechanism is the disposition grounded? Can we analyse it in counterfactual terms? Here problems to do with 'finks,' 'masks,' and 'mimics' abound. To answer this kind of question has proved difficult.¹³ However, the simple conditional analysis is entirely different from the dispositional view. It does not invoke any idea of any (stable) disposition. It operates directly with the counterfactual conditionals relevant to the particular action we want to assess from a normative point of view. So we need not answer the questions that have proved difficult to answer in relation to the dispositional view. We do not claim that they cannot be answered. Our claim is merely that we need not answer them. And this is so since, on the simple conditional analysis, no reference is made to any (general) ability in the explanation of what it means to have it in one's power to act otherwise from how one does. No such disposition is necessary to our power in a situation to act differently from how we do act. As we will see in the sequel, the possession of such a general ability is not sufficient for the power to act otherwise from how one does either. All this is good news from the point of view of the simple conditional analysis. It is simple, clear, and it does answer all the questions that need to be answered within a context of the consequentialist criterion of right and wrong action. And as we will see, even if there are sound arguments

¹¹ See Newlands 2013.

¹² See, e.g., Vihvelin 2004 and Fara 2008 and, most ambitiously and, most recently, Vihvelin 2013.

¹³ See Clarke 2009 about the problems.

against the simple conditional analysis as a general analysis of ‘could have,’ these objections all turn out to be irrelevant when we confine our interest to the consequentialist criterion of right and wrong action.

VI. The Regress Argument

Let us start with what has been taken to be a serious argument against the plausibility of the simple conditional analysis, the regress argument.¹⁴ It typically takes the following form. In order for it to be true that an agent could have acted differently it does not suffice that she would have acted differently, had she decided to do so. It is also required that she could have decided differently. But what does it mean that the agent could have decided differently? Suppose that we would answer this question by re-applying the conditional analysis, and thus claim that it means that the agent would have decided differently, had she decided to decide differently. Then the same problem would arise on the next level: could the agent have decided to decide differently, and what does that mean? If we keep re-applying the conditional analysis an infinite regress sets off. To stop it we would have to suggest a categorical analysis of ‘could have’ at some level. Hence, the simple conditional analysis fails.

This argument can be successfully dealt with if we confine our interest to a criterion of rightness and wrongness. The simple conditional analysis does not require that the agent could have decided differently, and in this context this is as it should be. We are looking for a suitable way of delineating the alternatives an action should be compared with, if we want to determine its deontic status. It makes sense to consider every counterfactual action of which it is true that, had the agent decided to perform that action, she would have done so. And it makes sense not to place any further requirements on the decisions in question. The regress does not get started.

But when we say of an agent that if she could have performed B rather than A, then there is a possible world in which she decides to perform B rather than A, do we not then *imply* that she can also decide to do B? We don’t. The simple analysis neither claims, nor implies that the agent must be able to decide differently, in order for her to be able to act differently.

All this does not preclude that, often when an agent could have acted differently on the simple conditional analysis, it is *also* true of her that, *as a matter of fact*, she could (somewhat earlier) have decided differently. To decide to do something is tantamount to forming an intention to do it, i.e., *to decide is to perform an action*. Thus, it makes sense to say that an agent can, or that an agent cannot, decide to do something else than she in fact decides to do. She could have decided differently if, and only if, had she decided to decide differently, she would have done so.

¹⁴ See, e.g., Broad 1952 and Chisholm 1964.

It is rare, we concede, that people decide to decide one way or another. This does not mean that people cannot do so, however, and there are interesting cases where it makes sense to decide what to decide. Suppose an agent knows that she has intransitive preferences, and suppose that she is about to meet with someone who is eager to exploit her weakness through pumping money from her.¹⁵ The agent will be offered successive choices where, if she decides to accept them, she will pay successively for transitions eventually leading back to the original state she was in. Knowing this, our agent now decides to decide not to accept any offers at this upcoming meeting. Consequently, she decides not to accept any offers at the meeting and saves her money.

Thus, it makes good sense to say that it can be true that an agent is also able to decide otherwise on the simple conditional analysis. And the fact that people seldom decide to decide something does not preclude that it is often true that *if* they had decided to decide differently, then they would have done so. But what about the cases in which they *cannot* decide differently? Is it not an indictment of the simple conditional analysis that it implies that agents in such cases can act otherwise as long as it is true that *if* they had decided to act otherwise, they would have done so?

It is not, if we restrict the analysis to the consequentialist criterion of right and wrong action. If we were to give a general account of freedom one might rightly wonder why we bother to insist upon a conditional analysis of freedom of action, while claiming that ability to decide otherwise is not required.¹⁶ But we merely want to stipulate a meaning of 'could have' that generates the most plausible version of consequentialism. For the present purposes, it is reasonable to delineate the set of alternatives to an action by looking at what would have happened *if* the agent had decided to act differently. If we also were to require that the agent *could* have decided differently, it seems that we would indeed, on pain of the regress argument, have to abandon the conditional analysis in favour of some kind of categorical analysis. However, we have already stated our reasons for not opting for such an analysis. It is doubtful whether it is even coherent, let alone such that we ever could have acted differently from how we do act. Thus, on such an analysis the consequentialist criterion may turn out to imply that no one ever acts wrongly. This implausible implication is avoided on the simple conditional analysis.

