

A Theoretical Argument against the Principle of Alternate Possibilities for Moral Responsibility

Heath White

Department of Philosophy and Religion
University of North Carolina – Wilmington
601 S. College Road, Wilmington, NC 28403-5601, USA.

Abstract

A theoretical approach to the practice of holding one another morally responsible shows that we should expect “could not have done otherwise” to excuse a faulty action only sometimes. Holding someone morally responsible for their actions is much like holding employees responsible for doing their jobs, or holding students responsible for learning material. In each case, “failures” can be excused by an inability to do otherwise, if the inability is not relevant to the purpose for which the person is being held responsible. But an inability to do otherwise will not be excused if it is an inability to successfully pursue that purpose.

This theory-driven rejection of the Principle of Alternative Possibilities fits well with a “quality of will” approach to moral responsibility. It also places a dialectical burden on the defender of PAP to come up with an alternative theory.

Keywords: moral responsibility; principle of alternate possibilities; Frankfurt; compatibilism; quality of will; excuses

Frankfurt is well-known for arguing, in “Alternate Possibilities and Moral Responsibility” and other places, that what he termed “the principle of alternate possibilities” is false. In Frankfurt’s original formulation, that principle is (PAP) A person is morally responsible for what he has done only if he could have done otherwise.¹

Frankfurt’s argument hinges on a particular kind of case in which an agent, Smith, may or may not do something terrible, say murder Jones. Black, waiting in the wings with some kind of mind-control device, will intervene to make Smith murder Jones in the event that Smith is on the verge of deciding against it. Otherwise, Smith goes ahead and murders Jones on his own, without any interference. We are supposed to have the intuition that although Smith could not do anything but murder Jones (that is, he could not have done otherwise), he is responsible for murdering Jones if he does so on his own, without any intervention from Black. This constitutes a counterexample to (PAP).

I believe Frankfurt’s conclusion is correct, but not everyone agrees. Incompatibilists, especially, are likely to raise the question of whether Smith is determined to murder Smith on his own, prior to actually making the decision. If so, they will say, the intuition Frankfurt appeals to is incorrect: no agent acting deterministically is responsible for his action. On the other hand, if Smith is not antecedently determined to murder Smith on his own, it is hard to construct the example so as to make Black knowledgeable enough that he can intervene in any case where Smith would not murder Jones on his own. In that case, Smith can do otherwise. Neither option provides a counterexample to (PAP).²

The arguments on both sides have been adjusted and developed in a number of ways over a lengthy period of time, and I do not expect further refinement of the Frankfurt-style examples to bring significant progress.³

¹ Frankfurt (1969) p. 829.

² See, e.g., Widerker (1995).

³ The literature on Frankfurt cases is by now voluminous. Many important essays are collected in Widerker and McKenna (2006).

What I wish to do in this paper is argue for Frankfurt's conclusion by a quite different method from his. Instead of appealing to intuitions about when Smith is and is not morally responsible—intuitions the incompatibilist can simply deny—I will argue by sketching a theory of the practice of holding each other morally responsible. Such a theory involves, roughly, claims about what holding each other morally responsible is and why it is done, and I will seek to provide theoretical understanding of the practice of holding each other morally responsible by assimilating it to other practices with a similar structure. Understanding the structure of such practices, I will argue, will give us a theory-driven reason to expect (PAP) to be false. For this reason, I refer to my argument as a “theoretical argument” against (PAP).

I am under no illusion that my argument will be persuasive to convinced incompatibilists. However, my preferred method of argument will place on them a burden of proof which Frankfurt-style examples do not. Frankfurt essentially appeals to an intuition which the incompatibilist can legitimately deny without much argument. Since the thrust of Frankfurt's case is to deny incompatibilism, he cannot depend on compatibilist intuitions, and the incompatibilist is within her dialectical rights to lean on her own intuitions in judging the effectiveness of Frankfurt's argument. My strategy, on the other hand, outlines a theory of the practice of holding each other morally responsible which gives us reason to reject (PAP). If the incompatibilist wishes to deny my conclusion, she will have some dialectical obligation to say what is wrong with the theory I outline, and to provide an alternative, incompatibilist theory or explanation of what is going on when we hold each other responsible. The bare appeal to incompatibilist intuition will no longer be enough. Thus my contribution does not aim at final victory for Frankfurt's conclusion, but it does aim at progress in the debate.

1. Accountability Practices

I will begin by drawing the reader's attention to a class of practices, institutions, or patterns of behavior—the exact ontology here is not too important—that I will call “accountability practices.” These are situations with a common structure: (i) there are criteria, laid down more or less clearly in advance, for some kind of behavior or performance; (ii) one's performance is measured against the criteria, e.g. one “passes” or “fails” (this amounts to different things in different cases, of course); (iii) this measurement normally carries some kind of consequence for future treatment. Employee performance reviews, student exams, and military readiness tests all qualify as accountability practices. For example, an employee knows more or less what is expected of him on the job, and during his review, or simply on the job, the employer evaluates him against those criteria. Depending on how he has performed vis-à-vis expectations, the employee may be fired, demoted, remanded for extra training, or on the other hand be given a raise or more responsibility or a promotion.

