Moral Uncertainty and the Principle of Equity among Moral Theories

Andrew Sepielli
Rutgers University – New Brunswick
Department of Philosophy

Abstract

Suppose your credence is divided between two moral theories – Theory T and Theory U. According to T, you have more reason to do Action A than you have to do Action B. According to U, you have more reason to do B than you have to do A. What is it rational to do in a situation in which A and B are the two possible actions? Many have argued that what it's rational to do depends on two things: (a) how your credence is distributed between the theories, and (b) how the difference in moral value between A and B if T is true compares to the difference in moral value between B and A if U is true. But this answer prompts a further question: How do we make the intertheoretic comparisons of value differences mentioned in (b)? The theories themselves seem not to provide the resources required to do so. In *Moral Uncertainty and Its Consequences*, Ted Lockhart argues that intertheoretic comparisons of value differences are possible if we adopt a principle he calls the "Principle of Equity among Moral Theories". I argue on several grounds that this principle is untenable, consider some rejoinders on Lockhart's behalf, and conclude that these rejoinders do not succeed.

Suppose your credence is divided between two moral theories – Theory T and Theory U. According to T, you have more reason to do Action A than you have to do Action B. According to U, you have more reason to do B than you have to do A.

Many real-life cases fall under this schema. For example, I might have some credence in a retributive theory of punishment and some credence in a non-retributive theory of punishment. According to the first theory, it may be better to subject a criminal to very harsh treatment than to rehabilitate him. According to the second theory, the reverse may be true. Or I might have some credence in a traditional consequentialist theory, and some credence in a non-consequentialist theory. The first theory might recommend killing one person to save five people, while the second theory might recommend against it

What is it rational for you to do when you're uncertain between conflicting moral theories? This is not the old question of what you should do, *given* some moral theory, when you are uncertain about the non-moral facts. The question I'm asking takes us back one step; what is it rational to do when you're uncertain regarding the theories themselves?[1] In this paper, I'm going to consider a problem that arises when we try to answer this question, and then evaluate one solution to that problem.

One possible answer to the question is: Act in accordance with the theory in which you have the highest credence. That is, if your degree of belief is highest in a theory according to which Action A is better than Action B, then you should do A rather than B. But we should be suspicious of this answer, since the parallel answer in the non-moral case seems so clearly mistaken. For suppose that I am deciding whether to drink a cup of coffee. I have a degree of belief of .2 that the coffee is mixed with a deadly poison, and a degree of belief of .8 that it's perfectly safe. If I act on the hypothesis in which I have the highest credence, I will drink the coffee. But this seems like a bad call, since the downside of death is so much greater than the

[1] This is a very under-addressed problem in moral philosophy. The only recent publications to address the issue are Hudson (1989), Oddie (1995), Lockhart (2000), Weatherson (2002), Sepielli (2006), Ross (2006), Guerrero (2007), and Sepielli (2009). A very similar debate – about so-called 'reflex principles' – occupied a central place in Early Modern Catholic moral theology. The most notable contributors to this debate were Bartolomé de Medina (1577), Blaise Pascal (1656-57), and St. Alphonsus Liguori (1755). The various positions are helpfully summarized in Prümmer (1957), *The Catholic Encyclopedia* (1913), and *The New Catholic*

upside of enjoying coffee.

Similarly, suppose I am deciding whether to do A or to do B. There's some chance that Theory T is correct, and a slightly greater chance that Theory U is correct. I also believe that, if T is correct, then A is saintly and B is abominable; but if U is correct, then B is slightly nasty and A is merely okay. Despite the fact that my credence is higher in the theory according to which there is more reason to do B , it still seems like I ought to do A instead, since A's "moral upside" is so much higher than B's, and its "moral downside" not nearly as low.

Here, then, is a more promising answer: Perform the action with the highest *expected moral value*. We get the expected moral value of an action by multiplying the subjective probability that some theory is true by the value of that action if it is true, doing the same for all of the other theories, and adding up the results.[2] This strategy is sensitive not only to my credences in the various theories, but also to what these theories say about *degrees of moral value* - the sizes of the "moral upsides" and "moral downsides" of actions, to which the "highest credence" strategy was insensitive.

