Philos Stud (2011) 155:307–324
DOI 10.1007/s11098-010-9574-4

Responsibility for necessities

Stephen Kearns

Published online: 5 June 2010

Abstract It is commonly held that no one can be morally responsible for a necessary truth. In this paper, I will provide various examples that cast doubt on this idea. I also show that one popular argument for the incompatibility of moral responsibility and determinism (van Inwagen's Direct Argument) fails given my examples.

Keywords Necessity · Moral responsibility · Free will · Determinism · Incompatibilism · Compatibilism · Direct Argument

The following idea is extremely intuitive and popular: (necessarily, for any proposition p) if it is necessary that p, then no one is even partly morally responsible for the fact that p. This idea plays a prominent role in discussions of the compatibility or otherwise of responsibility and determinism. Most notably, van Inwagen (1980, 1983), in arguing for the incompatibility of responsibility and determinism, makes use of the following inference rule: infer that no one is even partly morally responsible for p from the fact that p is necessary. He calls this inference rule 'Rule A'. Let us call the corresponding proposition that (necessarily) if it is necessary that p, then no one is even partly morally responsible for the fact that p, simply 'A'. After van Inwagen introduced this rule, it met with universal

S. Kearns (✉)
Department of Philosophy, Florida State University, 151 Dodd Hall, Tallahassee, FL 32306-1500, USA
e-mail: Stephen.a.kearns@gmail.com; skearns@fsu.edu

acceptance in the literature (as far as I am aware). Eleonore Stump and John Martin Fischer say that Rule A 'seems entirely uncontroversial' (in Stump and Fischer 2000, reprinted in Fischer 2006, p. 180). Ted Warfield says that it is '(nearly) as trivial and inconsequential as a rule of inference could be.' (1996, pp. 218–219). Indeed, philosophers often do not even address its validity in discussions of the Direct Argument, so obvious does it seem (see, for example, Haji 2008; McKenna 2008).

I think this universal consensus is mistaken. In this paper I will propose several counterexamples to A (the falsity of which in turn invalidates Rule A). Before this, I will set out van Inwagen's Direct Argument for the incompatibility of moral responsibility and determinism to show the typical role Rule A plays in arguments for incompatibilism. After I have discussed the counterexamples, I shall return to the Direct Argument.

1 The Direct Argument

The Direct Argument makes use of a non-responsibility operator, NR. Van Inwagen understands NRp to mean the following (type of) proposition: proposition p is true and no one at any time is even partly morally responsible for the fact that p. We may also relativize NR to agents and times. Understand $NRp_{a,t}$ to mean the following: proposition p is true and agent a, at t, is not even partly morally responsible for the fact that p. No point I make in this paper hangs on whether we relativize NR to agents and times or not. Any counterexample to Rule A understood as relativized to an agent and a time will also count as a counterexample to Rule A not thus relativized. Using this operator, van Inwagen introduces two rules of inference:

$$\text{Rule A: } \Box p \vdash NRp.$$

$$\text{Rule B: } NR(p \supset q),\ NRp \vdash NRq.$$

Rule A says that if a proposition is necessarily true then we can infer that no agent is (even in part) morally responsible for the fact that p. Rule B says that if no agent is (partly) morally responsible for a fact, p, nor for the fact that if p then q, then we may infer that no agent is (partly) morally responsible for the fact that q. Rule A and Rule B as relativized to agents and times tell us to make these exact inferences regarding certain agents and certain times. Armed with these rules, we may produce an argument for the conclusion that determinism is incompatible with moral responsibility. (In the following argument NR should be understood to be relativized to the agent, Bob. For ease of expression, I will drop the subscripts. Nothing of importance rides on this.)

Consider an agent, Bob, who donates money to a charity. Let us call the fact that Bob donates this money to charity 'C'. Let us call the conjunction of the laws of nature 'L' and an exact (and full) description of a state of the world at some time before Bob's birth 'P'. Determinism is understood to be the thesis that the laws of nature and an exact (and full) description of the world as it is at any one instant entails all other truths. Now consider the following argument:

1. □ (P & L ⊃ C)	Assumption (of determinism).
2. □ (P ⊃ (L ⊃ C))	1, and Logic.
3. NR (P ⊃ (L ⊃ C))	2, Rule A.
4. NR P	Premise.
5. NR (L ⊃ C)	3, 4, Rule B.
6. NR L	Premise.
7. NR C	5, 6, Rule B.

In effect, the argument is this. Assuming that determinism is true of the world in which Bob donates money to charity, then the past (before Bob's birth) and the laws of nature of that world together entail that Bob does just this. Bob is in no way morally responsible for the past before his birth, and is equally not morally responsible for the laws of nature. Rule A and Rule B together allow us to infer that Bob is thus not morally responsible for donating money to the charity.

Of particular interest given the topic of this paper is the role Rule A plays in the argument. This rule allows us to infer from the fact that it is necessary that if P then, if L, then C (i.e. the fact that it is necessary that if the distant past is as it is then, if the laws of nature obtain, then Bob donates money to charity) that Bob is not even partly morally responsible for the fact that if P then, if L, then C. The plausibility of this inference seems to rest almost entirely on the plausibility of the idea that no one is even partly morally responsible for any necessary truths. In the rest of this paper I will give various counterexamples to this thesis, thus undermining the validity of the move from premise 2 to premise 3 in the above argument. As we shall see in Sect. 7, showing that we are able to be morally responsible for necessary truths also undermines premise 6 (that no one is morally responsible for the laws of nature) and Rule B.

2 Murder!

Let us say that Stephen murders someone. Furthermore, it is completely uncontroversial that Stephen is morally responsible for the fact that he murders someone (except to those who think that moral responsibility is impossible—in this paper I am assuming that moral responsibility is at least *possible*). He does so knowingly and intentionally, he could have done otherwise, he is aware of the wrongness of his action, etc. Thus Stephen is responsible for the fact that he murders someone. This being so, it is also clear that he is responsible for the fact that he *actually* murders someone. However, the fact that he actually murders someone is necessarily true. It is true in every possible world that, in the actual world, Stephen murders someone. Therefore, Stephen is (partly) morally responsible for a necessary truth. A is false.

