

OBJECTIVE VS. SUBJECTIVE CONSEQUENTIALISM AND ACTUAL VS. EXPECTED UTILITY

I. Why 'Ought' and 'Wrong' Are Ambiguous

Consider the following:

- Someone might ask: "What ought I to do when I do not know what I ought to do?" (Zimmerman 2006, 359)
- I might figure out that, given my epistemic position, I ought to do x , and yet I can still wonder whether that's what I ought to do x . Thus I might ask someone in a superior epistemic position whether I ought to do x , and I am clearly not asking her whether, given my current epistemic position, I ought to do x , for I already know the answer to that question.
- Ewing says: "We may believe...that the soldiers who fight against us in a war are acting wrongly in fighting, yet every reasonable person will admit that, as long as they really think they ought to fight, they ought 'to obey their consciences' and fight" (as quoted in Zimmerman 2006, 335). Ewing seems to be claiming, as many others have claimed, that you ought to do what you think is right. In other words, you ought to do what it is that you believe that you ought to do.

Of course, none of the foregoing makes any sense if there is only one sense of ought. Thus we should assume that 'ought' and its correlatives (such as 'wrong') are ambiguous and have different senses.

II. The Different Senses of 'Ought' and 'Wrong'

Below, I'll attempt to disambiguate the word 'wrong', but, corresponding to these different senses of 'wrong', there will be different senses for its various correlatives: 'ought', 'obligatory', and 'permissible'. To get these different senses of the correlatives of 'wrong', we need only to keep in mind the following relations. An act is morally permissible if and only if it is not wrong. And an act is morally obligatory (or, in other words, what the agent morally ought to do) if and only if all of its alternatives are wrong.¹

First, let's assume that there is a sense of 'wrong' for which moral theories are concerned to give criteria. Call this: "'wrong' in the fact-relative sense." We can use this sense of 'wrong' to define two other senses of 'wrong'.

¹ The following is adapted from Parfit's *Climbing the Mountain*.

FRS S's performing x is wrong in the *fact-relative sense* if and only if the correct criterion of rightness along with all the other relevant facts entail that S's performing x is wrong.

MBRS S's performing x is wrong in the *moral-belief-relative sense* if and only if S believes that her performing x is wrong in the fact-relative sense.

ERS1 S's performing x is wrong in the *evidence-relative sense* if and only if, given the evidence available to S, S ought to believe that her performing x is wrong in the fact-relative sense.

The first two bullets in Section I seem to be trading on an ambiguity between ought in the fact-relative and evidence-relative senses. Thus, I might ask: "What I ought to do in the evidence-relative sense when I do not know what I ought to do in the fact-relative sense?" And, as the second bullet suggests, I may know what I ought to do in the evidence-relative sense, but still wonder whether this is what I ought to do in the fact-relative sense.

The third bullet in Section I, by contrast, seems to be trading on an ambiguity between ought in the fact-relative and moral-belief-relative senses. The thought seems to be that you ought, in the moral-belief-relative sense, to do what you believe that you ought to do in the fact-relative sense.

III. The Evidence-Relative Sense

Of these three senses, the evidence-relative sense is the trickiest.

A. Why ERS1 Is Unsatisfactory

ERS1 is unsatisfactory, as Parfit's *Mine Shafts* example shows:

Consider

Mine Shafts: A hundred miners are trapped underground, with flood waters rising. We are rescuers on the surface who are trying to save these men. We know that all of these men are in one of two mine shafts, but we don't know which. There are three flood-gates that we could close by remote control. The results would be these:

The miners are in

		Shaft A	Shaft B
We close	Gate 1	We save 100 lives	We save no lives
	Gate 2	We save no lives	We save 100 lives
	Gate 3	We save 90 lives	We save 90 lives