The first thing I wish to point out about accountability practices is that they exist for a purpose. They are not ends in themselves. Roughly speaking, employee reviews exist in order to maintain a competent workforce in a business. Student exams exist in order to measure academic achievement and, beyond that, to signal to graduate schools and potential employers how knowledgeable, talented, and hard-working the student is. The military tests the skills of its soldiers in order to maintain an effective fighting force. And so on. None of this is controversial.

Second, the purposes of the accountability practices give shape to the sorts of criteria they employ. For example, the military tests the skills of its soldiers in order to maintain an effective fighting force. This means that the criteria employed in the tests are the criteria relevant to military effectiveness: marksmanship, physical fitness, emotional control under conditions of stress and fatigue, maintaining one's equipment, and so on. The military does not evaluate its soldiers on how accurately they can throw a Frisbee or how many hot dogs they can eat at a sitting, such skills having not much military utility.

Therefore, and third, it is the case that usually or normally or as a general rule, failure to perform up to the standards of an accountability practice indicates that one is not furthering, or sufficiently contributing to, the purpose for which the practice is in place. For example, if a soldier does not pass a marksmanship test, this ordinarily indicates that he is not a sufficiently accurate shot to be part of a military force of the standard to which the test-setters aspire. If a student misses too many questions on her final exam, this ordinarily indicates that she does not know the class material well enough. If an employee fails to meet the performance standards his employer is looking for, then ordinarily this means that he is not competent at his job. And so on. In short, and schematically rendered, for any accountability practice, normally there is a good inference from “X has failed to measure up to standards” to “X is not furthering the ends for which the standards are in place.”

However, in unusual or abnormal or exceptional circumstances, this inference can be defeated. In particular, it can be undercut.⁴ That is, there may be reasons for “failure” which do not have anything to do with the purpose aimed at by the practice. For example, suppose a soldier, Marco, has to shoot with his rifle at a target a set distance away, and is expected to hit the target ten out of ten times. He may fail to hit the target no matter how many times he shoots at it, because some prankster has loaded his rifle with blanks. Or perhaps his corrupt officer has not issued him any ammunition, selling it on the black market instead, and Marco fails to hit the target for this reason. In these sorts of cases, the failure to hit the target tells us nothing about Marco’s marksmanship and thus his contribution to an effective fighting force. (The existence of pranksters in the ranks, or officers who steal ammunition, may tell us something about the effectiveness of the fighting force, but here the fault does not lie with the hapless Marco.)

Here we have the makings of a legitimate excuse for Marco. (From here on out, the noun “excuse” means “legitimate excuse” unless otherwise specified.) He can say that, although he failed to hit his targets, “It was because somebody loaded my rifle with blanks!” and maybe this will get him off the hook. (If he is supposed to load his rifle himself, obviously, it will not.) The reason it gets him off the hook, if it does, is that the excuse—“Somebody loaded Marco’s rifle with blanks”—constitutes an undercutting defeater for the inference from “Marco failed to hit the targets” to “Marco is not a sufficiently good soldier for this army,” which is itself an instance of the inference-schema from “X has failed to measure up to standards,” to “X is not furthering the ends for which the standards are in place.”⁵

So here is the general pattern with accountability practices. First, they exist for a purpose, not as ends in themselves. We evaluate each other and act on those evaluations, along various dimensions, in order to further certain goals beyond merely holding one another accountable for its own sake. Second, these goals determine what standards are used for evaluation. We measure each other’s performance against standards which are relevant to the purpose for which the evaluation is being conducted. Third, the purposes for which we evaluate one another also determine what sorts of excuses for failure we will accept. Excusing conditions are those which defeat the relevant instances of the inference-schema “X has failed to measure up to standards” to “X is not furthering the ends for which the standards are in place.”

Given this characterization of accountability practices, we can say a little more about inability to do otherwise as a putative excuse. Whether inability to meet the relevant standards (i.e. to do otherwise than fail to meet them) for some accountability practice is a legitimate excuse depends entirely on *why* one is unable to meet them. This will work as an excuse on those occasions when the inability to do otherwise is due to some factor irrelevant to the purpose of the evaluation, and it will fail as an excuse when the inability to do otherwise is relevant to the purpose of the evaluation. To return to the example of our soldier: if Marco doesn’t hit his targets because he has not been issued any ammunition, he can say, truly and with conviction, that he could not have hit the targets. This is an excuse, I claim, not because he was unable *per se*, but because he was unable for a reason that is irrelevant to his skills as a marksman. The fact that he was issued no ammunition means that another fact, that he failed to hit his targets, has no bearing on the truth of a third proposition, that he is an inept marksman, and thus the inference from “Marco failed to hit the targets” to “Marco is not a sufficiently good soldier for this army” is undercut. Now consider another soldier, Blinky, who fails to hit any targets and also claims, truly and with conviction, that he could not have hit them. He could not have hit them, in this case, because he is a terrible shot. Here, it seems clear to me, we have no excuse worth paying attention to. In this case, the inability to hit targets is no excuse for not hitting them, because the explanation of that inability is that the soldier does not further precisely those ends for which the practice of marksmanship testing is in place. The inference from “Blinky failed to hit the targets” to “Blinky is not a sufficiently good soldier for this army” is intact.

⁴ An *undercutting defeater* of the inference from A to B is a reason for supposing that A is not sufficiently indicative of the truth of B. The terminology derives from Pollock (1986), p. 39. There are other sorts of defeaters which are less applicable to the present issue.

