[Forthcoming in *Pacific Philosophical Quarterly*]

Subjective Rightness and Minimizing Expected Objective Wrongness

Kristian Olsen

Abstract: It has become increasingly common for philosophers to distinguish between objective and subjective rightness, and there has been lots of discussion recently about what an adequate theory of subjective rightness looks like. In this paper, I propose a new theory of subjective rightness. According to it, an action is subjectively right if and only if it minimizes expected objective wrongness. I explain this theory in detail and argue that it avoids many of the problems that other theories of subjective rightness face. I end by responding to some objections.

1. Introduction

Suppose a child has a fever. The child's mother is trying to decide whether to give him Advil or Tylenol. After carefully reading the instructions on each box, the mother chooses, somewhat arbitrarily, to give her child the Tylenol. However, unbeknownst to the mother, the Tylenol is defective. Tragically, the child has a severe allergic reaction to the Tylenol and dies. Had the mother given her child the Advil instead, he would have been fine. Question: has the mother acted wrongly? (That is, has the mother acted *morally* wrongly? I will always have moral wrongness [rightness, obligatoriness] in mind in this paper.)

One's intuitions may pull one in different directions in cases like this. On the one hand, the mother has killed her child, albeit accidentally. Upon realizing that the medicine she administered to her child resulted in his death, she will surely feel profound regret and may, in unbearable agony, think to herself, 'I never should have given my child that Tylenol. I should have given him Advil instead.' This thought may seem true, even if the 'should' here has a moral sense. It may therefore seem that the mother has indeed done something wrong. On the other hand, there was no way the mother could have predicted the tragedy that her action would cause. Upon seeing the mother after the tragedy, one might be inclined to say something like this to her: 'What happened was terrible, and I can understand why you are so distraught, but you should know that given your beliefs about your situation, you did nothing wrong.' This may also seem true, even if the 'wrong' here has a moral sense. It may therefore seem that the mother has not in fact acted wrongly.

A natural way of accommodating these intuitive data is to posit the existence of two different senses of the term 'wrong'—an objective sense and a subjective sense. An action is wrong in the objective sense when, roughly, it is wrong in light

of all the 'facts' of the agent's situation. On the other hand, an action is wrong in the subjective sense when, roughly, it is wrong in light of the agent's mental state. If we draw a distinction between these two senses of 'wrong,' then we can say that there *is* indeed a sense of 'wrong' according to which the mother has acted wrongly—namely, the objective sense; however, there is *also* a sense of 'wrong' according to which she has not acted wrongly—namely, the subjective sense.

While drawing a distinction between the objective and the subjective senses of 'wrong' is a common way of responding to the case at hand, it is not the only way. Some philosophers refuse to posit the existence of different senses of 'wrong' (and the related terms, 'right' and 'obligatory'). Some—the objectivists—insist that there is only one sense of 'wrong' ('right,' 'obligatory'), and it is objective.[1] Others—the subjectivists—insist that there is only one sense of 'wrong' ('right,' 'obligatory'), and it is subjective.[2] Still others suggest that there is only one sense of 'wrong' ('right,' 'obligatory'), but it is assessment-relative.[3]

I will not enter this debate here. I will simply assume that the 'sense-splitters,' as we might call them, are correct.[4] My concern in this paper will be with how we should understand the subjective sense of 'right' ('wrong,' 'obligatory') given that we also admit the existence of the objective sense of 'right' ('wrong,' 'obligatory'). Since a great many (perhaps the majority of) philosophers thinking about these issues will grant this assumption, what I say in this paper should be of interest to many.

In the next section, I discuss some initially attractive theories of subjective rightness[5] and argue that they are problematic. My discussion of these theories leads naturally to what I think is a much more plausible one: according to it, an action is subjectively right if and only if (iff) it minimizes expected objective

---

[1] Objectivists include Moore ([1912] 2005), Ross (1930), Thomson (1986), Zimmerman (1996), and Graham (2010).

[2] Subjectivists include Prichard ([1932] 2002), Ross (1939), Howard-Snyder (2005), and Zimmerman (2008).

[3] See Kolodny and MacFarlane (2010).

[4] Sense-splitters include Brandt (1959, 360-67), Russell (1966), Parfit (2011, chap. 7), Gibbard (2005), Smith (2010), Portmore (2011), Feldman (2012), and Dorsey (2012).

[5] What I should actually say here is that I discuss some attractive theories of the subjective sense of 'right.' However, for ease of exposition, I will often use phrases like 'subjective rightness' and 'subjectively right' to mean 'the subjective sense of "right"' and 'right in the subjective sense,' respectively. I will also use phrases like 'subjective wrongness,' 'subjectively obligatory,' and 'objective rightness.' I hope it's clear what I mean by these phrases.

wrongness. I discuss this theory in section 3. I argue that it circumvents the problems of the theories discussed prior to it, and it avoids the problems of its maximizing counterparts. In section 4, I discuss some important objections to my theory. Finally, in the appendix, I compare my theory to two additional recent theories of subjective rightness: Holly Smith's and Fred Feldman's. Comparing my theory to theirs is instructive because it helps to underscore some of my theory's central attractions.

2. The Path to Expected Objective Wrongness

If we are going to make a distinction between objective and subjective rightness, it seems promising to try to give a theory of subjective rightness in terms of objective rightness. Consider, for instance, the following theories:

SR1: An action is subjectively right iff the agent believes it is objectively right.[6]

SR2: An action is subjectively right iff it would be objectively right if the world were as the agent believed it to be.[7]

Each of these theories appears to make the correct conclusion about the feverish child case. As I said, it intuitively seems that there is a sense of 'wrong'—the subjective sense—according to which the mother does not act wrongly (i.e., she acts rightly) when she gives her child the Tylenol. And SR1 and SR2 each implies this, at least after we fill in the details of the case in plausible ways. For instance, it's plausible to assume that the mother believes that she is acting rightly when she gives her child the Tylenol. If this is a belief about the objective rightness of her action, then SR1 implies that her action is subjectively right. Likewise, it's plausible to hold that the mother believes that giving her child Tylenol is safe. Presumably, giving the Tylenol would have been objectively right if this belief were true. SR2 therefore seems to imply that giving the Tylenol is subjectively right as well.

However, SR1 and SR2 are implausible. While these theories have numerous problems, they share a common one: they fail to make the intuitively correct conclusion about Frank Jackson's (1991) Dr. Jill case:

Dr. Jill
Dr. Jill has a patient, John, who is suffering from a minor skin ailment. Jill can give him one of three drugs: A, B, or C. She is certain that drug A will cure him partially. She is also sure that one of B or C will cure him

[6] A version of this theory is suggested by Ross (1939, 146, 161, 190-91), Prichard ([1932] 2002), and Russell (1966, 35-36).

[7] A version of this theory is suggested by Ross (1939, 161), Brandt (1959, 365), and Graham (2014, 395).

