

Actualism, Possibilism, and Beyond¹

How is what an agent ought to do related to what an agent ought to prefer that she does? More precisely, suppose we know what an agent's preference ordering ought to be over the outcomes of performing the various courses of action open to her. Can we infer from this information how she ought to act, and if so, how can we infer it? One view (which, for convenience, I will call 'actualism') is that an agent ought to ϕ just in case she ought to prefer the outcome that would result from her ϕ -ing to the outcome of that would result from her not ϕ -ing.² Another view (which, for convenience, I will call 'possibilism') is that an agent ought to ϕ just in case all of her options (in the relevant domain) with maximally preferable outcomes involve ϕ -ing. I will discuss actualism and possibilism in parts 1 and 2, respectively. I will argue, in part 1, that actualism is very far from the truth. And I will argue, in part 2, that while the standard version of possibilism faces significant problems, there are much better versions of possibilism that avoid the objections to the standard view. Ultimately, however, I will argue that even the best forms of possibilism are not acceptable. Then, in part 3, I will offer a diagnosis of why the existing theories fail, and I will offer an alternative theory that is neither actualist nor possibilist in form, and that avoids the difficulties with the other theories.

1. Actualism

1.1 The Absurdity of Actualism

The main problem with actualism is that it's obviously false. This can be seen if we consider the following case:

¹ I am very grateful to my colleagues Kenny Easwaran, Mark Schroeder, and Gideon Yaffe for many invaluable discussions concerning this paper. I owe a special debt to Shieva Kleinschmidt for extremely helpful comments on several drafts of this paper. My greatest debt to Doug Portmore. Many of the key ideas in this paper arose in the course of a correspondence I had with him during the summer of 2010. Had it not been for this correspondence, this paper could not have been written.

² Examples of actualism are Sobel (1976) and Jackson and Pargetter (1986). Examples of possibilism are Feldman (1986) and Zimmerman (1996).

Arsenic and Old Ace: Absentminded Ace is looking after his three year old granddaughter Emily, who asks him for a glass of water. As a matter of fact, Ace is about to accidentally give Emily a glass containing an arsenic solution. When he does so, she will drink the contents of the glass and die. Ace has many options besides giving Emily the arsenic. He could instead give her a glass of water as she requested. Or he could give her a glass of Drano, or a glass of bleach, or a glass of ammonia. If he were to give her glass of water, she would drink it and be happy. If, on the other hand, he were to give her a glass of a glass of Drano, bleach, or ammonia, then while she would not drink enough of the liquid for it to be fatal, she would drink enough to suffer severe and irreversible damage to her mouth and esophagus.³

Of the options just described, it seems fairly obvious that there is only one that Ace ought to carry out: he ought to give Emily a glass of water. And there are many that he ought not to carry out: he ought not to give her arsenic, he ought not to give her Drano, he ought not to give her bleach, and he ought not to give her ammonia. The actualist, however, disagrees. The actualist says that, objectively speaking, Ace ought to give Emily Drano. For the nearest world in which Ace doesn't give Emily Drano is the actual world, where he gives her arsenic. Thus, Ace ought to prefer what would obtain if he gave her Drano (namely, that Emily is harmed but not killed) to what would obtain if he didn't give her Drano (namely, that Emily is killed). And so, objectively speaking, he ought to prefer the outcome of giving her Drano to the outcome of not giving her Drano, which implies, according to actualism, that he ought to give her Drano. For similar reasons, actualism implies that he ought to give her bleach, and it likewise implies that he ought to give her ammonia.

In this case, we can see two problems with actualism. First, there are many cases where it implies that agents ought to do really awful things: for any action ϕ , no matter how bad ϕ is, actualism entails that one ought to ϕ so long as ϕ -ing isn't quite as bad as what one actually does. In addition to implying that agents ought to do really awful things, there are simply too many things that actualism entails one ought to do. If there are a million options that are less bad than what one actually does, then actualism will entail, for each of these options, that one ought to do it. In the case just considered, we have seen that actualism entails, of each of three incompatible courses of action, that Ace ought to do it. But these aren't the only things it entails that Ace ought to do. Consider the course of action that consists in giving Emily Drano and then telling his boss where to go; or the course of action that consists in giving Emily Drano and then

³ A similar counterexample to actualism is presented in Wedgwood 2009.

lighting the cat on fire; or the course of action that consists in giving Emily Drano and then spending the rest his one's life dressed as Napoleon Bonaparte. Each of these options is such that the outcome of carrying it out is preferable to the outcome of giving Emily the arsenic solution. And so actualism entails, for each of these courses of action, that Ace ought to do it. Thus, actualism leads to a deontic explosion of obligations to do terrible things.

1.2 The Contextualist Response

How might the actualist respond to this objection? One response, suggested by Jackson and Pargetter (1986), is to distinguish between two questions: the question of *whether* an agent ought to ϕ at t , and the question as to *what* the agent ought to do at t . On the Jackson-Pargetter view, *whether* x ought to ϕ at t depends on whether ϕ -ing at t is better than what x would do if x didn't ϕ at t . But *what* x ought to do at t is whichever of the maximally relevantly specific options available to x at t would have the best outcome. Assume that, in *Arsenic and Old Ace*, the time of action is noon. While Jackson and Pargetter are committed to saying that Ace ought to give Emily Drano at noon (since doing so is better than what he would do otherwise, namely give her arsenic), they are not committed to saying that what he ought to do at noon is give Emily Drano. Rather, on their view, there is only one correct answer to the question of what Ace ought to do at noon, namely, that he ought to give her a glass of water. For, among the maximally relevantly specific options that are available to Ace at noon, giving her a glass of water is maximally preferable, as it would have the best outcome. Perhaps, therefore, we can explain away our intuition that Ace ought not to give Emily Drano at noon, as a result of our conflating the two questions that Jackson and Pargetter distinguish. Perhaps we correctly judge that it is not the case that giving Emily Drano is what Ace ought to do at noon, and from this we fallaciously infer that Ace ought not to give Emily Drano at noon.