VII. The Coma Argument

Peter van Inwagen presents a counter-example to the simple conditional analysis along the following lines. Suppose that Smith is in a coma at a hospital. The analysis seems to

¹⁵ See Andreou 2016.

¹⁶ See, e.g., Berofsky 2002: 188.

imply that Smith still can get out of bed. After all, it seems to be true that if Smith had decided to get out of bed, then he would, simply because if Smith did decide to get out of bed, he would not be in a coma. So the simple conditional analysis is faulty since it has the absurd implication that a person can get out of his bed at a time when he in fact is comatose.¹⁷

We agree that this shows that the simple conditional analysis fails as a general analysis of what it means for a person to be able to perform an action she does not in fact perform. But in the restricted use that we have put it to here in the consequentialist criterion, this objection does not apply. In this context, we are solely interested in specifying what it means to say of a person *who actually performed* an action that instead she could have performed a different one. The consequentialist criterion of rightness and wrongness applies only to actions. Since a person in a coma performs no action at all the criterion does not apply to her. Hence, it will never yield the implausible verdict that a person does something wrong while in a coma. So, we do not have to worry about the implications of the simple conditional analysis here.

But are there not cases where a person acts wrongly when she virtually does nothing? Imagine a person who is on guard when a fire breaks out. She is supposed to sound the alarm in case of a fire, but she just sits there. She hears the sound of fire, she smells smoke, and she actually sees flames next to her. However, she does absolutely nothing. Provided that the consequences of sounding the alarm would have been better, sounding the alarm is what she should have done in this situation. It was wrong of her not to undertake any active measure.

The consequentialist criterion does accommodate this intuition, since in such a situation the agent did act after all. To omit to do something while being conscious is to perform an act. Even if the agent did not act in the sense of performing either any overt physical act, or any internal mental act (to perceive is not to perform an act, we concede), technically speaking, her inaction, must in this context be seen as an act. It is hard to believe that it could take place in real life, and if it did take place, it must be because of some resolution the agent had made not to react in any way. But to *stay* in that passive stage is indeed to 'act,' in a sense relevant to consequentialism, and actually, in relation to all reasonable normative theories. The agent's inaction can be seen as intentional, given her resolution to be in this state.

This is different from the case where a person does not react in any manner because she has fallen asleep. To *be* asleep is not to 'act' in any sense relevant to the consequentialist—or, for that matter, to any reasonable—criterion of rightness and wrongness of actions. It is neither right, nor wrong to *be* asleep. To *allow* oneself to fall asleep is an action, however, and thus apt for moral evaluation.

¹⁷ Van Inwagen 1983: 119.

It might be thought that this allows for only an easy way of avoiding acting wrongly. Suppose that Jill is now at time t_0 . What she ought to do within a few minutes (at t_1) is to call her friend, arrange with a meeting later in the evening (at t_2), where she settles a dispute with him. Instead Jill takes a sleeping pill (at t_0). While asleep between t_1 and t_2 Jill does nothing wrong according to consequentialism, even if the consequences would have been better if she had arranged a meeting and settled the dispute between t_1 and t_2 . Is that counterintuitive?

We think not, since consequentialism still implies that Jill did *something* wrong, at some point in time, that led to the dispute not being settled at t_2 . Let us quite reasonably assume that if Jill had decided (before t_0) not to take the sleeping pill at t_0 , then she would not have taken the pill at t_0 , she would have called her friend at t_1 and she would have met with him at t_2 , and settled their dispute. Assuming that the consequences of settling the dispute are better than the consequences of not settling the dispute, Jill then acts wrongly according to consequentialism when she takes the sleeping pill at t_0 .

Yet, what if, had Jill not taken the pill at t_0 , she would as a matter of fact have called her friend at t_1 , but quarrelled with him at t_2 rather than settling the dispute with him? Well, in that case, her taking the pill at t_0 may very well have been morally all right. At least this is our firm intuition, which is defended by one of the authors elsewhere.¹⁸ Others may disagree.¹⁹

What both parties should (must) agree about, however, is that, *while asleep*, Jill committed no action at all and, hence, no wrong action.

We need to apply the criterion of rightness and wrongness of actions to particular times or time intervals, then. This allows us to say that, regarding the longest time interval in our example, stretching from t_0 to t_2 , what Jill ought to do is avoid taking the sleeping pill, call her friend, and settle the dispute with him. We may assume that this is something she would do, if she decided to do it. Hence, she acts wrongly when she takes the pill and sleeps through the interval. However, once asleep, there is no time at which she performs any moral mistake—this follows from the simple fact that she performs no action whatever during this time, let alone any wrong action.