This is, I think, a clear-cut case of a person's being responsible for a necessary truth. Though the proposition that Stephen actually murders someone is necessary, it is still true that, at least in some sense, Stephen *makes it the case* that he actually murders someone. The actual world is the way it is in part because of what Stephen

does. Furthermore, the idea that alternative possibilities are necessary for moral responsibility is compatible with this counterexample to A. Though the fact that Stephen actually murders someone could not have failed to obtain, there is a relevant action of Stephen's (the murder) that Stephen could have refrained from performing. If Stephen had not murdered someone, though it still would have been the case that he actually murders someone, he would not be morally responsible for the fact that he actually murders someone. Also, Stephen may still be considered the undetermined agent-causal source of his action even if we consider him responsible for this necessary truth (we may even consider him the undetermined source of this necessary truth!). Lastly, the fact that Stephen actually murders someone makes Stephen the appropriate subject (in the actual world) of certain moral reactive attitudes, such as indignation and resentment. This is enough for Stephen to be morally responsible for this fact.

Some may worry that, when described in a certain way, it seems strange to say that Stephen is responsible for actually murdering someone.[1] Let us call the actual world 'w1'. The claim that Stephen is responsible for the fact that he actually murders someone, then, is equivalent to the claim that Stephen is responsible for the fact that, in w1, he murders someone. It is, however, jarring (or, at the very least, not intuitive) to think of agents as being responsible for such facts. Indeed, it is unclear that we think of world-indexed facts much at all outside of the philosophy classroom, and as such, we cannot be expected to have clear intuitions about whether we are morally responsible for them.

My response to this worry is that, though there may be unnatural (or unusual) ways of *putting* the claim that Stephen is morally responsible for the fact that he actually murders someone, this does not show that the *claim itself* is in fact unintuitive or unusual. Rather, it simply shows that there are unfamiliar ways of making uncontentious points. In essence, my idea is that it is intuitive and natural to say that Stephen is morally responsible for actually murdering someone. (In general, the one type of world-indexed propositions that people *are* used to is the type expressed with sentences containing the word 'actually', or a synonym). Insofar as we want to say that Stephen is responsible for a *fact*, we may rewrite this to say that Stephen is responsible for the fact that he actually murders someone. Furthermore, this claim is equivalent to the odder sounding claim that Stephen is responsible for the fact that, in w1, he murders someone. It is odder sounding, however, only in the same way that the claim that grass instantiates the property of greenness sounds odder than the claim that grass is green. If we accept one (and we are happy with an ontology of possible worlds/properties), we should accept the other.

One last worry is that the way 'actually' works in the English language is not how it works in the language of philosophers. Thus it is simply *not* the case that the (ordinary English) claim that Stephen actually murders someone and the claim that, in w1, Stephen murders someone are equivalent. One basis for this claim might be that 'actually' in the English language is an indexical and so cannot rigidly pick out a particular possible world. But indexicals quite standardly rigidly pick out particular things ('I', in my mouth, rigidly designates me, 'here', in my mouth,

[1] I am grateful to an anonymous referee for making this point.

rigidly designates the place where I am located, etc.), so I see no tension in the ideas that 'actually' is an indexical and rigidly designates a particular possible world. (Also, one need not be a Lewisian about possible worlds to think that 'actually' is an indexical.) Furthermore, there is evidence that, whether an indexical or not, 'actually' does rigidly designate the actual world. It would be hard to make sense otherwise of sentences such as "If Peter were to drink that vodka, he'd be even more drunk than he actually is." I conclude, then, that we should take plausibility of the claim that Stephen is responsible for actually murdering someone at face value (see above for further reasons why we should accept this claim) and we should also accept that this claim expresses a necessary truth. It quickly follows that Stephen is responsible for a necessary truth.[2]

The most common reaction I receive to this case is to concede that it is indeed a counterexample to A. Compatibilists often leave it there, happy that another weapon has been removed from the incompatibilist's armory. Incompatibilists go onto make the following point. The fact that Stephen actually murders someone is a world-indexed necessity (that is, it is a necessary truth explicitly concerning one particular possible world). That makes it a special type of necessary truth. In the light of the counterexample, A can be amended as follows. A* states that (necessarily, for any proposition p) if it is necessary that p (and this necessary truth is not world-indexed) then no one is even in part morally responsible for the fact that p.

Furthermore, A* (and the corresponding Rule A*, which tells us to infer that no one is morally responsible for fact p from the fact that p is a non-world-indexed necessity) can be used in arguments for incompatibilism just as successfully as A (and Rule A) can. To see why this is so, consider the Direct Argument again. If some version of the Direct Argument is to work, it needs to validate the inference from the proposition that it is necessary that if P then, if L, then C to the proposition that no one is partly morally responsible for the fact that if P then, if L, then C. But Rule A* allows us to do just that, because the fact that if P then, if L, then C is a non-world-indexed necessity. Thus the Direct Argument can rest on the weaker Rule A* rather than Rule A.

This incompatibilist reply is too quick. Firstly, it is of interest independently of the question of the compatibility of responsibility with determinism that A is false. Secondly, the reply does not diagnose *why* A is false and why we should think that A* will fare any better. After all, the main intuitive argument for why we cannot be responsible for necessary truths seems to apply to both world-indexed and non-world-indexed necessary truths. Intuitively, it seems that we cannot be responsible for such truths because they would be true *no matter what we do*. But if this is not sufficient grounds to show that we cannot be responsible for world-indexed necessary truths, why would it be sufficient grounds to show that we cannot be responsible for non-world-indexed necessary truths? More is needed than the bald assertion that while A may be false, A* is true.

[2] For another discussion of the idea that one can be responsible for world-indexed propositions, see Fischer (1996). In this paper, Fischer discusses the idea that *all* propositions are world-indexed. Though this is not the tack I take, if one accepted this and accepted that morally responsibility is possible, one would have to accept that one is responsible for necessary truths. This is because, as Fischer points out, all truths are necessary according to this account. To my mind, Fischer's criticisms of this idea are decisive.