On the face of it, this solution does not appear coherent. If Ace ought to give Emily Drano at noon, then how could this fail to be (or to be part of) what he ought to do at noon? Jackson and Pargetter suggest a way of avoiding this charge of incoherence. The suggestion is that 'ought' claims are implicitly relativized to sets of options. When we ask whether Ace ought to give Emily Drano, the relevant set of options consists of just two alternatives: {giving Emily Drano; not giving Emily Drano}. But when we ask what she ought to do at noon, the relevant set of options will be {giving Emily Drano; giving Emily bleach; giving Emily ammonia; giving

Emily the arsenic solution; giving Emily water}. Thus, there is no conflict between the claim that Ace ought to give Emily Drano at noon and the claim that giving Emily Drano is no part of what Ace ought to do at noon. For the former claim is implicitly relativized to the first, more coarse-grained partition of options, whereas the latter claim is implicitly relativized to the second, more fine-grained partition of options.⁴

As I see it, however, this contextualist maneuver doesn't really solve the problem we discussed in the last section, for it retains the implausible implication that Ace ought to give his granddaughter Drano (or, in the formal mode, that utterances of the sentence 'Ace ought to give his granddaughter Drano' are true). Further, even if it isn't strictly incoherent, it still fails to respect the intuitive connection between question of *whether to ϕ at t* and the question of *what to do at t* . Moreover, it introduces a further problem, which can be seen if we consider the following case.

President Muffley: It is now noon. President Merkin Muffley had accidentally activated the Doomsday Machine. Before him are two buttons, A and B. At 12:01, he will have exactly three options: he can press button A, he can press button B, or he can press neither button. If and only if he presses neither button at 12:01, at 12:02 he will have the opportunity to press button A. The outcomes for the various courses of action open to him are as follows:

- Press button A 12:01: 30% of the world destroyed
- Press button B at 12:01: 70% of the world destroyed
- Press neither button at 12:01; press button A at 12:02: 0% of the world destroyed
- Press neither button at either time: 100% of the world destroyed

Muffley knows that he himself will die right after 12:02, and he is indifferent to the fate of the world after his death. Thus, he has made up his mind about what to do by tossing coins. As a result, he will press neither button at 12:01, and regardless of what he does at 12:01, he will press neither button at 12:02.

On Jackson and Pargetter's contextualist view, when we ask *whether Muffley ought to press only button B at 12:01*, the relevant options are *pressing only button B at 12:01* and *not pressing only button B at 12:01*. And since the former would result in 70%, rather than 100%, of the world being destroyed, the following claim is true:

- (i) Muffley ought to press only button B at 12:01.

⁴ For other views on which 'ought' claims are relativized to sets of options, see Cariani (forthcoming) and Snedegar (forthcoming).

For similar reasons, if we ask whether Muffley ought to press only button A at 12:01, the answer will again be affirmative, so the following is true:

(ii) Muffley ought to press only button A at 12:01.

But now suppose we ask not *whether* Muffley ought to perform a given action, but rather *what* he ought to do at a given time. Suppose we ask, first, what he ought to do at 12:01. Here the relevant options will be all the maximally relevantly specific things he could do at 12:01, namely pressing button A, pressing button B, and pressing neither button. Among these, pressing button A is the option that would in fact have the best outcome, since it would result in only 30% of the world being destroyed, whereas pressing button B would result in 70% destruction, and pressing neither button would result in 100% destruction. And so the following claim is true.

(iii) What Muffley ought to do at 12:01 is press only button A.

Next, suppose we ask what Muffley ought to do at 12:02. Since, at 12:01, he won't press either button, he will have two options at 12:02, namely pressing only button A and pressing neither button. And, of these options, pressing button A would have the best outcome. And so the following is true:

(iv) What Muffley ought to do at 12:02 is press only button A.

Finally, suppose we ask not what Muffley ought to do at a particular time, but rather what he ought to do over the course of his life. In this case, the relevant options will be the four courses of action indicated in the description of the case. And, among these, the one with the best outcome consists in pressing neither button at 12:01, and then pressing button A at 12:02. And so the following will be true.

(v) What Muffley ought to do over the course of his life is press neither button at 12:01, and then press button A at 12:02.

Thus, we arrive at very different answers depending on which normative question we ask. What role, therefore, should the answers to these various questions play in deliberation?

It is generally agreed that rational agents intend to do what they believe they ought to do. As John Broome (1999 and forthcoming) puts the point,

Enkrasia: Rationality requires, of every agent, that if she believes she ought to ϕ , and she believes that it's up to her whether she ϕ s, then she intends to ϕ .

But if we accept the Jackson-Pargetter view, then we must reject this principle of Enkrasia. For on the Jackson-Pargetter view, if Muffley knows the facts of the case as we have described them, then he should believe both that he ought to press only button A at 12:01 and that he ought to press only button B at 12:01. It is hardly plausible, however, that rationality both requires him to intend to press only button A at 12:01 and requires him to intend to press only button B at 12:01. For then rationality would require him to have inconsistent intentions, and so rationality would require him to be irrational. Thus, anyone who accepts the Jackson-Pargetter view must deny that the principle of Enkrasia applies to our beliefs of concerning *whether we ought to act in certain ways*, such as Muffley's beliefs in (i) and (ii).