But it might be retorted that these cases with a person who is asleep or comatose do not really capture the underlying worry here. We escape these counter-examples because it is clear that such a person does not perform any action. But what about the following slightly modified example. A young person plays in her garden. This person is sufficiently developed to perform actions apt for moral evaluation. However, she has not yet acquired any advanced philosophical concepts. She has little formal education and knows nothing about philosophy and the problem of how to define knowledge.

¹⁸ See Tännsjö 1985.

¹⁹ See Smith 1986.

Suppose that the act of her thinking about the content of the Gettier problem at this specific time would have had better consequences than her actual act of playing around. Doesn't she act wrongly then, on our understanding of the consequentialist criterion? It seems to be true that *if* she had decided to think about the content of the Gettier problem, then she would have done so. This task is not that difficult, so it seems reasonable that if an agent has the concepts that are necessary to *decide* to perform it, then she also has the concepts that are necessary for her to succeed in performing it.²⁰

We concede that she should have thought about the content of the Gettier problem. Given her age and circumstances, her thoughts about it would of course have been extremely shallow, but even so, if her thinking about it would have had better consequences, she should have thought about it. This does not mean that she would have formed any intelligent view about the problem.

But isn't the closest possible world where she decides to think about the content of the problem a world in which she has previously taken some kind of philosophy course? Then it is not only true that she would have thought about the content of the problem, had she so decided, but also that she would have thought about it in a rather competent manner after all. In that case, it seems that there are a lot of things that she can do on the simple conditional analysis, like, for instance, repeat a proof for the completeness of first order predicate logic. In the closest possible world in which she decides to do so, she has taken at least a one semester course in logic. So, if that action would have had the best consequences, that is what she should have done, even though she in fact has not taken any such course. However, in this context such worlds are *not* the closest, since they involve a certain kind of backtracking. When we evaluate actions from a moral point of view, worlds where the agent is very different from how she actually is, where she has previously taken philosophy courses, and so forth, are not admissible.

When we claim that the only admissible possible worlds are those where the agent keeps her traits of character, her wishes, and skills, we must acknowledge that sometimes no possible world exists where the agent makes a certain different decision, viz., a decision requiring a different character, different wishes and skills (and hence different previous actions). Does this mean that all claims about such alternative decisions are vacuously true? We take no stand on this issue. However, we claim that under the circumstances, where there is no such possible world where the agent as she is makes the required decision, the agent could not perform the action in question. But doesn't that imply that an agent with an evil character never acts wrongly when she tortures or kills people because there is no possible world where she has the same character and decides to abstain from these actions? It does not, we submit, because

²⁰ This kind of example was suggested to us by Jens Johansson and Olle Risberg, among others.

there *is* a possible world in which such an agent acts against her character by deciding to abstain and sticks with that decision. In that world the agent has the same traits of character, wishes, etc., but at this one occasion she acts “out of character,” so to speak.²¹

This is not to say that *every* world with a different past is inadmissible. We have taken a neutral stance towards most metaphysical problems but we have assumed that there are possibilities and that there is a possible world where a person decides to perform, and performs, another action than the one she actually performs. If determinism is true, we have noted, this implies either that the person can perform an action which is at variance with some law of nature, or that she can perform an act such that, had she performed it, the past would have been different (or both). One of the authors has defended the backtracking solution.²² Most adherents of backtracking in this context maintain that the differences in the past would have been miniscule throughout the whole history.²³ This is roughly the tack we want to adopt, should the backtracking solution prove to be correct.

But then, how can this young girl think about the content of the Gettier problem? She will certainly not think of it under that label, since she has not heard of Gettier. But she might for example look at her non-functioning watch (she would do so if she decided to do so) and note that it shows 1 p.m. (we assume that it does). Now, she knows for independent reasons that the present time is indeed 1 p.m. She might think (and would do so, we assume, if she decides to ponder about the question): What if someone looked at my watch and didn’t know it was not functioning? That person would still believe it is 1 p.m.! We think this counts as thinking about the content of the Gettier problem, and given certain decisions on her part, this is what she would and hence should do.

Thus, on the simple conditional analysis, our young agent could have thought about the content of the Gettier problem, but she would have done so in a very shallow (childish) and yet clever manner. If that action would have had better consequences than her actual act of playing around, she acts wrongly. An ordinary person, with no experience of any courses in logic, could not, however, on the simple conditional analysis, have repeated a proof of the completeness of first order predicate logic. Assume that, as a matter of fact, she overhears that there is such a proof. In the closest world in which she decides to repeat the proof, she has not taken any course previously. However, given that she knows that there exists such a proof, it is possible for her, being the kind of person she is, with the talents (and lack of talents) she possesses, to make a childish attempt to recite it. Now she fails due to the

²¹ See Tännsjö 1995: 123–124 on this.