Perhaps more can be provided, based on certain claims I made above. As I said, it seems that, in some sense, Stephen makes it the case that he actually murders someone. Thus agents are (sometimes) able to make it the case that world-indexed necessary facts are the case. We may therefore provide the following argument for A* that does not apply equally to A. An agent is in part morally responsible for a fact only if the agent makes this fact obtain. No agent can make a non-world-indexed necessary fact obtain. Therefore, no one is in part morally responsible for any non-world-indexed necessary fact (that is, A* is true).

I think both premises of this argument are false. In the next three sections, I provide three further counterexamples to A (and A*) that also cast doubt on the second premise. In the section after this, I will criticize the first premise of the argument by providing cases that are counterexamples to both it and A (and A*).

3 Hey Jude

Paul McCartney composes the melody of Hey Jude. He does so knowingly and intentionally. He is morally responsible for various facts concerning Hey Jude. For instance, he is morally responsible for the fact that Hey Jude starts in the way it does and for the fact that Hey Jude has a beautiful melody. After all, he created Hey Jude. However, these facts about Hey Jude for which McCartney is responsible are necessary truths. It is necessarily true that Hey Jude starts in the distinctive way it does and that it has a beautiful melody. Thus McCartney is (partly) morally responsible for certain necessary truths. A is false.

If this counterexample is successful, it does three things. Firstly, it shows that A is false. Secondly it shows that A* is false (as the fact that Hey Jude starts a certain way is a non-world-indexed necessary truth). Thirdly, it shows that agents can make it the case that certain non-world indexed necessary facts obtain (which defeats the argument given above for A*). Let us look at the case more closely.

The above counterexample relies on two main ideas. Firstly, McCartney is responsible for the fact that Hey Jude starts in a certain way (and is beautiful). Secondly, it is necessary that Hey Jude starts this way (and is beautiful). Both of these ideas are at the very least *prime facie* plausible, and I believe this plausibility stands up to scrutiny.

Consider the first idea—that McCartney is responsible for the way in which Hey Jude starts (and for its beauty). McCartney knowingly and intentionally creates Hey Jude, including the way it starts. Let us say that he also agent-causes this creative action, and that, at the time of composing Hey Jude, he has many other alternatives open to him (he could have written a different song, he could have refrained from writing anything, etc.). As with Stephen's making it the case that he actually murders someone, McCartney can even be said to make it the case that Hey Jude starts in the way it does, and that it has the melody (and beauty) it has. These factors are very strong reasons to conclude that McCartney is responsible for the way in which Hey Jude starts (and for the fact that it is beautiful).

One objection to the idea that McCartney is responsible for the fact that Hey Jude starts the way it does is the following. There are many types of action that

McCartney performs when he composes Hey Jude (he composes a song, he creates something beautiful, he creates Hey Jude, etc.) and many facts for which he may be held responsible. Perhaps McCartney isn't responsible for the fact that Hey Jude starts in the way it does, or that Hey Jude is beautiful, but is instead responsible for other related facts, such as the fact that he composes Hey Jude. Furthermore, these facts are contingent and thus pose no problem for A or A*.

I think this reply fails, for two reasons. Firstly, the fact that Hey Jude starts in the way it does can intuitively be seen as a consequence of McCartney's composing it. Thus if McCartney is responsible for composing Hey Jude, he can sensibly be held responsible for the fact that it starts in the way it does. Secondly, it is perfectly appropriate to praise McCartney for the fact that Hey Jude starts in a certain way, and for the fact that Hey Jude has beautiful melody. This is even more obvious if we allow that it is appropriate to praise McCartney for composing Hey Jude. Imagine two people, one of whom praises McCartney for composing the beautiful melody of Hey Jude and one of whom praises McCartney for the fact that Hey Jude is beautiful. It is not plausible that one of these people is appropriate in praising McCartney and the other is not. Both exemplify perfectly fitting attitudes (given, again, that morally responsibility is possible at all) and the practical difference between them is negligible.

The second idea that the counterexample relies upon is that it is a necessary truth that Hey Jude starts in the way it does (and also that Hey Jude is beautiful). It is plausible that the beauty of Hey Jude supervenes on its melody. Thus if Hey Jude has its melody essentially, it has its beauty essentially. The fact that Hey Jude's melody starts in the way it does also supervenes on its melody. Thus if Hey Jude has its melody essentially, then Hey Jude has its start essentially. Furthermore, it is plausible that Hey Jude simply *is* its melody. Thus it is necessary that Hey Jude has the melody it does. If Hey Jude had had a different melody, *it wouldn't have been Hey Jude* (in some sense, then, one *can't* take a sad song and make it better!). Perhaps some elements of Hey Jude aren't essential to it (its key or its tempo, for example), but many of its distinctive features are essential to it. I take it that these points are relatively obvious.

Of course, someone who believes that objects (even abstract objects like Hey Jude) have only the barest of essences will disagree (I am thinking of people who would find it acceptable to claim that the number three could have been an apple). However, such people are in an extreme minority. Hey Jude is, in essence, a certain melody and it could not have been a different melody. Thus it is a necessary truth that Hey Jude starts in a certain way, and is beautiful. Even so, it can be appropriate to hold McCartney responsible for these exact facts. This shows that A is false. It also shows that A* is false. And insofar as McCartney can be said to *make it the case* that Hey Jude starts in a certain way, it also shows that the idea that agents cannot make non-world-indexed necessary facts obtain is false. This will be further brought out in the next example.