> completely and the other will kill him, but she doesn't know which is which. Jill is a committed objective consequentialist; so, she believes that she ought, objectively, to do the best that she can. In the present case, she knows that giving John drug A will *not* have the best consequences (she knows this because she knows that giving him one of B or C will have better consequences). Jill therefore thinks that giving A is objectively wrong.[8]

What should Jill do? Presumably, Jill ought, objectively, to give whichever of B or C is the perfect cure, but she ought, subjectively, to give drug A. However, neither SR1 nor SR2 is consistent with this. After all, Jill does not believe that giving drug A is objectively right; in fact, she believes it is objectively *wrong*. SR1 therefore implies that giving A is subjectively wrong as well. Moreover, Jill may not have any false beliefs. She believes—truly—that giving A will be an imperfect cure, and she believes—also truly—that one of B or C will be the perfect cure and the other will be the killer. Of course, Jill *lacks* certain information: she doesn't know which of B or C is the perfect cure and which is the killer. But, she may not have any false beliefs; the world might be exactly as she thinks it is. And if that's the case, then since giving A is objectively wrong in the world Jill is actually in, SR2 implies that it is also subjectively wrong.[9]

At this point, a natural move to make is to abandon SR1 and SR2 and replace them with a theory of subjective rightness that makes use of expected value. This is Jackson's strategy. The expected value of an alternative is the sum, for each of its possible outcomes, of the outcome's actual value times the probability that it (the outcome) will obtain if the alternative is performed. More precisely, where A is an alternative, O_1, O_2, …, O_n are the possible outcomes of A, $V(O_i)$ is the value of outcome O_i, and $prob(O_i/A)$ is the probability of O_i conditional on A, the expected value of A = $\{[prob(O_1/A) \times V(O_1)] + [prob(O_2/A) \times V(O_2)] + \ldots + [prob(O_n/A) \times V(O_n)]\}$.

There are different ways of understanding the probability involved in expected value: we might understand it as being subjective, epistemic, or objective. Subjective probability has to do with degrees of belief (or credences). So, the subjective probability that, say, I will win the lottery is the degree to which I believe (or my credence) that I will win. Epistemic probability has to do with the degree to which one's evidence supports something, or the degree of belief that one is justified in having. So, the epistemic probability that I will win the lottery is the degree to which I am justified in believing that I will win. Objective probability has to do with chances that exist as objective features of the world—chances that

---

[8] Regan (1980, 264-65) and Parfit (2011, 159-60) give similar cases.

[9] Some additional problems for SR1 and SR2 are discussed by Zimmerman (2008, chap. 1) and Smith (2010).

exist 'in nature,' as it were. So, the objective probability that I will win the lottery is the chance that I will win, regardless of what I think that chance is or what my evidence suggests it is.

I will assume that the type of probability involved in expected value is epistemic probability. In general, it seems to me that a person's subjective obligations are determined by her evidence, not her actual beliefs. However, not much turns on this. If one prefers to use subjective (or objective) probability to calculate expected value, most of what I will say should still be of interest: one can, without loss, simply substitute talk of epistemic probability in what follows with talk of subjective (or objective) probability.

Let me now say a bit more about epistemic probability. I shall assume that for any proposition and any person, we can assign a precise degree—between 0 and 1—to which the person is justified in believing the proposition. If a person is fully justified in believing that a proposition is true, then the epistemic probability of the proposition for the person is 1. On the other hand, if a person is fully justified in believing that a proposition is false, then the epistemic probability of the proposition for the person is 0. If a person is just as justified in believing that a proposition is true as she is justified in believing that it is false, then the epistemic probability of the proposition for the person is .5. Of course, determining the precise conditions in virtue of which a person is justified in believing a proposition is an important philosophical task. Unfortunately, it's one that I will not be able to address here. I do want to note, however, that *being justified in believing* a proposition is crucially different from *justifiably believing* the proposition. The latter implies that one in fact believes the proposition in question. The former does not. If one is justified in believing a proposition, this merely implies, roughly, that *if* one *were* to believe the proposition, one *would* believe it justifiably. It does not, however, imply that one believes it already.

Now consider the following theory:

SR3: An action is subjectively right iff it maximizes expected value,

where

> An action maximizes expected value $=_{df.}$ there is no alternative to it that has a higher expected value.

One nice thing about SR3 is that it makes the correct conclusion about the Dr. Jill case, at least after we make some plausible assumptions about it. Recall that Jill has three alternatives: give drug A, give drug B, or give drug C. Jill knows the outcome of giving A: John will be partially cured. So, let's assume that the epistemic probability for Jill that this outcome will occur if she gives drug A is 1. On the other hand, Jill knows that the outcome of giving drug B will be either that John is perfectly cured or that he is killed; however, she doesn't know which

outcome B will have (her evidence, we'll assume, is equally divided on the matter). Let's assume, then, that for each of these outcomes, the epistemic probability for Jill that it will occur if she gives drug B is .5. The same things apply, *mutatis mutandis*, for drug C. If we further assume that the value of a partial cure is 90, the value of a perfect cure is 100, and the value of death is -500, the expected values of Jill's alternatives are calculated as follows:

Alternative	Relevant possible outcomes	Epistemic probability of outcome/ alternative	Value of outcome	Expected value of alternative
Give drug A	Partial cure	1	90	$90 \times 1 = 90$
Give drug B	Perfect cure Death	.5 .5	100 -500	$(100 \times .5) + (-500 \times .5) = -200$
Give drug C	Perfect cure Death	.5 .5	100 -500	$(100 \times .5) + (-500 \times .5) = -200$

As the chart indicates, only giving drug A maximizes expected value. SR3 therefore implies that giving A is subjectively right but giving either of B or C is subjectively wrong, just as it intuitively seems.

However, while SR3 makes the intuitively correct conclusion about the Dr. Jill case, it has an important problem. Consider someone who accepts (and has good reason to accept) a non-consequentialist theory of objective rightness, say Kantianism. According to a Kantian view of objective rightness, an action is objectively right iff its maxim is in fact universalizable. Now suppose our person has two options: X and Y. He has overwhelming evidence that X (alone) has a universalizable maxim. Suppose, however, that Y (alone) maximizes expected value. If SR3 is true, then it is subjectively right for our person to do Y and subjectively wrong for him to do X. But that is counterintuitive. Given that our person is justified in believing that X (alone) has a universalizable maxim, and given that he has good reason to accept Kantianism about objective rightness, it seems subjectively right for him to do X and subjectively wrong for him to do Y. The problem, then, is that SR3 is tailored too closely to consequentialism. A theory of subjective rightness should be relativized to the theory of objective rightness that agents are justified in accepting. SR3 doesn't do this.

A possible move to make at this point would be to disjunctivize SR3 so that it includes every theory of objective rightness that a person could be justified in accepting. Such a theory might look something like this:

SR4: An action is subjectively right iff either
(a) the action maximizes expected value and the agent is justified in believing consequentialism about objective rightness, or

(b) the action maximizes expected universalizability and the agent is justified in believing Kantianism about objective rightness, or
(c) the action maximizes expected prima facie rightness and the agent is justified in believing Rossianism about objective rightness, or
(d) the action maximizes expected virtue exemplification and the agent is justified in believing virtue ethics about objective rightness, or

…and so forth, for the rest of the theories of objective rightness that a person could be justified in accepting.