Suppose next that on the evidence available, and as we justifiably believe, it is equally likely that the miners are all in Shaft A or all in Shaft B. If we closed either Gate 1 or Gate 2, we would have a one in two chance of doing what would be right in the fact-relative sense, because our act would save all of these hundred men. If we closed Gate 3, we would have no chance of doing what would be in this sense right. But this is clearly what we ought to do, since by closing Gate 3 we shall be certain to save ninety of these men. (Parfit, *Climbing the Mountain*)

On ERS1, it would be wrong for us close Gate 3, because, given the available evidence, we ought to believe that closing Gate 3 is wrong in the fact-relative sense. Yet this is clearly what we ought to do given the available evidence. Although the only way that we have any chance of doing what's right in the fact-relative sense is to open either Gate 1 or Gate 2, we ought (in the evidence-relative sense) not do so, because we don't have any idea which of these two is the right one, and, given the great risk and high stakes involved, we should play it safe and open Gate 3, ensuring that we save most of the men.

B. A Better Account

1. The Account

ERS S's performing x is wrong in the *evidence-relative sense* if and only if, given the evidence available to S, S ought to believe that x 's expectable deontic value is lower than that of some available alternative.

2. Clarifications

a. Ought to Believe

The 'ought' in 'S ought to believe' is not the moral ought, but the epistemic ought. Thus there is no circularity problem.

b. Deontic Value

The deontic value of an act is a measure of how morally good or morally bad it is. The deontic value of an act is positive if and only if it

is morally permissible and negative if and only if it is morally impermissible.

Some immoral acts are morally worse than others. Clearly, this is the case where one immoral act has worse consequences than another. But it may even be that some immoral acts are morally worse than others even though its consequences are no worse than those others. For instance, it might be that intentionally causing harm is morally worse than negligently causing harm even when the consequences in each case are equally bad.

Some permissible acts are morally better than others. Clearly, this is the case where one permissible act has better consequences than another. But it may even be that some permissible acts are morally better than others even though its consequences are no better. For instance, it might be that the act of sacrificing an arm to save a life (where this is necessary to save the life) is morally better than the act of sacrificing a \$100 to save a life (where this is necessary to save life), and this might be true even if the consequences of the former are worse than the consequences of the latter.

c. Expectable Deontic Value

Suppose that there are a number of possible deontic values that a given act alternative, a_i , may, for all you know, have: $DV_1(a_i)$, $DV_2(a_i)$, $DV_3(a_i)$, ..., $DV_n(a_i)$. Suppose also that for each of these possible deontic values there is a probability, given your epistemic position, that it has this deontic value: $\Pr[DV_1(a_i)]$, $\Pr[DV_2(a_i)]$, $\Pr[DV_3(a_i)]$, ..., $\Pr[DV_n(a_i)]$. Also, let's assume that for each act alternative, a_i , there is a set of possible future courses of action (FCA) that are available to you if and only if you perform this act: FCA_1 , FCA_2 , FCA_3 , ..., FCA_n . Each of these FCAs has an expected deontic value (eDV): $eDV(FCA_1)$, $eDV(FCA_2)$, $eDV(FCA_3)$, ..., $eDV(FCA_n)$. And, for each of these FCAs, there is a probability that you'll complete that FCA given your performance of a_i : $\Pr(FCA_1/a_i)$, $\Pr(FCA_2/a_i)$, $\Pr(FCA_3/a_i)$, ..., $\Pr(FCA_n/a_i)$. The expectable deontic value (EDV) of an act, a_i , is equal to: $\sum_{(j=1-n)}\{DV_j(a_i) \times \Pr[DV_j(a_i)]\} + \sum_{(j=1-n)}\{eDV(FCA_j) \times \Pr(FCA_j/a_i)\}$.

d. ERS Illustrated

Let's consider a couple of examples.