⁵ This way of identifying excuses is somewhat, though not exactly, similar to the diagnosis in Austin (1957). Austin does not define “excuse” precisely but his summary statement is “we admit that [the action] was bad but don’t accept full, or even any, responsibility.” (p. 2) This more or less describes the situation when an undercutting defeater is present, but Austin also construes excuses so as to include claims that an action was unintentional, or accidental, or a side effect of some other intended action, or “he was looking at the whole thing quite differently,” none of which would qualify as excuses in the sense I identify.

To take an example more immediately familiar to my readers, consider two students who each fail a logic test. Jack fails because he hasn't learned any logic all semester. Jill fails because her professor hastily passed out a take-home test at the end of class, left the room before Jill could protest that she had not received a test, and remained incommunicado until the test was due. Both students could not have answered the questions correctly—Jack because he can't do logic, Jill because she never got a test and couldn't get one no matter how she tried. Jill has an excuse for her non-performance, but Jack has none, and this is because the explanations of their respective failures differ. Jack fails to complete his test because he has failed to successfully pursue the ends for which the test exists, namely mastery of logic. Jill fails to complete her test for reasons which have nothing to do with her mastery of logic. The facts pertinent to Jill's case, but not Jack's, constitute an undercutting defeater for the inference from "J failed the test" to "J hasn't mastered the material." It is reasonable, then, to flunk Jack but give Jill a make-up exam.

If something along these lines is the correct theory of accountability practices, it follows that we have a quite general, theory-driven reason to expect that, for any accountability practice, some cases of inability to do otherwise will not count as excuses for failure to perform within that practice. Take any practice P, and suppose that it exists to further some end E. E is the reason for the standards upheld in P, and it gives shape to the excuses for not meeting those standards. Suppose some agent "fails" with respect to these standards, and moreover, could not have done anything but fail. It is easy to see that such an inability might be, broadly, of one of two kinds: either the failure to pursue E successfully, or something else. In the latter kind of case, inability to do otherwise will count as an excuse within the practice P. In the former kind of case, it will not.

To illustrate: suppose logic exams exist in order to measure progress toward the end of learning logic. Then some students will not be able to pass the test because they do not know enough logic, and some will not pass the test for some other reason. The latter have an excuse, the former do not. Suppose, again, that the military tests soldiers' marksmanship so that it has a competent fighting force. Some soldiers will not pass the marksmanship test because they are militarily incompetent, and others will not pass for some other reason. The latter have an excuse, the former do not. And so on for any kind of accountability practice: there is a quite general and theory-driven reason to expect that at least some kinds of inability to do otherwise will not count as excuses.

2. Moral Responsibility as an Accountability Practice

The next stage of the argument is to claim that the practice of holding one another morally responsible (hereafter, "the moral responsibility practice") is an accountability practice. We will not get very far with this claim, however, before we know what the moral responsibility practice is.

I take this practice to include *judgments*, say that an action was morally permissible or impermissible, for example. The practice also includes the *responses* people undertake on the basis of these judgments. These may be intentional actions, like scolding, shunning, refusing to associate, or on the other hand public praising or awarding a medal. The responses may also include reactive attitudes like resentment or gratitude, which themselves have various behavioral manifestations. Sometimes people make the judgment without the ordinarily associated response, for various reasons: there is no opportunity for response, or the object of judgment is too distant in time or space to worry about, or one feels one lacks moral standing to undertake the response, or one chooses to forgive.

That way of putting things is a cognitivist take on the practice of moral responsibility. A non-cognitivist, more Strawsonian take might hold that the reactive attitudes we feel toward others constitute or are expressed in moral judgments. Although my own views are more cognitivist, this particular dispute doesn't make much difference for present purposes. The important thing is to get a grip on the activities that go into the practice of holding each other morally responsible, which is to say the whole matrix of praise and blame: cognitive judgments, affective reactions, and overt behavior, in whatever mix or proportion or explanatory ordering.

Claiming that this network of thoughts, feelings, and actions is an accountability practice has a certain initial plausibility. What is quite obvious is that holding each other morally responsible fits into the general framework of accountability practices. That is, (i) there are moral standards for behavior, more or less publicly known. (ii) Individuals' actions are evaluated against these standards, and (iii) this moral evaluation ordinarily carries consequences for how one will be treated in the future. If this structural similarity is not a coincidence—and I do not think it is—several implications follow directly from the points made in the previous section. It means that the practice of moral praise and blame—the moral responsibility practice—exists for a purpose beyond itself.

That is, we engage in moral evaluations and reactions for some further end. It also means that both the moral standards we evaluate each other against, and the excuses we will accept for not living up to them, are determined by this further purpose. Also, normally, failure to meet these standards implies failure to further the ends for which the practice of holding one another responsible is in place. Finally, it means that the inability to do otherwise will sometimes be a legitimate excuse for not living up to moral standards, and sometimes not, depending on whether the explanation for this inability is relevant to the purpose for which moral standards are in place, or more precisely, whether the putative excuse defeats the normal implication.