And yet this approach confronts us with a difficulty – something I will call the *Problem of Intertheoretic Comparisons of Moral Value*. In order for me to determine whether A or B has the higher expected value, I'll need to know how the difference between A and B if T is true compares to the difference between B and A if U is true.

Encyclopedia (2002). Thanks to Des Hogan for clueing me in to this body of work.

[2] Stated slightly more formally: Expected Moral Value = Σ_i p(Theory$_i$) · (Value of A according to Theory$_i$).

The problem is that there doesn't seem to be any way of making this sort of comparison. In particular, neither T nor U will be helpful, since no moral theory contains information about how its own value differences compare to the value differences that may obtain if it is mistaken.[3]

This puts us in a tough situation. It seems plausible that what it's rational to do under moral uncertainty depends on the relative sizes of theories' value differences, but we're without any way of determining what these are. We have two options. First, we could opt for a theory of rationality under moral uncertainty that is *not* sensitive to degrees of moral value. While this suggestion deserves more attention than I can offer here, we should at least notice how the "coffee" example and that example's moral analogue militate against it. The right theory – whether it's expected value maximization or something else – *should* be sensitive to degrees of moral value. The second option, of course, is to solve the Problem of Intertheoretic Comparisons of Value.

In *Moral Uncertainty and its Consequences*, Ted Lockhart diagnoses this problem and attempts to solve it. He suggests that we compare moral value across theories via a stipulation he calls the Principle of Equity among Moral Theories (PEMT):

> The maximum degrees of moral rightness of all possible actions in a

[3] There are some interesting parallels between this problem and the problem in welfare economics of interpersonal comparisons of well-being.

> situation according to competing moral theories should be considered equal. The minimum degrees of moral rightness of possible actions in a situation according to competing theories should be considered equal unless all possible actions are equally right according to one of the theories (in which case all of the actions should be considered to be maximally right according to that theory).[4]

The idea is that, if I have credence in theories T, U, and V, I set the value of the best action according to theory T equal to the value of the best action according to theory U equal to the best action according to theory V; same goes for the worst action according to each theory.

So how does the PEMT fare as a solution to the Problem of Intertheoretic Comparisons of Value? Lockhart claims that the PEMT makes these comparisons possible, and that it is attractive in its own right. I think he is wrong on both counts.

First, it is incompatible with the intuitive claim that moral theories disagree not only about what to do in different situations, but about which situations are "high stakes" situations and which are "low stakes" situations, morally speaking. A momentous decision from the perspective of traditional Christian ethics may be a relatively unimportant decision from the utilitarian perspective. But according to PEMT, all moral theories have the same amount "at stake" in every situation.[5]

[4] Lockhart (2000), p. 84.

[5] A version of this objection also appears in Ross (2006), p. 762, n. 10.

Second, the PEMT is arbitrary. Consider: It's not difficult to find a method of comparing values of actions across theories. I could, for example, declare by fiat that the difference in moral value between lying and not lying, on a Kantian deontological theory, is equal to the moral value of 23 utils, on a utilitarian theory. But if there's no principled reason for that "rate of exchange", I haven't solved anything. And, similarly, if there's no principled reason to use the PEMT, rather than some other possible method, Lockhart hasn't solved it either.

Lockhart recognizes that the PEMT may appear *ad hoc*, and tries to provide a reason why it, rather than some other principle, is the correct method of comparing values of actions across theories. He says:

> The PEMT might be thought of as a principle of fair competition among moral theories, analogous to democratic principles that support the equal counting of the votes…in an election regardless of any actual differences in preference intensity among the voters.[6]

Lockhart is right that the PEMT is analogous to this voting principle. But while the latter makes good sense, the former does not. One cannot be unfair to a moral theory as one can be unfair to a voter. And yet, presumably, fairness is why we count votes equally, regardless of preference intensity. Insofar as we care only about maximizing voters' preference satisfaction, equal counting of votes seems like quite a *bad* policy. Rather, we would want to weight peoples' votes according to the intensity of their preferences regarding the issue or candidates under consideration. Similarly, insofar as we care about maximizing expected value, it seems quite bizarre to treat moral theories as though they had equal value at stake in every case. If some act would be nightmarish according to one theory, and merely okay according to another, it seems right to give the first theory more "say" in my decision.