Of course, more needs to be said about SR4 before it can be fully understood. In particular, more needs to be said about the concepts of expected universalizability, expected prima facie rightness, etc. These concepts are intended to closely model the concept of expected value, but it remains to be seen how exactly this modeling will look. I won't attempt to do that here, for even if it can be done, SR4 will remain an ugly monstrosity. Many will rightly be suspicious of radically disjunctive theories like it. At the very least, it would be nice if the theory could be stated in a more succinct way.

The road has now been sufficiently paved for me to offer my own theory of subjective rightness, which I will do in the next section. This road-paving has been necessary because it helps to highlight the allure of the theory I will offer. As we'll see, the theory I propose solves the problems of the theories discussed so far.

3. Minimizing Expected Objective Wrongness

It's part of commonsense morality that wrongness comes in degrees. Some actions are only slightly wrong, while others are very wrong. For instance, suppose I have three options: I can tell the truth, I can tell a relatively inconsequential lie, or I can kill someone. It would seem (unless the case is unusual) that my first option is morally right, while my latter two options are morally wrong. But it seems that my second option is only a minor wrong whereas my third option is extremely wrong—it seems much more wrong to kill someone than it is to tell an inconsequential lie.[10] This contrasts with obligatoriness. Obligation seems to be all-or-nothing. It doesn't make much sense to say that I am only slightly obligated to do something or that I am very obligated to do it, nor does it make sense to say that I am more obligated to do one thing than another.[11]

---

[10] Although, curiously, it sounds awkward to say that my third option is wrong*er* than my second. However, I don't think this threatens my contention that wrongness comes in degrees. Other terms that express degreed notions sound awkward when 'er' is attached to them. For instance, 'blameworthy' plausibly expresses a degreed notion, but we wouldn't say that one person is blameworthier than another. The same goes for 'good'—we wouldn't say that one person is gooder than another.

[11] It's an interesting question whether rightness/permissibility comes in degrees. Some might hold that there is a symmetry between rightness and wrongness: rightness comes in

If wrongness comes in degrees, as it seems to, then this opens the door for a new theory of subjective rightness—namely, one that is very much like SR3, but it appeals to the degreed notion of objective wrongness instead of the degreed notion of value. However, before I can state this theory precisely, I need to set the stage a bit.

First, I will assume that each of a person's alternatives in a situation can be assigned a precise degree of objective wrongness greater than or equal to 0. For an alternative to be objectively wrong to degree 0 is for it to be not wrong at all (i.e., objectively right). The greater the degree of objective wrongness above 0 that an alternative is assigned, the more wrong it is. So, in the situation just mentioned, telling the truth is objectively wrong to degree 0 (since it is objectively right); telling the inconsequential lie is, perhaps, objectively wrong to degree 10; and killing a person is, perhaps, objectively wrong to degree 100. Of course, it will be very difficult to actually assign precise degrees of objective wrongness to a person's alternatives in any realistic scenario. However, I will assume that this can, in principle, be done. I am also obviously assuming here that an agent's alternatives can be ranked cardinally in terms of their objective wrongness. Later on, I will discuss what my theory might look like if this assumption is denied.

Second, not only will I assume that a person's alternatives can be assigned specific degrees of objective wrongness, I will also assume that their *expected* objective wrongness can be calculated. The expected objective wrongness of an alternative is the sum, for each possible degree of objective wrongness, of that degree of wrongness times the probability that the alternative is wrong to that degree. More precisely, where A is an alternative and $W_1, W_2, \ldots, W_n$ are the possible degrees of objective wrongness of A, the expected objective wrongness of A = {[prob(A is objectively wrong to degree W_1) × W_1] + [prob(A is objectively wrong to degree W_2) × W_2] + … + [prob(A is objectively wrong to degree W_n) × W_n]}. (Again, the sort of probability I have in mind is epistemic probability.) Obviously, it will be difficult to actually calculate the expected objective wrongness of a person's alternatives in realistic situations; however, I will assume that the expected objective wrongness of a person's alternatives can always, in principle, be determined. I will discuss, later on, what my theory might look like if this assumption is rejected.

---

degrees just as wrongness does. However, I take it that the claim that one option can be more right than another is much more contentious than the claim that one option can be more wrong than another. For an interesting discussion of the possibility of degrees of rightness, see Tom Hurka's 5/4/2013 post on the PEA Soup blog (www.peasoup.typepad.com). Hurka also discusses the possibility of degrees of wrongness on the blog on 4/29/2013.

In order to better understand expected objective wrongness, it may be helpful to see how the expected objective wrongness of some alternatives are calculated in a specific case. Let's return to the case of Dr. Jill. Recall that Jill knows that giving drug A won't be best and thus, since she's a consequentialist about objective rightness, knows that it is objectively wrong. But she also knows that giving A will partially cure her patient, whereas one of her other options will kill her patient. So, Jill presumably knows that while giving A is objectively wrong, it's only mildly wrong (say, to degree 10). Let's assume, then, for purposes of illustration, that Jill's epistemic probability that giving A is objectively wrong to this degree is 1. Given this, the expected objective wrongness of giving A is $1 \times 10 = 10$.

Let's turn now to Jill's second option: giving drug B. She knows that this option will either be the best thing or the worst thing she can do in her situation. So, since she's a consequentialist about objective rightness, she presumably knows that giving B will either be objectively wrong to degree 0, or objectively wrong to some degree greater than 10 (say, 100). Jill, however, has no idea which outcome drug B will have. So, let's assume, again for purposes of illustration, that Jill's epistemic probability that giving B is objectively wrong to degree 0 is .5, and her epistemic probability that it is objectively wrong to degree 100 is also .5. Given this, the expected objective wrongness of giving B is $(.5 \times 0) + (.5 \times 100) = 50$.

What I just said about giving drug B also applies, *mutatis mutandis*, to Jill's third option: giving drug C. We'll assume, then, that the expected objective wrongness of this option is also 50.

The following chart summarizes all of this information:

Alternative	Relevant possible degrees of objective wrongness	Epistemic probability the alternative is wrong to this degree	Expected objective wrongness of alternative
Give drug A	10	1	$10 \times 1 = 10$
Give drug B	100 0	.5 .5	$(100 \times .5) +$ $(0 \times .5) = 50$
Give drug C	100 0	.5 .5	$(100 \times .5) +$ $(0 \times .5) = 50$

As I've said, it intuitively seems that it is subjectively right for Jill to give drug A and subjectively wrong for her to give either of B or C. But notice that the expected objective wrongness of giving A is *less* than the expected objective wrongness of giving either of B or C. This suggests the following theory of subjective rightness:

SR_{MEOW}: An action is subjectively right iff it *minimizes* expected objective wrongness,

where

An action minimizes expected objective wrongness $=_{df.}$ there is no alternative[12] to it that has a lower expected objective wrongness.