Nor can the principle of Enkrasia apply to our beliefs concerning *what we ought to do at a given time*, such as Muffley's beliefs in (iii) and (iv). For if this principle did apply to such beliefs, then it would imply that, if Muffley were aware of the relevant facts, then rationality would require him to press button A at 12:01, and rationality would also require him to intend to press button A at 12:02. But Muffley knows that if he presses button A at 12:01, then he won't have the option of pressing button A at 12:02. Thus, if the principle of enkrasia applied to such beliefs, then rationality would require Muffley to have intentions that he knows he can't jointly satisfy. And so rationality would require that Muffley be irrational.

It seems, therefore, that on the Jackson-Pargetter view, if the principle of Enkrasia applies at all, it can apply only to our beliefs about what we ought to do over the course of our lives, not to our ordinary beliefs about whether we ought to act in a given way, or about what we ought to do at a given time. Note, however, that as Jackson and Pargetter (1986) themselves admit (p. 251-252), when it comes to the question of what an agent ought to do over the course of her life, their view has the exact same implications as the standard possibilist view. Thus, with respect to the only kinds of ought claims that satisfy the principle of Enkrasia, the Jackson-Pargetter view collapses into the standard form of possibilism. And, as I have argued elsewhere, following John Broome, it is precisely the 'ought' claims that satisfy the principle of Enkrasia that are fundamentally normative, in the sense that they directly guide the deliberations of rational agents.⁵ It follows, therefore, that with respect to the ought-claims that are fundamentally

⁵ See Ross (2010) and (Unpublished) and Broome (Unpublished).

normative, the Jackson-Pargetter view collapses into the standard form of possibilism. It is to possibilism, therefore, that we should now turn.

2. Possibilism

We have defined possibilism as the view that an agent ought to ϕ just in case all her maximally preferable options (within the relevant domain of options) involve ϕ -ing. Hence we can distinguish among different versions of possibilism, corresponding to different specifications of the relevant domain of options, as follows.

Possibilism concerning a domain D of options is the view that, for any option ϕ in domain D, an agent x ought at t to ϕ just in case all the maximally preferable options among the options in D available to x at t involve ϕ -ing.

In the next three sections, we will consider three versions of possibilism corresponding to three alternative specifications of the relevant domain of options.

2.1 Possibilism Concerning Performable Options

The standard form of possibilism involves a very broad conception of the relevant domain of options. On this conception, the relevant options are any options available to the agent, in the broadest sense of the term, or what we may call the *performable options*. How exactly the class of performable options is to be defined is a complicated matter that I will not pursue in detail, but to a first approximation, we may say that, at time t, an option ϕ is performable for x just in case there is some schedule of intentions, beginning at t, such that if x's intentions followed this schedule, then x would ϕ . Or, stated more simply, ϕ -ing is performable for x at t just in case, if x had the right intentions at the right times from t forward, then x would ϕ .

Possibilism concerning performable options is the view x ought to ϕ just in case all of x's maximally preferable performable options involve ϕ -ing. While this view may be prima facie plausible, it seems to get the wrong results in certain cases in which an agent's future actions are not currently under her deliberative control. Consider, for example, the following case.

The Cabinet of Dr. Caligari: Ace's granddaughter Emily needs medicine. If she doesn't get her medicine, she will be very sick for a few days, and then she'll fully recover. However, the medicine she needs is in the cabinet, right next to the arsenic solution. And the cabinet is guarded by Dr. Caligari. Ace knows that if, and only if, he goes to the cabinet, Dr. Caligari will inject him with the dreaded Nepticide Serum. Anyone who is

injected with this serum acquires a strong desire to kill his or her granddaughter. After being injected with this serum, one remains psychologically normal aside from this desire. One does not, for example, lose the psychological capacity to form or retain non-murderous intentions. However, if Ace were to go to the cabinet with the intention of giving Emily the medicine, then, as a matter of fact, he would change his mind after being injected with the serum, and he would end up giving Emily the arsenic solution, thereby killing her.

It seems clear that, in this case, Ace ought not to go to the cabinet. However, possibilism concerning performable options implies otherwise. For going to the cabinet and then giving Emily the medicine is a performable option for Ace—for if he had the right intentions at the right times (and, in particular, if he continued to intend to do so even after being injected with the serum) then he would do so. Moreover, this will be his maximally preferable performable option. And since this option involves going to the cabinet, it follows that all his maximally preferable performable options involve going to the cabinet. And so it follows from possibilism concerning performable options that Ace ought to go to the cabinet. And this result is very implausible.

2.2 Simple Securitism

So what's going wrong in *The Cabinet of Dr. Caligari*? Possibilism says Ace should carry out the best performable option, which is to go to the cabinet and give Emily the medicine. This seems wrong, because it seems that, in the relevant sense, this option isn't now really open to Ace: even if he now fully intended to do it, he wouldn't do it, because he'd change his mind later. Perhaps, therefore, we ought to focus not on the domain of performable options, in the sense defined above, but rather on the narrower class of options that one would actually carry out if one were to intend, at the time under consideration, to carry them out. Inspired by Doug Portmore, I will call these the agent's *directly securable options*. More precisely, let us say that ϕ is a directly securable option for x at t just in case there is some set S of intentions such that,

- (i) If, at t , x had all the intentions in S , then x would carry out all these intentions;
- (ii) Carrying out all the intentions in S would involve ϕ -ing;
- (iii) At t , x has the psychological capacity to have the intentions in S .

Thus, while the options that are now *performable* for an agent at a time are the ones she would carry out if she had the right intentions *now and in the future*, the options that are now *directly*

securable for an agent are the ones she would carry out if she had the right intentions *now*. We may now define *simple securitism* as the possibilist view as applied to directly securable options. That is, simple securitism is the view that, at t , x ought to ϕ just in case all the maximally preferable, directly securable options that are available to x at t involve ϕ -ing. A view of this kind was proposed by Holly Smith (then Holly Goldman) in her brilliant paper “Doing the Best One Can” (1978).