4 Torturing babies

Karen likes to make certain actions wrong to perform. For instance, she decides to make eating cakes wrong by creating a machine that kills thousands of people if

anyone eats a cake. She then informs everyone that she has done this. It is now wrong to eat a cake, as it would lead to the death of many people. Furthermore, Karen is morally responsible for the fact that eating cakes is wrong. Karen then decides to build a similar machine that kills thousands of people if anyone tortures a baby for their own enjoyment. She then informs everyone that she has done this. Similarly, then, she is morally responsible for the fact that it is wrong to torture babies for one's own enjoyment as she has ensured that doing so will lead to the death of thousands of people. However, it is necessarily true that torturing a baby for one's own enjoyment is wrong. This is because there are *other* reasons that doing so is wrong—reasons that necessarily apply. In a sense, then, Karen overdetermines the fact that it is wrong. She is thus (partly) morally responsible for a necessary truth (that torturing babies for one's own enjoyment is wrong). A is false. As the fact that torturing babies simply for one's own amusement is a non-world-indexed necessary truth, A* is also false.

This counterexample relies on two main ideas. First, it is necessary that torturing babies simply for one's own amusement is wrong. Second, Karen is (in part) morally responsible for this necessary fact. Let us examine both claims.

I imagine that almost everyone will acquiesce to the first claim. There are, however, a few people who will not. There are the psychopaths who do not think that torturing babies is wrong. We may (safely?) ignore such people. There are those philosophers who think that moral claims are not truth-apt (non-cognitivists) and those that think that (atomic) moral claims are truth-apt but all false (error-theorists). I cannot here defend the view that there are some moral truths, so I shall simply take it as read. This does not mean, however, that there is nothing to say to such philosophers. Firstly, many non-cognitivists accept a way in which (certain) moral claims are true (see, for example, Blackburn 1993). Such quasi-realists will probably be happy with my claim that it is necessarily true that torturing babies for one's own enjoyment is wrong. Secondly, similar examples to that above can be constructed concerning, not wrongness, but what one has sufficient reason to do or avoid doing (or concerning what it is rational to do or avoid doing). Unless one is a non-cognitivist or an error theorist about reasons claims and rationality claims such examples are sufficient to prove my point.

Let us assume, then, that it is a necessary truth that torturing a baby merely for one's own pleasure is wrong. I further claim that, given the case as specified above, Karen is in part morally responsible for this fact. Just as she is responsible for the fact that it is wrong to eat cake, she is similarly responsible for the fact that it is wrong to torture babies merely for one's own pleasure. Why is this? The story should be somewhat familiar by now. She acted intentionally, she knew what she was doing, she had many alternatives open to her, she agent agent-caused her actions, etc. Furthermore, in a sense, she *made it* wrong to torture babies merely for pleasure. Additionally, it seems perfectly appropriate to praise or blame her for the fact that it is wrong to torture babies merely for pleasure. Lastly, the case is exactly parallel to that of Karen's making it wrong to eat cakes, which is a clear case of her being responsible for a relevantly similar fact. Thus, she is responsible for this fact.

Some might object that Karen is not in fact morally responsible for the fact that it is wrong to torture babies (for one's own enjoyment) because she does *not* make this

fact obtain. There is *already* a perfectly good (and complete) explanation of why torturing babies for one's own enjoyment is wrong, namely that it causes an innocent and defenseless human extreme agony for no discernable benefit.

This objection seems to rest on the idea that if a fact has a complete explanation that obtains independently of an agent's activity, then the agent's activity does not explain the fact and thus the agent is not (even in part) morally responsible for this fact. This idea is false, as illustrated by Fischer and Ravizza's Erosion* example. Fischer and Ravizza first explain their example, "Erosion", as follows:

> In this example, Betty plants her explosives in the crevices of the glacier and detonates the charge at *T1*, causing an avalanche that crushes the enemy fortress at *T3*. Unbeknownst to Betty and her commanding officers, however, the glacier is gradually eroding (in a certain way). Had Betty not placed the dynamite in the crevices, some ice and rocks would have broken free at *T2*, starting an avalanche that would have crushed the enemy campy at *T3*. (Fischer and Ravizza, 1999, p. 250)

Fischer and Ravizza use this example to show the falsity of Transfer NR, which states that if no one is even partly morally responsible for P and no one is even partly morally responsible for the fact that if P then Q, then no one is even partly responsible for Q. (A is to Rule A what Transfer NR is to Rule B). In response to a modification of Transfer NR, they modify Erosion (calling the new case "Erosion*"). In Erosion*, Betty causes an avalanche at the exact same time as the erosion does. Betty's actions and the erosion *overdetermine* the destruction of the enemy camp.

This example helps us see the following. Betty, in Erosion*, is morally responsible for the destruction of the enemy camp. We shall assume that she agent-causes the detonation that leads to its destruction, that she could have done otherwise, that she does so knowingly and intentionally, etc. It is entirely clear (again, except to people who think moral responsibility is impossible) that she is morally responsible for the camp's destruction. It is equally clear that Betty makes it the case (through her actions) that the camp is destroyed. Despite this, however, there is a complete explanation for the destruction of the camp that is completely independent of Betty's actions (namely the erosion's causing an avalanche). Betty can be said to have made it the case that the enemy camp is destroyed even though something completely independent of Betty also makes it the case that the enemy camp is destroyed.

I suggest that the same is true in the case of Karen's being responsible for the fact that it is wrong to torture babies for fun. To be sure, there is a complete explanation independent of Karen's actions that explains why torturing babies merely for fun is wrong. However, Karen's actions provide another, equally legitimate explanation of why doing so is wrong. Isn't it simply *obvious* that the fact that torturing babies for fun would lead to the deaths of thousands of people a very strong consideration against torturing babies for fun? Indeed, outside some very freakish circumstances, it is a *sufficient* reason not to torture babies for fun. The people in Karen's world have two sufficient reasons not to torture babies for fun—one which necessarily applies and one which contingently applies. Either way, given that Karen's actions

result in one of these sufficient reasons applying, she can correctly be said to have made it the case that it is wrong to torture babies for fun. And given this (plus the fact that she is in all other ways a paradigm example of a free agent), she is morally responsible for this necessary truth.

