Maximalism and Moral Harmony

DOUGLAS W. PORTMORE
Arizona State University

Maximalism is the view that an agent is permitted to perform a certain type of action (say, baking) if and only if she is permitted to perform some instance of this type (say, baking a pie), where φ-ing is an instance of ψ-ing if and only if φ-ing entails ψ-ing but not vice versa. Now, the aim of this paper is not to defend maximalism, but to defend a certain account of our options that when combined with maximalism results in a theory that accommodates the idea that a moral theory ought to be morally harmonious—that is, ought to be such that the agents who satisfy the theory, whoever and however numerous they may be, are guaranteed to produce the morally best world that they have the option of producing. I argue that, for something to count as an option for an agent, it must, in the relevant sense, be under her control. And I argue that the relevant sort of control is the sort that we exercise over our reasons-responsive attitudes (e.g., our beliefs, desires, and intentions) by being both receptive and reactive to reasons. I call this sort of control *rational control*, and I call the view that φ-ing is an option for a subject if and only if she has rational control over whether she φs *rationalism*. When we combine this view with maximalism, we get *rationalist maximalism*, which I argue is a promising moral theory.

The performance of one option can entail the performance of another. For instance, I have both the option of baking and the option of baking a pie, and my baking a pie entails my baking, for I can't bake a pie without baking.[1] More precisely, for any two of my options φ and ψ, my φ-ing entails my ψ-ing if and only if I don't have the option of φ-ing without ψ-ing. Such instances of performance entailment are ubiquitous. Kissing passionately entails kissing. Driving under 55 mph entails driving under 100 mph. Marrying an unmarried man entails marrying a bachelor. Typing the word 'the' entails typing the letter 't'. Stretching a t_1 and then going for a run at t_2 entails going for a run at t_2.

[1] Note, then, that options needn't be mutually exclusive. I can have both the option of baking and the option of baking a pie even though these two are not mutually exclusive. Only *alternative* options must be mutually exclusive. Thus, two options, φ and ψ, are alternative options if and only if both φ-ing and ψ-ing is not an option. By contrast, two options, φ and ψ, are *distinct* options if and only if it is not the case that each entails the other. So, since baking doesn't entail baking a pie, these two are distinct options, though they are not alternative options.

Given that our options include both those that entail others and those that are entailed by others, a moral theory needs to be able to assess both types. If I have the options of (1) baking a pie, (2) baking an apple pie, and (3) baking an apple pie and then taking it to the family who just moved in across the street, it won't do for a moral theory to tell us only whether some proper subset of these options are permissible. If these are all equally options, then a moral theory—a theory concerning the moral permissibility of an agent's options—owes us an account of the permissibility of each and every one.

On the view known as *maximalism*, an agent is permitted to perform a certain type of action (say, baking) if and only if she is permitted to perform some instance of this type (say, baking a pie), where φ-ing is an instance of ψ-ing if and only if φ-ing entails ψ-ing but not vice versa.[2] But, of course, if there are options that are so specific that they don't have any instances, then we can't appeal to the permissibility of their non-existent instances to determine their permissibility. Thus, on maximalism, we must distinguish between two types of options: maximal and non-maximal options. An option φ is a *maximal option* if and only if it is maximally specific—that is, if and only if there is no option ψ such that ψ-ing entails φ-ing but not vice versa. Otherwise, it's a *non-maximal option*. So walking at t_1 won't be a maximal option if walking fast at t_1 is an option. And walking fast at t_1 won't be a maximal option if walking fast at t_1 and then jogging at t_2 is an option. It seems, though, that we will eventually arrive at an option that isn't entailed by any other option. And that's a maximal option.[3]

---

[2] See, for instance, Brown 2015, Bykvist 2002, and Gustafsson 2014. Note, however, that my use of the term 'maximalism' is broader than theirs. They use the term to refer specifically to consequentialist theories that hold that whereas the permissibility of a maximal option is to be assessed directly in terms of the optimality of its consequences, the permissibility of a non-maximal option is to be assessed in terms of whether it is entailed by some permissible maximal option. By contrast, I use the term more generally to refer to any moral theory that holds that whereas the permissibility of a maximal option is to be assessed directly in terms of some right-making feature F, the permissibility of a non-maximal option is to be assessed in terms of whether it is entailed by some permissible maximal option.

[3] One may doubt that there will always be some maximal option—that is, an option that is not entailed by any other option. For instance, one may think that I have all the following options: (Opt_1) thinking of a number greater than 1, (Opt_2) thinking of a number greater than 2, (Opt_3) thinking of a number greater than 3, and so on and so forth, *ad infinitum*. Thus, one may conclude that for any option Opt_n, there will always be another option Opt_{n+1} that entails it and, so, no maximal option in this situation. But even if, for every number n, there is always a greater number (e.g., n+1), it doesn't follow that I always have the option of thinking of that greater number. My abilities (and, hence, my options) are limited in a way that the set of numbers is not. Nevertheless, if there are situations in which an agent has no maximal options, then I just need to replace talk of maximal options with talk of normatively maximal options, where an option is a normatively maximal option if and only if every option that entails it is normatively equivalent to it—that is, equivalent to it in terms of all normatively relevant factors. See Brown 2015. Still, one may further worry that there won't always be a normatively maximal option. Suppose, for instance, that there is no limit to how large a number I can

We must distinguish between these two types of options, because, on maximalism, they are to be assessed differently.

> *Maximalism:* (Max1) For any subject S and any non-maximal option ν, S's ν-ing is permissible if and only if there exists an option φ such that S's φ-ing is permissible and S's φ-ing entails S's ν-ing. And, (Max2) for any subject S and any maximal option μ, S's μ-ing is permissible if and only if S's μ-ing has feature F.

Maximalism is neutral on what sorts of things count as options. Perhaps, it is only voluntary acts. But, perhaps, the non-voluntary formations of beliefs and other attitudes also count as options. And maximalism allows for options to be conjunctive. An agent could, for instance, have the option of asserting that p while believing that p is false. What's more, maximalism is neutral on what 'F' stands for.[4] Indeed, we can substitute for 'has feature F' anything that would render Max2 coherent, including 'maximizes utility', 'accords with Kant's categorical imperative', or 'contains only beliefs for which S has sufficient evidence and contains all and only those acts that maximize expected utility'. As this last example illustrates, F can be something quite complicated, involving the assessment of different types of options according to different criteria—e.g., evaluating beliefs in terms of evidence and acts in terms of expected utility.

---

think of and that, for whatever number I do think of, God will give me and my loved ones precisely those many days in heaven. In that case, no matter what number n I think of, there will be an alternative option, Opt_{n+1}, that would be better in terms of the normatively relevant considerations. And so, in this case, there seems to be no normatively maximal option. But this is a problem for maximalism only if we think that I must always have at least one permissible option. After all, if there is no permissible normatively maximal option that entails my opting for Opt_n, then maximalism will just imply that, for any n, I am not permitted to opt for Opt_n. But, to my mind, that seems exactly right. For it seems that I will do something impermissible no matter what I do, because no matter what I do there will be an option that I had more moral reason to perform, could have just as easily performed, and yet didn't perform. Thanks to Ángel Pinillos and Brad Armendt for pressing me on this.

[4] Now, we could restrict the variable F to only those features that have the property of being preserved under coarsing (call this PropPUC), where F has PropPUC if and only if, necessarily, if a maximal option is F, then every option it entails is also F. To illustrate, the feature of containing no instances of murder has PropPUC. For, necessarily, if a maximal option contains no instances of murder, then every option it entails will contain no instances of murder. But many features (such as that of having optimal consequences) do not have PropPUC—see Brown 2015. In any case, I want to acknowledge that, if we were to restrict F to only those features with PropPUC, then the above formulation of maximalism would be equivalent to the view that, for any subject S and any (maximal or non-maximal) option φ, S's φ-ing is permissible if and only if S's φ-ing has feature F. But I'll put no such restriction on F.

What motivates maximalism is the thought that if I'm permitted to perform a certain type of action (say, baking), I must be permitted to perform some instance of this type (say, baking a pie). For if I'm not permitted to bake anything—not a pie, not a cake, not cookies, not anything, then I'm not permitted to bake. Moreover, if I'm not permitted to bake a pie and then eat either some or none of it (and assume that, if I bake a pie, I must then eat either some or none of it), I'm not permitted to bake a pie. Suppose, for instance, that someone will kill me and everyone I love if either I bake a pie and then eat some of it or bake a pie and then eat none of it. In that case, it would be impermissible for me to bake a pie. And this is because there is no instance of baking a pie that I'm permitted to perform. Note, then, that, as I'm understanding things, S's φ-ing counts as an instance of S's ψ-ing if and only if S's φ-ing entails S's ψ-ing but not vice versa. Thus, both baking-a-pie-and-then-eating-some-of-it and baking-a-pie-and-then-eating-none-of-it count as instances of baking a pie.