Simple securitism gets what are intuitively the right results in all the cases we have considered so far. In *Arsenic and Old Ace*, it implies that Ace ought to give his granddaughter water, and that he ought not to give her Drano (since all his maximally securable, directly securable options involve giving her water and not giving her Drano). In *President Muffley*, it implies that Muffley ought to press neither button at 12:01, and that he ought to press button A at 12:02, since all his maximally preferable, directly securable options involve doing both these things. And in *The Cabinet of Dr. Caligari*, it implies that Ace ought not to go to the cabinet. However, there are some cases where it intuitively gets the wrong results. Here is one such case.

General Ripper: Brigadier General Jack D. Ripper has accidentally activated the Doomsday Machine, and he’s the only one who can deactivate it. It is now noon, and at 12:01 he will receive a prompt saying “Do you want to deactivate the Doomsday Machine?” If he presses the button to answer “yes,” then at 12:02 he will receive a second prompt saying “Are you *sure* you want to deactivate the Doomsday Machine?” If Ripper proceeds to press the second button, thereby answering “yes” to the second prompt, then the Doomsday Machine will be deactivated, and the world will be saved. But if he fails to press both buttons, then the machine will not be deactivated, and the entire world will be destroyed. Furthermore, if he presses the first button but not the second, then, right before the world is destroyed, he will undergo a mild electric shock. It so happens that Ripper would prefer that the world be destroyed. Hence, even if he were now to intend to press both buttons, he would change his mind before 12:02, and so, while he might press the first button, he would not press the second.

It seems obvious, in this case, that Ripper is under a moral obligation to press both buttons. After all, it was he who accidentally activated the Doomsday device, he’s the only one who can deactivate it, and pressing both buttons is only way he can deactivate it. Moreover, pressing both buttons is something he could easily do. It is highly implausible that the mere fact that he has a preference for the destruction of the world would get him off the hook from any obligation to save the world.

But simple securitism implies otherwise. For at noon, Ripper has no securable options that involve pressing both buttons, since, at noon, his intending to do so would not result in his doing so. And, among his directly securable options, those that involve pressing the first button would not have better outcomes than those that do not. It is not the case, therefore, that all his maximally preferable, directly securable options involve pressing the first button, and so simple securitism does not imply that he ought to do so. To the contrary, it implies that he ought to press *neither* button. For his only directly securable options involve either pressing neither button or pressing only the first button. And, among these options, it would be preferable for him to press neither button so as to avoid the electric shock.

It seems, therefore, that simple securitism has unacceptable implications in cases like *General Ripper*, where what would seem to be obligatory courses of action (such as pressing both buttons) are not directly securable because the agent has morally impermissible attitudes (such as preferring the destruction of the world).

2.3 Scrupulous Securitism

In an early draft of his excellent forthcoming book, *Commonsense Consequentialism: Wherein Morality Meets Rationality*, Portmore advocated a version of simple securitism. However, when I presented cases similar to *General Ripper*, he has revised his view. On his revised view—which we may call *scrupulous securitism*—what an agent ought to do is not whatever is involved in all her maximally preferable *directly securable* options, but rather whatever is involved in all her maximally preferable *scrupulously securable* options. He defines such options as follows:

A set of actions α_j is, as of t_i , scrupulously securable by S if and only if there is a time t_j that either immediately follows t_i or is identical to t_i , a set of actions α_i (where α_i may, or may not, be identical to α_j), and a set of background attitudes B such that the following are all true: (1) S would perform α_j if S were to have at t_j both B and the intention to perform α_i ; (2) S has at t_i the capacity to continue, or to come, to have at t_j both B and the intention to perform α_i ; and (3) and S would continue, or come, to have at t_j B (and, where α_i is not identical to α_j , the intention to perform α_i as well) if S both were at t_i aware of all the reason-constituting facts and were to respond at t_j to these facts/reasons in all and only the ways that they prescribe, thereby coming to have at t_j all those attitudes that she has decisive reason to have and only those attitudes that she has sufficient reason to have.

Given some plausible assumptions, scrupulous securitism gets the intuitively correct verdict in *General Ripper*. In particular, it implies that Ripper ought to press both buttons. For, while

there is no *directly securable* option that involves pressing both buttons, there is plausibly a *scrupulously securable* option that involves pressing both buttons. This is true because, plausibly, there is a set of background attitudes B (which will include a preference for saving the world) such that (1) Ripper would press both buttons if he had B and intended to press both buttons, (2) Ripper has the capacity to have B and the intention to press both buttons, and (3) Ripper would have B if he were aware of, and appropriately responsive to, all the reasons for the attitudes in B. Since pressing both buttons is scrupulously securable for Ripper, and since doing so is the only way Ripper can save the world, we may infer that all of Ripper's maximally preferable, scrupulously securable options involve pressing both buttons. And so it will follow, from scrupulous securitism, that Ripper ought to press both buttons.

However, there are other cases in which Scrupulous Securitism gets what is intuitively the wrong result. Here's an example

Ice Cream of Doom: Vanessa has been kidnapped by Martians, and is playing a high-stakes game called "ice cream of doom." She is asked which, out of chocolate or vanilla, is her preferred flavor of ice cream. If she successfully says either "chocolate is my preferred flavor" or "vanilla is my preferred flavor," then she will be released and given one million dollars, regardless of whether her answer is true. But if she fails to utter either sentence, then she and every other human being will be vaporized. Marvin the mindreading Martian is in attendance, and he hates liars. And so if she begins to utter a falsehood, he will paralyze her larynx, preventing her from completing her sentence, and thus ensuring the destruction of humanity.