²² Tännsjö 1998: 143–144.

²³ This claim is challenged by Ginet (1990: 107–110).

complexity of the task, however. So even if that action is the one with the best consequences, it is not the action she should have performed.

However, all this does not preclude, of course, that we assess also a temporal interval, including where the agent both much earlier abstains from taking a logic course when she was offered to attend it, and then ends up in a situation where there would have been the best consequences had she attended it. But then it is her act of abstaining from taking that course that is wrong. And provided she did not in fact take the course, she doesn't act wrongly now, when she doesn't repeat the proof.

But the fact that the consequentialist criterion implies that our young agent acts wrongly when she does not think about the content of the Gettier problem may cause some to call for an amendment. They find it implausible that an agent should perform an act which she, as a matter of fact, cannot perform intentionally. She hasn't heard of Gettier and cannot, for this reason, think of his problem under this label. A similar example that comes to mind is an agent who fails to open a safe because she doesn't know the combination. On the simple conditional analysis, she could have opened the safe, because if she had decided to put in the specific combination that is in fact the correct one, she would have done so and thereby opened the safe. So, if the consequences of her opening the safe would have been better, then she acts wrongly when she in fact fails to open the safe, according to the consequentialist criterion.

We take no definitive stance on this issue here, but will return to the question in the upcoming section on unintentional action.

VIII. The Phobia Argument

Suppose a person has a pathological fear of blood and thus shuns everything of the colour red. This person is offered a choice between a green and a red apple. She takes the green apple. It may seem as though, due to her phobia, she could not take the red apple. However, it may well be true that *if* she had decided to take the red apple, she would have done so. The simple conditional analysis implies, then, that she could have taken the red apple. Hence the implication it delivers is counter-intuitive.²⁴

Even A. J. Ayer, who came close to accepting the conditional analysis, at least in the 1950s, thought that it needed to be revised in light of this objection.²⁵ But there are many possibilities we need to consider before we draw this conclusion.

First of all, a strong aversion to red objects may be seen as a strong evaluative stance, so strong that it usually trumps other values. This is compatible with the fact that in some rare situations, this person would decide to take the red apple, and take it

²⁴ Lehrer (1980) has, among others, claimed that this argument is decisive against the simple conditional analysis. See Lehrer 1980: 188.

²⁵ Ayer 1954: 282.

too. Perhaps she would do so if the life of her children were at stake. So it is not that the phobia renders her unable to decide to take the red apple, it just renders her extremely *unwilling* to make such a decision. Thus, in most, or even all, actual situations when she is offered a choice between something red and something of a different colour, she ends up not deciding to take the red object.

But that does not imply that she can't decide to take the red object. She never decides to take the red because she never wants to. And that doesn't imply anything about her abilities to do it. If this is so, the phobia argument against the simple conditional analysis fails since it then seems reasonable to claim that she could have taken the red apple. And, in particular, if the consequences of her doing so would have been better, she acted *wrongly* when she not did do so.

Here is another possible diagnosis allowing us to escape the objection. It may be true of this person that even if she were to decide to have the red apple, such a decision would not be effective. She would not act on it. This is how the phobia works; it renders ineffective certain decisions. We could compare with the decision the addict makes not to use the drug, only to find that, when the opportunity to use the drug is there, she does use it. It is debatable whether this kind of addiction exists. Is it not more plausible to understand the behaviour even in this case as an instance of idiosyncratic and extremely strong values? Perhaps there is, after all, some price that would make this person act on her decision?

Since this is debatable, we do not take a stand here. If genuine addiction exists, we may see it as an instance of the phenomenon here adumbrated. But then, on the simple conditional analysis, we are allowed to say that people who suffer from addiction or phobia cannot act otherwise from how they act. In the closest possible world where they decide to act otherwise, they don't. So our agent doesn't act wrongly by not taking the red apple, even if that act would have had better consequences.

However, a way of pressing the argument further is as follows. The phobia does not operate as a very strong value judgement, it could be held, nor does it make the decision to take the red apple inefficacious; had she decided to take it she would have taken it; it is rather that the phobia renders this person unable to make that decision. Because of her phobia, she cannot decide to take the red apple. However, in the closest possible world where she decides to take the red apple, she takes it. In that world her phobia is absent, or does at least not block her decision and her follow-through. How does the simple conditional analysis handle this case?

Well, it implies that the person could have taken the red apple, since, had she decided to do so, she would have taken it. This scenario seems to be the one that has been taken to show that the simple conditional analysis needs amendment. Psychologically speaking, the example seems perhaps a bit far-fetched. It is not likely that we meet in reality with this kind of person. If her phobia renders her unable to

make the decision, it is also likely that it renders her unable to perform the action. Her decision to decide, in the former case, and her decision to act, in the latter case, are likely to *both* be rendered inefficacious by her phobia. Still, the scenario remains a logical possibility.