CASE 1: Assume that you promised your ex-girlfriend that you would attend her wedding, but that now you are realizing that the probability that you would complete a future course of action in which you get drunk and punch the groom in the nose were you to accept the wedding invitation is quite high relative to your epistemic situation. Let a_1 be the act of keeping your promise to attend the wedding, and let a_2 be the act of breaking your promise to attend the wedding. Assume that the expectable deontic values are as follows:

$$\begin{aligned} \text{EDV}(a_1) &= \{(10 \times 1) + [(10,000 \times .9) + (11,000 \times .1)]\} = 10,110 \\ \text{EDV}(a_2) &= [(-10 \times 1) + 10,900] = 10,890 \end{aligned}$$

“10” is the deontic value of your keeping your promise, and “-10” is the deontic value of your breaking your promise. Assume that if you keep your promise, there is, given your epistemic position, a 90% chance that you’ll undertake a future course of action that involves your getting drunk at the wedding and punching the groom in the nose and that there is only a 10% chance that you’ll undertake a future course of action that involves your behaving yourself at the wedding. The deontic values of these two FCAs are 10,000 and 11,000, respectively. Assume that sum of the expected deontic values of all the possible FCAs that you might undertake if you were to break your promise to attend the wedding is 10,900.

Also, assume that, given your epistemic position, you ought to believe that a_1 ’s expectable deontic value is lower than a_2 ’s. This is true even though you obviously don’t have anything close to the precise sort of information about deontic values and probabilities that I give above. In this case, then, you ought, in the evidence-relative sense, to do what’s wrong in the fact-relative sense: that is, break your promise (i.e., a_2). In this instance, doing what you know that you ought to do in the fact-relative sense is just too risky, for there is, given your epistemic position, a very good chance that, if you keep your promise, you’ll do something much worse than breaking a promise: you’ll assault the groom.

CASE 2: Imagine that you are faced with the situation in Parfit’s *Mine Shafts* example. Suppose, then, that there are three act alternatives available to you, a_1 - a_3 , corresponding to each of the three gates that you could close. Assume that the expectable deontic values are as follows:

$$\text{EDV}(a_1) = [(10,000 \times .5) + (0 \times .5)] + 9,500 = 14,500$$

$$\text{EDV}(a_2) = [(0 \times .5) + (10,000 \times .5)] + 9,500 = 14,500$$

$$\text{EDV}(a_2) = [(9,000 \times .5) + (9,000 \times .5)] + 10,000 = 19,000$$

Also, assume that, given your epistemic position, you ought to believe that both a_1 's expectable deontic value and a_2 's expectable deontic value are lower than a_3 's. In this case, then, you ought, in the evidence-relative sense, to do what you know to be wrong in the fact-relative sense: that is, close gate 3 (i.e., a_3). In this instance, trying to do what you ought to do in the fact-relative sense is just too risky, for there is, given your epistemic position, a very good chance that, if you close either gate 1 or gate 2 (one of which is what you ought to do in the fact-relative sense), you could end up saving none of the trapped miners.

CONCLUSION: It seems that ERS does an excellent job of accounting for the risk of wrongdoing and is, therefore, the most plausible account of 'wrong' in the evidence-relative sense.

3. Some Advantages

a. Practicality

ERS is practical. Although it will, in most instances, be impossible for one to know whether the act that one is considering performing has lower expectable deontic value than some available alternative (for the relevant information—i.e., the possible deontic values and associated probabilities of all the various act alternatives—will simply be unattainable), it is always possible for one to know whether, *given one's epistemic position*, one ought to believe that the act that one is considering performing has lower expectable deontic value than some available alternative. What one ought to believe, given one's epistemic position, is simply a function of the available evidence and, therefore, necessarily attainable. Thus, it is always possible for one to know what one ought to do in the evidence-relative sense.

Contrast ERS with Parfit's:

"Expectabilism: When the rightness of our acts depends on the goodness of their effects or possible effects, we ought [in the evidence-relative sense] to act, or at least try to act, in the ways whose outcomes would be expectably-best."