5 Fermat's last stand

Anna believes she can refute Fermat's Last Theorem (FLT). She believes this rationally and on the basis of extremely good evidence. Of course, as we now know, it is impossible for her to refute FLT because it is necessarily true. Anna does not know this however. Anna deliberates about whether or not to refute FLT and then intentionally (and knowingly, etc.) leaves FLT unrefuted. I wish to claim that, in this case, Anna is morally responsible for leaving FLT unrefuted (and thus for the fact that she leaves FLT unrefuted) even though it is impossible for her to refute it. She is morally responsible for the necessary fact that she leaves FLT unrefuted. She is thus (partly) morally responsible for a necessary fact. A is false, as is A*.

Before I defend the idea that Anna is responsible for the necessary fact that she leaves FLT unrefuted, I wish to consider the question of whether Anna can be said to have made this fact the case. My own inclination is to say that she does make it the case. I shall spell out why in what follows.

In all of the previous cases, I have claimed that an agent makes some necessary fact obtain. It would be nice to have a general test for when an agent makes a fact obtain and when she does not. I propose what we may call the "Why-Test". This says that an agent makes a fact obtain if some relevant fact about the agent's actions can be cited in an explanation of why this fact obtains. Consider the counterexamples to A that I have proposed thus far. All of these seem to pass this test. Why is it the case that Stephen actually murdered someone? A perfectly acceptable answer to this question is "Because Stephen murdered someone." Why does Hey Jude start in the way it does? Because McCartney wrote it like that. Why is Hey Jude such a great song? Because McCartney has considerable songwriting talent and used that talent when composing Hey Jude. Why is it wrong to torture babies for fun? Because Karen has ensured that if anyone does so, thousands of people will die. (Of course, there is an entirely independent answer to this last question that is also perfectly acceptable). The Why-Test also rules out an agent's making other necessary truths the case. The question "Why does $2 + 2 = 4$?" does not have any acceptable answer that appropriately refers to the activity of any agent (with the possible exception of God). The same applies to "Why is water identical to H_2O?" and "Why is it the case that, if someone is a bachelor, then that person is unmarried?" No agent makes these necessary facts obtain.

Now consider our current case. Does Anna make it the case that she leaves FLT unrefuted? Let us apply the Why-Test. Does the question "Why does Anna leave FLT unrefuted?" have an answer that appropriately refers to Anna's actions? It seems to me that it does. One good answer to this question is: "Because she decided she didn't want to refute it". Of course, there are other very good answers to the question that do not refer to Anna's actions (e.g. "Because FLT is necessarily

true"), but as long as some do, then this is good evidence that her actions go some way to explaining why the fact obtains. Indeed, this may be a case in which the truth of the proposition that Anna leaves FLT unrefuted is overdetermined. Either way, the Why-Test suggests that Anna makes it the case that she leaves FLT unrefuted. Insofar as this is a reasonable test, my conclusion is reasonable. (If one does not accept this conclusion, one need not immediately conclude that Anna is not morally responsible for leaving FLT unrefuted. As we shall see in the next section, an agent may still be morally responsible for a (necessary) truth even if she does not *make* it true.)

Why think that Anna is responsible for the fact that she leaves FLT unrefuted? The reasons are familiar. Anna intentionally and knowingly leaves FLT unrefuted, despite rationally believing she could have refuted it. Indeed, she could have tried very hard to refute it. She agent-causes her actions. It is perfectly appropriate (in this instance) to hold certain reactive attitudes towards her for leaving FLT unrefuted. She *makes it the case* that she leaves FLT unrefuted. All of this is very good evidence that she is responsible for leaving FLT unrefuted.

A further reason to think she is responsible is that her case is in many ways parallel to other cases in which a person is morally responsible for a fact. Consider John Locke's classic case in which a person is carried into a room while he sleeps and is then locked in (see Locke 1690, book 2, Chap. 21, Sect. 10). When he awakes, the person decides to stay in the room of his own free will. It is plausible that this person is responsible for staying in the room, despite the fact that he is unable to do other than stay in the room. He is thus responsible for the fact that he stays in the room.

I hope the parallels between the above case and Anna's are clear. The man is unable to render a certain fact false, but is still responsible for this fact. Of course, part of the reason why he is so responsible may be due to the fact that he could have (unsuccessfully) attempted to leave the room, but this is equally true of Anna (mutatis mutandis)—she could have attempted to refute FLT. What seems irrelevant in the case of the man locked in the room is the fact that there are other possible worlds in which he is not locked in the room. Seeing as these worlds are inaccessible to him, their existence *does no work* in showing that he is responsible for staying in the room. If someone were of the (admittedly rather odd) opinion that it is necessary that the man stays in the room (though not necessary that he voluntarily stays in the room), it would still be perfectly reasonable for this person to conclude that the man is responsible for staying in the room. After all, how would the mere addition of inaccessible possible worlds in which he does not stay in the room help ground his responsibility? Given this, it would be unsurprising to find similar cases in which a person is responsible for a fact that she cannot render false and she cannot render it false precisely because this fact is necessarily true. I suggest that Anna's case is one such case. Her inability to refute FLT rests on FLT's necessity, while the man's inability to exit the room rests on the fact that the room is locked. Either way, if the man can be held responsible for staying in the room, Anna can be held responsible for leaving FLT unrefuted.

Not everyone agrees with my intuitions about the responsibility of Anna and Locke's man. Van Inwagen (1978) describes a case in which he fails to ring the

police after having seen an altercation outside his window (p. 204). Unbeknownst to him (in the case in question), however, the telephone system is down and he would have been unable to have called the police even if he had tried. In van Inwagen's judgment, he is not responsible for failing to phone the police in such a situation. I presume van Inwagen would similarly say that the man in the locked room is not responsible for staying in the room, and that Anna is not responsible for leaving FLT unrefuted. For what it's worth, I think van Inwagen *is* responsible for failing to ring the police even though, if he were to try, he would fail. He intentionally and knowingly fails to ring the police, he easily could try to call them, he has excellent evidence that he would succeed if he tried, etc. Furthermore, the telephone system's being down does not play any role at all in van Inwagen's actions. His failure to call the police, though ensured by the system's being down, is not caused by it. Even if intuitions here are not clear enough to establish the truth either way, it seems obvious that *if* van Inwagen is responsible for failing to call the police then Anna is responsible for leaving FLT unrefuted. Insofar as van Inwagen's case can reasonably be seen as one in which he is responsible, the same can be said of Anna's case.