So, on maximalism, if I'm permitted to bake, this is because I'm permitted to perform some instance of baking, such as baking a pie. And if I'm permitted to bake a pie, this is because I'm permitted to perform some instance of pie-baking, such as baking an apple pie. And if I'm permitted to bake an apple pie, this is because I'm permitted to perform some instance of apple-pie-baking, such as baking an apple pie and then taking it to the family who just moved in across the street. And so on and so forth. But, of course, if this just goes on forever, we'll end up with an infinite regress. But eventually we will arrive at an option that is not entailed by any other option—that is, a maximal option.[5] And when we do, we won't be able to derive its permissibility from that of some other permissible option that entails it. Given that it's a maximal option, there will be no other option that entails it. Thus, maximalism must include not only Max1 but also Max2. Max2 tells us that a maximal option is permissible if and only if it has some right-making feature F. And what's distinctive and interesting about maximalism is that it tells us that we don't evaluate non-maximal options in the same way. Instead of evaluating non-maximal options in terms of whether or not they have feature F, we evaluate them in terms of whether there is some other permissible option that entails it. Thus, Max1 tells us that a non-maximal option is permissible if and only if there is some permissible option that entails it.

[5] But, again, see the caveat discussed in note 3.

Now, the point of this paper is not to defend maximalism. That has been done elsewhere.[6] Instead, the aim is to argue for a certain account of what our options are and to show that when we combine this account with maximalism, we arrive at view that can accommodate the idea that a moral theory ought to be morally harmonious—that is, ought to be such that the agents who satisfy the theory, whoever and however numerous they may be, are guaranteed to produce the morally best world that they have the option of producing. In section 1, I argue that, for something to count as an option for a subject, it must, in the relevant sense, be under her control. And I argue that the relevant sort of control is the sort that we exercise over our reasons-responsive attitudes (e.g., our beliefs, desires, and intentions) by being both receptive and reactive to reasons—forming, revising, sustaining, and/or abandoning these attitudes in light of our awareness of facts (or what we take to be facts) that count for or against them.[7] I call this sort of control *rational control*, and I call the view that φ-ing is an option for a subject if and only if she has rational control over whether she φs *rationalism*. In section 2, I show that when we combine rationalism with maximalism, arriving at what I call *rationalist maximalism*, we get a moral theory that, unlike most moral theories, is morally harmonious. And so I conclude, in section 3, that rationalist maximalism is a promising moral theory.

1. What are our options?

[6] See Brown 2015, Feldman 1986, Goldman 1978, Portmore 2011, Portmore 2013, Portmore 2016a, Portmore 2016b, and Zimmerman 1996.

[7] The notion of a *reasons-responsive attitude* is, perhaps, the same as T. M. Scanlon's notion of a *judgment-sensitive attitude*, an attitude that is sensitive to the subject's judgments about reasons (1998, 20). But Scanlon's notion is, if not distinct, misleading, for we can respond to reasons without having any judgments about what our reasons are. "We respond to reasons when we are aware of facts that give us these reasons, and this awareness leads us to believe, or want, or do what these facts give us reasons to believe, or want, or do" (Parfit 2011, 493). Thus, we can respond to reasons while neither knowing that this is what we are doing nor having any judgments about our reasons (Parfit 2011, 461).

Reasons-responsive attitudes include all and only those mental states that a rational subject will tend to have, or tend not to have, in response to reasons (or apparent reasons)—facts (or what are taken to be facts) that count for or against the attitudes in question. So beliefs are clearly reasons-responsive attitudes, for a rational subject will, for instance, tend to believe that it will rain in response to her awareness of facts that constitute decisive reasons for her believing this, such as the fact that a reliable weather service has predicted that it will rain. Although reasons-responsive attitudes include many mental states, they exclude feelings of hunger, nausea, tiredness, and dizziness, which are not responsive to reasons. Suppose, for instance, that I have too quickly consumed a good-sized meal and am still feeling hungry, as there has not yet been sufficient time for my brain to receive the relevant physiological signals from my stomach. Even if I am aware that I've eaten more than enough to be satiated, my hunger is not responsive to this awareness. Instead, it is responsive only to the physiological signals that supposedly take about twenty minutes to travel from the stomach to the brain.

There is some set of events such that, if one of its members, viz., φ, is more highly favored by the relevant considerations than any other member, then S ought to φ. I call this the set of S's options. As noted above, some of these options will be maximal and some will be non-maximal. And, if maximalism is correct, these two types of options are to be assessed differently. But before we can even get to assessing them, we need to know what they are. They are, I believe, all and only those things are, in some relevant sense, under S's control. This would explain why, for instance, I ought to spend the weekend writing a lecture even though the relevant considerations (e.g., those concerning what would be most beneficial for me and for others) favor my spending it writing a literary masterpiece. Unfortunately, given my lack of literary talent, whether I write a literary masterpiece is not under my control. And, thus, it can't be something that I ought to do. It is only those metaphysically possible actions over which I exert control that can be things that I ought to do. And, of those, my spending the weekend writing a lecture is, or so we'll assume, the most highly favored by the relevant considerations. Thus, it is what I ought to do.

The thought, then, is that S's having control at *t* over whether she φs at *t'* is both necessary and sufficient for φ-ing at *t'* to be, as of *t*, an option for her, which in turn is necessary for her to be obligated, as of *t*, to φ at *t'* (*t* < *t'*).[8] I'm assuming, then, that the statement "S is, as of *t*, obligated to φ at *t'*" implies that "S's φ-ing *t'* is, as of *t*, an option for her."[9] And it's important to make the relevant time indices explicit, because what one has control over—and, thus, what constitutes an option and an obligation for one—can vary over time. To illustrate, suppose that, late Saturday, a travelling friend gives me a letter to mail for him on Monday morning and that I promised him right then that I would do so. In that case, I have, as of late Saturday, an obligation to mail the letter come Monday morning. And this fact is what explains why I have the further obligation, as of late Saturday, to find a safe place to store the letter until Monday. Nevertheless, suppose that I neglect to store it in a safe place and that, consequently, a

[8] As I think of them, temporally-indexed options and obligations refer to properties that are possessed by the subject at certain times. Thus, the phrase "S is, as of *t*, obligated to φ" is equivalent to "S has at *t* the property of being obligated to φ." And, likewise, the phrase "φ-ing is, as of *t*, an option for S" is equivalent to "S has at *t* the property of having φ as an option."

[9] This assumption has been disputed. Walter Sinnott-Armstrong has, for instance, argued against this assumption—see his 1984. For some of the reasons why I find his arguments unpersuasive, see Streumer 2003. In any case, the reader may note that my conclusions in this paper are conditional upon this assumption.

housemate burns the letter on Sunday night, using it as kindling to start the logs in the fireplace. So, as of Sunday night, I no longer have the option of mailing the letter that my friend gave me on Saturday. After all, that letter no longer exists. And since, as of Sunday night, I do not have the option of mailing the letter, I do not, as of Sunday night, have an obligation to do so.

Since I'm claiming that an agent's having control at *t* over whether she φs at *t'* is both necessary and sufficient for φ-ing at *t'* to be, as of *t*, an option for her, I need to defend both the idea that it is necessary and the idea that is sufficient. To see that it is necessary, consider the implausibility of the following view, which denies this.

> *Schedulism:* A subject S has, as of *t*, the option of φ-ing at *t'* if and only if whether she φs at *t'* depends (in the right way) on her having certain intentions at certain times—that is, it depends on her having a certain schedule of intentions from *t* on.[10]

On this view, an agent's having control at *t* over whether she φs at *t'* is not necessary for φ-ing at *t'* to be, as of *t*, an option for her. To illustrate, suppose that no matter what I do now and no matter what I intend now to do tomorrow, I'm not going to get up and exercise when the alarm wakes me at 5 AM. Instead, I'm going to hit the snooze button several times, getting up at 6 AM, leaving myself with no time to exercise. Still, on schedulism, I have now the option of getting up and exercising at 5 AM so long as (a) I will at 5 AM have the capacity to form the intention to get up and exercise and (b) would get up and exercise if I were to form that intention. The fact that no matter what I do now or intend now to do later, I won't form the intention to get up and exercise when the alarm wakes me is, on schedulism, irrelevant. But this seems a mistake. In order for my getting up and exercising at 5 AM tomorrow to be at present an option for me, I must have control now over whether I will get up and exercise at 5 AM tomorrow. But, if I'm not going to get up and exercise at 5 AM tomorrow no matter what my present actions and intentions are, then I don't now have control over whether I'll get up and

[10] This is adapted from my definition of *personally possible*—see my 2011, 166. And this in turn was adapted from Jacob Ross's definition of a *performable option*—see his 2012, 81. Proponents of schedulism include Fred Feldman (1986) and Michael J. Zimmerman (1996).

exercise at 5 AM tomorrow. After all, to have control now over what I'll do at 5 AM tomorrow is to have the present ability to affect what I'll do then. But, in this case, I have, at present, no way of affecting whether I'll get up and exercise at 5 AM tomorrow, or so we're assuming.[11]

To better understand where schedulism goes wrong, consider the following case.