This view is either too strong or too weak, depending on whether it requires us to act, or only to try to act, in these ways. On the one hand, if it requires us to act in these ways, then it is too strong, for the evidence as to which acts would make things go expectably best will often be unattainable. On the other hand, if it requires only that we try to act in these ways, then it seems too weak, as we may try to act so as to produce the expectably best outcome while ignoring relevant evidence that we should have taken into account.

b. Neutrality

Expectabilism seems either to ignore important deontic factors or to assume that the degree to which an act can be morally good/bad is solely a function of the goodness/badness of its consequences. ERS takes all deontic factors into account and is neutral with respect to whether the degree to which an act can be morally good/bad is solely a function of the goodness/badness of its consequences.

IV. Which Sense Is Most Salient to Our Moral Deliberations and Assessments of Blameworthiness?

A. Which Sense Is Most Salient to Our Moral Deliberations?

In our moral deliberations, we should focus on ERS, for, as we saw above, there are many instances, in which avoiding doing what we know to be wrong in the fact-relative sense would lead us to act in ways that are just too risky.

B. Which Sense Is Most Salient to Our Assessments of Blameworthiness?

Acting in a way that is wrong in the fact-relative sense provides neither (1) a necessary nor (2) a sufficient condition for being blameworthy. (1) If I were to close gate 2 in the *Mine Shafts* case, and if the miners happen, unbeknownst to me, to be in shaft B, then I will have acted rightly in the fact-relative sense, but I would still be blameworthy for having taken such a big risk. (2) Were I to close gate 3 in the *Mine Shafts* case, I would be acting wrongly in the fact-relative sense, but I would not be blameworthy.

Acting in a way that is wrong in the moral-belief-relative sense provides neither (1) a necessary nor (2) a sufficient condition for being blameworthy. (1) I can be blameworthy for acting in a way that I do not believe to be wrong in the fact-relative sense when my belief that what I'm doing is morally right is the result of negligence or culpable ignorance. (2) Were I to close gate 3 in the *Mine Shafts* case, I would be acting in a way that I believe to be wrong in the fact-relative

sense, but I would not be blameworthy. Also, consider the case of Huck Finn. He acted in a way that he believed to be wrong in helping a runaway slave to escape, but he doesn't seem blameworthy.

Acting in a way that is wrong in the evidence-relative sense provides a sufficient condition for being blameworthy, but probably not a necessary condition for being blameworthy. To see that it doesn't constitute a necessary condition for being blameworthy, consider the case where I break my promise to my ex-girlfriend to attend her wedding. I act in a way that that's right in the evidence-relative sense, but it seems that my ex-girlfriend can legitimately blame me for breaking my promise, as I could have gone to the wedding and behaved myself. The fact that I have the moral failing of having the disposition to get drunk and be belligerent at such functions does not excuse me from blame for not keeping my promise.

V. Subjective and Objective Consequentialism

A. The Distinction

Objective Consequentialism: S 's performance of x is wrong if and only if there is some available alternative act that would produce more value.

Subjective Consequentialism: S 's performance of x is wrong if and only if, given the evidence available to S , S ought to believe that x 's expectable value is lower than that of some available alternative.

B. Which is more plausible?

The answer will depend on which sense of 'wrong' we're trying to give an account of. Clearly, objective consequentialism gives the more plausible account of 'wrong' in the fact-relative sense. And, clearly, subjective consequentialism gives the moral plausible account of 'wrong' in the evidence-relative sense.

VI. Various Versions of the Epistemic Objection to Objective Consequentialism

A. The Standard Version of the Epistemic Objection

If objective consequentialism is true, then nearly 100% of the acts that people perform are morally wrong. To see this, consider the following sort of cases: (1) Feldman's safe-cracking case, (2) Miller's choking-neighbor case, (3) Howard-Snyder's beating-Karpov-at-chess case, (4) my talking-to-John-after-class-giving-rise-to-the-next-Hitler case, and (5) Wiland's the proverbial-monkey-at-

the-typewriter case. If objective consequentialism is true, then even the best of us act wrong in nearly 100% of the cases. Indeed, the best of us probably don't act wrongly any more often than the worst of us do, although their wrong acts are certainly more likely to be morally worse than our wrong acts.