As a matter of fact, Vanessa prefers vanilla ice cream to chocolate ice cream. However, Vanessa could change her preference to chocolate ice cream. Moreover, Vanessa isn't very knowledgeable about ice cream, and if she became aware of all the relevant facts, then she would come to prefer chocolate ice cream. This reversal in her preferences, however, would be merely permissible. That is, it would be perfectly permissible for Vanessa, upon learning all the relevant facts, to continue to prefer vanilla ice cream. For the objective reasons do not favor either ice cream preference over the other, and in light of these reasons, either preference would be permissible.

It seems obvious that Vanessa is under no obligation to utter, or to begin uttering, "chocolate is my preferred flavor." For, given her permissible preference for vanilla ice cream, her beginning to utter this sentence would result in the destruction of humanity. By contrast, her uttering "vanilla is my preferred flavor" would result in her receiving a million dollars. It seems, therefore, that it must be at least permissible, if not required, for Vanessa to utter the latter sentence. Scrupulous securitism, however, implies the opposite. For there is no scrupulously securable option in which Vanessa utters "vanilla is my preferred flavor." For it so happens that

if Vanessa were aware of, and responded appropriately to, all the facts, then she would prefer chocolate ice cream, and if she had this preference and intended to say “vanilla is my preferred flavor,” then she would be prevented from completing the sentence. By contrast, there *is* a scrupulously securable option in which Vanessa utters “chocolate is my preferred flavor.” For if she were aware, and responded appropriately to, all the relevant facts, then she’d prefer chocolate ice cream, and with this preference, she could successfully utter “chocolate is my preferred flavor.” Moreover, successfully uttering “chocolate is my preferred flavor” is her maximally preferable, scrupulously securable option, since it would result in her being released and receiving million dollars. Thus, scrupulous securitism implies that Vanessa ought to utter “chocolate is my preferred flavor.” And since uttering this sentence would involve beginning to utter this sentence, scrupulous securitism implies that Vanessa ought to begin to utter this sentence.

There would seem to be a natural way to solve this problem. All we need to do is substitute the word “could” for the word “would” in clause (3) of Portmore’s definition of a scrupulously securable option, so that it reads “(3) and S *could* continue, or come, to have at t_j ...” On this revised conception, a scrupulously securable option is defined not as an option that would be directly securable given *the set of background attitudes the agent would actually have* if she responded appropriately to the relevant reasons; rather it is defined as an option that would be directly securable given *some set of background attitudes the agent could have* if she responded appropriately to the relevant reasons.

If we make this move, then Vanessa’s uttering “vanilla is my preferred flavor” will count as scrupulously securable, and so we will avoid the implication that Vanessa ought to utter “chocolate is my preferred flavor.” The problem will reemerge, however, if we consider a modified version of *Ice Cream of Doom*, in which Vanessa would receive a million dollars for saying “vanilla is my preferred flavor,” but she would receive a million and one dollars for saying “chocolate is my preferred flavor.” In this modified version of the case, the mere fact that Vanessa would receive one dollar more if she said “chocolate is my preferred flavor” would hardly seem to obligate her to prefer chocolate—it seems like a reason of the wrong kind to require preferring chocolate. So it seems Vanessa could permissibly prefer Vanilla even in the modified case. And if she could permissibly prefer Vanilla, then it seems she could permissibly answer “vanilla is my preferred flavor.” But according to scrupulous securitism, even in the

modified form we are now considering, Vanessa is obligated to answer “chocolate is my preferred flavor.” For uttering this sentence is a scrupulously securable option, and its outcome is preferable to that of any other scrupulously securable option.

So far we have considered problems that arise for specific versions of securitism. *General Ripper* raises a special problem for simple securitism, and *Ice Cream of Doom* raises a special problem for scrupulous securitism. In the next section, we will consider a more general problem, the problem of *nonratifiability*, facing both simple and scrupulous securitism.

2.4 The Problem of Nonratifiability

The problem of nonratifiability is illustrated by the following case.

Satan’s School for Girls: It is now June 6, 2011. Sally is a fine, upstanding teenage girl whose moral character is impeccable. However, she is about to be kidnapped by Satanists, and brought up in Satan’s School for Girls. As a result, her moral character will be severely corrupted, and she will come to desire, more than anything else, to sacrifice her firstborn child to the Prince of Darkness. At the appointed hour, on June 6, 2016, Sally will have the opportunity to kill her firstborn child, and will be able to do so either by cutting off the child’s head with an axe, or by bludgeoning the child to death with a club. At that very moment, Child Services personnel will be arriving on the scene, and so if she refrains from killing her child, the child will be taken into protective custody and live happily ever after. If, on June 6, 2011, before being kidnapped by the Satanists, Sally were to intend to cut off the head of her firstborn child, then she would follow through with this intention five years later. And if she were to intend to bludgeon her child to death, then she would likewise follow through with this intention five years later. But if she were to intend not to kill her firstborn child, then she would change her mind after becoming a Satanist, and end up cutting off her child’s head. The above three subjunctive conditionals are true not by any fault of Sally’s, but purely in virtue of the fact that she is about to be kidnapped and brought up at Satan’s School for Girls.

In this case, both forms of securitism have a very peculiar implication: they imply that on June 6, 2011, Sally ought to cut off the head of her firstborn child five years later, and yet they imply that, no matter what she may do in the mean time, it will be true five years later that she ought not to cut off the head of her firstborn child then. For on June 6, all of Sally’s maximally preferable, *directly* securable options involve cutting off the head of her firstborn child. And it is likewise true that all of her maximally preferable, *scrupulously* securable options involve cutting off the head of her firstborn child. For, given that she’s about to be brainwashed, there is no combination of intentions and background attitudes she could have that would prevent her from

killing her firstborn child five years later, and, among the two ways of killing her firstborn child, cutting off its head would cause less pain. Therefore, both simple securitism and scrupulous securitism imply that on June 6, 2011, Sally ought to cut off the head of her firstborn child five years later. However, on June 6, 2016, it will be true that all her maximally preferable, *directly* securable options involve *not* cutting off the head of her first born child. And it will likewise be true that all her maximally preferable, *scrupulously* securable options involve *not* cutting off the head of her firstborn child. For, at that time, refraining from killing her firstborn child will be both directly securable and scrupulously securable. And so simple securitism and scrupulous securitism both imply that, on June 6, 2016, Sally ought not to cut off the head of her firstborn child.