> *Curing Cancer:* A five-year-old boy named Saru is playing with a computer keyboard. Millions of lives depend on his writing and sending an email to the National Institutes of Health that explains how to cure cancer in such a way that those who read it will take it seriously.[12]

Schedulism implausibly implies that Saru has the option of curing cancer (or, at least, the option of effecting the cure of cancer). After all, to do so, he need only write and send this email, which in turn just involves his making a certain series of keystrokes. And Saru would make each keystroke in that series if he were to have certain intentions at certain times. To illustrate, suppose that such an email would start with: "The cure for cancer is…." If at t_1 Saru were to intend to hit Shift + T, he would do so at t_2, thereby typing an uppercase T. And, having done that, he would have the capacity to form at t_3 the intention to hit the H key. Moreover, if he were to form this intention at t_3, he would then hit the H key at t_4, thereby typing a lowercase H. And similar assumptions apply for all the remaining keystrokes in the series. Thus, schedulism implies that Saru has, as of t_0, the option of curing cancer. And since the relevant considerations favor Saru's doing so, it seems that the proponent of schedulism must insist that Saru is obligated to cure cancer. This, of course, is absurd. Saru cannot control whether he types out the cure for cancer. For, as we'll plausibly assume, Saru will not make the required series of keystrokes no matter what he does now or intends now to do later. Even if he does happen to

---

[11] Admittedly, if I set the alarm this evening, I will, when the alarm goes off, have control at that point over whether I get up an exercise. After all, I will at that point be able to affect whether I exercise. For let's suppose that if I jump right out of bed when the alarm goes off, I'll exercise, but if I hit the snooze button, I won't. But I'm concerned with whether I have the option now (this evening) of exercising tomorrow morning, not whether I will have the option of doing so tomorrow morning. And I'm claiming that if I'm not going to exercise tomorrow morning no matter what acts and attitudes I form and perform this evening, then I don't have control this evening over whether I'm going to exercise tomorrow morning. And if I don't have control now over whether I'm going to exercise tomorrow morning, then I don't now have the option of exercising tomorrow morning.

[12] This example is modeled after a similar one given in Wiland 2005.

hit Shift + T at t_2, he wouldn't (or so I'll stipulate) follow up by making each of the other ten thousand or so specific keystrokes that are required to complete the task of writing and sending such an email. And even if Saru were to intend to type out (or to try to type out) the cure for cancer, he would fail (or so I'll stipulate). And although God could perhaps form the complex intention to make the entire series of specific keystrokes required, Saru does not have the capacity to form such a complex intention. Thus, Saru has no way at t_0 to effect the typing out of such an email. No present intention or action that he has the capacity to form or perform would result in his curing cancer. Thus, he does not now control whether he cures cancer. And the problem with schedulism is that it holds that Saru has, at present, the option of curing cancer even though he does not now control whether he does so. It seems, then, that we should hold that an agent's having control at t over whether she φs at t' is necessary for φ-ing at t' to be, as of t, an option for her.

And, as noted above, I believe that having control is sufficient as well. To see this, consider the implausibility of the following view, which denies this.

> *Tryism:* A subject S has, as of t, the option of φ-ing at t' if and only if whether she φs at t' depends (in the right way) on whether she tries at t to φ at t'.

To see both that tryism implies that an agent's having control at t over whether she φs at t' is insufficient for φ-ing at t' to be, as of t, an option for her and that this view is problematic for this very reason, consider the following case.

> *Stupid Mistake:* A genius named Albert took a math test and missed one of the easiest problems because he overthought things and, consequently, overlooked its simple solution.[13]

[13] This is inspired by John Maier's example of a professional golfer who misses what should have been a gimme putt, which he uses to question whether its being the case that S would φ if she were to try to φ is necessary for S's having the option to φ. See his 2014.

According to tryism, Albert didn't have the option of providing the correct answer to this easy problem. After all, it wasn't for a lack of trying that he didn't provide the correct answer. Rather, he tried to provide the correct answer and failed because he overthought things, consequently overlooking the most obvious solution. So, if anything, he tried too hard. And given that he was going to try and fail, it was false that "he would provide the correct answer if he were to try do so." So tryism implies that providing the correct answer wasn't even an option for him.

But that's strange. Even Albert would admit that he could and should have got the correct answer to such an easy problem. And we should too. For it seems that he had both the ability and opportunity to provide the correct answer. In other words, it seems that he had control over whether he was to provide the correct answer. The fact that he failed to exercise this control to good effect doesn't mean that he lacked control. It just means that he messed up. Indeed, these sorts of stupid mistakes cause us the most frustration precisely because we think that we could and should have got things right. When we get a test back and find that the solution was one that would never have occurred to us, we are not nearly as frustrated (if at all) as when we find that the solution was so obvious that it should have occurred to us. This suggests, then, that an agent's having control at t over whether she φs at t' is sufficient for φ-ing at t' to be, as of t, an option for her and that we should reject views, such as tryism, that deny this.

Of course, ultimately how plausible my claim that an agent's having control at t over whether she φs at t' is both necessary and sufficient for φ-ing at t' to be, as of t, an option for her will depend on what we take the relevant sort of control to be. So I'll turn now to this issue. As I see it, there are two main contenders: (1) voluntary control and (2) rational control. I'll explain each in turn.

Voluntary control is the sort of control that we exert directly over our voluntary actions (such as raising one's arm) and indirectly over those things that we manipulate via such actions (such as the movement of the car that one's driving). To better understand this kind of control, it's important to note that voluntary actions have three key features. For if φ is something that I can directly and voluntarily do, it follows that: (1) I can φ at will, and, thus, I can φ simply by

trying, deciding, intending, choosing, or otherwise willing to φ; (2) I can φ for any reason that I take to be sufficient for doing so, and, thus, I can φ to win a bet, to help others, or to please my partner; and (3) I can choose when to φ, and, thus, can choose to φ now, five minutes from now, a day from now, or only on Tuesdays.[14] Taking these features into account, I offer the following tentative account of voluntary control: S has at *t* direct voluntary control over whether she φs at *t'* if and only if S has, as of *t*, the ability to intentionally φ at *t'* (as well as the ability to intentionally refrain from φ-ing *t'*) for any reason that she takes as counting sufficiently in favor of her doing so. I offer this account, not as something that I'm committed to in its details, but only as an approximation that will be sufficient for our purposes. Thus, voluntary control just is whatever sort of control that we exercise over our voluntary actions and those things that we manipulate via such actions, and this holds regardless of whether I have the details exactly right.

Even with only this rough and ready account to work with, it's clear enough that we don't typically exert voluntary control over our reasons-responsive attitudes. For the sort of control that I have, say, over whether I believe that Aristotle went for a swim on his 30th birthday is clearly distinct from the sort of control that I have, say, over whether I touch my nose. I can do the latter but not the former at will. Thus, I can do the latter but not the former to win a bet. And whereas I can choose when to touch my nose, I cannot choose when (or even whether) to believe that Aristotle went for a swim on his 30th birthday.[15]

The fact that we don't typically exert voluntary control over our reasons-responsive attitudes along with the plausible thought that people ought to form certain reasons-responsive attitudes and can be held accountable if they fail to do so has led several philosophers to conclude that the sort of control that's relevant to determining our obligations and responsibilities is not voluntary control but rational control.[16] By 'rational control', I just mean whatever sort of control that we exert directly over our reasons-responsive attitudes and

[14] See McHugh 2012, McHugh 2014, and McHugh Forthcoming.

[15] For a further defense of the claim that we don't typically exert voluntary control over our reasons-responsive attitudes as well as a response to those who have argued to the contrary, see McHugh 2012, McHugh 2014, and McHugh Forthcoming.

[16] See Graham 2012 (13, note 22), Hieronymi 2006, McHugh Forthcoming, Scanlon 1998, Smith 2005, and Smith 2015. Whereas Angela M. Smith and I use the term 'rational control', Pamela Hieronymi uses the term 'evaluative control' and Conor McHugh uses the term 'attitudinal control'. The basic idea, though, is the same.

indirectly over those things that we influence via such attitudes (such as our voluntary actions). We exercise control over our reasons-responsive attitudes by being both receptive and reactive to reasons—forming, revising, sustaining, and/or abandoning these attitudes in light of our awareness of facts (or what we take to be facts) that count for or against them.

Some may hesitate to call this a kind of control given that control must be active and forming an attitude in response to reasons seems less than fully active. But even if forming an attitude in response to reasons is not as active as, say, performing a voluntary action, it does involve thinking, and thinking is active. Compare, then, the formation of an attitude with the feeling of a sensation (such as hunger or dizziness). The feeling of a sensation is completely passive. We simply suffer our sensations. And, because of this, we can't be asked to justify them. We can be asked only to explain them. By contrast, the formation of an attitude is active. For we shape our attitudes by attending to, reflecting upon, and responding to our reasons. And it is because our attitudes reflect the extent to which we have attended to, reflected upon, and responded to our reasons that we can be asked to justify them—to give our reasons for them. So even if the formation of an attitude in response to reasons isn't active in the same way that a voluntary act is, it is, nevertheless, active. So, we do exert a kind of control over our reasons-responsive attitudes: it's just that it is rational control as opposed to voluntary control.[17]

But what does this sort of control amount to? Here's a tentative proposal: S has at t rational control over whether she φs at t' if and only if she has, as of t, the capacity to respond appropriately to the relevant reasons and whether she φs at t' depends (in the right way) on whether and how she responds at t to these reasons ($t < t'$).[18] Here, too, I offer this only as a rough approximation and not as something that I'm committed to in all its details. Rational control just is whatever sort of control we exert by attending to, reflecting upon, and responding to reasons, regardless of whether I've got the specifics right.