This fact has spurred a number of objections to objective consequentialism: (1) it violates the ought-implies-can principle, (2) it fails to be practical and action-guiding, (3) it yields a counterintuitive account of blameworthiness, and (4) it absurdly implies that even the best of us act wrongly in nearly 100% of the time.

B. Does objective consequentialism violate the ought-implies-can principle?
Howard-Snyder's Argument and Moore's Reply

Imagine that the fate of the world depends upon Howard-Snyder's beating chess grandmaster Anatoly Karpov. Now here's Howard-Snyder's Argument, as interpreted by Moore:

1. If objective consequentialism is true, then she ought to beat Karpov.
2. But she cannot beat Karpov.
3. 'Ought' implies 'can'. If an agent cannot do something, then it is not the case that she ought to do that something.
4. Therefore, objective consequentialism is false.

In defense of (2), she offers the following sub-argument:

- I. An agent can do something only if whether she does that something is under her control such that, if she sets out to do that something, she will do that something and she will do so precisely because she set out to do that something.
- II. Beating Karpov is not under her control. If she sets out to beat Karpov and happens to perform a sequence of moves that results in her beating Karpov, that would be a matter of extraordinary luck.
- III. Therefore, she cannot beat Karpov.

And, in defense of (1), Howard-Snyder gives the following sub-argument:

- i. If objective consequentialism is true, then, of all those acts that she can perform, she ought to perform the one that would produce the best consequences.
- ii. Beating Karpov is an act that she can perform. After all, there is a series of moves, each of which she can perform, that would result in her beating Karpov.

- iii. Beating Karpov is, of all the acts that she can perform, the one that would produce the best consequences.
- iv. Therefore, if objective consequentialism is true, then she ought to beat Karpov.

But, as Moore points out, (III) and (ii) contradict each other, and Howard-Snyder cannot have it both ways. She cannot say that she both can and cannot beat Karpov in the same sense.

It seems important, then, to disambiguate two senses of 'can', for it seems that in one sense Howard-Snyder can beat Karpov and in another sense she cannot. Now Moore suggests that the sense in which she can beat Karpov is the sense in which it is physically possible for her to do so. He says, "Her beating Karpov is consistent with the physical laws of nature. Call this sense, 'can₁'" (2007, 87). But this seems to conflict with determinism. If determinism is true, then it may well be that her beating Karpov is inconsistent with the physical laws of nature. It is preferable, I think, to remain neutral as to whether the sense of 'can' under which she can beat Karpov is compatible with determinism. So let us say, then, that there is a sense of 'can' under which Howard-Snyder can beat Karpov, Tiger Woods can make a hole-in-one, and I can publish in the *Philosophical Review* even though, in none of these cases, is this something that is under the agent's control in the sense explained in (I) above. Let's call this sense, 'can₁', and let's call the sense of can given in (I) above, 'can₂'.

It's clear, then, that if the foregoing interpretation of Howard-Snyder is correct, then Howard-Snyder is illegitimately equivocating between these two senses of can.

Perhaps, though, the following is a more charitable interpretation of the argument. This is the interpretation that was suggested to Moore by an anonymous referee (viz., me):

- a. If objective consequentialism is true, then (since Howard-Snyder can₁ beat Karpov) she ought to beat him.
- b. But Howard-Snyder cannot₂ beat Karpov.
- c. 'Ought' implies 'can₂': if an agent cannot₂ do something, then it's not true that she ought to do it.
- b. [sic] Therefore, it's not true that Howard-Snyder ought to beat Karpov. [from (b) and (c)]
- e. Therefore, objective consequentialism is false. [from (a) and (d)] (Moore 2007, 87)

This argument doesn't commit the fallacy of equivocation, so it is valid. But, as Moore argues, we should reject (c). If objective consequentialism is to be

formulated in terms of can_1 , then so should the ought-implies-can principle. Arguably, ' can_1 ', not ' can_2 ', is the 'can' of ordinary language, and we should formulate the ought-implies-can principle in term of the 'can' of ordinary language, for the ought-implies-can principle is justified by appeal to ordinary language. We would not say that when Tiger Woods sinks a hole-in-one, he does something he cannot do. Yet sinking a hole-in-one is not something he can_2 do.