She is able to take the red apple, then, but she is not able to *decide* to take it. It is not only that she is now, when she actually takes the green apple, unable to make another decision because it is too late. She was also unable to make the decision to take the red apple instead of the green one a few seconds ago. Even if she had earlier decided to decide to take the red apple, and had committed herself to doing so, once the opportunity surfaced, she would still have decided *not* to take the red apple. Her decision to make a different decision would have been inefficacious. This is where the phobia kicks in.

But then her *decision* not to take the red apple was morally permissible, even if the decision to take the red apple would have had better consequences. To decide to take the red apple was not an option for her on the simple conditional analysis. However, her *action* of not taking the red apple, but to take the green instead, is wrong according to the consequentialist criterion.

Should we really say of this person that she acted wrongly when she took the green apple? We think we should. In our attempt to make the best possible sense of the consequentialist criterion, we have made the choice to focus on how we control our actions with our decisions. The simple conditional analysis cast in terms of decisions captures how we think about alternatives when we subject actions to moral evaluation. Whether or not we control our decisions is not of decisive importance in this context. Maybe we do not have ultimate control over any of our decisions because they are completely determined by the laws of nature together with the remote past, or because they are mere chance events. But this does not mean that we never act wrongly. So it may well be that a person does not have any control over her decision in a particular situation because she suffers from phobia. But this does not deprive her of the ability to *act* otherwise, in the relevant sense. Whether or not this person can, in any sense, *decide* to act otherwise is not crucial to the deontic status of the act she in fact performs. The only thing that matters is whether it is true that *if* she had decided to act otherwise, she would have done so.

So, if we speak abstractly of criteria of right and wrong action, it is indeed plausible to say of our agent that she acted wrongly when she took the green apple. If we speak of whether she should be blamed, or even punished, for not having taken the red apple, things may be different, however.

It is of note, that on the simple conditional analysis, the consequentialist criterion yields different verdicts in this example here depending on how the phobia operates. This is appropriate in the present context. In our criterion we have chosen to focus on morally evaluating actions and to delineate the proper set of alternatives, we have

chosen to focus on what would have happened if the agent had decided to act differently.

However, there is one complication here that shows that the simple conditional analysis needs to be slightly amended, after all. It becomes most obvious on the account where the phobia is taken to render the person's decision to take the red apple inefficacious. We stated that in this case, her actual act of taking the green apple is morally permissible, even if taking the red apple would have had better consequences. But that might not be the plausible verdict, since there may be some *other* decision such that if she had made *that* decision, then she would have taken the red apple. Maybe if she had decided to shut her eyes and just reach for the nearest apple, then she would have taken the red apple. If any such conditional is true, then it seems that there is something that she could, and should, have done instead of taking the green apple. So, she acts wrongly in this case too.

In light of this, we suggest that 'could have' in the consequentialist criterion should be understood as follows:

(The amended conditional analysis) An agent who did A in S could have done B instead of A in S if and only if, there is a decision such that had the agent made that decision, she would have done B, and not A, in S.

In standard cases, the relevant decision is the decision to perform the action B. However, in some situations like the one adumbrated above, an agent needs to decide to perform a different alternative action in order to perform the relevant action B, without any intention to perform B. It is of interest to see that this slight amendment also means that in the scenario where the phobia is taken to render our agent unable to *decide* to take the red apple, there might be some other decision that she is able to make and that would have led her to taking the red apple. This doesn't, of course, change our stance that being able to *decide* otherwise is not required, in this context, for the agent to be able to act otherwise. Still, it is worth noting.

IX. Coercion and Manipulation

Do cases of coercion and manipulation pose a problem for the simple conditional analysis?

First of all, think of a situation when someone says, 'Give me your money, or I'll kill you.' You hand over your money. If it is true that, had you decided to refuse, you would have refused, then this was indeed an option for you. A. J. Ayer has discussed the case and he concurs. However, he also adds:

[I]f the circumstances are such that no reasonable person would be expected to choose the other alternative, then the action that I am made to do is not one for which I am held to be morally responsible.²⁶

But note that this is not a revision of the simple conditional analysis as such. It is a moral point made in *addition* to it; we should not *hold* people responsible, when they act wrongly in circumstances such as this one. We suppose the idea is that one should not *blame* them for what they have done. And, of course, it is highly likely that the consequences of blaming them in this situation would be bad, so it is indeed a good idea not to blame them. In particular, this seems to be the case if, as we do, we assess matters from the point of view of consequentialism.

What if what we meet with is rather a clever demon who has wired your brain in a way that allows her to implant desires in you and mould your decisions and who takes advantage of this possibility. So you do A because you decide to do A and you decide to do A because you want to do A, but your want is instilled in you. Now it is true that, had you instead decided to do B, you would instead have done B—but once again merely as an instrument of the demon. Does this mean that the consequentialist criterion on the simple conditional analysis yields an implausible verdict in this case?