One might further worry that my claim about Anna does not tally with what is perhaps the most developed account of responsibility for omissions currently available, namely Fischer's and Ravizza's (see especially Fischer and Ravizza 1999).[3] Before I discuss how their account conflicts with my claim, it is worth seeing how certain statements Fischer and Ravizza make may be used to *support* the idea that Anna is responsible for leaving FLT unrefuted.

Van Inwagen uses his example to support the claim that one can be responsible for an omission only if one could have done what one omitted to do. Fischer and Ravizza (rightly in my opinion) reject this idea. There are cases, according to them, in which an agent is morally responsible for omitting to do something even though she is unable to do what she omits. One such case they mention is given by Randolph Clarke, and runs as follows:

> Sam promises to babysit little Freddy. But Sam forgets. No one makes Sam forget; it just slips his mind. Consequently, he fails to show up to babysit little Freddy. Unbeknownst to Sam, a mad scientist is monitoring his thoughts. Had Sam been going to remember his promise, the scientist would have intervened and prevented him from remembering it. The scientist would not have intervened in any other way. As it happened, the scientist did not intervene at all; there was no need to. (Clarke 1994, p. 203, quoted in Fischer and Ravizza 1999, p. 128)

This case (which is clearly in the mold of Harry Frankfurt's famous counterexamples to the Principle of Alternative Possibilities, see Frankfurt (1969)), is one in which Sam is morally responsible for an omission even though he is unable to do what he omits. Fischer and Ravizza concur that Sam is responsible for his omission (they also mention similar examples given in Frankfurt

[3] I am grateful to an anonymous referee for highlighting several of the points that follow.

(1994) and McIntyre (1994) in which an agent is responsible for an unavoidable omission).

One lesson we might take from Clarke's case (and others like it) is that, just as Sam is unable to babysit Freddy yet is still morally is responsible for failing to do so, Anna is unable to refute FLT, and Locke's man is unable to leave the room, yet both are still responsible for their respective omissions. That is, one very simple way of keeping our judgments about these cases consistent and coherent is to judge that the agents in *all* such cases are morally responsible. This is again to bring out that the similarity Anna's case bears to certain other cases lends support to the idea that Anna is morally responsible for leaving FLT unrefuted. In particular, Clarke's case makes it easier to head off at the pass the following kind of reply to Anna's case and Locke's man's case: "We should not say that Anna is responsible for leaving FLT unrefuted, but rather that she is responsible merely for not *trying* to refute FLT. Similarly, Locke's man is not responsible for staying in the room, but only for not trying to leave the room." We may ask of anyone who gives this response why the same does not apply to Sam. He is morally responsible for *not babysitting*, not merely failing to *try* to babysit. Given this, it is (I claim) reasonable to conclude that Anna is responsible not merely for failing to try to refute FLT, but also for leaving it unrefuted.

Fischer and Ravizza do not take this lesson from Clarke's case (nor, I should point out, does Clarke, though for different reasons). They believe that while Sam is morally responsible for his omission, Locke's man is not morally responsible for his (they state this explicitly in footnote 20, p. 137, of their 1999). They would also say the same of Anna's omission. Their reason stems in part from certain other cases in which it is (they claim) intuitive that an agent is not responsible for an unavoidable omission. In one such case, "Sharks" (see Fischer and Ravizza 1999, p. 124), John sees a drowning child as he walks on the beach. He decides not to save the child, despite thinking himself perfectly able to do so. Unbeknownst to John, sharks are patrolling the area and would have eaten John had he attempted to save the child. In another case, "Sloth" (p. 124), John decides against saving the child. The child dies seconds after this decision, which makes it effectively impossible for John to save the child in any case. Fischer and Ravizza judge John not morally responsible for his omissions in each of these cases.

The essence of Fischer and Ravizza's account of responsibility for (complex) omissions[4] centers around two questions. First, are the agent's bodily movements moderately responsive to reasons? And second, is the relevant consequence-universal[5] sensitive to these bodily movements? If the answer is "yes" to both questions the agent is morally responsible.[6] If the answer is "no" to either question,

[4] The authors distinguish bodily omissions from complex omissions. Bodily omissions are wholly constituted by an agent's bodily movements (thus a failure to raise one's left hand may be constituted by placing one's left hand by one's side). A complex omission involves a bodily movement and some other state of the world (such as FLT being unrefuted). Fischer and Ravizza call these states 'consequence-universals' (for more on these, see p. 96).

[5] See note 4.

[6] Assuming that certain other conditions are met (such as that the agent takes responsibility for the mechanism that leads to the bodily movement).

the agent is not responsible. According to Fischer and Ravizza, the difference between cases like that featuring Anna and cases like Clarke's is that, in the latter cases, the agent's consequence-universal is sensitive to the agent's bodily movements, while in the former cases, it is not. Thus in Clarke's case, Sam's not babysitting Freddy is sensitive to Sam's bodily movements, while in my case, Anna's leaving FLT unrefuted is completely insensitive to Anna's bodily movements.

What I take Fischer and Ravizza's account to show is that, once one solidifies one's judgments in the way they have about certain cases, one is able to build a coherent and elegant theory that explains and captures these judgments. Reflective equilibrium in action! However, I believe that one may just as reasonably conclude, as I have, that *all* of these cases are ones in which the agent is truly responsible for her omission (*ceterus paribus*). One may begin with solid intuitions concerning certain cases (such as Clarke's), plus allegiance to certain plausible principles such as the idea that, if a fact has no causal relevance to what an agent does (or refrains from doing), then it has no bearing on her moral responsibility for her action or omission. One then further notes that one's intuitive cases show that one need not be able to do something in order to be responsible for *not* doing it. Cases can then be built, such as "Sharks", in which one is unable to do something, one refrains from doing it, but what makes one unable to do it is not causally relevant to one's refraining from doing it. Thus even though the sharks are not a purely counterfactual threat (as is the mad scientist in Clarke's case), they still have no bearing on John's responsibility, as they are causally irrelevant to John's omission. Given this (and that John and Sam are equal in other regards), our principle implies that we should hold both responsible. Given the commitments with which we started, the fact that Sam's omission is sensitive to his bodily movements and John's is not seems irrelevant—what *makes* John's omission insensitive to his bodily movements is, in the actual sequence, completely irrelevant to his omission.