Thus, simple securitism and scrupulous securitism both imply that, prior to the time of action, it may be *obligatory* for an agent to act in a given way even though it will inevitably be *impermissible* for the agent to act in this way at the time of action. Thus, both theories make what we may call *nonratifiable prescriptions*: prior to the time of action, they make prescriptions which they will inevitably reverse at the time of action, regardless of what the agent does in the interim. It's hard to deny that this feature threatens the plausibility, if not the very coherence, of these theories.⁶

3. Beyond

In this concluding part of the paper, I will begin, in section 3.1, by offering a diagnosis of why the theories we have considered thus far go wrong. I will then present an alternative theory of obligation that avoids these difficulties.

3.1 Diagnosis: Why the Existing Theories Fail

So far we have considered various theories of obligation that attempt to understand what an agent ought to do in terms of a preferability ordering over the outcomes of her options. We have seen that each of these theories faces serious problems. I will now offer a diagnosis. An adequate theory of obligation must capture the following intuition:

⁶ The problem of nonratifiability is also faced by actualism. It is avoided by possibilism concerning performable options, as well as by the view presented in Goldman (1976).

The Core Idea: What is obligatory for an agent is that, at all times, she does the best she can do at that time, holding fixed what is not up to her at that time, but not holding fixed what is up to her at that time.

The various theories we have considered fail, because they fail to capture some aspect or other of this core idea. To see how this works, let's label the various aspects of the core idea.

The Core Idea Broken Down: What is obligatory for an agent is that, (i) *at all times*, she does (ii) *the best she can do* at that time, (iii) *holding fixed what is not up to her at that time*, but (iv) *not holding fixed what is up to her* at that time.

Actualism fails because it fails to capture aspect (ii) of the Core Idea. Actualism does not identify what an agent ought to do with whatever is involved in doing the best she can do; rather, it identifies what an agent ought to do with whatever is preferable to what she would do otherwise. And as a result, actualism has horrific implications in cases such as *Arsenic and Old Ace*, where it implies that the agent ought to carry out the second worst option, no matter how bad it may be.

The standard form of possibilism, what we have called *possibilism concerning performable options*, fails because it fails to capture aspect (iii) of the Core Idea. That is, in evaluating what an agent ought to do at a given time, it fails to hold fixed the facts that are not up to the agent at the time in question. In particular, in evaluating what an agent ought to be doing at a time *t*, it fails to hold fixed facts about what the agent will do at future times *irrespective* of the actions, intentions, and other attitudes of the agent at *t*. And so it fails to recognize that, in deciding how to act at a given time, there are sometimes facts about our future actions that we ought to treat as facts about the world rather than as objects of choice. As a result, possibilism concerning performable options gets the wrong result in *The Cabinet of Dr. Caligari*. In this case, it treats certain ends as realizable (namely, giving Emily the medicine) even though Ace would not realize these ends regardless of what his current actions and attitudes might be.

Simple Securitism fails because it fails to capture aspect (iv) of the core idea. That is, in evaluating what an agent ought to do at a given time, it holds fixed facts that are up to the agent at the time in question. In particular, while it doesn't hold fixed facts about the agent's present intentions, it does hold fixed all the facts about the agent's other present attitudes that do not depend on her present intentions. As a result, it gets the wrong result in *General Ripper*. Here it holds fixed the fact that Ripper prefers the destruction of the world, and so it implies that he

ought to press neither button, since pressing both is not securable given this preference. But it seems that Ripper's present preferences are no less up to Ripper than Ripper's present intentions. After all, Ripper *ought* to prefer that the world be saved, and so, in the sense of 'can' that is relevant to normative evaluation, it seems to follow that Ripper *can* prefer that the world be saved. And if Ripper can prefer that the world be saved, it seems he must be able, in the relevant sense, to press both buttons. It seems, therefore, that simple securitism is wrong to exclude this course of action from consideration in evaluating what Ripper ought to do.

Finally, both simple securitism and scrupulous securitism fail because they fail to capture aspect (i) of the Core Idea. That is, they fail to capture the idea that we are obligated to do the best we can at all times. Since the maximally preferable securable option at a given time, t , may involve failing to do the best one can do at future times, both forms of securitism imply that, at t , one can be obligated to fail to do the best one can do at future times. As a result, they get the wrong result in *Satan's School for Girls*, where they imply that, on June 6, 2011, Sally is obligated to kill her firstborn child five years later.

3.2 Toward an Adequate Theory of Obligation

I propose the following theory, which, for reasons that will become clear, I will call *momentwise wide-scope securitism* (MWSS).

For any x and t , at t , x ought to be such that, for all t' from t forward, x satisfies the following conditional:

For all ϕ , if whether x ϕ s does not causally depend on the intentions x has after t' , and if every maximally preferable option that is directly securable for x at t' involves ϕ -ing, then x ϕ s.

Roughly, this principle states that, at any given time, an agent is obligated, at every future time, to be currently satisfying a wide-scope version of simple securitism. I will now illustrate this theory and argue that it captures each of the four aspects of the Core Idea. Recall that we stated this idea as follows.

The Core Idea Broken Down: What is obligatory for an agent is that, (i) *at all times*, she does (ii) *the best she can do* at that time, (iii) *holding fixed what is not up to her at that time*, but (iv) *not holding fixed what is up to her* at that time.