The crucial thing here seems to be how we regard the notion of agency. What does it mean to perform an action in the first place? We should perhaps not say that you perform any action at all, if what you do is in this manner dictated by a demon. But then there is nothing to capture for the consequentialist criterion and the simple conditional analysis. If you do not perform any action, then it is not meaningful to ask whether you could have acted otherwise, and whether what you did was wrong. Then we should instead focus exclusively on the normative status of the actions performed by the *demon*. To ponder whether *you* did the right thing is superfluous.

However, given that your actions track your decisions, and your decisions track your wishes, and given that the actions of the demon are counterfactually dependent on what you do, our intuition is to claim that you do act after all. Who knows, this may be our *ordinary* position in the world! It is only when the demon meddles with the links between your wishes, your decisions, and your actions, that the criterion doesn't apply to what you do.

Here so-called Frankfurt-style cases come to mind.²⁷ Jones decides to, and does, murder Smith. Black is waiting in the wings, ready to ensure that, were Jones to decide not to murder Smith, Jones would end up, somehow or other, murdering Smith anyway. Does this mean that Jones acts wrongly, even in the absence of any alternative action?

²⁶ Ayer 1954: 279.

²⁷ See Frankfurt 1969.

It does not. To decide to murder Smith, and then to murder him, and to decide not to murder Smith, and then murder him (regardless of the mechanism here operating) are different actions. According to consequentialism, it is all right if Jones intentionally murders Smith, since, regardless of whether he performs this action, Smith will be murdered. According to deontology, however, the difference could be of the utmost importance. Jones acts wrongly when deliberately he murders Smith. He should have acted differently. He should have decided not to murder Smith. The outcome would have been the same, but he would have avoided violating an important deontological constraint. The focus here, however, *is* on consequentialism. So what Jones is facing are two different alternatives, but each of them is, according to consequentialism, all right for him to perform.

X. Unintentional Action

J. L. Austin famously objected to the simple conditional analysis as follows:

I miss a very short putt and kick myself because I could have holed it. It is not that I should have holed it if I had tried: I did try, and missed.²⁸

At the beginning of this article we noted that, on the simple conditional analysis, we need not assume that a person has any (general) ability to perform a kind of action in order to be able to perform an instance of it in a particular situation. No reference to any general ability is *necessary* in the analysis. One could see Austin's example as an illustration of the lack of *sufficiency* of such a general ability as well. No (general) ability is in itself sufficient for a person to be able to perform an alternative action to the one she does perform. You may have a general ability to hole putts in situations such as this one and yet, for all that, fail to do it in a particular situation (where your general ability fails you). But this need not be the end of the story. In this example we would say that Austin might very well be right when he claims that (in the situation) he could have holed his putt. But we can account for this in terms of the simple conditional analysis, now that we have expanded it to cover also unintentional action.

There are some movements such that, had he decided to undertake them, he would have holed the putt. But he did not make *this* decision. Furthermore, *intentionally* there was probably no way for him to hole the putt. This is so because he was ignorant about the exact movements he should have opted for instead of the ones he performed. He performed an action that resulted in his not holing the putt. He would have succeeded, if he had acted slightly differently, but his success would not have been intentional. His action that would have holed the putt would have been

²⁸ Austin 1966: 308.

intentional in the sense that the *movements* he made would have been intentional (the ones he decided to make). However, the *holing* would not have been intentional.

Suppose it was important that he holed it. Suppose that he could not have holed it intentionally. In the closest world, where deliberately he performs the exact movement it takes to hole the putt, he doesn't know that this is a unique way of holing it. Should we yet say that he *should* have made the correct movement, which would have resulted in his holing it? Should we say that what he actually did was wrong?

Again, we think we should, if we take our normative theory to provide a consequentialist criterion of right and wrong action. After all, there was a better alternative open to Austin, unbeknownst to him, and had he gone for it, the world would have been a better place (we assume). Yet, we should keep in mind that right and wrong action is one thing, questions about praise and blame (not to speak of punishment) is quite another thing. When a person does the wrong thing without knowing that what she does is the wrong thing, it makes poor consequentialist sense to blame her for what she does. To blame and praise people is also to act and, at least according to consequentialism, we should only blame or praise people where this has the best consequences. And it is plausible to assume that, when we blame people for wrongdoings they have committed without any intention to act wrongly, our blame is counterproductive. Even the deontologist may want to keep the possibility open that there exists blameless wrongdoing.²⁹

On this point normative theorists tend to disagree, however. This is so also with normative theorists of a consequentialist bent. Some want to introduce some kind of 'epistemic filter' in their statement of the criterion of right and wrong action, in order to narrow down the distance between questions of right- and wrongdoing on the one hand, and questions of praise and blame on the other.³⁰ An agent is not morally obliged to perform actions, which she can only perform unknowingly, they claim. Thus, this epistemic filter constitutes a moral restriction, and not a metaphysical one.