In conclusion, we should reject the notion that objective consequentialism conflicts with the ought-implies-can principle.

C. Howard-Snyder Argument in Terms of Blameworthiness and Miller's Two Arguments

1. Howard-Snyder's Argument in Terms of Blameworthiness

There is one more way we might interpret Howard-Snyder's argument against objective consequentialism:

- a. If objective consequentialism is true, then we are morally blameworthy for performing actions that we can_1 but cannot $_2$ do.
- b. But it is not the case that we are morally blameworthy for performing actions that we can_1 but cannot $_2$ do.
- c. Thus, objective consequentialism is not true. (Moore 2007, 89)

EVALUATING THE ARGUMENT: Premise (a) is false. Objective consequentialism says nothing about when agents are morally blameworthy. Of course, Howard-Snyder might claim that there is a conceptual connection between wrongness and blameworthiness, but if there is such a connection, the connection is between blameworthiness and 'wrong' in the evidence-relative sense, whereas objective consequentialism gives us a criterion for wrongness in the fact-relative sense.

2. Miller's First Argument

Imagine that while I'm having my breakfast, my next-door neighbor is, unbeknownst to me, choking. It is possible for me to run next door, break into his house, and perform the Heimlich maneuver on him, thereby saving his life. But since I have absolutely no reason to think that he is choking, the thought of my doing so never even crosses my mind. And so he dies. While saving him is something I can_1 do, it is not something that I can_2 do. So consider the following argument:

- a. If objective consequentialism is true, then I did the wrong thing when I failed to rescue my neighbor from choking.
- b. If I did the wrong thing when I failed to rescue my neighbor from choking, then I ought to be punished.
- c. But it is not true that I ought to be punished.
- d. Thus, objective consequentialism is not true. (Moore 2007, 90)

EVALUATING THE ARGUMENT: There's an equivocation here between the fact-relative and the evidence-relative senses of wrongness. Premise (a) is true only if 'wrong' refers to 'wrong' in the fact-relative sense, whereas premise (b) is true only if 'wrong' refers to 'wrong' in the evidence-relative sense.

3. Miller's Second Argument

Miller claims that a good person can't be someone who is constantly making one moral mistake after another, acting wrongly time and time again. "He asks us to imagine a plumber who makes mistake after mistake. Such incompetence indicates a bad plumber even if in other respects the plumber is friendly, hardworking, etc. This is analogous to our evaluation of moral agents" (Miller 2007, 90).

- a. If objective consequentialism is true, then people will often do the wrong thing.
- b. If people will often do the wrong thing, then even the best possible humans are morally bad people.
- c. But it's not true that even the best possible humans are morally bad people.
- d. Thus, objective consequentialism is not true. (Moore 2007, 90)

EVALUATING THE ARGUMENT: Premise (b) is false. Miller defends (b) by drawing an analogy between plumbers and moral agents. He claims that just as a plumber can't make mistake after mistake and still count as a good plumber, a person can't perform wrong act after wrong act and still count as a morally good person. But, as Moore points out, the analogy with a plumber doesn't seem apt. It makes more sense to compare us with physicians.

D. Wiland's the-Proverbial-Monkey-at-the-Typewriter Objection

E. The Impracticality Objection and Three Possible Responses

- 1. The Move from Actual Value to Expected Value and Why This Doesn't Work

2. Distinguishing between a Criterion of Rightness and a Decision Procedure
3. Partners in Guilt