---

[17] The ideas in this paragraph are taken from Hieronymi 2006. See also Hieronymi 2008 and McHugh Forthcoming.

[18] As formulated, this seems to presume that the relevant sort of control is regulative control rather than guidance control, the difference being that you can have the latter, but not the former, with respect to φ even when not–φ-ing is not an option. But, again, I'm not committed to this or to any other detail concerning this account. If we have only guidance control, and not regulative control, with respect to our reasons-responsive attitudes, then rational control should be thought of as a kind of guidance control. For more on the distinction between these two types of control and its relevance, see Fischer & Ravizza 1998. For how we might think of rational control as a kind of guidance control, see McHugh 2013 and McHugh 2014.

Given these two types of control, we have the following two competing accounts of what our options are.

> *Voluntarism:* S's having at t voluntary control over whether she φs at t' is both necessary and sufficient for φ-ing at t' to be at t an option for her ($t < t'$).

> *Rationalism:* S's having at t rational control over whether she φs at t' is both necessary and sufficient for φ-ing at t' to be at t an option for her ($t < t'$).

Of these two views (and I see no plausible third view), I believe that rationalism is the clear winner. There are at least three reasons for this. First, voluntarism forces us to deny that people ought ever to form, revise, sustain, or abandon their reasons-responsive attitudes. But ordinarily we think that people often ought to do so. We think, for instance, that parents ought to want what's best for their children, that agents ought to intend to take what they believe to be the necessary means to their ends, and that those who believe that the Earth is no more than a few thousand years old ought to abandon this belief in light of the overwhelming scientific evidence to the contrary. Now, since wanting, intending, and believing are reasons-responsive attitudes that are under our rational control, the rationalist can accept such commonsense normative judgments. But, since these attitudes are not under our voluntary control, the voluntarist cannot. The voluntarist must deny, for instance, that those who believe that the Earth is no more than a few thousand years old ought to abandon this belief. For the voluntarist must deny that they have the option of abandoning this belief given that they cannot do so at will. The best the voluntarist can do, then, is to claim that these people ought to perform (at will) whatever voluntary acts would cause them to abandon this belief. But, in many instances, they can't even claim that. For, in many instances, agents ought not to perform the acts that would cause them to abandon such a belief. To illustrate, suppose that, given Chen's penchant for conspiracy theories, the only thing that would cause her to abandon her belief that the Earth is no more than a few thousand years old is to read a book by some quack claiming that the Bible was written and propagated by the CIA for the purposes of controlling the masses. And

let's assume that Chen ought not to read this book both because it contains a lot of dangerous misinformation that she is liable to believe and because she promised her mother that she would stop feeding her hysteria by reading such books. Here, then, is a case where Chen ought to abandon her belief and the voluntarist can't even claim that she ought to do what will cause her to abandon this belief. The rationalist, by contrast, has no problem accounting for the fact that Chen ought to abandon her belief. For Chen has, we'll presume, rational control over whether she does so. That is, she has, we'll presume, the capacity to respond appropriately to the reasons that she has for abandoning this belief—specifically, those stemming from the scientific evidence of which she is aware. And if she were to respond appropriately to these reasons, she would thereby abandon her belief. It's just that, in this case, she culpably fails to respond appropriately to her reasons.

Second, it's not just that voluntarism forces us to reject our commonsense normative judgments about reasons-responsive attitudes. It also forces us to reject some of our commonsense normative judgments about acts. For many of the acts that we ought to perform are *mixed acts*—acts that have both a voluntary component and a non-voluntary component.[19] Such acts include acting in good faith, offering a sincere apology, expressing one's gratitude, and diverting a trolley onto to the side track with the intention of saving the five on the main track.[20] To perform such acts, we must have certain reasons-responsive attitudes. For instance, we can't act in good faith without having the intention to follow through with our part of the bargain. We can't offer a sincere apology without feeling contrite. We can't express our gratitude without feeling grateful. And we can't divert a trolley with the intention of having one effect and not another without intending to have the one but not the other. So whereas we may have voluntary control over whether we say the words "I'm sorry" or "Thank you," we don't have voluntary control over whether we offer a sincere apology or express our gratitude. For

[19] Alex King (2014) is, as far as I know, the first to point out that we often presume that we ought to perform such mixed acts even though we cannot perform them at will. I'm indebted to her 2014 for many of the ideas presented in this paragraph.

[20] One type of mixed act is what Ralph Wedgwood (2011) calls *thick act-types*: act-types that incorporate specific intentions. Examples include "diverting the trolley in order to kill the person on the side track" and "diverting the trolley in order to save the five people on the main track." And, as Wedgwood points out, we must be able to prescribe such mixed acts (and not merely the physical acts of flipping or not flipping a switch) if we are to give a plausible account of the Doctrine of Double Effect. So, if Wedgwood is right about this, and I'm right in thinking that such mixed acts essentially include a non-voluntary component, we would have to abandon hope of giving a plausible account of the Doctrine of Double Effect if we were to endorse voluntarism.

we don't have voluntary control over whether we feel contrite or grateful, which is essential to performing such mixed acts. The voluntarist must, therefore, deny that we have the option of performing such mixed acts (at least, where we don't already have the relevant attitudes) and, so, deny both that we ought to perform such acts and that we can appropriately be held accountable when we fail to do so. This seems like a substantial cost for those who accept voluntarism. Fortunately, rationalism doesn't come at such a high price. Since we have rational control, for instance, over both whether we utter the words "I'm sorry" and over whether we feel contrite, the rationalist holds that we have the option of performing such mixed acts and, thus, can be obligated to perform them and responsible for failing to perform them.

Third, voluntarism implausibly implies that, in determining whether lower animals, young children, and the criminally insane have the sorts of options for which they can be obligated to perform and accountable for failing to perform, what's relevant is whether they have voluntary control as opposed to rational control over their actions. But such beings are often not responsible for their actions despite having voluntary control over them. Consider, for instance, that although both cats and psychopathic killers have the capacity to form an effective intention to kill their victims quickly (after all, they do so when it's in their self-interest to do so), we don't hold them obligated to do so or accountable for failing to do so when it's not in their self-interest to do so. For they lack the capacity to recognize the relevant moral reasons for dong so and so lack the capacity to respond appropriately to these reasons. That is, they are incapable of recognizing that the fact that killing quickly causes less suffering counts in favor of their killing quickly. And because of this, their awareness of the fact that their killing quickly causes less suffering could not lead them (qua rational agents) to kill quickly.[22] And this is why we treat them very differently than, say, some contract killer who fails to respond appropriately to the moral reasons that she has for killing quickly despite her having the capacity to do so. So although psychopathic killers and contract killers both have voluntary control over whether

[22] To respond appropriately to the relevant moral reasons, the cat/psychopath would have to be aware of the fact that its victim will suffer less if killed quickly, and this awareness must then be what (non-deviantly) causes the cat/psychopath to kill its victim quickly.

they kill quickly, we hold that only contract killers are responsible for failing to do so, because only they have the capacity to respond appropriately to the moral reasons for killing quickly.

To sum up, then, we're interested in an account of when a subject S has the option to φ in the sense that's relevant to determining whether S has an obligation to φ and would be accountable for failing to φ, and such an account must get the extension of both variables—'S' and 'φ'—right if it's to be plausible. And the problem with voluntarism is that it gets neither right. First, voluntarism holds that the extension of 'S' is all and only those subjects who can do things at will (that is, voluntarily). But it seems that although lower animals, young children, and the criminally insane can do things (such as kill quickly) at will, they are not the sorts of subjects that have, say, the option to kill quickly in the sense that's relevant to determining their obligations and responsibilities. Rather, it seems that the relevant subjects are, as rationalism would have it, all and only those who have the capacity to recognize and respond appropriately to the relevant reasons.

Second, voluntarism holds that the extension of 'φ' is all and only those things over which we exert voluntary control—that is, our voluntary acts and those things that we manipulate through our voluntary acts. But our options seem to extend well beyond our voluntary acts and those things that we manipulate through them. It seems, for instance, that we can have obligations and responsibilities with respect to things that are not under our voluntary control, such as whether we form certain reasons-responsive attitudes, whether we perform certain mixed acts, and whether we overlook the most obvious solution to an easy math problem. And this is as rationalism would have it.

So, at this point, I hope to have convinced the reader that rationalism provides a plausible account of what are options are. I will now show that when we combine rationalism with maximalism, we end up with a view—viz., rationalist maximalism—that offers our best hope for finding a moral theory that is morally harmonious.

2. The Principle of Moral Harmony

The principle of moral harmony holds that a moral theory must be morally harmonious—that is, it must be such that the agents who satisfy the theory, whoever and however numerous they

may be, are guaranteed to produce the morally best world that they together have the option of producing. And although many philosophers accept this principle, they often disagree on how exactly it is to be formulated.[23] But before we can properly evaluate some of the different proposed formulations, it's important for us to understand the general motivation for the principle. The motivation stems from our intuition that there has been a moral failure in following sort of case, which I borrow, with some modification, from David Estlund (Forthcoming).