SR_{MEOW} is an attractive theory.[13] For one thing, as we've just seen, it makes the correct conclusion about the Dr. Jill case, at least if we fill in the details of the case as I have. Note also that SR_{MEOW} will make the correct conclusion about the case on many other plausible ways of filling in its details. Suppose, for instance, that Jill is certain that giving drug A is objectively wrong to degree 15 instead of to degree 10. Then the expected objective wrongness of this option will be $15 \times 1 = 15$, which will still minimize expected objective wrongness. Or suppose that Jill doesn't know for sure how objectively wrong giving drug A is. Suppose her evidence is divided as to whether it is objectively wrong to degree 10 or is objectively wrong to degree 20—so, let's suppose, her epistemic probability that giving drug A is objectively wrong to degree 10 is .8, and her epistemic probability that giving drug A is objectively wrong to degree 20 is .2. Then the expected objective wrongness of this option is $(10 \times .8) + (20 \times .2) = 12$, which again continues to minimize expected objective wrongness. The point is, SR_{MEOW} makes the intuitively correct conclusion about the Dr. Jill case on many ways of precisifying it. It therefore avoids the problems of SR1 and SR2.

SR_{MEOW} also avoids the problems of SR3. SR3, recall, ties subjective rightness to expected value. The problem with this, we've seen, is that it is customized too closely to those who accept consequentialism about objective

---

[12] There's an interesting question about what a person's alternatives are in this context. See Zimmerman (2008, chap. 3), Smith (2010, 97-98), Feldman (2012, 159), and Hedden (2012). Unfortunately, I can't explore this issue here.

[13] I developed SR_{MEOW} after reading Markosian (2009) and Graham (2010). In the former, Markosian argues that instead of saying that we are obligated to *maximize* prima facie rightness, a Rossian should say that we are obligated to *minimize* prima facie wrongness. In the latter, Graham argues that a morally conscientious person is, in part, concerned to avoid acting wrongly, where the strength of this concern is directly proportional to the degree of wrongness of the actions she is concerned to avoid. While neither Markosian nor Graham is attempting to give a theory of subjective rightness, Markosian got me thinking about the advantages of using minimization in moral theorizing, and Graham got me thinking about degrees of wrongness—ideas that eventually led me to SR_{MEOW}. I should also note that Krister Bykvist suggests an idea similar to SR_{MEOW} in an unpublished paper, though Bykvist is talking about moral conscientiousness and rational preference, not subjective rightness. Thanks to an anonymous reviewer for making me aware of Bykvist's work.

rightness; intuitively, a theory of subjective rightness should be relativized to the theory of objective rightness that an agent is justified in accepting. SR_{MEOW}, however, is fully relativized in this way. According to it, the subjective rightness of an action is determined by the agent's evidence about its possible degrees of objective wrongness. The theory makes no substantive assumptions about what determines objective wrongness. Rather, the theory allows agents to determine this on their own, and it makes the subjective rightness of their actions a function of whatever theory of objective rightness/wrongness they are justified in accepting.

Moreover, SR_{MEOW} does not need to resort to radical disjunctivization in order to relativize the subjective rightness of an action to the theory of objective rightness the agent is justified in accepting. SR_{MEOW} therefore avoids the problem of SR4.

Finally, SR_{MEOW} avoids the problems of its maximizing counterparts. Consider, for instance, this view:

SR_{MEOO}: An action is subjectively right iff it maximizes expected objective obligation.

The problem with this theory is that objective obligation does not plausibly come in degrees. SR_{MEOO} is thus a non-starter. However, consider this view:

SR_{MEOR}: An action is subjectively right iff it maximizes expected objective rightness.

While I am not convinced that rightness comes in degrees either, I am less confident about this than I am about obligation's not coming in degrees. However, even if rightness comes in degrees, SR_{MEOR} is problematic. In particular, it has difficulty with the Dr. Jill case. Recall that Jill knows that giving drug A (the partial cure) is objectively wrong. So, she presumably knows that it is not objectively right and thus that it has a degree of objective rightness of 0. On the other hand, Jill's evidence is equally divided as to whether giving B will kill her patient or will perfectly cure him. So, her evidence is presumably equally divided as to whether giving drug B is not right (and thus has a degree of objective rightness of 0) or is very right (and thus has a high degree of objective rightness, say 100). The same things apply, *mutatis mutandis*, for drug C. Given all of this information, we can calculate the expected objective rightness of Jill's alternatives as follows:

Alternative	Relevant possible degrees of objective rightness	Epistemic probability the alternative is right to this degree	Expected objective rightness of alternative
Give drug A	0	1	$0 \times 1 = 0$
Give drug B	100 0	.5 .5	$(100 \times .5) +$ $(0 \times .5) = 50$
Give drug C	100 0	.5 .5	$(100 \times .5) +$ $(0 \times .5) = 50$

SR_{MEOR} implies that it's subjectively wrong for Jill to give drug A, and it's subjectively right for her to give either of B or C. But that's the incorrect result.

In response, a defender of SR_{MEOR} might say that I have miscalculated the expected objective rightness of Jill's alternatives. In particular, one might contend that giving drug A has a high degree of objective rightness—say, a degree of 90. Then the math works out in SR_{MEOR}'s favor: giving A maximizes expected objective rightness. However, at this point I lose my (already weak) grip on the notion of degrees of rightness entirely. How can an action have a high degree of objective rightness when it is *not* right? Shouldn't all actions that are objectively *wrong* have a degree of objective *rightness* of 0? Note that a proponent of SR_{MEOW} needn't say seemingly strange things about degrees of wrongness to handle the Dr. Jill case. In particular, a proponent of SR_{MEOW} can hold, as it intuitively seems, that all actions that are objectively *right* have a degree of objective wrongness of 0.

In order to avoid these problems, a maximizer might try to appeal to a right-like notion instead of rightness itself. In particular, she might appeal to moral value, or what Zimmerman (1996) calls 'deontic value.' Consider, then, this view:

SR_{MEDV}: An action is subjectively right iff it maximizes expected deontic value.

The problem with this view, however, is that it's not clear what deontic value is supposed to be. It is clearly not the same type of value that consequentialists traditionally appeal to when formulating their theory. It also can't merely be a placeholder for whatever feature in fact makes actions objectively right. If it were, then SR_{MEDV} would have the same problem as SR3—it would be tailored too closely to a particular objective theory. I suspect, then, that deontic value is an altogether new concept, and a rather mysterious one at that. Perhaps a defender of SR_{MEDV} can satisfactorily explain what this concept is supposed to be, but my point is that we don't need to appeal to a new, unfamiliar concept in order to give an adequate theory of subjective rightness. Once the intelligibility of objective

rightness/wrongness is granted, we already have all the resources we need to construct a plausible theory of subjective rightness.

I would object to many other theories of subjective rightness in a similar way. Many theories of subjective rightness end up appealing, in one way or another, to a vague normative notion. As a result, many theories of subjective rightness are guilty of explaining the obscure with obscurity. In the appendix, I discuss two additional theories of subjective rightness (Holly Smith's and Fred Feldman's) and argue that they suffer from precisely this problem. However, SR_{MEOW} does not have this difficulty, at least if the notion of objective rightness/wrongness is not an obscure one, and any theorist of subjective rightness will surely grant that it isn't.