To see how MWSS captures aspect (i), consider *Satan's School for Girls*, where the other forms of securitism fail to capture (i). The other forms of securitism imply that, on June 6th, 2011, Sally is under an obligation to cut off the head of her firstborn child five years later, since, on June 6th, 2011, all her maximally preferable, directly securable options involve so doing. MWSS avoids this implication. For what this theory implies is that, on June 6, 2011, Sally ought to be such that (if whether she cuts off the head of her firstborn child five years later does not depend on the intentions she will have *after* June 6, 2011, and if all her maximally preferable, directly securable options involve cutting off the head of her firstborn child five years later, then she cuts off the head of her firstborn child five years later). However, whether Sally cuts off the head of her firstborn child five years later clearly *does* depend on the intentions she will have after June 6, 2011, and so the antecedent of this conditional is not satisfied. Thus, Sally can satisfy the conditional without satisfying the consequent. And so MWSS does not imply that she is under any obligation to satisfy the consequent by cutting off the head of her firstborn child.

To the contrary, MWSS implies that, on June 6th, 2011, Sally is under an obligation to refrain from killing her firstborn child five years later. For MWSS implies that, on June 6th, 2011, Sally ought to be such that, on June 6th, 2016, she satisfies the conditional (if whether Sally refrains from killing her firstborn child on June 6th, 2016 does not depend on her intentions after that time, and if every maximally preferable option that is directly securable for Sally on June 6th, 2016 involves refraining from killing her firstborn child, then she refrains from killing her firstborn child). And, given the description of the case, come what may, Sally will satisfy the antecedent of this conditional. Thus, the only way Sally can satisfy this conditional is by refraining from killing her firstborn child on June 6th, 2016. MWSS therefore implies that, on June 6th, 2011, Sally is under an obligation to refrain from killing her firstborn child five years later. Thus, because of the manner in which MWSS quantifies over times—because, we might say, of its *momentwise* character—it captures aspect (i) of the Core Idea.

To see how MWSS captures aspect (ii), consider *Arsenic and Old Ace*, where actualism fails to capture (ii). In this case, MWSS correctly implies that Ace ought to now give his granddaughter water. For it implies that Ace ought to satisfy the following conditional (if whether Ace now gives his granddaughter water does not depend on his future intentions, and if all his *maximally* preferable, directly securable options involve now giving his granddaughter water, then he now gives his granddaughter water). And, given the description of the case, come what may he will satisfies the antecedent of this conditional.⁷ And so he can satisfy the conditional only by giving his granddaughter water. The same does not hold, however, if we substitute “giving his granddaughter Drano” for “giving his granddaughter water”—for then Ace will permissibly satisfy the corresponding conditional by failing to satisfy the second conjunct of its antecedent: it is not the case that all his *maximally preferable*, directly securable options involve giving his granddaughter Drano. And so, unlike actualism, MWSS does not imply that Ace ought to give his granddaughter Drano. Thus, because of its *maximizing* character, MWSS captures aspect (ii) of the core idea.

To see how MWSS captures aspect (iii), consider *The Cabinet of Dr. Caligari*, where possibilism concerning performable options fails to capture (iii). Unlike the latter view, MWSS implies that Ace ought now to refrain from going to the cabinet. For it implies that he ought to satisfy the following conditional (if whether Ace now refrains from going to the cabinet does not depend on his future intentions, and if all his maximally preferable, *directly securable* options involve so refraining, then he so refrains). And since, given the description of the case, he cannot fail to satisfy the antecedent of this conditional, MWSS implies that he ought to *refrain from going to the cabinet*. The same does not hold, however, if we substitute “going to the cabinet” for “refraining from going to the cabinet”—for again Ace will permissibly satisfy the

⁷ Here, for the sake of simplicity, I am assuming that giving his granddaughter water is a basic action that Ace could do right away. We can get the same result if we understand it as a complex action involving a sequence of steps: we would simply need to consider each component of the action in turn, and then show how MWSS implies an obligation to perform each step.

corresponding conditional by failing to satisfy the second conjunct of its antecedent: it is not the case that all his maximally preferable, *directly securable* options going to the cabinet. And so, MWSS does not imply that Ace ought to go to the cabinet. Thus, because of its *securitist* character, MWSS captures aspect (iii) of the core idea.

Finally, to see how MWSS captures aspect (iv), consider *General Ripper*, where simple securitism fails to capture (iv). Unlike simple securitism, MWSS implies that, at noon, Ripper is under an obligation to press both buttons. It implies that he is obligated to press the first button because it implies that he is obligated to satisfy the following conditional: (if, at 12:01, whether Ripper presses the first button does not depend on the intentions he has after 12:01, and if every maximally preferable option that is directly securable for Ripper at 12:01 involves pressing the first button, then he presses the first button). Now whether Ripper satisfies the antecedent of this conditional will depend on his background attitudes. For, on the one hand, if he prefers the destruction of the world, then pressing *both* buttons will not be directly securable for him, and so it will not be the case that all his maximally preferable, directly securable options involve pressing the *first* button. Hence, he will not satisfy the antecedent of the conditional. But, on the other hand, if he does *not* prefer the destruction of the world, then pressing both buttons *will* be directly securable for him, and so he will satisfy the antecedent of the conditional. Consequently, there are two ways in which Ripper could satisfy the entire conditional. First, he could prefer the destruction of the world, and hence fail to satisfy its antecedent. Or he could *not* prefer the destruction of the world, and press the first button, thereby satisfying both the antecedent and the consequent. But, of these two ways of satisfying the conditional, only the second is permissible, since (we may assume) it is impermissible for Ripper to prefer the destruction of the world. Thus, since MWSS implies that Ripper ought to satisfy this conditional, and since the only way he could permissibly satisfy this conditional is by pressing the first button, MWSS implies that Ripper ought to press the first button.