The coherence, and therefore the credibility, of these claims increases if we can postulate a plausible purpose for which moral standards are in place. Why do we bother holding each other morally responsible for our behavior? What is the ultimate point of this practice? Here I will follow the lead of a number of other writers. What we are interested in, in the practice of moral responsibility, is the quality of will others hold toward us. There are a number of different strands in the “quality of will” literature, but what follows is a fairly generic version that will do the work needed here. What we most want, in the contexts relevant to moral evaluation, are high-quality relationships with others. These relationships consist, mostly, in the attitudes of the parties toward one another. When others have good wills toward us, we are inclined to reciprocate, strengthening and bolstering that relationship. When others have bad wills toward us, we are likewise inclined to reciprocate. The basic picture could be elaborated in multiple ways: maybe the value of relationships is instrumental, or maybe it is an intrinsic part of the good life. Maybe our inclinations to reciprocate toward others of good (or bad) will are rationally self-interested, or a product of evolution, or simply a brute fact about human nature. These details, though, don’t make a lot of difference for present purposes.

Once the basic idea is in place—that the practice of holding each other morally responsible is aimed at fostering high-quality relationships and therefore is sensitive to the quality of will others show toward us—moral excuses look like this. A number of different types of action (or omission) would normally imply or express the agent’s indifference or hostility toward one. That is, an inference from the fact of the action, to the agent’s bad (or insufficiently good) will toward one, would normally be justified. An excuse for such behavior will be some fact about the action or circumstances that undercuts this inference.

So, for example, a hard shove is normally good grounds for concluding that the person shoving has a bad will toward the person being shoved. However, if the shove was intended to get the shoved person out of the path of a runaway car, then there is no good inference to the bad will of the shover: quite the contrary. So “I was saving your life by getting you out of the way of the oncoming car” is a legitimate moral excuse for a hard shove. Likewise, my stepping on your toes may give no grounds for thinking I have a bad will toward you, if it was not intentional and I apologize profusely and immediately. On the other hand, if I seem indifferent to the fact that I stepped on your toes, albeit accidentally, or worse, if I did it intentionally because I enjoy causing others pain, the normal inference about my bad will is confirmed and I have no excuses.

Exactly the same thing will go for the candidate excuse, “I could not have done otherwise.” This will succeed as an excuse if it defeats inferences to the bad will of an agent, and it will fail otherwise. In the cases where it fails as an excuse, we will have a counterexample to (PAP). Moreover, we should *expect* there to be such counterexamples, based on the overall theory here of what is going on in the practice of holding each other morally responsible, what its point is, and what would count as an excuse. That is, we will have theory-driven counterexamples to (PAP), not merely appeals to moral intuition.

When we go looking for examples of such theory-driven counterexamples to (PAP), they are not hard to find. Suppose Brad and Angelina are members of the same department, Brad wants to go to a conference during the semester, and Angelina promises to teach his classes while he is gone. When the time comes, however, Angelina doesn’t show. This causes confusion among the students, disruption in Brad’s plans for the semester and an awkward discussion with his chair. Naturally, his first reaction is anger at Angelina for letting him down.

When he confronts her about not keeping her promise to him, she replies, “Well, I couldn’t teach your classes like I promised. I just wasn’t able. There was nothing else I could do.” This piques Brad’s curiosity, if nothing else, so he asks for more details. And here it matters a great deal how the story continues.

Suppose Angelina explains that she was unable to teach Brad’s class because she was hit by a car and was in the hospital. Or the police mistakenly detained her. Or she was kidnapped by a disgruntled former student. Any of these would be legitimate excuses for not keeping her promise to take Brad’s class in his absence.

They are legitimate because they all go to show that what prevented Angelina from keeping her promise to Brad was not anything in her attitudes toward him. Her will toward him remains as good as ever.

Alternatively, however, suppose Angelina has a quite different style of explanation for why she could not teach Brad's class. Suppose she had just read a recent article of his and was so vehemently in disagreement with the conclusion that she could not bring herself to assist him with classes as she had promised. Or suppose she discovered a talk by a visiting speaker that she really, really, *really* wanted to hear, unfortunately scheduled simultaneously with the classes of Brad's that she was supposed to teach, and she couldn't bring herself to miss the talk in order to keep her promise to Brad. In these sorts of cases, I take it, Angelina has no legitimate excuse for failing to keep her promise to Brad, and his anger is perfectly justified.

Both the legitimate and the illegitimate excuses Angelina might offer are cases where she did not keep her promise to teach Brad's classes and could not have done otherwise. Yet we treat the two sets of cases quite differently, and this calls out for explanation. The theory sketched here, taking the practice of holding each other morally responsible to be an accountability practice, explains the difference well. In the first set of cases, the inference from "Angelina didn't keep her promise" to "Angelina bears Brad insufficiently good will" is defeated by the addition of further premises describing the circumstances. In the second set of cases, it is not. When someone fails to keep an important promise because they don't like your philosophical positions, or because they would rather go hear a lecture, that just is a case of insufficiently good will.

We might use this theory to redescribe the original Frankfurt cases also. There is ordinarily an excellent inference from "Smith murdered Jones" to "Smith bore Jones ill will." So when Smith murders Jones on his own initiative, without interference from Black, we have been described a paradigm case of bad will on the part of Smith toward Jones. That is why we are inclined to blame Smith for the murder, even though he could not do otherwise. On the other hand, when Smith murders Jones because Black intervenes, we have been described a case in which Smith's will toward Jones was not bad—that is why Black needed to intervene—and so in that unusual context the murder does not indicate any bad will on Smith's part toward Jones. That is why we are inclined not to blame Smith in the case of intervention.