In sum, SR_{MEOW} solves the problems of its predecessors and is superior to its maximizing counterparts. For this reason, it is an appealing theory, one that deserves to be taken seriously. However, like any theory, it has its own problems to confront. I will now discuss some of them.

4. Problems

When formulating SR_{MEOW}, I made various assumptions. I assumed that agents have good reason to think that there are degrees of objective wrongness. I assumed that these degrees of wrongness can be cardinally ranked. And I assumed that for each particular degree of objective wrongness that an action might have, there is a precise degree to which the agent is justified in believing that the action is wrong to that particular degree. However, each of these assumptions can be challenged. In this section, I discuss what implications the denial of these assumptions has for SR_{MEOW}.

4.1. Problem 1: Agents Who Reject Degrees of Wrongness

First, suppose an agent denies, and has good reason for denying, that objective wrongness comes in degrees. Such a person holds that objective wrongness is an all-or-nothing affair. Either an action is objectively wrong or it is not; it doesn't make sense to say that one wrong action is more wrong than another. Instead, all objectively wrong actions are equally objectively wrong.

Does this type of agent pose a problem for SR_{MEOW}?

I don't think so. The agent in question does think, and has good reason for thinking, that actions can be either wrong or not wrong. Thus the agent in question does think, and has good reason for thinking, that wrongness comes in at least *two* degrees: one degree for actions that are not wrong (i.e., 0), and one degree for actions that are wrong (we can arbitrarily assign this any positive number—say, 1). And then SR_{MEOW} says that it is subjectively right for the agent to minimize expected objective wrongness, where wrongness comes in only these two degrees. And that, it seems to me, is the correct thing to say about such a case.

To illustrate, suppose the agent we have been discussing is a hardcore deontologist who believes, and has good reason to believe, that it's objectively right to keep a promise, it's objectively wrong to kill a person, and it's just as wrong to kill one as it is to kill five. Suppose now that our agent has two options: A and B. She knows that if she does A, she will either keep her promise or kill one, but her evidence is equally divided as to which will happen. On the other hand, she knows that if she does B, she will either keep her promise or kill five, and her evidence suggests that there is a slightly better chance that B will result in her keeping the promise than her killing the five (suppose the probability is 51% for the former and 49% for the latter).

With this information, we can calculate the expected objective wrongness of the agent's alternatives as follows:

Alternative	Relevant possible degrees of objective wrongness	Epistemic probability the alternative is wrong to this degree	Expected objective wrongness of alternative
A	0 1	.5 .5	(0 × .5) + (1 × .5) = .5
B	0 1	.51 .49	(0 × .51) + (1 × .49) = .49

Note that I have said that the relevant possible degrees of objective wrongness are 0 and 1, but since the agent holds that all wrong actions are equally wrong, I could have just as easily said that the degrees are 0 and 10, or 0 and 20, or 0 and any other positive number. The result would be the same: SR_{MEOW} implies that it's subjectively right for the agent to do B and subjectively wrong for her to do A. This seems to me to be the correct conclusion. Given the agent's strange (though, we are imagining, *justified*) deontological beliefs, she subjectively ought to take the better chance of keeping her promise even though doing so could kill more. If that sounds odd, one should fault the agent's repugnant deontological beliefs rather than SR_{MEOW}.

4.2. Problem 2: Agents with Merely Ordinal Wrongness Rankings

Next, consider an agent who holds, and has good reason to hold, that there are many degrees of objective wrongness, but she holds that these degrees cannot be cardinally ranked. So, suppose an agent holds that it's wrong to kill one and it's wrong to kill five, and, furthermore, it's *more* wrong to kill five, but we cannot say just *how much* more wrong it is. On this agent's view, only an ordinal ranking of actions in terms of their wrongness is possible.

Does this type of agent pose a problem for SR_{MEOW}?

I don't think so, though I do concede that for such an agent, the expected objective wrongness of her actions must be calculated in a somewhat different way.[14] As we've seen, the expected objective wrongness of an action is the sum, for each possible degree of objective wrongness, of that degree of wrongness times the probability that the action is wrong to that degree. This, of course, presumes that we can assign numbers to the degrees of objective wrongness of actions. These numbers will correspond to the cardinal ranking of these degrees for agents who have good reason for thinking that this type of ranking can be made. However, for agents who have good reason for thinking that only ordinal rankings of the degrees of wrongness can be made, these numbers will have to correspond to the ordinal ranking of these degrees. In particular, for such agents, the expected objective wrongness of an action is the sum, for each ordinal wrongness ranking of which the action could be a part, of the action's position in that ranking times the agent's probability that that ordinal ranking obtains. More precisely, where A is an alternative, R_1, R_2, …, R_n are the possible ordinal wrongness rankings of which A is a part, $P_A(R_i)$ is the position A has on ranking R_i, and $prob(R_i)$ is the probability that ranking R_i obtains, the expected objective wrongness of A = $\{[prob(R_1) \times P_A(R_1)] + [prob(R_2) \times P_A(R_2)] + \ldots + [prob(R_n) \times P_A(R_n)]\}$. (Once again, I have epistemic probability in mind.)

To illustrate, consider an agent who has good reason for believing that actions can only be ranked ordinally in terms of their objective wrongness, and suppose she has three options: A, B, and C. The agent believes, and has good reason to believe, that C is the most objectively wrong, but she is unsure whether A is more objectively wrong than B or whether B is more objectively wrong than A (her evidence is equally divided on the matter). For such an agent, there are two relevant possible ordinal wrongness rankings: $\{A <_w B <_w C\}$ and $\{B <_w A <_w C\}$, where '$A <_w B$' means that A is less objectively wrong than B. Focus now on the former ordinal wrongness ranking: $\{A <_w B <_w C\}$. On this wrongness ranking, action A is the least wrong—it is therefore right, and I will thus say its 'position' on the wrongness ranking is 0. Action B, on the other hand, is the *first* runner-up for being least wrong, and I will therefore say that its position on the wrongness ranking is 1. Action C is the *second* runner-up for being least wrong, and I will therefore say that its position on the wrongness ranking is 2. With regard to the second ordinal wrongness ranking I mentioned ($\{B <_w A <_w C\}$), B's position is 0, A's position is 1, and C's position is 2.[15] The following chart illustrates how the expected objective wrongness of these alternatives is calculated:

---

[14] The remainder of this section has benefited from William MacAskill's (2014) work on how to accommodate merely ordinal value rankings in decision theory.

[15] What if there are ties? Consider, for instance, this wrongness ranking: $A <_w B =_w C$, where '$B =_w C$' means that B and C are equally wrong. What are the positions of A, B, and C on this ranking? What we should say, I think, is that A's position is 0, B's position

Alternative	Relevant possible ordinal wrongness rankings of which the alternative is a part	Epistemic probability that this wrongness ranking obtains	Alternative's position in the ordinal wrongness ranking	Expected objective wrongness of alternative
A	$A <_w B <_w C$ $B <_w A <_w C$	.5 .5	0 1	$(.5 \times 0) + (.5 \times 1) =$ .5
B	$A <_w B <_w C$ $B <_w A <_w C$	.5 .5	1 0	$(.5 \times 1) + (.5 \times 0) =$ .5
C	$A <_w B <_w C$ $B <_w A <_w C$	.5 .5	2 2	$(2 \times .5) + (2 \times .5) = 2$

SR_{MEOW} implies that it is subjectively right for the agent to do either of A or B; doing C, on the other hand, is subjectively wrong. This seems to me to be the correct conclusion in the case.