MWSS also implies that Ripper ought to press the second button. For it implies that he ought to satisfy the following conditional: (if, at 12:02, whether Ripper presses the second button does not depend on the intentions he has after 12:02, and if every maximally preferable option that is directly securable for Ripper at 12:02 involves pressing the second button, then he presses the second button). Now, so long as Ripper presses the first button at 12:01, he will satisfy the antecedent of this conditional. And we have just seen that, according to MWSS, Ripper ought to press the first button, and so he ought to satisfy the antecedent of this conditional. And so it follows that the only way he can permissibly satisfy the conditional is by satisfying the consequent, and hence by pressing the second button. Thus, unlike simple securitism, MWSS implies that Ripper ought to press both buttons, and hence that he ought to save the world.

Recall that *scrupulous* securitism likewise implies that Ripper ought to press both buttons. However, it does so at a cost. For, as we saw in section 2.3, the very features of scrupulous securitism that enable it to get the right result in *General Ripper* give rise to problematic implications in *Ice Cream of Doom*, such as the implication that Vanessa ought to begin uttering “chocolate is my preferred flavor.” By contrast, MWSS avoids this implication. What MWSS implies is that Vanessa ought to satisfy the following conditional: (if whether Vanessa begins to utter “chocolate is my preferred flavor” does not depend on her future intentions, and if all her maximally preferable, scrupulously securable options involve beginning to utter “chocolate is my preferred flavor,” then she begins to utter “chocolate is my preferred flavor”). However, since Vanessa permissibly prefers vanilla, she permissibly satisfies the conditional by failing to satisfy its antecedent. Hence, MWSS does not imply that she is under any obligation to satisfy the consequent.

In addition to avoiding the implication that Vanessa *ought* to begin uttering “chocolate is my preferred flavor,” MWSS implies that it would be *impermissible* for Vanessa to begin uttering this sentence while simultaneously preferring vanilla. For in doing so she would fail to satisfy the conditional (if whether Vanessa begins to utter “vanilla is my preferred flavor” does

not depend on her future intentions, and if all her maximally preferable, scrupulously securable options involve beginning to utter “vanilla is my preferred flavor,” then she begins to utter “vanilla is my preferred flavor”). For in preferring vanilla, and beginning to utter “chocolate is my preferred flavor,” she would satisfy the antecedent of this conditional but not its consequent.

Note that it is the *wide scope* character of MWSS that enables it to get the right results in *Ice Cream of Doom*. For, in this case, Vanessa could permissibly (prefer chocolate ice cream and begin uttering “chocolate is my preferred flavor”), just as she could permissibly (prefer vanilla ice cream and begin uttering “vanilla is my preferred flavor”). Thus, the reason why it is impermissible for her to (prefer vanilla ice cream and begin uttering “chocolate is my preferred flavor”) is not that this would involve an impermissible action or an impermissible attitude, but rather because it would involve acting in a way that it is inappropriately related to one’s attitudes. Consequently, the correct principle must concern not simply what actions we ought to perform, but rather how our actions ought to be related to our attitudes. The correct principle, therefore, cannot be one in which the ‘ought’ takes narrow scope within a conditional—it cannot, e.g., state that *if* you have such and such attitudes, *then* you *ought* to act in such and such a way. Rather, it must be one in which the ‘ought’ takes wide scope over any such conditional—it might, e.g., state that you *ought* to be such that (*if* you have such and such attitudes *then* you act in such and such a way). And MWSS has just this kind of form, as it states that one is obligated to satisfy certain conditionals, where the satisfaction of the antecedents of these conditionals depends on the agent’s attitudes.

I conclude, therefore, that if we want a theory that captures all four aspects of the Core Idea, then we need a theory that (i) quantifies over all future times, (ii) requires maximization, (iii) concerns securable rather than performable options, and (iv) is wide scope in form. In short, we will need a theory along the lines I have proposed.

References

- Broome, John. (1999). "Normative Requirements." *Ratio* 12: 398-419.
- Broome, John. (Unpublished). *Rationality Through Reasoning*. Unpublished manuscript.
- Cariani, Fabrizio. (forthcoming). "'Ought' and Resolution Semantics," possibly forthcoming in *Nous*.
- Feldman, Fred. (1986). *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht: Riedel.
- Goldman, Holly. (1976). "Dated Rightness and Moral Imperfection." *Philosophical Review* 85(4): 449-87.
- Goldman, Holly. (1978). "Doing the Best One Can." In A. I. Goldman and J. Kim (editors) *Values and Morals*, pp. 185-214. Dordrecht: Reidel.
- Jackson, Frank and Robert Pargetter. (1986). "Oughts, Options and Actualism." *The Philosophical Review* 95: 233-255.
- Portmore, Douglas. (Forthcoming). *Commonsense Consequentialism: Wherein Morality Meets Rationality*, forthcoming with OUP.
- Ross, Jacob (2010) "The Irreducibility of Personal Obligation" *Journal of Philosophical Logic* 39: 307-323.
- Ross, Jacob (Unpublished) "Rationality, Normativity, and Commitment" possibly forthcoming in *Oxford Studies in Metaethics*, volume 7.
- Sobel, J. Howard. (1976). "Utilitarianism and Past and Future Mistakes." *Nous* 10(2): 195-219.
- Snedegar (forthcoming) "Contrastive Semantics for Deontic Modals." Possibly forthcoming in *Contrastivism in Philosophy*, edited by Martijn Blaauw, Routledge.
- Wedgwood, Ralph (2009) "Against Actualism" *PEA Soup*, posted on September 11, 2009. Accessed on March 28th, 2011 from <http://peasoup.typepad.com/peasoup/2009/09/against-actualism.html>
- Zimmerman, Michel. (1996) *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.