3. Summary So Far

At this point, let me summarize the argument, before proceeding to objections. Accountability practices are widespread and relatively easy to understand. They have a common structure with three elements: (i) there are criteria for some kind of behavior or performance; (ii) one's performance is measured against the criteria; (iii) this measurement normally carries some kind of consequence for future treatment. The moral responsibility practice for actions has the same elements. There is a *prima facie* case, then, for supposing that whatever else is true of accountability practices generally, is true of the moral responsibility practice also.

Some other broad truths about accountability practices include, I argued, that they exist for a purpose beyond themselves and that these further ends drive the particular criteria they employ. The criteria are selected to be such that, normally, there is a good inference from an agent's failure to meet the criteria for performance to the agent's failure to further or contribute to the ends for which the practice is in place. This inference can be defeated, however, and in cases where it is defeated, the circumstances which explain the defeat count as legitimate excuses for failure to meet the behavioral criteria of the practice. "I could not have done otherwise" is a form of excuse that sometimes does and sometimes does not defeat this inference, so there will—quite generally, for any accountability practice—be cases of not being able to do otherwise which are excuses and other cases which are not.

If holding each other morally responsible in fact follows the pattern of other accountability practices, we can predict that there will be cases of not being able to do otherwise which do constitute excuses for behavior that would ordinarily count as a moral failure, and cases which do not. This is in fact exactly what we do see. If we also adopt a hypothesis about what the point of the practice of holding each other morally responsible is, we can also predict what sorts of cases will fall on which side of the line. I have suggested that the point of the practice of holding each other morally responsible is to track and reinforce positive relationships with others, wherein what we principally care about in such relationships is the quality of others' wills toward us. This predicts that when the reason an agent cannot do otherwise than perform some putatively bad action defeats the inference from the agent performing the bad action to the agent having a bad will, the reason will count as an excuse—in *this* case, that one cannot do otherwise means that one is not appropriately blamed for one's action.

On the other hand, when one cannot do otherwise than perform some putatively bad action for a reason which does not defeat that inference, i.e. precisely because of the poor quality of the agent's will, the reason does not count as an excuse. And, I have argued, what we predict on this hypothesis is exactly what we see when we look at which invocations of "couldn't have done otherwise", intuitively, count or do not count as excuses. In those cases where the excuses fail, we have theory-driven counterexamples to (PAP).

4. Objections and Replies

The main line of argument is now complete and it is time to answer some objections to it. I will begin with what I take to be weaker objections and proceed to more serious ones.

4.1. The defender of (PAP) could always defend her position by a Moorean tactic. Suppose she accepts the conditional: *if* the moral responsibility practice had the characteristics of an accountability practice, then we should expect (PAP) to be false in just the ways I have argued. However, she remains convinced that (PAP) is true. Therefore, she concludes, the moral responsibility practice does not, in fact, have all the characteristics of an accountability practice.

Logically, there is no flaw in this reasoning. It lacks other intellectual virtues, however. At this point, the opponent of (PAP) has offered a fairly coherent, albeit sketchy, theory of what goes on when we hold each other responsible, and why that practice includes just the exceptions it does. The defender of (PAP) who adopts the Moorean tactic has offered no similarly compelling theory. Until she does, she cannot say that the competing understandings of moral responsibility offered by the two sides are on a level. There are probably no knockdown arguments in philosophy,⁶ but the dialectical burden on the defender of (PAP) has increased.

4.2. The defender of (PAP) could also point out that the theory offered here is incomplete in certain respects. In particular, not every circumstance which would defeat the ordinary inference, from (schematically) "X has failed to measure up to standards" to "X is not furthering the ends for which the standards are in place," counts as an excuse. For example, suppose Jordan knows her logic perfectly well, but she deliberately misses all the questions on her logic exam as part of a feud with her parents. That extra fact would undercut any inference from "Jordan failed the test" to "Jordan hasn't mastered the material." However, it doesn't get Jordan off the hook: even knowing the circumstances, her professor could fairly give Jordan an F. Thus the analysis offered here of what counts as a legitimate excuse is incomplete.

I think this is right: the theory I sketched above would have to be improved to accommodate cases like Jordan's. On the other hand, I don't see that the defender of (PAP) can make much hay out of this point. What Jordan's example shows is that a fact which defeats a certain kind of inference is not sufficient to constitute a legitimate excuse. The way I have suggested we understand excuses in the context of accountability practices casts the net too wide: not everything the theory says is an excuse, is one. But what the defender of (PAP) needs to show is rather different: that a fact which *does not* defeat a certain kind of inference—namely, the fact that one could not have done otherwise precisely because of the poor quality of one's will—is sufficient to constitute an excuse. That is, the defender of (PAP) needs to show that the theory casts the net of excuses too narrowly: there are more excuses than it says there are. In showing what the defender of (PAP) needs to show, Jordan's example is just irrelevant.