> *Slice and Patch Go Golfing:* Unless a patient's tumor is removed very soon, she'll die (though not painfully). Immediate surgery and stitching by the only two available doctors, Dr. Slice and Dr. Patch, is the one thing that will save her life. But if there is surgery without stitching, her death will be agonizing. And if there is stitching without surgery, her death will likewise be agonizing. It would even be cruel for one of them to show up to the hospital knowing that the other won't, for this would only needlessly get the patient's hopes up, making her death psychologically agonizing. So, it seems that Dr. Slice, who doesn't know how to stitch, ought to show up to the hospital to perform the surgery if and only if Dr. Patch will be there to stitch her up afterwards. And Dr. Patch, who doesn't know how to perform the surgery, ought to be there to give her stitches if and only if Dr. Slice will be there to perform the surgery. Unfortunately, Slice and Patch are each going golfing whether the other attends to the patient or not, because neither cares whether the patient is cured. And each knows this about the other. Predictably, then, the patient dies (though not painfully) while Slice and Patch each enjoy a pleasant round of golf.

As Estlund (Forthcoming) points out, "many of us respond to this case with the intuition that there is some moral violation here, but the puzzle is to find an agent who has committed it." It seems that Dr. Slice was under no obligation to be at the hospital to perform the surgery

[23] Proponents include, for instance, Baier 1958, Castañeda 1974, Parfit 1984 (94), Pinkert 2015, Regan 1980, and Zimmerman 1996. Of course, not all philosophers accept the principle. Critics include Feldman 1980 and Kierland 2006.

given that, as he knew, no one was going to be there to stitch the patient up afterwards. And it seems that Dr. Patch was under no obligation to be at the hospital to stitch up the patient given that, as he knew, no one was going to be there to perform the surgery. And this is exactly what most moral theories imply. For instance, act-utilitarianism implies that Dr. Slice was under no obligation to show up to the hospital, because his showing up to the hospital would have produced less utility than his going golfing did. After all, had he showed up to the hospital, the patient's death would have been agonizing. As it was, it was painless. Likewise, act-utilitarianism implies that Dr. Patch was under no obligation to show up to the hospital, because his showing up to the hospital would have produced less utility than his going golfing did. So, on act-utilitarianism, neither violated a moral obligation in not showing up to hospital. Moreover, it seems that other plausible moral theories will have the same implication. At least, they will insofar as it seems plausible to suppose that there is no obligation to do what will only make things worse.

Still, it seems that there must have been some moral violation here. After all, the group consisting of Slice and Patch could have brought about a significantly better world: the one in which the patient lives. If Slice had shown up to the hospital and performed the surgery and Patch had shown up to hospital and stitched her up afterwards, the patient would have lived. So how could it be that they each did all that they were morally required to do when they failed to bring about the morally best world that the two of them were capable of bringing about? Of course, to even pose this rhetorical question is to presuppose that the satisfaction of a moral theory by some group must guarantee that they end up in the morally best world that they are together capable of bringing about. Thus, it seems that if we are to accommodate the intuition that there has been a moral failure in this case, we must accept the principle of moral harmony. And, perhaps, we should formulate this principle as follows:

(MH) A moral theory, T, is correct if and only if the agents who satisfy T, whoever and however numerous they may be, are guaranteed to produce the morally best world that that they could together bring about.[24]

But however plausible the principle of moral harmony is, MH is an implausible formulation of it. For, if MH were true, no moral theory could be correct. Donald Regan proved this back in 1980 with his excellent book *Utilitarianism and Co-operation*.[25] Let me explain his reasoning. First, Regan shows that no moral theory that is exclusively act-orientated can satisfy MH, where a theory is exclusively act-orientated if and only if it requires only that agents perform and refrain from performing certain voluntary acts.[26] To illustrate, consider that, in *Slice and Patch Go Golfing*, neither Slice nor Patch perform any immoral act. After all, there is nothing immoral about going golfing when there is nothing better that one could be doing. And, given that each is going golfing regardless of what the other does, neither can do anything to save the patient's life. Each individual can only make things worse for the patient by showing up to the hospital, giving her false hope and, thereby, making her death psychologically agonizing. Thus, if we are to claim that there has been a moral violation in this case, the violation must lie, not with their voluntary actions, but with something else. In this case, the violation seems to lie with neither of them caring whether the patient is cured and/or with each of them intending to go golfing regardless of what the other does. But to refrain from caring whether the patient is cured is not to refrain from performing a voluntary act. Nor is intending to go golfing regardless of what the other does a voluntary act.[27] Thus, the only way for a moral

[24] This is based on Donald Regan's definition of 'adaptability': "a theory T is adaptable if and only if the agents who satisfy T, whoever and however numerous they may be, are guaranteed to produce the best consequences possible as a group, given the behaviour of everyone else" (1980, 6). The difference is that whereas Regan is concerned only with consequentialist moral theories, I'm concerned with moral theories in general. Thus, the morally best world could, if, say, Kantianism is correct, be the world in which the group commits the fewest and/or least significant violations of the categorical imperative possible as opposed to the world in which the group produces the best consequences possible.

[25] He proved that no theory can be adaptable unless we ignore the direct consequences of applying that theory's decision procedure. See his 1980, chapter 10.

[26] This is not Regan's definition, for he provides no definition—see 1980, 109. But I believe that this definition captures (at least, sufficiently well for our purposes) the notion that he had in mind.

[27] To see that the formation of an intention is not something one can do at will, consider the toxin puzzle (Kavka 1983). I will receive a million dollars tomorrow morning if and only if, at midnight tonight, I intend to drink some toxin tomorrow afternoon. Drinking the toxin will not kill me, but it will make me terribly ill for several days. Whether I receive the million dollars tomorrow

theory to pass MH is for it to require that agents not only perform certain voluntary acts, but also form (non-voluntarily) certain attitudes, such as the desire that the patient is cured and the intention to attend to the patient's needs should the other also be willing to do so.

Second, Regan shows that any moral theory that is not exclusively act-orientated will violate MH. For any moral theory that's not exclusively act-oriented will have to require "something more" of agents than just the performance (or non-performance) of certain voluntary acts. And, as Regan notes, "there is always the possibility that there will be a mad telepath…who will blow up Macy's [or half of the planet] in response to that 'something more'" (1980, 181). And, of course, a world in which half of the planet is blown up is not going to be the morally best world that Slice and Patch can bring about. For they could just refrain from this something more, in which case billions of lives would be spared. And it's better that one patient dies painlessly than that billions die in a horrible explosion. (Assume that the other half of the planet will be fine after the one half is blown up and that the patient resides on the side of the planet that won't be blown up.) So consider a revised version of *Slice and Patch Go Golfing*, which I'll call *The Mad Telepath.* In this case, everything is the same as in original but for the addition of a mad telepath who will blow up half of the planet if either Slice or Patch form the desire to cure the patient (or even if either of them forms the intention to attend to the patient's needs). So, unlike *Slice and Patch Go Golfing*, this is a case where Slice and Patch would not bring about the morally best world by desiring to cure the patient.

To illustrate the problem with MH, note that in order for Slice and Patch to cure their patient they must perform the set of acts consisting in Slice's showing up to the hospital and performing the surgery and Patch's showing up to the hospital and stitching up the patient afterwards. Call this set of acts 'SET'. And consider the following two theories. The first is an exclusively act-orientated theory (call it 'ACTS-ONLY') that holds that an agent's only moral obligation is to voluntarily perform her portion of SET if and only if the other agent has fitting

---

morning depends only on what I intend to do at midnight tonight, not on whether I drink the toxin tomorrow afternoon. Realizing this, I'm unable to intend at midnight tonight to drink the toxin tomorrow afternoon. For I see no reason to drink the toxin. I know that, come tomorrow afternoon, I'll either have the million dollars or I won't. And in neither case will I have any reason to drink toxin. Indeed, I have decisive reason not to drink the toxin. Given this, I'm unable to form the intention to drink the toxin. And this shows that forming an intention is not typically something that I can do at will—that is, by forming the intention to form that intention.

attitudes, including the desire to cure the patient.[28] The second theory, by contrast, is not exclusively act-oriented. Call it 'MORE-THAN-ACTS', because it requires more than just the performance of certain voluntary acts. It holds that an agent has exactly two moral obligations: an obligation to voluntarily perform her portion of SET if and only if the other agent has fitting attitudes and an obligation to non-voluntarily form fitting attitudes, including the desire to cure the patient.