Once we have come this far, other cases, such as Locke's locked room, ought to be judged similarly. And in all of these cases, we have exemplary agents who intentionally and knowingly *make it the case* that they omit to do something, in which they reasonably believe they are able to do this thing and in which their ability to do this thing is removed by something that is not causally relevant to their omission. The difference between such cases to which Fischer and Ravizza point looks, from this way of thinking, too insubstantial to ground a difference in responsibility between such cases. One may then easily extend this way of thinking to the case of Anna. Anna intentionally and knowingly *makes it the case* that she leaves FLT unrefuted. What makes her unable to refute FLT (that FLT is necessary) plays no causal role in Anna's omission, and plays no role at all in the manner in which she goes about this omission (i.e. knowingly, intentionally, while under the impression she can refute FLT, etc.). Thus even though her leaving FLT unrefuted is insensitive to how her body moves, she should still be considered responsible for this omission.

By no means do I take myself to have refuted, or even *objected* to, Fischer and Ravizza's account. Furthermore, the above thoughts merely point to, but do not express, an alternative account of omissions. All I hope to have done is to show that

their account is not the only one available and that, on other reasonable accounts, the idea that Anna is responsible for leaving FLT unrefuted is left standing. Fischer and Ravizza show that one need not accept all of the intuitions or principles that lead to this result, but even so, I believe I have done enough to show that that A and A* are far from obvious.

6 Accomplice to murder!

The preceding four counterexamples to A involve cases in which an agent can be said to have made it the case that a certain necessary fact obtains. I have suggested that, in such cases, agents can be morally responsible for these facts. In this section, I wish to consider cases in which an agent does not make a fact obtain, but is nonetheless responsible for it. These cases will show that the first premise in the argument for A* that we considered in Sect. 1 is false. This premise states that an agent is in part morally responsible for a fact only if the agent makes this fact obtain.

Counterexamples to this idea are easy to come by, at least when the fact in question is contingent. Just consider cases in which someone lets something happen that they could have prevented. Imagine a person sees a priceless vase falling. She could dive and catch it, but instead simply watches it smash. She does not make it the case that the vase breaks, but she is morally responsible for the fact that it does. Thus this premise is false.

The obvious response is to concede the point and to modify the premise accordingly. Perhaps the following will do: an agent is in part morally responsible for a *necessary* fact only if the agent makes this fact obtain. The problem with this premise is that, without attempting to justify it, it seems ad hoc. Why is this principle limited to necessary truths? Someone who makes this claim may try to justify it as follows. An agent is in part morally responsible for a fact only if the agent either makes this fact obtain or could have prevented this fact from obtaining. It is impossible for someone to be able to prevent a necessary fact from obtaining. Therefore, an agent is in part morally responsible for a necessary fact only if she makes this fact obtain.

The problem with this argument is that it is false that an agent is in part morally responsible for a fact only if either the agent makes this fact obtain or the agent could have prevented this fact from obtaining. A modification of Erosion* (or, indeed, Erosion) shows this. Imagine Betty is accompanied by Bill, who allows Betty to detonate the bomb. Bill could have prevented her from doing so and knew that the detonation would lead to the destruction of the village. Bill is morally responsible (at least in part) for the fact that the village is destroyed. Despite this, Bill neither made it the case that the village is destroyed nor was able to prevent the village being destroyed (because the avalanche would have destroyed the village even if Betty had not detonated the bomb). Thus he is responsible for a fact that he neither made true nor could have prevented from being true.

Once we have seen this, it is simple to invent cases in which an agent is responsible for a necessary fact without having made this fact obtain or having been

able to prevent its obtaining. Take our first counterexample to A—Stephen is morally responsible for the fact that he actually murders someone, even though this fact is necessary. Imagine now that Daniel could have prevented Stephen from murdering someone, but let him commit the murder anyway. Daniel is, in part, morally responsible for the fact that Stephen murdered someone. He is thus also partly morally responsible for the fact that Stephen actually murders someone. Consider our third counterexample to A—Karen is responsible for the fact that it is wrong to torture babies for fun because she ensures that horrible consequences would result from doing so. Now imagine Julia could have prevented Karen from ensuring that horrible consequences would result from torturing babies for fun. Julia is in part responsible for the fact that Karen makes it wrong to torture babies for fun, and is thus in part responsible for the fact that it is wrong to torture babies for fun. Similar modifications (which I will leave to the reader) can be made to the other two counterexamples to A. I conclude, therefore, that there are not only cases in which an agent is morally responsible for a necessary fact, but cases in which an agent is morally responsible for a necessary fact that the agent does not make obtain.

7 The Direct Argument revisited

I have argued in this paper that it is quite possible for an agent to be (at least partly) morally responsible for a necessary fact. Furthermore, an agent need not even have made a necessary fact obtain in order to be thus responsible. Aside from the inherent interest in this result, perhaps the most important conclusion we can draw from my examples is that the Direct Argument fails to show that moral responsibility and determinism are incompatible. As we saw above, this argument relies on the exact idea that one cannot be responsible for necessary truths. Even if it instead were to rely on the idea that one cannot be responsible for non-world-indexed necessary truths, the argument still fails. Both Rule A and Rule A* allow invalid inferences and thus must be rejected (and with them the Direct Argument).