The gist of the analogy, though, is that we should somehow treat moral theories equally. But even granting that some "equalization" of moral theories is appropriate, Lockhart's proposal seems arbitrary. Why equalize the maximum and minimum value, rather than, say, the *mean* value? And especially, why equalize the maximum and minimum value with regard to particular situations, rather than the maximum and minimum *conceivable* or *possible* rightness? This is all to make a more general point: It seems as though we could find other ways to treat theories equally, while still acknowledging that the moral significance of a situation can be different for different theories. Thus, even if we accept Lockhart's "voting" analogy, there is no particularly good reason for us to use PEMT rather than any of the other available methods.

Third, the PEMT is nearly useless. It requires that all theories have highest-possible-valued acts of equal value, and lowest-possible-valued acts of equal value. But it tells us nothing about how to assign values to the acts that are intermediately ranked according to the various theories. Lockhart recognizes this, and rather halfheartedly suggests that we employ the "Borda count" method as a solution. On this method, we assign values to options equal to their numerical ranking on an

[6] Lockhart (2000), p. 86.

ordinal worst-to-best scale. So, suppose I am deciding which Bob Dylan album to listen to. My worst option is *Dylan*; my second worst option is *Street Legal*; my second best option is *Blonde on Blonde*; and my best option is *John Wesley Harding*. The Borda count method would assign utilities of 1, 2, 3, and 4, respectively, to these options.

Lockhart recognizes the flaws of this method, perhaps the most interesting of which is that it has the consequence that the value of any of the options depends on *how many* other options there are. Anyhow, it seems clear that if the defender of PEMT needs to use Borda counting to make use of his principle, then his principle isn't particularly useful.

Fourth, PEMT has the consequence that the expected values of actions will depend on which other actions are possible in a situation. This is a violation of something akin to the "Independence of Irrelevant Alternatives".[7] Suppose that your credence is divided between Theory 1, according to which A is better than B, and Theory 2, according to which B is better than A. Now, if A and B are the only two options in a situation, then by PEMT, the value of A on Theory 1 must be equal to the value of B on Theory 2; *mutatis mutandis* for B on Theory 1 and A on Theory 2. For illustration's sake, let's just assign absolute values to these action-Theory pairs:

Theory 1: Value of A = 10; Value of B = 0

Theory 2: Value of A = 0; Value of B = 10

But now suppose an additional action becomes available that is better than A according to Theory 1, and, say, ranked between A and B according to Theory 2. This action – call it “C” – will then be the best action on Theory 1, and neither the best nor the worse action according to Theory 2. So the value assignments will have to change slightly:

Theory 1: Value of A = ?; Value of B = 0; Value of C = 10
Theory 2: Value of A = 0; Value of B = 10; Value of C = ?

As we observed earlier, Lockhart gives us no good way of assigning values to sub-optimal but super-minimal actions, so the value of A according to Theory 1 and the value of C according to Theory 2 will have to stay “up in the air” for now. But one thing’s for sure: The value of A according to Theory 1 must be less than 10, since the value of C according to that theory is 10, and the theory ranks C over A. However, A’s value on Theory 1, before C got added to the mix, was 10. So the addition of C had the consequence that A’s value on one theory, and therefore A’s expected value, were reduced, *vis a vis* B’s. This is an unwanted result. Violations of the Independence of Irrelevant Alternatives, if they’re to be countenanced at all, should be explained by the concrete, particular features of the available actions. PEMT leads to violations of this principle in virtue of merely formal features of the choice situation – the number of actions, and the rankings of the actions according to the various theories.

[7] Thanks to Brian Weatherson for suggesting this argument.