The problem with ACTS-ONLY (a problem that other exclusively act-orientated theories share) is that it implies that, in *Slice and Patch Go Golfing*, neither Slice nor Patch violate any moral obligation. Thus, ACTS-ONLY does not satisfy MH. For, as *Slice and Patch Go Golfing* illustrates, Slice and Patch can each satisfy ACTS-ONLY and yet fail to bring the morally best world that that they could together bring about—the one where their patient is cured. Now, this problem is avoided by MORE-THAN-ACTS, for to comply with this theory Slice and Patch must each form fitting attitudes. Moreover, each has to perform his portion of SET if the other forms fitting attitudes, which the other must form if he is to satisfy MORE-THAN-ACTS. But, unfortunately, MORE-THAN-ACTS has its own problem (one that it shares with other theories that are not exclusively act-orientated). It implies that, in *The Mad Telepath,* Slice and Patch are each required to form a desire that would result in the mad telepath's blowing up half of the planet. And the world where the mad telepath blows up half of the planet is far from the morally best world that they could together bring about. Together they could bring about the world where only the patient dies by refraining from forming the desire to cure the patient. Thus, MORE-THAN-ACTS does not satisfy MH.

We see, then, that there is no way for a moral theory to satisfy MH. A moral theory will either have to be exclusively act-orientated or not. If, on the one hand, it is exclusively act-

[28] What makes it fitting to desire that the patient is cured is the fact that the patient's being cured is itself desirable. Thus, it can be fitting for Slice (and Patch) to form the desire that the patient is cured even though a mad telepath is going to destroy half of the planet if he does so. The mad telepath's response makes it instrumentally bad (and, thus, unfortunate) to desire that the patient is cured, but it doesn't change the fact that the patient's being cured is itself desirable and thus something fitting to desire. And it's crucial to make the obligation to perform one's portion of SET conditional on one's having fitting attitudes (attitudes that are appropriate) as opposed to fortunate attitudes (attitudes that are instrumentally valuable), for, otherwise, there would be no obligation for Slice to form the desire to cure the patient in *Slice and Patch Go Golfing*, because no good would come from his forming this desire given that Patch is going golfing regardless. And a theory must require an agent like Slice to form this fitting desire if it's going to be such that the agents who satisfy the theory, whoever and however numerous they may be, are guaranteed to produce the morally best world that they have the option of producing in cases like *Slice and Patch Go Golfing*.

orientated, then there will be instances, such as in *Slice and Patch Go Golfing*, where a group of agents who all satisfy the theory fail to bring about the morally best world that that they could together bring about, because they fail to have certain fitting attitudes (such as the desire to cure the patient) that they're permitted by that theory to refrain from forming. If, on the other hand, a moral theory is not exclusively act-orientated, then there will be instances, such as in *The Mad Telepath*, where a group of agents who all satisfy the theory fail to bring about the morally best world that they could bring about, because they are required by that theory to form certain fitting attitudes (such as the desire to cure the patient) that will result in a mad telepath's destroying half of the planet. So either way there will be instances in which the moral theory is satisfied by everyone in the group and yet the group fails to bring about the morally best world that they could bring about. For there is no way for a moral theory to require agents to have certain fitting attitudes (such as the desire to cure the patient) while also requiring them not to have those same attitudes.

Clearly, then, MH is too strict a formulation of the principle of moral harmony. I think that MH goes wrong in insisting that a moral theory must be such that if, in *The Mad Telepath*, Slice forms the desire to cure the patient, he must have thereby violated a moral requirement, since his forming this desire results in the production of a suboptimal world—specifically, the one in which the mad telepath destroys half of the planet. This, I believe, is a mistake, because Slice doesn't have voluntary control over whether he refrains from forming this desire. And this means that he cannot refrain from forming this desire for whatever reason he takes to be sufficient reason to do so. So, even if he takes the fact that his forming this desire would lead to the destruction of the planet as sufficient reason for him to refrain from forming this desire, he cannot refrain from forming this desire for this reason. Indeed, given his lack of voluntary control, he can refrain from forming this desire only for the sort of reason that would make it fitting for him to so refrain, such as the reason that he would have to so refrain if the patient's being cured were undesirable. But the patient's being cured *is* desirable. Thus, the only way that he can refrain from forming this desire is by failing to respond appropriately to the decisive

reason that he has to form this desire—i.e., the fact that the patient's being cured is desirable.[29] And I don't think that morality can require us to fail to respond appropriately to our reasons just because our so failing would prevent some disaster. That is, I don't think that morality can require us either to form attitudes for which we lack sufficient reason or to refrain from forming attitudes for which we have decisive reason.[30]

To see why, consider the following example.

> *Hating and Saving Rocks:* Unless I both hate professional wrestler Dwayne Johnson (a.k.a. The Rock) because of his Samoan ancestry and intend to kill him with my bare hands in a fair fight, an evil demon will destroy the third rock from the sun (that is, the planet Earth).

Note that the fact that Dwayne Johnson is of Samoan ancestry is no reason to hate him. And I can't just voluntarily hate him for the fact that he is of Samoan ancestry. I cannot even hate him for the fact that my hating him would save the planet. Given that I don't have voluntary control over whether I hate him, I cannot hate him for whatever "reason" I take to be sufficient reason for hating him. I can hate him only for the reason that I think him despicable or otherwise deserving of hatred—that is, only because I think that it is fitting to hate him. The problem is that I don't think that he is despicable. Certainly, neither the fact that he is of Samoan ancestry nor the fact that an evil demon will destroy our planet if I don't hate him make him despicable. So I can't hate him—at least, not insofar as I respond appropriately to my reasons.[31] What's more, I can't form the intention to kill him with my bare hands in a fair fight insofar as I respond appropriately to my reasons. For I know that I cannot take him in a fair fight given his

---

[29] I'm assuming (somewhat controversially) that the fact that Slice's being in the state of desiring his patient's continued life would have undesirable consequences is not a reason for Slice to refrain from desiring his patient's continued life and that only facts about the desirability or undesirability of his patient's continued life can constitute reasons (or, at least, reasons of the right kind) for or against his desiring his patient's continued life. For more on this, see Way 2012.

[30] As I'll use the terms, S has *decisive reason* to φ if and only if S's reasons are such as to make S obligated to φ, and S has *sufficient reason* to φ if and only if S's reasons are such as to make S permitted to φ.

[31] Again, I'm assuming (somewhat controversially) that the relevant reasons for having an attitude are only fittingness reasons (i.e., reasons having to do with the fittingness of that attitude) and not pragmatic reasons (i.e., so-called "reasons" that supposedly stem from the instrumental value in having that attitude). Thus, I respond appropriately to my reasons by having fitting attitudes, not by having attitudes that are instrumentally valuable. And, again, see Way 2012 for more on this issue.

massive physique and superior fighting abilities.[32] Moreover, even if I could, I would not intend to kill him, for I have no good reason to kill him and have many good reasons not to. And so, if I respond appropriately to my reasons, I will not intend to kill him with my bare hands in a fair fight. Thus, responding appropriately to my reasons precludes me from forming the attitudes that I must form in order to prevent the destruction of our planet.

Now, it would be very strange to think that morality could require me to respond inappropriately to my reasons given that what makes me the sort of subject to which moral obligations and responsibilities apply is that I'm the sort of subject who's capable of responding appropriately to my reasons—that is, a rational agent. It seems nonsensical for some moral requirement to apply to me because I have the capacity to respond appropriately to my reasons when I can fulfill that requirement only by failing to respond appropriately to my reasons. That is, it seems nonsensical for the very capacity in virtue in which I'm obligated to φ is the one that, if exercised flawlessly, leads me not to φ.[33] Such would be the nature of a moral requirement for me either to hate Johnson because of his Samoan ancestry or to intend to kill him with my bare hands in a fair fight. Thus, it seems a mistake to think, as MH supposes, that morality can require us to respond inappropriately to our reasons by either forming attitudes for which we lack sufficient reason (e.g., my forming a hatred for Johnson) or refraining from forming attitudes for which we have decisive reason (e.g., Slice's refraining from forming the desire that the patient is cured).

If this is right, the principle of moral harmony should not insist that the correct moral theory guarantees that we produce the morally best world that it's possible for us to produce (where this may only be possible by, say, our failing to respond appropriately to our reasons), but should instead insist only that the correct moral theory guarantees that we produce the morally best world that is compatible with each of us responding appropriately to our reasons—call this *reasonable moral harmony*.

---

[32] I'm assuming here that I cannot rationally intend to φ while knowing that I will not φ. The idea that there is a consistency requirement to the effect that the set of the propositional contents of all of one's beliefs and intentions must be logically consistent is a popular, but not uncontroversial, view. For a defense of this view, see Ross 2009.

[33] For a defense of this idea, see my 2011, 38–51.