In sum, the threat of merely ordinal rankings of degrees of wrongness does not pose a deep problem for SR_{MEOW}. Proponents can leave their theory as it is, though they should admit that there are two somewhat different ways of calculating the expected objective wrongness of an alternative: one for those who have good reason for thinking that the wrongness of actions can be ranked cardinally, another for those who have good reason for thinking that the wrongness of actions can only be ranked ordinally. Of course, more will need to be done to fully defend this proposal than I have offered in these brief remarks. However, I hope it is clear that proponents of SR_{MEOW} needn't despair over the possibility of merely ordinal rankings of degrees of objective wrongness. There is a prima facie plausible way of accommodating for the possibility.

4.3. Problem 3: Agents with Imprecise Justificatory States

Finally, consider again an agent who believes, and has good reason to believe, that the objective wrongness of actions comes in degrees and that actions can be ranked cardinally in terms of their degrees of objective wrongness; however, suppose that there is no sharp degree to which the agent is justified in believing that an action of hers has any particular degree of objective wrongness. Consider, for instance, the degree of wrongness of 14.59. For our agent, there is no precise degree to which she is justified in believing that an alternative of hers is objectively

is 1.5, and C's position is also 1.5. We need to say this because the sum of the positions on each possible wrongness ranking where A, B, and C are the sole members should be the same, regardless of whether there are ties. More generally, an action's position on a wrongness ranking is equal to the number of alternatives that are less wrong than it plus .5 for each member with which it is tied.

wrong to degree 14.59; at most, there is an *interval* or *span* of degrees to which she is justified in believing that an alternative is wrong to this exact degree.

Does this type of agent pose a problem for SR_{MEOW}?

It would seem so since the expected objective wrongness of this agent's alternatives looks impossible to calculate. Recall that the expected objective wrongness of an alternative is the sum, for each possible degree of objective wrongness, of this particular degree of wrongness times *the* degree to which the agent is justified in believing that the alternative is wrong to that particular degree. But for the agent we are imagining, there simply is no such thing as *the* degree to which she is justified in believing propositions about the precise degrees of objective wrongness of her alternatives. So, what should a proponent of SR_{MEOW} say about such agents?

One way of responding to this would be to say that the case is impossible: for any agent and any proposition, there is some precise degree to which the agent is justified in believing that proposition. Of course, it's very difficult, perhaps impossible, for agents to *know* the precise degree that they are justified in believing certain propositions, but, one might hold, there is nonetheless some precise degree to be known. While I am sympathetic with this response, I acknowledge that not everyone will be. Some will insist that *God Himself* wouldn't be able to determine the precise degree to which a person is justified in believing certain propositions (such as the proposition that one's alternative is objectively wrong to degree 14.59). And the reason for this is not that God has some intellectual defect or shortcoming (that would be incoherent); rather, it's because there simply is no such degree to be known. Since some will hold this view, I'd like to sketch a way of modifying SR_{MEOW} to respond to it. I hope to show that a version of the theory can survive even if it turns out that an agent's justificatory state is imprecise in the way described.[16]

First, I'd like to restate the problem in a slightly different way. Earlier, when I explained how to calculate the expected objective wrongness of an alternative, it was clear that I was presuming that for any agent, his or her justificatory state can be represented by a completely determinative function that takes as an input any proposition and outputs a sharp degree of belief that the agent is justified in having toward that proposition (call such a function a 'justificatory state function'). But the worry is that the justificatory state of some agents—perhaps every actual agent—cannot be represented in this way. At most, the justificatory state of these agents can be represented by a *set* of justificatory state functions.

---

[16] There is now a burgeoning literature on the topic of whether our credences or our justificatory states are always (or ever) sharp and, if they aren't, how this would impact Bayesian decision theory. See Joyce (2010), Elga (2010), and White (2010).

Here's a more picturesque way of thinking about this. Consider a hypothetical politician who is extremely opinionated. He has beliefs about everything. Ask him any question, and he'll give you a definite answer. So, if you ask him how justified you are in believing that one of your alternatives is objectively wrong to degree 14.59, our politician will have an answer. '.6947,' he might say. Now, the claim currently under consideration is that it is impossible to represent the justificatory state of some agents with any single opinionated politician. To represent the justificatory state of some agents, we'll need to appeal to a whole *congress* of such politicians.[17] So, then, how justified is a person in believing that one of her alternatives is objectively wrong to degree 14.59? Perhaps one of the opinionated politicians who partially represents her justificatory state will say '.6948.' Perhaps another will say '.6946.' Another might say '.6945.' And so forth. This captures the intuitive idea that there is a *range* or *set* of degrees to which a person is justified in believing that one of her alternatives is objectively wrong to some precise degree.

If an agent's justificatory state really cannot be represented by only one justificatory state function, then SR_{MEOW} will need to be modified. However, there is a natural way of doing this. We are currently supposing that if we are going to represent some agent's justificatory state with a justificatory state function, we can do so only by appealing to many such functions. But recall that each justificatory state function is, by definition, completely determinative: each takes as an input any proposition and outputs a sharp degree of belief that the agent is justified in having toward that proposition. So, we *can* calculate the expected objective wrongness of an agent's alternatives for each of the justificatory state functions that represent her justificatory state. We can, then, modify SR_{MEOW} as follows:

> **SR_{MEOW}***: An action is subjectively right iff it minimizes expected objective wrongness on each of the justificatory state functions that together represent the agent's justificatory state.

In more picturesque terms: an action is subjectively right iff it minimizes expected objective wrongness according to each of the extremely opinionated hypothetical politicians that together represent the agent's justificatory state.

Given SR_{MEOW}*, what shall we say about subjective *wrongness*? I think that a proponent of SR_{MEOW}* should accept the following:

> An action is subjectively wrong iff it fails to minimize expected objective wrongness on each of the justificatory state functions that together represent the agent's justificatory state.

[17] Joyce (2010, 288) uses a similar analogy.

If a proponent of SR_{MEOW}* accepts this, she will be committed to saying that when an alternative minimizes expected objective wrongness on some but not all of the justificatory state functions that together represent the agent's justificatory state, the alternative is neither subjectively right nor subjectively wrong. This seems to me to be exactly the right thing to say about such actions. Given the agent's justificatory state, these actions neither definitively minimize nor definitively fail to minimize expected objective wrongness, and we should thus hold that their subjective moral status is indeterminate.