4.3. More sharply, an incompatibilist might object that the practice of holding each other morally responsible is sharply distinct from other accountability practices, because it is not a practice which exists for an end beyond itself. To be sure, she can allow, we evaluate employees and students and recruits for larger social or organizational ends. But evaluating one another morally is importantly different, she will say, in that the practice of doing so is an end in itself. What we are interested in is not, say, relationships, but Accountability (the Platonic capital is apt): the correct acknowledgement of, and response to, moral deeds and misdeeds. So the criteria for moral failure are not set by any further purpose of this practice, and it is just constitutive of moral failure—a built-in aspect of Accountability—that one could have done otherwise than one did. There is a strong analogy between this position about moral responsibility and a certain hardcore Kantian view of retributive punishment. On such a view, the Justice system (more Platonic capitals) exists for its own sake, not for any further social end, and indeed, Justice must be done though the heavens fall.

⁶ Nozick (1983) p. 5

It is only a partial reply to this objection that it is perfectly obvious that both the institution of criminal punishment, and the related institution of holding one another morally responsible, serve important social ends. That these institutions or practices help ensure peace, order, and productive social relations is clear to anyone. However, the objection we are considering is not the view that holding one another morally responsible has no instrumental value. It is the view, rather, that an important part of the value of this practice is final (i.e. non-instrumental) value, and moreover, that important aspects of this practice—like which excuses are legitimate and whether being not able to do otherwise is always one of them—are set by the non-instrumental aspects of the practice, rather than by the fact that the practice is oriented toward furthering some social end.

A more thorough reply in this vein would, I think, take a long time, beyond the scope of what I can offer here. So I will have to make some mere assertions more briefly. The practice of holding each other morally responsible, and the institution by which we hold each other legally responsible, are both retributive, broadly speaking. What we know about the history, the logic, and even the biology of retribution and the associated reactive attitudes all strongly indicates that our practices of both holding each other morally responsible as individuals, and holding each other legally responsible as a society, are not exercises in Platonic contemplation. They are, rather, institutions for the maintenance of a livable social order, with very distant biological origins and a long history of development.⁷

In case all that seems unpersuasive, or overly empirical, here is a different sort of counterargument. In cases of forgiveness, we depart from the practice of holding each other morally responsible in an important way. Speaking roughly, forgiveness involves judging that someone has acted wrongly but not responding to them in the ordinary, sanctioning way. Moreover, forgiveness is generally understood to be acting in such a fashion that a broken relationship is restored.⁸ Finally, forgiveness is widely thought to be morally beautiful, something of important moral value.

Thinking of the moral responsibility practice as valuable for its own sake, versus thinking of it as existing for the sake of maintaining positive relationships, produce two different interpretations of the value of forgiveness. If Accountability is its own final value, then forgiveness always involves a tradeoff of moral values. We forgive, and that is morally valuable, yes, but it comes at the moral cost of someone not being Held Accountable, treated as they deserve. On the other hand, if we think of the practice as existing for the sake of maintaining positive relationships, forgiveness is a method of accomplishing the end of that practice by other means. There is no tradeoff of value, and forgiveness retains its moral beauty at no moral cost. (There is generally a cost to the forgiver, which I don't mean to ignore. But this will be a financial cost, or an emotional cost, or some such. What I mean is that the moral order has suffered no violation; nothing that morally ought to have happened has been neglected.)

It seems to me that the ordinary view of forgiveness is much more consonant with the latter interpretation than the former. Suppose a man cheats on his wife, they separate, and later he repents, she forgives him, and the couple reconciles. I am imagining a case in which the reconciliation is wholehearted and genuine, not one which is reluctant or forced by circumstances. In such a circumstance it would be churlish to think, "Well, that's nice in its way, but it's really too bad that she didn't give him what he really deserves." The moral mistake made in such a judgment is precisely that it takes blaming the cheating husband for his wrongdoing to be an end in itself, and the resulting attitude is indistinguishable from vindictiveness. In reality, the whole apparatus of moral judgment and blame exists to facilitate relationships between parties of good will toward one another.

4.4. Even if they can be convinced that holding one another is an accountability practice as I have described it, some defenders of (PAP) may claim that the sorts of counterexamples to (PAP) I have described are impossible.

⁷ For the biology of reactive attitudes, see e.g. Bloom (2013). For an evolutionary story about their function, see Trivers (1971). For the socio-economic rationale for retribution, see Posner (1980). For a stricter rational-choice approach, see Elster (1998). For some other historical-cum-psychological context, see Miller (2006). Each of these citations must do duty for a rather large literature.

⁸ "[F]orgive" refers to the act of giving up a feeling, such as resentment, or a claim to requital or compensation.... [F]orgiveness is a dyadic relation involving a wrongdoer and a wronged party, and is thought to be a way in which victims of wrong alter their and a wrongdoer's status by, for instance, acknowledging yet moving past a transgression.... The standard definition of forgiveness makes clear that its main purpose is the re-establishment or resumption of a relationship ruptured by wrongdoing." Hughes (2015).

My counterexamples are those where one could not have done otherwise precisely because of the poor quality of one's will, in the forms of hostility or callousness; that is, they are forms of psychological determinism. And a defender of (PAP) might argue that there just is no such thing as psychological determinism, or that no one believes there is. Nobody will accept that Angelina could be so outraged by an article, or so seduced by a lecture, that she metaphysically could not keep her promise to Brad. And if intuition is not enough to make the case, there is this argument: we *do* take "could not have done otherwise" as an excuse (that is, (PAP) is true) and therefore we would not hold people responsible if they were psychologically determined. But since we do hold Angelina responsible for breaking her promise to Brad, that is evidence that we don't believe in any such kind of psychological determinism.