(RMH) A moral theory, T, is correct if and only if the agents who satisfy T, whoever and however numerous they may be, are guaranteed to produce the morally best world that is compatible with each of them responding appropriately to their reasons.[34]

RMH allows us to account for our intuition that there must have been some moral failure in *Slice and Patch Go Golfing*. The moral failure lies with the fact that Slice and Patch failed to have the desires and intentions that they were morally required to have. They were morally required to desire that the patient lives and to intend to attend to her so long as the other is also willing to do so. And if they both had had these attitudes, they would have each been such that the other was required to attend to the patient. Thus, abiding by a moral theory that passes RMH ensures that Slice and Patch produce, in *Slice and Patch Go Golfing*, the morally best world that they are capable of producing: the one in which the patient lives. But RMH does not imply that abiding by the correct moral theory ensures that Slice and Patch prevent, in *The Mad Telepath*, the destruction of half of our planet. Abiding by a moral theory that passes RMH ensures only that Slice and Patch produce the morally best world that's compatible with their responding appropriately to their reasons—that is, the morally best world that they can reasonably be expected to bring about. And for them to respond appropriately to their reasons, they must, or so I'm assuming, form the desire that the patient is cured and form the intention to attend to her so long as the other also desires to cure the patient. So, unfortunately, the morally best world that's compatible with their responding appropriately to their reasons is, in *The Mad Telepath*, the world in which the mad telepath destroys half of the planet. But this is no reason to reject RMH. After all, we should not expect the correct moral theory to ensure that no

[34] My formulation of the principle of moral harmony is distinct from those that have been proposed by Pinkert (2015), Regan (1980), and Zimmerman (1996). I don't have space here to explain why I reject each of these alternative proposals, but see Kierland 2006 and Forcehimes & Samrau 2015 for criticisms of each of these proposals. Now, Kierland would reject my proposal as well on the grounds that it violates the principle that 'ought' implies 'can'. But it violates this principle only if we assume that the relevant sense of 'can' is to be analyzed in terms of voluntary control as opposed to rational control, contrary to what I've argued for here. And Forcehimes & Samrau 2015 would probably question the need to appeal to the principle of moral harmony at all and suggest that we would do better to locate Slice's and Patch's moral failures with whatever past acts that they performed that led them to have their problematic attitudes. But see my example involving Chen and her belief that the Earth is no more than a few thousand years old as well as Smith 2015 for why I think this sort of tracing strategy won't work.

disaster ever befalls us. For no matter what attitudes a moral theory requires of us, a mad telepath can always wreak havoc upon us for fulfilling those requirements by responding appropriately to our reasons and thereby forming the appropriate attitudes. Moreover, we should think that the correct moral theory will never require us to respond inappropriately to our reasons given that our being rational is what subjects us to moral obligations and responsibilities in the first place. So, we should expect only reasonable moral harmony from the correct moral theory.

According to RMH, a moral theory, T, will be correct if and only if the agents who satisfy it while otherwise responding appropriately to their reasons, whoever and however numerous they may be, are guaranteed to produce the morally best world that is compatible with each of them responding appropriately to their reasons. And the only sort of theory that can do that in a plausible way is one that, like rationalist maximalism, holds that the permissibility of some particular act depends on whether it is part of some permissible whole that includes both the performance of various voluntary acts and the formation of various reasons-responsive attitudes. Indeed, the only sort of theory that I can see satisfying RMH in a plausible way is some version of rationalist maximalism that takes 'has feature F' in Max2 to stand for something like the following: 'includes all and only those attitudes that are fitting and includes all and only those acts that would, if the agent were to have all and only fitting attitudes, actualize the morally best world that would be actualized by any of the acts available to the agent'.[35]

Let me explain. Only those theories that endorse rationalism will satisfy RMH. To satisfy RMH, a theory must require not only that agents perform certain voluntary acts (such as showing up to the hospital and then cutting [or stitching up] the patient), but also that they

[35] If we interpret rationalist maximalism as a theory about *moral* permissibility and make this substitution for 'has feature F', rationalist maximalism will entail that a maximal option is morally permissible only if it includes all those attitudes that are fitting. This may seem problematic since, in many instances, the fittingness of an attitude will have nothing to do with morality. For instance, the fitting response to the perception of a bright flash of lighting is often the belief that one will hear a loud clap of thunder very shortly. Nevertheless, there wouldn't necessarily be anything morally impermissible about a maximal option that failed to include this fitting attitude. I suggest, then, that when we make this substitution for 'has feature F', we also modify Max2 so that we get: "(Max2) for any maximal option μ, S's μ-ing is *both rationally and* morally permissible if and only if S's μ-ing includes all and only those attitudes that are fitting and includes all and only those acts that would, if the agent were to have all and only fitting attitudes, actualize the morally best world that could be actualized by any of the acts available to the agent."

non-voluntarily form certain attitudes (such as the desire to cure the patient). This was the lesson of *Slice and Patch Go Golfing*, a case that showed that no exclusively act-orientated theory will satisfy RMH. And the only way for a theory to require that agents not only voluntarily perform certain acts but also non-voluntarily form certain attitudes is to embrace rationalism's idea that the sort of control that's relevant to determining an agent's obligations and responsibilities is rational, not voluntary, control, for this is the only sort of control that we have over both acts and attitudes.

For another, only those theories that endorse maximalism will satisfy RMH in a plausible way. For, unless a theory adopts maximalism, it's going to generate conflicting obligations: specifically, both an obligation to refrain from φ-ing and an obligation to both φ and ψ.[36] To illustrate, let's consider the case that I call *Only Slice Goes Golfing*. In this case, Patch desires to cure the patient. And so he'll show up to the hospital and stitch up the patient if he knows that Slice desires to cure the patient and plans on showing up and performing the surgery. Nevertheless, Patch knows that Slice doesn't desire to cure the patient. Consequently, he stays at home and watches television while monitoring Slice's desires via some telepathic or technological means. And now let's compare a version of maximalism that is not exclusively act-orientated with a version of non-maximalism that is not exclusively act-orientated. Call these two "FA-versions," because they both require agents to have fitting attitudes (hence, "FA").

> *FA-Maximalism:* (Max1) For any subject S and any non-maximal option *ν* (which could be any non-maximal set of acts and/or attitudes), S's *ν*-ing is permissible if and only if there exists an option φ such that S's φ-ing is permissible and S's φ-ing entails S's *ν*-ing. And, (FA-Max2) for any subject S and any maximal option *μ*, S's *μ*-ing is permissible if and only if S's *μ*-ing includes all and only those attitudes that are fitting and includes all and only those acts that would, if the agent were to have all and only fitting attitudes, actualize the morally best world that would be actualized by any of the acts available to S.

[36] For more on this, see Portmore 2016a and Portmore 2016b.

> *FA–Non-Maximalism:* (1) For any subject S and any attitude α, S is permitted to form α if and only if it would be fitting for S to form α. And, (2) for any subject S and any act χ, S is permitted to perform χ if and only if S's χ-ing would, given the agent's actual (fitting or unfitting) attitudes, actualize the morally best world that would be actualized by any of the acts available to S.

Now, on FA-maximalism, Slice is obligated both to form the desire to cure the patient and to show up to the hospital and perform the surgery, because Slice's only permissible maximal options are ones in which he both forms the desire to cure the patient and shows up to the hospital and performs the surgery. And since this entails his showing up hospital and performing the surgery, he's obligated to show up to the hospital and perform the surgery. So there are no conflicting obligations here. But, on FA–non-maximalism, there are. FA–non-maximalism implies that Slice is obligated to form the desire to cure the patient, show up to the hospital, and perform the surgery. For he's obligated to form this desire and is obligated to perform these actions if he forms this desire. But FA–non-maximalism also implies that Slice is obligated not to show up to the hospital. Given that he doesn't desire, and isn't going to form the desire, to cure the patient, Patch isn't going to show up to the hospital. And given that Patch isn't going to show up to the hospital, it would be better if Slice didn't show up either. So on FA–non-maximalism we get conflicting obligations. He's obligated to form and perform the set consisting of his forming the desire to cure the patient, showing up to the hospital, and performing the surgery. Yet he's obligated to refrain from showing up to the hospital.[37]

Thus, it seems that rationalist maximalism is uniquely well-suited to accommodate the idea that the correct moral theory must be reasonably morally harmonious.[38] A theory must

---

[37] Admittedly, not everyone (e.g., Jackson and Pargetter 1986) finds such conflicting obligations problematic, but see Kiesewetter 2015 (pp. 929–934) and Portmore 2011 (pp. 181–183) for why we should deny that there can be such conflicting obligations.

[38] Rationalist maximalism also has the benefit of avoiding Feldman's objection to the principle of moral harmony. Feldman asks us to imagine "that, a group of adults has taken a group of children out to do some ice skating. The adults have assured the children and their parents that, in case of accident, they will do everything in their power to protect the children. …A lone child is skating in the middle, equidistant from the adults. Suddenly, the ice breaks, and the child falls through. There is no time for consultation or deliberation. Someone must quickly save the child. However, since the ice is very thin, it would be disastrous for more than one of the adults to venture near the place where the child broke through. For if two or more were to go out, they would all fall in and all

accept rationalism so as not to be exclusively act-orientated and thus accommodate RMH. And it must endorse maximalism in order to avoid conflicting obligations. Thus, insofar as we think that a moral theory must both be reasonably morally harmonious and avoid conflicting obligations, we have reason to think that rationalist maximalism (or something very much like it) is the correct moral theory.

3. Conclusion

Rationalist maximalism is, I've argued, the most plausible version of maximalism. For, as I've argued, rationalism is the most plausible account of our options, and so the one that we should combine with maximalism. And I've done more that just argue that rationalist maximalism is the most plausible version of a certain class of moral theories (viz., maximalist theories). I've also argued that, insofar as we think that a moral theory must both be reasonably morally harmonious and avoid conflicting obligations, we have reason to think that it is the correct moral theory.