Fifth, the PEMT leads to inconsistent results when applied. Suppose my credence is divided between two moral theories. According to Theory T, the value of an action is a positive linear function of the number of instantiations of some property P that the action causes. According to Theory U, the value of an action is a positive linear function of the number of instantiations of some property Q that it causes. Now imagine two situations. In one situation, I can either cause 100 instantiations of P and no instantiations of Q, or 10 instantiations of Q and no instantiations of P. In the other situation, I can either cause 100 instantiations of Q and none of P, or 10 instantiations of P and none of Q.

Situation 1

100 P’s + 0 Q’s OR 10 Q’s + 0 P’s

Situation 2

100 Q’s + 0 P’s OR 10 P’s + 0 Q’s

If PEMT is true, then in both situations, the maximum possible moral value according to Theory T must be equal to the possible moral value according to Theory U.

But this is impossible. Since the moral values assigned by T and U are positive linear functions of the number of instantiations of P and Q, respectively, the theories’ respective value functions are:

V_T = (# of instantiations of P) · (W) + X

V_U = (# of instantiations of Q) · (Y) + Z

In the *first situation*, the highest possible value according to T is 100W + X, since the best possible action according to T is the one that causes 100 instantiations of P. By PEMT, this must be equal to the highest possible value according to U. This value is 10Y + Z, since the best possible action according to U is the one that causes 10 instantiations of Q. In the *second situation*, the highest possible value according to T is 10W + X. If PEMT is correct, this value must be equal to the highest possible value according to U, or 100Y + Z.

But PEMT cannot hold in the second situation if it held in the first situation. Why not? Well, the highest possible value according to T in the second situation (10W + X) is *lower* than the highest possible value according to T in the first situation (100W + X), because W is a positive number. On the other hand, the highest possible value according to U in the second situation (100Y + Z) is *higher* than the highest possible value according to U in the first situation (10Y + Z), because Y is a positive number. So if the highest possible values according to the two theories were equal in the first situation, they cannot also be equal in the second. PEMT fails.

Now, that example was simplistic in two respects. First, the theories were monistic; each assigned moral relevance to only a single factor – the promotion of some property. Many moral theories, however, are pluralistic; they assign moral relevance to several factors. Second, one of the acts in each situation promoted *only* what was valuable according to one theory, while the other act in each situation promoted *only* what was valuable according to the other theory. This is, of course, rarely the case in real life. Lockhart might respond, then, that even if I have shown

PEMT to be inapplicable to monistic theories in rather stark choice situations, I have not shown it to be inapplicable to the more complex scenarios in which we typically find ourselves.

Consider, then, a modified version of that example. Suppose my credence is divided between two moral theories. According to Theory T, the value of an action is a positive function of the number of instantiations of some property P that the action causes, and the number of instantiations of some property Q that it causes. Another theory, Theory U, also assigns value to instantiations of P and Q, but weights P less heavily, and Q more heavily, than T did.

We can imagine the two theories' value functions as:

$$V_T = (\#\text{ of instantiations of } P) \cdot (W1) + (\#\text{ of instantiations of } Q) \cdot (W2) + X$$

$$V_U = (\#\text{ of instantiations of } P) \cdot (W1 - A) + (\#\text{ of instantiations of } Q) \cdot (W2 + B) + Z$$

Now imagine two situations. In one situation, I can either cause 100 instantiations of P and 10 instantiations of Q, or 50 instantiations of Q and 5 instantiations of P. In the other situation, I can either cause 100 instantiations of Q and 10 of P, or 50 instantiations of P and 5 of Q.

Situation 1

100 P's + 10 Q's OR 50 Q's + 5 P's

Situation 2

100 Q's + 10 P's OR 50 P's + 5 Q's

It should not be difficult to see how, if PEMT holds in the first situation, it cannot also hold in the second situation. If we take it as given that PEMT holds in the first situation, then Theory U must have more available value in the second situation than Theory T. This is because Theory U weights the production of Q more heavily, and the production of P less heavily, than Theory T does, and there is more Q and less P available in the second situation than in the first.

The lessons of these examples could be applied to still richer cases. All one needs to generate an impossibility result for the PEMT are at least two theories, and at least two scenarios each allowing at least two possible acts. PEMT must hold in the first scenario and in the second. This is impossible unless the difference between the highest possible value in the first situation and the highest possible value in the second situation, according to one theory, is *precisely the same* as the difference between the highest possible value in the first situation and the highest possible value in the second situation, according to the other theory. But this will almost never be the case.