5. Concluding Remarks

Of course, other problems for SR_{MEOW} remain to be solved. For instance, consider an agent who has good reason for thinking that some of her alternatives can be ranked cardinally in terms of their degrees of objective wrongness, but others can only be ranked ordinally. Or consider an agent who has good reason for thinking that some of her options can be ranked on one cardinal scale in terms of their degrees of objective wrongness, and others can be ranked on a different cardinal scale, but there is no single cardinal scale on which they can all be ranked together. How is expected objective wrongness to be calculated in these cases? I have also not addressed the topic of action-guidingness. Some claim that in order for a theory of subjective rightness to be plausible, it must be sufficiently action-guiding. But is SR_{MEOW} action-guiding, and if it isn't, is that really a problem? In addition, some hold that there is an intimate connection between subjective rightness/wrongness and praiseworthiness/blameworthiness. Does SR_{MEOW} establish the proper connection between these concepts, and what is that connection supposed to be anyway? All of these are important questions that I have not attempted to answer. I therefore recognize that I have not given a complete defense of SR_{MEOW} in this paper. However, I do hope I have demonstrated that SR_{MEOW} has significant advantages over many of its competitors and that it has the resources to address several potential problems. At the very least, I hope I have shown that SR_{MEOW} is a promising theory worthy of continued discussion.[18]

Department of Philosophy
University of New Hampshire

[18] Many people have helped me bring this paper to its current form. Thanks especially to Fred Feldman, Pete Graham, Chris Meacham, and Timm Triplett for giving me comments on various drafts of the paper. Thanks also to an anonymous referee from this journal who gave me excellent feedback that greatly improved the paper.

Works Cited

Brandt, Richard B. (1959). *Ethical Theory: The Problems of Normative and Critical Ethics*. Englewood Cliffs: Prentice-Hall.

Bykvist, Krister. (Unpublished). 'Objective versus Subjective Moral Oughts.'

Dorsey, Dale. (2012). 'Objective Morality, Subjective Morality and the Explanatory Question.' *Journal of Ethics and Social Philosophy* 6(3): 1-24.

Elga, Adam. (2010). 'Subjective Probabilities Should Be Sharp.' *Philosophers' Imprint* 10(5): 1-11.

Feldman, Fred. (2012). 'True and Useful: On the Structure of a Two Level Normative Theory.' *Utilitas* 24: 151-71.

Gibbard, Allan. (2005). 'Truth and Correct Belief.' *Philosophical Issues* 15: 338-50.

Graham, Peter. A. (2010). 'In Defense of Objectivism about Moral Obligation.' *Ethics* 121: 88-115.

———. (2014). 'A Sketch of a Theory of Moral Blameworthiness.' *Philosophy and Phenomenological Research* 88: 388-409.

Hedden, Brian. (2012). 'Options and the Subjective *Ought*.' *Philosophical Studies* 158: 343-60.

Howard-Snyder, Frances. (2005). 'It's the Thought that Counts.' *Utilitas* 17: 265-81.

Jackson, Frank. (1991). 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.' *Ethics* 101: 461-82.

Joyce, James M. (2010). 'A Defense of Imprecise Credences in Inference and Decision Making.' *Philosophical Perspectives* 24: 281-323.

Kolodny, Niko, and John MacFarlane. (2010). 'Ifs and Oughts.' *Journal of Philosophy* 107: 115-43.

MacAskill, William. (2014). *Normative Uncertainty*. Ph.D. Thesis, University of Oxford.

Markosian, Ned. (2009). 'Rossian Minimalism.' *Journal of Ethics and Social Philosophy* 4(1): 1-17.

Moore, G. E. ([1912] 2005). *Ethics*. Edited by W. H. Shaw. Oxford: Clarendon Press.

Parfit, Derek. (2011). *On What Matters*. Vol. 1. Oxford: Oxford University Press.

Portmore, Douglas W. (2011). *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.

Prichard, H. A. ([1932] 2002). 'Duty and Ignorance of Fact.' In *Moral Writings*, edited by Jim MacAdam, 85-110. Oxford: Oxford University Press.

Regan, Donald. (1980). *Utilitarianism and Co-Operation*. Oxford: Clarendon Press.

Ross, W. D. (1930). *The Right and the Good*. Oxford: Clarendon Press.

———. (1939). *The Foundations of Ethics*. Oxford: Clarendon Press.

Russell, Bertrand. (1966). 'The Elements of Ethics.' In *Philosophical Essays*, 13-59. New York: Simon and Schuster.

Smith, Holly M. (2010). 'Subjective Rightness.' *Social Philosophy and Policy* 27: 64-110.

Thomson, Judith Jarvis. (1986). 'Imposing Risks.' In *Rights, Restitution, and Risks*, edited by William Parent, 173-91. Cambridge: Harvard University Press.

White, Roger. (2010). 'Evidential Symmetry and Mushy Credence.' In *Oxford Studies in Epistemology*, edited by Tamar Szabó Gendler and John Hawthorne, 161-87. Oxford: Oxford University Press.

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

———. (2008). *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press.

Appendix

In this appendix, I discuss two additional theories of subjective rightness: Holly Smith's (2010) and Fred Feldman's (2012). Smith and Feldman agree with me about many things. They both think that we should draw a distinction between objective and subjective rightness. They both think that the Dr. Jill case is an important test case. They both agree that the move to expected value cannot adequately handle the Dr. Jill case. And they both maintain that subjective rightness should be relativized to the objective moral theory of an agent. However, while Smith and Feldman share my general approach, the theories of subjective rightness they endorse are very different from mine. Smith's and Feldman's theories are therefore two of SR_{MEOW}'s most important competitors. As I will demonstrate, however, their theories have a similar shortcoming, a shortcoming that SR_{MEOW} avoids. Comparing their theories to SR_{MEOW} will therefore make the attractions of SR_{MEOW} all the more apparent.

Let's begin with Smith's theory. Smith thinks that in order to give a theory of subjective rightness, we first need to say what it is for a moral principle to be objective and what it is for a moral principle to be subjective. She gives a precise account of what it is for a moral principle to be objective (2010, 93), but the details don't matter much for my purposes. It will suffice for me to note that principles like 'an action is objectively right iff it maximizes utility' and 'an action is objectively right iff its maxim is universalizable' count as objective moral principles on Smith's account. According to Smith, a subjective moral principle is always relativized to an objective one. She seems to think that for any objective moral principle, there

is a set of subjective moral principles that 'correspond' to it. Each of the principles in this set is, Smith suggests, 'appropriate' to the objective one (94-99, 102, 106). Smith also gives a detailed account of what it is for a moral principle to be subjective (93-94), but for my purposes, there is again no need to discuss the details. It will suffice for me to say that, relative to the utilitarian principle of objective rightness just mentioned, the following principles of subjective rightness are plausibly appropriate to it: an action is subjectively right if it maximizes utility; an action is subjectively right if it maximizes expected utility; and an action is subjectively right if it maximizes minimum utility.

Smith notes that for most sets of subjective moral principles, some of the principles in it will be action-guiding. She also holds that for any set of subjective principles, there is a 'hierarchy' of the principles within it (2010, 92, 94, 99, 102). As we've seen, Smith holds that for each objective moral principle, there is a corresponding set of subjective principles that is appropriate to it. And Smith's idea seems to be that some of the principles in this set are 'higher up,' 'more important,' or 'more appropriate' to the objective principle than others are. Consider, for example, a utilitarian theory of objective rightness. As I've said, the set of subjective principles appropriate to it plausibly includes principles like (1) an action is subjectively right if it maximizes utility, (2) an action is subjectively right if it maximizes expected utility, and (3) an action is subjectively right if it maximizes minimum utility. (1) is plausibly at the top of the hierarchy of principles in this set—it is the 'most important' one, or the 'most appropriate' to objective utilitarianism. And (2) is plausibly 'higher up' than (3).