We see no good reason to enforce any such restriction. The crucial thing is to ensure that our criterion of wrong action does not get unnecessarily narrow. The criterion should allow all comparisons that may in some situation turn out to be relevant to the agent's decision-making. As a limiting case, the criterion should provide an omniscient creature with all the resources needed for moral guidance in each particular situation. Questions of blame and punishment are only indirectly related to questions about the wrongness of our actions. In particular, this is so, if our point of departure is consequentialism.

²⁹ See Tännsjö 2015: chap. 5 about this.

³⁰ See Dancy 2000: 66. The idea stems from Bergström 1996, to which Dancy makes a reference.

XI. Conclusion

The simple conditional analysis makes best sense of the consequentialist criterion of right and wrong action. It allows us to compare actions to be assessed for their deontic status with the relevant class of alternatives (those actions the agent would have performed, had she decided to act differently).

Our focus in the paper has been the consequentialist criterion of right and wrong action. As a matter of fact, there seems to be room for some generalisation here, however. The simple conditional analysis seems to meet the needs of many competing criteria of right and wrong action as well, such as the idea that an action is wrong if it means that a deontological constraint is violated, if it means that a moral right is violated, and so forth. All these criteria require that there exist alternatives to the actions that are performed; otherwise all the actions we perform will turn out trivially to be right. And the simple conditional analysis does not in itself preclude that there are such alternatives available to the agent.

This is not to deny that there are other areas of morality, such as Kantian retributivism, where we *may* need a different analysis of what it means to be able to act otherwise.³¹ If the claim is made that, if deliberately you perform wrong actions, you acquire ‘guilt’ in a sense that renders punishment morally necessary, then it may well be true that some different, perhaps indeterminist, analysis of what it means to be able to act otherwise is required. Such views, which make *further* requirements regarding what it means to have the ability to act otherwise than we do, run the risk of being vacuous, however. We hold no definite view on this. The discussion of this question must be the matter for another occasion.

Notes

Jens Johansson read and commented upon the penultimate draft of this paper and made valuable suggestions. Earlier versions were presented by Torbjörn Tännsjö at the joint Uppsala/Stockholm seminar on 11 March 2016, and by Maria Svedberg at the Agency Workshop at the University of California, Riverside, on 14 February 2017. We thank all the participants for valuable comments.

References

Andreou, C. 2016. “Dynamic Choice,” in *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/dynamic-choice/>.

³¹ See Ryman 2012 about this.

- Austin, J. L. 1966. "Ifs and Cans," in *Free Will and Determinism*, ed. B. Berofsky, 295–322. New York: Harper and Row.
- Ayer, A. J. 1954. "Freedom and Necessity," in *Philosophical Essays*, 271–284. London: McMillan.
- Bergström, L. 1996. "Reflections on Consequentialism," *Theoria* 62(1–2): 74–94.
<https://doi.org/10.1111/j.1755-2567.1996.tb00531.x>
- Berofsky, B. 2002. "Ifs, Cans and Free Will: The Issues," in *The Oxford Handbook of Free Will*, ed. R. Kane, 182–202. New York: Oxford University Press.
- Broad, C. D. 1952. *Ethics and the History of Philosophy*. London: Routledge & Kegan Paul.
- Campbell, C. A. 1951. "Is Free Will a Pseudo-Problem?," *Mind* 60: 446–465.
- Chisholm, R. M. 1964. "J. L. Austin's Philosophical Papers," *Mind* 73(289): 1–26.
<https://doi.org/10.1093/mind/LXXIII.289.1>
- Chisholm, R. M. 1966. "Freedom and Action," in *Freedom and Determinism*, ed. K. Lehrer, 11–44. New York: Random House.
- Chisholm, R. M. 1971. "Reflections on Human Agency," *Idealistic Studies* 1(1): 33–46.
<https://doi.org/10.5840/idstudies1971116>
- Chisholm, R. M. 1976. "The Agent as Cause," in *Action Theory*, ed. M. Brand and D. Walton, 199–211. Dordrecht: D. Reidel. https://doi.org/10.1007/978-94-010-9074-2_12
- Clarke, R. 2009. "Dispositions, Abilities to Act and Free Will: The New Dispositionalism," *Mind* 118(470): 323–351. <https://doi.org/10.1093/mind/fzp034>
- Dancy, J. 2000. *Practical Rationality*. Oxford: Oxford University Press.
- Fara, M. 2008. "Masked Abilities and Compatibilism," *Mind* 117(468): 843–865.
<https://doi.org/10.1093/mind/fzn078>
- Fischer, J. M. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell.
- Frankfurt, H. G. 1969. "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66(23): 829–39. <https://doi.org/10.2307/2023833>
- Ginet, C. 1966. "Might We Have No Choice?," in *Freedom and determinism*, ed. K. Lehrer, 87–104. New York: Random House.
- Ginet, C. 1990. *On Action*. New York: Cambridge University Press.
<https://doi.org/10.1017/CBO9781139173780>
- Hobbes, T. 1962. *The English Works of Thomas Hobbes*, vol 5, ed. W. Molesworth. London: Scientia Aalen. Originally published in 1654.
- Hume, D. 1955. *An Enquiry concerning Human Understanding*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press. Originally published in 1743.
- Lamb, J. W. 1977. "On a Proof of Incompatibilism," *Philosophical Review* 86(1): 20–35.
<https://doi.org/10.2307/2184160>
- Lehrer, K. 1968. "Cans without Ifs," *Analysis* 29(1): 29–32.
<https://doi.org/10.1093/analys/29.1.29>