Still, Lockhart has a response available. I asked you to imagine moral theories as represented by single value functions. But perhaps this is not the only way, or even the best way, to understand moral theories. Instead, moral theories might specify different value functions for different situations. For example, the value a hedonistic utilitarian theory assigns to an action that produces some number of hedons might depend on the situation; it might be *situation-relative*.[8]

If situation-relativity is possible, then Lockhart can reply to my impossibility arguments as follows (here I imagine his reply to the first, more simplistic, impossibility argument): It is false that the maximum value in Situation 2 according to Theory T must be lower than the maximum value in Situation 1 according to Theory T, simply because fewer instantiations of P may be produced. Similarly, it is false that the maximum value in Situation 2 according to Theory U must be higher than the maximum value in Situation 1 according to Theory U, simply because more instantiations of may be produced. Both theories could, after all, have different value functions corresponding to the different situations. If that's so, then PEMT could hold in both the first situation and the second.

An interesting approach, but still, I think, a multiply flawed one.

First, the mere possibility of situation-relative value is not enough to rescue PEMT. It is not sufficient simply for theories' value functions to vary depending on the situation. They must vary in precisely the way that ensures that PEMT will hold in every situation. But why wouldn't theories' value functions vary in one of countless other ways that are *not* amenable to PEMT? Absent some kind of answer to this question, the defender of PEMT can find help from situation-relativity only by employing it in a suspiciously *ad hoc* way.

But let's put this worry aside, and assume that theories' value functions vary across situations such that PEMT is preserved. This has some counterintuitive implications. Suppose my credence is divided between Theories T and U. According to

[8] Ruth Chang suggested this clever response on Lockhart's behalf.

T, the rightness of an action in some situation is some positive function of the number of instantiations of P it produces. According to U, the rightness of an action in that situation is a positive function of the number of instantiations of Q it produces. Now, imagine that in that situation, I can either create some instantiations of P and slightly fewer instantiations of Q, or else some instantiations of Q and slightly fewer instantiations of P. By PEMT, the value according to Theory T of taking the first option must be equal to the value according to Theory U of taking the second option. So far, so good.

Now, suppose I start by believing that I can create 10 instantiations of P, but later come to believe that I can create 100 instantiations of P. That is, I start believing that I'm in one situation, and then come to believe that I'm in another, P-richer situation. All of my other beliefs about the number of instantiations of P and Q that I can produce remain constant, and the rest of my relevant beliefs correspond to the facts as laid out in the previous paragraph.

It's natural to think, "Okay, I thought the situation was such that only a few instantiations of P were possible. Now I think the situation is such that many more instantiations of P are possible. So, according to the P-favoring theory (Theory T), more value should be possible than I'd previously thought." But this sort of thinking is disallowed by the type of situation-relativity that necessarily preserves PEMT. For whatever the situation turns out to be, T's and U's value functions for that situation will be such that the value of the best action according to T must be equal to the value of the best action according to U. So if the situation turns out to be particularly P-poor, T's value function will "expand" so that the best action according to T has as much value as the best action according to U. If the situation turns out to be particularly P-rich, T's value function will "contract" to preserve the same.

We can more easily see how odd this kind of situation relativity is by imagining an agent who's deciding whether to find out how many instantiations of P are possible on option one. If she knows for sure that some act will produce more instantiations of P than any other act, it seems as though she should take no effort whatsoever to find out exactly how many P instantiations this is, since it will have no effect on the maximum value of this act as compared with Theory U's favored act. But this just seems incredible. Suppose she's got some credence in utilitarianism and some credence in deontology, and is deciding whether to kill one person to save X people. Let's stipulate that utility is maximized if X is 2 or greater, and that she knows this. Is it really a complete waste of time for her to find out whether X is 2 or 2 million? Again, tough to believe.