I do not believe psychological determinism is such an incredible phenomenon, nor do I believe that the folk are so incredulous about it. Many people, I would say, are psychologically incapable of performing certain great evils, like murdering their children. Tremendous acts of supererogation—like selling all one has and giving it to the poor—are probably about equally rare for the same psychological reason. It should also be pointed out that many libertarians today would say that psychological determinism is, in fact, possible, maybe even frequent.⁹ These libertarians' core claim is only that an action for which one is responsible have, somewhere in its causal ancestry, some choice which was not determined, psychologically or in any other way. For example, if Angelina has developed an ingrained habit of running off madly after celebrity lecturers, she may have gotten herself psychologically unable to keep her promise to Brad when it comes down to it. Her ability to do otherwise, the contemporary libertarian might claim, was forfeited some time ago.

If the contemporary libertarian is willing to go this far, however, she has all but given up on this objection. For we might hold that Hume at least got right about causation that, in principle, anything can cause anything. If one's psychology can be determined by prior free choices, as she admits, then she ought also to admit that one's psychology could, in principle, be determined by early upbringing or bad diet or ... whatever. She has no apriori ground on which to rule out the possibility of psychological determinism.

This point highlights a subtler problem with the objection. Even if it were true that we could not be psychologically determined, this hardly seems like a necessary truth. That is, even if the objection is correct that *we* are never psychologically determined, there are possible beings much like us, who would have a concept of moral responsibility much like ours and a practice of holding one another morally responsible much like ours, who sometimes are psychologically determined. So that "there is no such thing as psychological determinism" is at best a contingent claim, whereas (PAP), if true, is presumably necessarily true. And it is no good providing merely contingent defenses for allegedly necessary truths.

4.5. The defender of (PAP) has one more move, and I think it is the strongest one. She might take on everything I have said about the practice of holding one another moral responsible as an accountability practice, whose purpose is the maintenance of positive relationships, where this purpose drives the norms and exceptions recognized within the practice. However, she might claim that I have the nature of the desired relationship wrong. A necessary condition of a good will, she might argue, is that it be freely chosen, or more precisely, that its attitudes and choices be ones which the agent could have done otherwise than adopt. Similarly, a will is only bad if its attitudes and choices are ones which the agent could have done otherwise than adopt. Without that "could have done otherwise" condition, she might claim, no will has any moral valence, and no relationship has any moral value, positive or negative.

She might defend this view by pointing to cases of manipulation or indoctrination. If someone loves us (or hates us) because they are somehow manipulated or indoctrinated into doing so, she might say, we don't hold them morally responsible for their attitudes or actions: *that* kind of good (or bad) will doesn't count, morally speaking. Why would it not count? Intuitively, because there is no real relationship to be had with that person; they are just a pawn in a larger game. By extension, any kind of psychologically determined good or bad will wouldn't count morally. What would make a relationship real, and valuable, would be the existence of good will from an individual who was not merely a pawn in a larger game, that is, who could have directed his will differently. It is this kind of relationship that we are interested in, and which sets the standards and excuses available in the moral responsibility practice. And this set of excuses, quite logically, includes (PAP).

⁹ E.g. Kane (1996) and elsewhere, Ekstrom (2000), Stump (2003) ch. 9 attributes this view to Aquinas, Mele (2006), Pawl and Timpe (2009).

I have a lot of respect for this position and I think it deserves more development than I can give it here, from someone more sympathetic to it than me. Nevertheless, there are a few things to be said against it, I believe. The position turns on the claim that relationships with a person, where the will of that person is caused by forces outside the agent, have no moral value one way or the other. They are not even real relationships at all, one might say. This position seems both strange, and subject to counterexamples, or at least cases which give some impression of being counterexamples.

It is strange because it seems an open, empirical question how our wills are set, for good or ill. However, it does not seem an open question whether we have any morally valuable relationships with other people. The incompatibilist position offers hostages to fortune for no particularly good reason.¹⁰

It is subject to counterexamples because there seem to be a number of cases where we can think of genuinely morally important relationships, where at least one party to these relationships is under quite obvious and salient influences to form just that kind of relationship. On the positive side, consider the love of a mother for her child. This is one of the most morally beautiful of relationships, honored the world over. Yet the causes of mother love have a great deal to do with biology and the contingent fact that mothers normally raise their children from birth through young adulthood. This quite naturally forms a deep bond which would not exist, at least not to the same degree, if someone else raised the child, or if (per impossibile) the child had been born to someone else. It might be replied, with some justice, that just because choices are easy does not mean they aren't real; none of this shows that mothers could not have done otherwise than love their children. After all, some don't. On the other hand, I wonder what the "cash value," to use William James' locution, of the suggestion that the typical mother has a choice here amounts to. Phenomenologically, for many mothers anyway, the attitudes adopted toward one's children are rarely voluntarily chosen.