Of course, even if rationalist maximalism is the correct moral theory, that doesn't mean that we must reject wholesale all alternative moral theories. For rationalist maximalism doesn't provide a substantive account of what we should aim at, morally speaking. Should we, for instance, aim solely at making the world contain as much happiness as possible or also at respecting people's autonomy even when doing so will make the world contain less happiness? Thus, even if we should, as I propose, adopt a version of rationalist maximalism according to which 'has feature F' in Max2 stands for something like 'includes all and only those attitudes that are fitting and includes all and only those acts that would, if the agent were to have all and only fitting attitudes, actualize the morally best world that could be actualized by any of the acts available to the agent', we still need to know which of the various possible worlds that the agent could actualize is the morally best one from her perspective. Is it, for instance, the one

---

would be in profound trouble. In fact, let us suppose, no one goes to the aid of the child" (1980, 171–172). Feldman believes that each adult was morally obligated to quickly head out to the hole in the ice to rescue the child. But, of course, if each adult did this, disaster would ensue. Thus, Feldman concludes that the principle of moral harmony is false. But if we accept rationalist maximalism, we should say that the relevant moral obligation was not merely to head out to the hole in the ice to rescue the child, but to do so while intending to stop abruptly should there be any indication that others are heading out toward the hole as well. And it is not the case that if each adult satisfied this moral requirement, disaster would ensue.

where she commits one murder so as to prevent five others from each committing one comparable murder or is it the one in which there are more murders but none that have been committed by herself? Theories like Kantianism and utilitarianism can help us answer such questions.

Still, even if we are not to reject such theories wholesale, we must, if rationalist maximalism is correct, reject them in their current forms. As things stand, such theories assess only acts and assess each act (whether maximal or non-maximal) according to the same standards.[39] But if rationalist maximalism is correct, this is a mistake. If we are to accept rationalist maximalism, we should think that moral theories must require us to form certain reasons-responsive attitudes, and not just to perform certain voluntary acts. What's more, they must evaluate each non-maximal act in terms of whether its performance is entailed by that of some permissible maximal option, not in terms of whether it maximizes happiness or accords with Kant's categorical imperative.[40]

References

Baier, K. (1958). *The Moral Point of View*. Ithaca, NY: Cornell University Press.

[39] I take Kantianism to be the view that, for any subject S and any act *x*, S's *x*-ing is permissible if and only if S's *x*-ing accords with Kant's categorical imperative. Now, I readily acknowledge that Kant did provide an account of the virtues and this account can be used to assess attitudes. The problem, then, is that Kantianism, as I've just defined it, holds that we can assess acts independently of our assessing attitudes. But I've argue that we need to assess them together as maximal options that include both acts and attitudes.

[40] For helpful comments and discussions on precursors to this paper, I thank audiences at Syracuse University, Arizona State University, Australian National University, University of Colorado, Boulder, University of Maryland, College Park, the 2014 Rocky Mountain Ethics Congress, the Arizona Center for the Philosophy of Freedom, and the 2014 Pacific Division Meeting of the American Philosophical Association. I especially thank Chrisoula Andreou, Brad Armendt, David Boonin, Elizabeth Brake, Campbell Brown, Cheshire Calhoun, Kendra Chilson, Eric Chwang, Stew Cohen, Yishai Cohen, Brad Cokelet, Peter de Marneffe, Dale Dorsey, Andrew Forcehimes, Josh Gert, Peter A. Graham, Pat Greenspan, Johan E. Gustafsson, Chris Heathwood, Brian Hedden, Jonathan Herington, Adam Hosein, Alex King, Victor Kumar, Frank Jackson, Seth Lazar, Eden Lin, Thomas Mautner, David McElhoes, Conor McHugh, Gene Mills, Dan Moller, Christopher Morris, G. Shyam Nair, Howard Nye, Graham Oddie, N. Ángel Pinillos, Felix Pinkert, Andrew Reisner, Steve Reynolds, Melinda Roberts, Luke Semrau, Dan C. Shahar, Walter Sinnott-Armstrong, Nicholas Southwood, Steve Sverdlik, Christian Tarsney, Larry Temkin, Travis Timmerman, Michael Tooley, Jean-Paul Vessel, Steve Wall, Ralph Wedgwood, Stephen White, and several anonymous referees. Work on this paper was supported by the RSSS Visiting Fellows Program, School of Philosophy, Australian National University.

Brown, C. (2015). "Consequentialism, Maximalism, and the Structure of Acts." Unpublished manuscript.

Bykvist, K. (2002). "Alternative Actions and the Spirit of Consequentialism." *Philosophical Studies* 107: 45–68.

Castañeda, H.-N. (1974). *The Structure of Morality*. (Springfield, Ill.: Charles Thomas Publisher).

Estlund, D. (Forthcoming). "Prime Justice." In K. Vallier and M. Weber (eds.), *Political Utopias*. Oxford: Oxford University Press.

Feldman, F. (1986). *Doing the Best We Can: An Essay in Informal Deontic Logic.* Dordrecht: D. Reidel Publishing Company.

———. (1980). "The Principle of Moral Harmony." *The Journal of Philosophy* 77: 166-179.

Fischer, J. M., and M. Ravizza. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Forcehimes, A. T. & Semrau, L. (2015). "The Difference We Make: A Reply to Pinkert." *Journal of Ethics & Social Philosophy*, www.jesp.org, (September Discussion Note).

Graham, A. (2014). "A Sketch of a Theory of Moral Blameworthiness." *Philosophy and Phenomenological Research* 88: 388–409.

Goldman, H. S. [now H. M. Smith]. (1978). "Doing the Best One Can." In A. I. Goldman and J. Kim (eds.), *Values and Morals*, pp. 185–214. Dordrecht: D. Reidel Publishing Company.

Gustafsson, J. E. (2014). "Combinative Consequentialism and the Problem of Act Versions." *Philosophical Studies* 167: 585–596.

Hieronymi, (2008). "Responsibility for Believing." *Synthese* 161: 357–373.

———. (2006). "Controlling Attitudes." *Pacific Philosophical Quarterly* 87: 45–74.

Jackson, F. and R. Pargetter (1986). "Oughts, Actions, and Acualism." *Philosophical Review* 95: 233–255.

Kavka, G. (1983). "The Toxin Puzzle." *Analysis* 43: 33–36.

Kierland, B. (2006). "Cooperation, 'Ought Morally', and Principles of Moral Harmony." *Philosophical Studies* 128: 381–407.

Kiesewetter, B. (2015). "Instrumental Normativity: In Defense of the Transmission Principle." *Ethics* 125: 921–946.

King, A. (2014). "Actions that We Ought, But Can't." *Ratio* 27: 316–27.

Maier, J. (2014). "Abilities." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall Edition), URL = <http://plato.stanford.edu/archives/fall2014/entries/abilities/>.

McHugh, C. (Forthcoming). "Attitudinal Control." *Synthese*.

———. (2014). "Exercising Doxastic Freedom." *Philosophy and Phenomenological Research* 88: 1–37.

———. (2012). "Epistemic Deontology and Voluntariness." *Erkenntnis* 77: 65–94.

Parfit, D. (2011). *On What Matters. Vol. 2.* Oxford: Oxford University Press.

———. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Pinkert, F. (2015). "What If I Cannot Make a Difference (and Know It)." *Ethics* 125: 971-998.

Portmore, D. W. (2016a). "Maximalism vs. Omnism about Reasons." Unpublished manuscript.

———. (2016b). "Maximalism vs. Omnism about Permissibility." Unpublished manuscript.

———. (2013). "Perform Your Best Option." *The Journal of Philosophy* 110: 436–459.

———. (2011). *Commonsense Consequentialism: Wherein morality meets rationality*. New York: Oxford University Press.

Regan, D. (1980). *Utilitarianism and Co-operation*. New York: Oxford University Press.

Ross, J. (2012). "Actualism, Possibilism, and Beyond." In M. Timmons (ed.) Oxford Studies in Normative Ethics: Volume 2, pp. 74–96, Oxford: Oxford University Press.

———. (2009). "How to Be a Cognitivist about Practical Reason." *Oxford Studies in Metaethics* 4: 243–282.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, Mass.: Belknap Press.

Sinnott-Armstrong, W. (1984). "'Ought' Conversationally Implies 'Can'." *Philosophical Review* 93: 249–61.

Smith, A. M. (2015). "Attitudes, Tracing, and Control," *Journal of Applied Philosophy*. Advance online publication available at http://dx.doi.org/10.1111/japp.12107.

———. (2005). "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115: 236–271.

Streumer, B. (2003). "Does 'Ought' Conversationally Implicate 'Can'?" *European Journal of Philosophy* 11: 219–228.

Way, J. (2012). "Transmission and the Wrong Kind of Reason." *Ethics* 122: 489–515.

Wedgwood, R. (2011). "Defending Double Effect." *Ratio* 24: 384–401.

Wiland, E. (2005). "Monkeys, Typewriters, and Objective Consequentialism." *Ratio* 18: 252–360.

Zimmerman, M. J. (1996). *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.