With that, I'm now in a position to state Smith's theory of subjective rightness:

> **SR_{HS}**: An action is subjectively right for a person to perform (relative to her theory of objective rightness T) iff it is prescribed by the highest action-guiding subjective moral principle in the hierarchy of moral principles that is appropriate for T. (2010, 100)

I'd like to raise two problems for Smith's theory. First, Smith suggests that for each principle of objective rightness, there is a set of subjective principles that is 'appropriate' to it. But the concept of appropriateness at issue here is mysterious. Clearly the principle that an action is subjectively right if it maximizes utility is appropriate to a principle of objective utilitarianism. So is the principle that an action is subjectively right if it maximizes expected utility. But the set of subjective principles appropriate to objective utilitarianism (or any other objective moral principle) needs to be enormous—otherwise Smith won't be able to guarantee (as she wants to) that for any agent in any circumstance, there is a subjective moral principle appropriate to her objective principle that is action-guiding. But it's hard to imagine a sense of 'appropriate' that generates these massive sets of subjective principles.

To further press the issue, note that Smith suggests that for any set of subjective principles, there is a 'bottom level' principle like this: an action is subjectively right if the agent believes she can do it (2010, 109n73). This is the subjective principle that agents may need to use in circumstances of extreme ignorance. But is this principle really appropriate to any objective principle? Is it appropriate to objective utilitarianism, for instance? It's hard to see what sense of 'appropriate' will allow us to answer 'yes.' So, I think one should worry about Smith's use of appropriateness in her theory. The notion is not much clearer than the notion of subjective rightness itself. Because of this, giving an account of subjective rightness in terms of appropriateness has little payoff; Smith is merely substituting one mysterious notion for another.

Even more troubling is Smith's reliance on a hierarchy of subjective principles. Recall: Smith thinks that for any set of subjective principles, some are 'higher up,' 'more important,' or 'more appropriate' to the associated objective principle than others. But what exactly makes one subjective principle 'higher up' than another? This is a difficult question, and one that Smith does not attempt to answer; she sets it aside for another occasion (2010, 92-93). This isn't necessarily bad, but it *is* an important question, and one that I think is going to be very difficult to answer. I suspect that providing an account of what makes one subjective principle 'higher up' than another in a hierarchy will be just as difficult as giving an account of subjective rightness in the first place.

Feldman's (2012) theory has similar difficulties. Feldman first focuses on giving an account of subjective rightness for those who accept utilitarianism about objective rightness. On his view, an action is subjectively right for such a person iff it is an 'outcome' of the Utilitarian Decision Procedure. We can determine whether an alternative of ours is an outcome of the Utilitarian Decision Procedure by following these steps:

1. First, we need to consider all of the actions we think we can perform (described in action-guiding ways).
2. Next, we need to consider what we take the values of these actions to be.
3. Then, we need to take into consideration our views about the morality of risk as it applies to our situation.
4. Finally, we need to identify the actions that 'seem most nearly consistent with the general policy of maximizing utility where possible while avoiding things that put people at excessive risk of serious harm.'
5. The actions identified in (4) are the 'outcome' of the Utilitarian Decision Procedure. (166-67)

Thus far, Feldman has only given an account of subjective rightness for those who accept utilitarianism about objective rightness. However, he thinks this account can be extended. Consider, for instance, a virtue ethicist who holds that an action is objectively right iff it maximizes the virtue/vice quotient (an action

maximizes the virtue/vice quotient iff the balance of virtue over vice it manifests is at least as great as that of any alternative). According to Feldman, the subjectively right actions for proponents of this view are the ones that are the outcome of the Virtue Ethics Decision Procedure, which looks like the Utilitarian Decision Procedure except that talk of 'value' is replaced with talk of 'virtue' and 'vice.' The same goes, *mutatis mutandis*, for those who accept Kantianism, Rossianism, etc. about objective rightness. Given this, we can state Feldman's theory in a more general way as follows:

> **SR_{FF}**: An action is subjectively right iff it is an outcome of the decision procedure associated with the theory of objective rightness that the agent accepts,

where

> The decision procedure associated with the theory of objective rightness that an agent accepts is a decision procedure that looks relevantly like the Utilitarian Decision Procedure above.

In my view, the primary problem for Feldman's theory has to do with the notion of consistency that appears in the Utilitarian Decision Procedure (and, presumably, in the decision procedures associated with every other theory of objective rightness). Feldman claims that for those who accept utilitarianism about objective rightness, subjectively right actions are ones that seem 'most nearly consistent' with the policy of maximizing utility while avoiding things that put people at risk of serious harm. But what exactly does Feldman mean by 'consistent' in this context? I doubt that he has in mind the ordinary philosophical notion of logical consistency. The notion of consistency that Feldman has in mind is one that comes in degrees, or, at least, is one that can be more or less approximated. However, it's not clear that logical consistency is degreed in this way. None of Dr. Jill's options is logically consistent with both the policy of maximizing utility and the policy of avoiding excessive risk of serious harm. Is drug A *more* logically consistent with these things than either drug B or drug C? It's not clear to me that it is (or that it would seem to Jill that it is). So, I suspect that Feldman has some other, non-logical notion of consistency in mind here. But if Feldman's notion of consistency is not the ordinary logical one, then what is it? Feldman never addresses this question, though I suspect he is ultimately appealing to some type of appropriateness or fittingness. Feldman's theory of subjective rightness therefore has a problem much like Smith's: it makes use of an opaque normative notion that is at least as mysterious as the notion of subjective rightness itself. And even if Feldman does have in mind a non-mysterious notion of consistency (such as logical consistency), he still needs to tell us how exactly *degrees* of this type of consistency are determined. In particular, he needs to tell us what exactly it takes for an action to be more or less consistent with the policy of maximizing utility

while avoiding excessive risk of serious harm. Until we know that, not much ground will have been gained.

Notice, however, that SR_{MEOW} avoids these problems. It makes use of normative concepts that, I am assuming, are readily understood. The central normative concept it appeals to is the concept of objective wrongness, and this concept is already a fixture in the moral landscape. It therefore seems to me to be preferable to give a theory of subjective rightness in terms of objective wrongness instead of in terms of Smith's appropriateness or Feldman's consistency, or some other normative concept that is difficult to grasp. And even if one comes to doubt the intelligibility of objective rightness/wrongness, note that Smith's and Feldman's theories make use of it too (in addition to the other, vaguer concepts I've highlighted). So, attacking the coherence of objective rightness/wrongness will undermine their theories as well as mine. The advantage of the theory I've offered is that it doesn't employ a normative concept in addition to objective rightness/wrongness. My point is that once the intelligibility of objective rightness/wrongness is granted, we have all the resources we need to construct a plausible theory of subjective rightness. For these reasons, the theory I have offered is preferable to Smith's and Feldman's.