- Lehrer, K. 1980. "Preferences, Conditionals, and Freedom," in *Time and Cause*, ed. P. van Inwagen, 187–201. Dordrecht: D. Reidel.
https://doi.org/10.1007/978-94-017-3528-5_11
- Lewis, D. K. 1979. "Counterfactual Dependence and Time's Arrow," *Noûs* 13(4): 455–476.
<https://doi.org/10.2307/2215339>
- Lewis, D. K. 1981. "Are We Free to Break the Laws?," *Theoria* 47(3): 113–121.
<https://doi.org/10.1111/j.1755-2567.1981.tb00473.x>
- Moore, G. E. 1912. *Ethics*, Oxford: Oxford University Press.
- Newlands, S. 2013. "Spinoza's Modal Metaphysics," *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/spinoza-modal/>.
- O'Connor, T. 1995. "Agent Causation," in *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. T. O'Connor, 173–200. New York: Oxford University Press.
- O'Connor, T. 1996. "Why Agent Causation?," *Philosophical Topics* 24(2): 143–158.
<https://doi.org/10.5840/philtopics199624218>
- O'Connor, T. 2002. "Libertarian Views: Dualist and Agent-Causal Theories," in *The Oxford Handbook of Free Will*, ed. R. Kane, 337–355. New York: Oxford University Press.
- Perry, J. 2004. "Compatibilist Options," in *Freedom and Determinism*, ed. J. K. Campbell, M. O'Rourke, and D. Shier, 231–54. Cambridge, MA: MIT Press.
- Perry, J. 2008. "Can't We All Just Be Compatibilists? A Critical Study of John Martin Fischer's *My Way*," *Journal of Ethics* 12: 157–166.
<https://doi.org/10.1007/s10892-008-9030-2>
- Ryman, P. 2012. "On Free Will as Categorical and Conditional Freedom." Uppsala University, Dissertation.
- Saunders, J. T. 1968. "The Temptations of 'Powerlessness,'" *American Philosophical Quarterly* 5: 100–108.
- Smith, H. M. 1986. "Moral Realism, Moral Conflict, and Compound Acts," *The Journal of Philosophy* 83(6): 341–345. <https://doi.org/10.2307/2026095>
- Strawson, G. 1994. "The Impossibility of Moral Responsibility," *Philosophical Studies* 75(1): 5–24. <https://doi.org/10.1007/BF00989879>
- Svedberg, M. 2014. "The Consequence Argument: An Essay on an Argument for the Incompatibility of Free Will and Determinism." Stockholm University, Dissertation.
- Tännsjö, T. 1985. "Moral Conflict and Moral Realism," *The Journal of Philosophy* 82(3): 113–117. <https://doi.org/10.2307/2026350>
- Tännsjö, T. 1995. "Blameless Wrongdoing," *Ethics* 106(1): 120–127.
<https://doi.org/10.1086/293781>
- Tännsjö, T. 1998. *Hedonistic Utilitarianism: A Defence*. Edinburgh: Edinburgh University Press.

- Tännsjö, T. 2015. *Taking Life: Three Theories on the Ethics of Killing*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190225575.001.0001>
- van Inwagen, P. 1975. "The Incompatibility of Free Will and Determinism," *Philosophical Studies* 27: 185–99. <https://doi.org/10.1007/BF01624156>
- van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- van Inwagen, P. 2000. "Free Will Remains a Mystery," *Philosophical Perspectives* 14: 1–19.
- Vihvelin, K. 2004. "Free Will Demystified: A Dispositional Account," *Philosophical Topics* 32(1/2): 427–450. <https://doi.org/10.5840/philtopics2004321/211>
- Vihvelin, K. 2013. *Causes, Laws and Free Will: Why Determinism Doesn't Matter*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199795185.001.0001>
- Wiggins, D. 1973. "Towards a Reasonable Libertarianism," in *Essays on Freedom of Action*, ed. T. Honderich, 31–62. London: Routledge & Kegan Paul.41