Let me consider one final response available to Lockhart. His formulation of the PEMT states that "the maximum degrees of moral rightness of all possible actions in a situation according to competing moral theories should be considered equal." It's the *in a situation* part that has generated controversy so far. But perhaps PEMT can be modified. Why not instead say that the maximum *conceivable* degrees of moral rightness according to competing moral theories should be considered equal? Call this the "Conceivability PEMT".

This PEMT doesn't suffer from all of the problems of the first, but it suffers from some of them. First, it does seem a bit counterintuitive, in the following respect: There may be some theories, like utilitarianism, according to which there just isn't a maximum conceivable value. If infinite utility is possible, and value is an unbounded function of utility, then an infinite amount of value is possible, too. We might just stipulate that utilitarianism's value function must be bounded, but this gives rise to two problems. First, what could possibly be the argument for setting the bound at one place rather than another? In the absence of such an argument, requiring a bound introduces a significant element of arbitrariness into the proceedings. Secondly, as Frank Jackson and Michael Smith demonstrate in a recent paper, interpreting theories as bounded value functions leads us to bizarre conclusions.[9] At the very, very least, unbounded utilitarianism is a live option, and one for which the Conceivability PEMT dos not allow.

The Conceivability PEMT is just as arbitrary as Lockhart's version – why equalize the maximum and minimum, rather than the mean, two and a half standard deviations from the mean, and so on? If anything, the very possibility of this new PEMT ought to make the original PEMT seem even more arbitrary. Why go with the old version rather than this new one? And insofar as the old PEMT is a viable option, it ought to make *this* PEMT seem more arbitrary.

This PEMT is, if anything, even more noticeably useless than the original version. Unless one of my possible actions in some situation is *the best or worst conceivable action* according to one of my theories – and let's face it, when's *that* ever going to be the case? – the Conceivability PEMT will say nothing about it. Lockhart's PEMT at least had something to say about two actions in every situation for

[9] See Jackson and Smith (2006).

every moral theory.

For what it's worth, the Conceivability PEMT doesn't generate violations of the Independence of Irrelevant Alternatives, and isn't vulnerable to the kind of "impossibility arguments" that I made against the original PEMT. [10]

References

The Catholic Encyclopedia (1913), available at http://www.newadvent.org/cathen/

Guerrero, Alexander. “Don’t Know, Don’t Kill: Moral Ignorance, Culpability, and Caution,” *Philosophical Studies*, 136 (2007), 59-97.

Jackson, Frank and Michael Smith. “Absolutist Moral Theories and Uncertainty,” *Journal of Philosophy* 103 (2006), pp. 267-283.

Liguori, St. Alphonsus. *Theologia Moralis, 2nd ed.*(1755), trans. R.P. Blakeney. London: Reformation Society, 1852.

Lockhart, Ted. *Moral Uncertainty and Its Consequences*. Oxford University Press, 2000.

de Medina, Bartolomé. *Expositio in 1am 2ae S. Thomae*, 1577.

The New Catholic Encyclopedia, 2nd ed. Catholic University Press, 2002.

Oddie, Graham. "Moral Uncertainty and Human Embryo Experimentation," in *Medicine and Moral Reasoning*, edited by K.W.M. Fulford. Grant Gillett and Janet Martin Soskice. Oxford University Press, 1995.

Pascal, Blaise. *The Provincial Letters* (1656-57), trans. A.J. Krailsheimer. Penguin Books, 1967.

Prümmer, Dominic M. *Handbook of Moral Theology*. P.J. Kennedy and Sons, 1957.

Ross, Jacob. “Rejecting Ethical Deflationism,” *Ethics* 116 (2006), pp. 742-768.

Sepielli, Andrew. “What to Do When You Don’t Know What to Do,” in *Oxford Studies*

[10] Thanks to Ruth Chang and Brian Weatherson for helpful feedback on this paper.

in Metaethics, Volume 4, edited by Russ Shafer-Landau Oxford University Press, 2009.

-----. “Review of *Moral Uncertainty and Its Consequences*,” *Ethics* 116 (2006), pp. 601-603.

Weatherson, Brian. “Review of *Moral Uncertainty and its Consequences*,” *Mind* 111 (2002), pp. 693-696.