On the negative side, many of the fighters for the barbaric organization ISIS have no doubt been subject to various forms of indoctrination, and in part due to this indoctrination they probably sincerely believe that beheading innocents, etc. is the righteous course. Maybe this indoctrination is so effective, and the live alternatives for them so unpalatable, that for at least some of the fighters, no other allegiance is possible. None of that, I take it, makes their actions a whit less morally objectionable. Incompatibilists might deny that indoctrination *could* be that effective (objection #4) or they might claim that, if it is, the fighters bear no moral blame for their actions (the present objection). Both replies strike me as implausible, but I don't know what else to say at this point.

It is true that in traditional "manipulation" cases, involving nefarious neurosurgeons or meddling hypnotists who cause their victims to perform bad actions (e.g. Frankfurt's interfering Black), we might not hold the victims of manipulation morally responsible for their deeds. However I believe there is a different explanation available than the one offered by the defender of (PAP). There is an important difference between the cases of mothers and ISIS fighters, on the one hand, and traditional manipulation cases on the other. In the latter cases, it is easy to think of the will of the agent as being circumvented or interfered with, in ways that allow it to remain plausible that the agent, "deep down" or "really", does not bear his victim ill will. In these cases, the bad action proceeds from a psychologically superficial set of motives. Whereas when we think of cases like mother love or violent fanaticism (even if the external influences involved do not rise to the level of determinism) there is no agent "below the surface" who "really" feels otherwise than they act. That is, there is no deeper level of understanding on which mothers do not really love their children, or on which ISIS fighters do not really hate their victims.

Of course, we don't have to think of the traditional manipulation cases this way. If what Black does is turn Smith into an evil person, who murders Jones precisely because he hates him and wants him dead, then it seems to me that Smith has no excuses for his behavior. We *might* pardon Smith if reforming his character is as simple as flipping a switch, and we can flip it; this would accomplish the ends of the practice of holding each other morally responsible, although it will not restore Jones. But the fact that the newly-vicious Smith wanted Jones dead so badly that he could not but kill him, is no reason not to hold Smith morally responsible for the murder.

So, to sum up: the defender of (PAP) might argue that it is only good will that could have been otherwise that interests us, in the accountability practice of holding one another morally responsible, because only that sort of good will leads to the positive relationships that have real value, and moreover that this goal underwrites (PAP).

¹⁰ In several places, John Martin Fischer has made analogous arguments to the effect that our view of ourselves as morally responsible should not be held hostage to discoveries in physics. For example, see Fischer and Ravizza (1999) pp. 15-16.

I do not think we can know this value claim, because while we know our relationships have value we do not know that they include any agent-originated features. Moreover, there is some reason to think that the value claim is false, since many of our firmest moral convictions include moral attitudes toward people whose wills toward us are, at least in significant measure, not able to be different. Finally, the value claim is not well-defended by appeal to traditional manipulation cases, since those can be explained in other ways.

5. Conclusion

I have advanced a “theoretical” argument against (PAP). The heart of the argument is a theory (sketch) of the moral responsibility practice, casting it as one member of a larger family of accountability practices, such that it is not only possible but to be expected that each member of this family will cover situations where “I could not have done otherwise” is no excuse for behavior that violates the standards of the practice. I have examined and rejected a number of objections to this view; in no case, so far as I can tell, does the defense of my view turn purely on the appeal to controversial compatibilist intuitions. The defenders of (PAP) will surely not agree with everything I have to say, but they now bear the dialectical burden of saying what exactly *is* going on with the moral responsibility practice, if it is not an accountability practice as I have suggested.

Works Cited

- Austin, J. L. (1957). A Plea for Excuses. *Proceedings of the Aristotelian Society*, New Series, 57, 1-30.
- Bloom, Paul. (2013). *Just Babies: The Origins of Good and Evil*. New York: Crown Publishers.
- Ekstrom, Laura Waddell. (2000). *Free Will: A Philosophical Study*. Boulder: Westview Press.
- Elster, Jon. (1998). *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.
- Fischer, John Martin and Mark Ravizza. (1999). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, Harry G. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66, 829-839.
- Hughes, Paul M. (2015). Forgiveness. *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/spr2015/entries/forgiveness/>.
- Kane, Robert. (1996). *The Significance of Free Will*. New York: Oxford University Press.
- Mele, Alfred R. (2006). *Free Will and Luck*. New York: Oxford University Press.
- Miller, William Ian. (2006). *Eye for an Eye*. Cambridge: Cambridge University Press.
- Nozick, Robert. (1983). *Philosophical Explanations*. Cambridge, MA: Belknap Press.
- Pawl, Tim and Kevin Timpe. (2009). Incompatibilism, Sin, and Free Will in Heaven. *Faith and Philosophy* 26, 398-419.
- Pollock, John. (1986). *Contemporary Theories of Knowledge*. Savage, MD: Rowman and Littlefield.
- Posner, Richard A. (1980). “Retribution and Related Concepts of Punishment. *The Journal of Legal Studies* 9, 71-92.
- Stump, Eleonore. (2003). *Aquinas*. London: Routledge.
- Trivers, Robert L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology* 46, 35-57.
- Widerker, David and Michael McKenna (eds.) (2006). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Burlington, VT: Ashgate.
- Widerker, David. (1995). Libertarianism and Frankfurt’s Attack on the Principle of Alternative Possibilities. *Philosophical Review* 104, 247-61.