Given the examples above, there are two other interesting points to make about the Direct Argument. Firstly, the fact that we can be morally responsible for necessary truths also shows that premise 6 is false. This premise says that no one is even in part responsible for the truth of the laws of nature. This premise can be cast into doubt in three separate ways. Firstly, what makes premise 6 plausible is the fact that the laws of nature are physically necessary and we cannot be responsible for something that is physically necessary. But as soon as we allow that we can be responsible for *metaphysically necessary* truths, then the intuitive pull of premise 6 is lost (indeed, metaphysically necessary truths *are* physically necessary). Secondly, van Inwagen allows any proposition to be a law of nature if it is entailed by any law of nature or set of laws of nature (see van Inwagen 1983, p. 64). As metaphysically necessary truths are entailed by every law of nature, they themselves are laws. Thus our being responsible for some necessary truths makes us also responsible for some laws of nature. Thirdly, examples like those I have given above can be modified to directly attack premise 6.

To see this, consider the following case (based on Anna's case above). Jon has excellent evidence that he is able to reduce a certain substance's temperature to absolute zero. As we know, this is in fact physically impossible. It is a law of nature that nothing can have that temperature (this is, of course, unknown to Jon). Jon thinks hard about whether to reduce the temperature of the substance to absolute zero and decides against it. He intentionally, knowingly, sanely (etc.) leaves the temperature of the substance above absolute zero. Furthermore, it is quite reasonable to suppose he is *morally responsible* for leaving the temperature of the substance above absolute zero. It is however, *physically necessary* that Jon leaves the temperature of the substance above absolute zero. Jon is thus responsible for a (mere) physical necessity. Consider also a case based on Karen's situation. Wylie likes to keep certain substances at room temperature. He (intentionally, knowingly, etc.) builds a machine that ensures that these substances are always at room temperature. He is thus morally responsible for the fact that they are kept above 1°C. He is also responsible for them being above, for example, −27°C, and being above −273.15°C. This is all part of his plan. He *makes it the case* that these substances are above these temperatures. However, −273.15°C *is* absolute zero. Thus Wylie overdetermines the fact that certain substances are above this temperature. His efforts are enough to keep them above it, as are the laws of nature. Wylie is morally responsible for the physically necessary fact that these substances are above absolute zero.

The second point about the Direct Argument that I wish to make is that the falsity of A not only invalidates Rule A, but also Rule B. This is quite easy to see. Though we could use any of the examples we have looked at in previous sections to make this point, let us consider Anna's case. Anna is responsible for the fact that she leaves FLT unrefuted. She is *not*, however, responsible for the fact that it is *necessary* that she leaves FLT unrefuted. One may be responsible for the truth of a necessary truth without being responsible for its necessity. She is also not responsible for the necessary fact that if it is necessary that she leaves FLT unrefuted then she leaves FLT unrefuted. She does not make this fact obtain, nor allows it to obtain, nor allows anyone to make it obtain. This is enough to show that Rule B is invalid. There are facts P and Q such that Anna is not responsible for P, not responsible for if P then Q, but *is* responsible for Q.

The Direct Argument has been dealt a hefty blow. If A is false, then Rule A is invalid, Rule B is invalid and premise 6 is false. Perhaps some will say that my counterexamples to A, though suggestive, do not *show* that A (or at least A*) is false. Even if this is right, however, they are strong enough to make the Direct Argument dialectically ineffective against compatibilists. That is, they cast enough doubt on Rule A (and Rule B, and premise 6) for the incompatibilist to be unable to rely on them to rationally persuade people of incompatibilism.

8 Conclusion

I have not given a knockdown objection to A (or at least to A*) in this paper. I'm sure that all of the examples I provide can be resisted in various ways (and I look

forward to discovering what those ways are). Despite this, what I hope to have done is to convince some people that we can indeed be responsible for necessary truths and to show those people who are not so convinced that it is not *obvious* that we cannot be. A very reasonable case can be made for the idea that agents can be responsible for necessary truths. Even if it is still reasonable to believe that no agent can be thus responsible, gone are the days when we can simply assume this to be true. Furthermore, this shows that the Direct Argument for the incompatibility of moral responsibility and determinism fails. It is either unsound or at the very least not rationally persuasive.

Acknowledgements Thanks to Ofra Magidor, Julia Markovits, Al Mele, Carl Ginet and an anonymous referee for comments.

References

Blackburn, S. (1993). *Essays in quasi-realism*. New York: Oxford University Press.
Clarke, R. (1994). Ability and responsibility for omissions. *Philosophical Studies, 73*, 195–208.
Fischer, J. (1996). A new compatibilism. *Philosophical Topics, 24*, 49–66.
Fischer, J. (2006). *My way: Essays on moral responsibility*. New York: Oxford University Press.
Fischer, J., & Ravizza, M. (1999). *Responsibility and control*. New York: Cambridge University Press.
Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy, 66*, 829–839.
Frankfurt, H. (1994). An alleged asymmetry between actions and omissions. *Ethics, 104*, 620–623.
Haji, I. (2008). Reflections on the incompatibilist's direct argument. *Erkenntnis, 68*, 1–19.
Locke, J. (1690). In P. H. Nidditch (Ed.), *An essay concerning human understanding*. Oxford: Oxford University Press (1975).
McIntyre, A. (1994). Compatibilists could have done otherwise: Responsibility and negative agency. *Philosophical Review, 103*, 453–488.
McKenna, M. (2008). Saying good-bye to the direct argument the right way. *Philosophical Review, 117*(3), 349–383.
Stump, E., & Fischer, J. (2000). Transfer principles and moral responsibility. *Philosophical Perspectives, 14*, 47–56. (reprinted in Fischer 2006).
van Inwagen, P. (1978). Ability and responsibility. *The Philosophical Review, 87*(2), 201–224.
van Inwagen, P. (1980). The incompatibility of responsibility and determinism. In M. Bradie & M. Brand (Eds.), *Bowling Green studies in applied philosophy* (Vol. 2, pp. 30–37). Bowling Green, OH: Bowling Green State University Press.
van Inwagen, P. (1983). *An essay on free will*. Oxford: Clarendon.
Warfield, T. (1996). Determinism and moral responsibility are incompatible. *Philosophical Topics, 24*, 215–226.