

## Should You Look Before You Leap?

Andrew Sepielli is an assistant professor of philosophy at the University of Toronto, and is presently writing a book on pragmatism and moral objectivity entitled *Foregrounding Values*.

I want morality to be practical, followable, action-guiding. That's why I'm unsatisfied with the traditional formulation of utilitarianism, for example, which enjoins us to do the action that yields the greatest utility. Since I am never sure which action that is, I can never follow utilitarianism so understood; I can never use it as a guide to action.

Utilitarians are aware of this problem. Their typical response draws a distinction between two types of rightness - subjective and objective. What's subjectively right depends on beliefs, evidence, or probabilities. What's objectively right needn't depend on anything like that; it depends on the way the world generally is and will be. Utilitarianism as described above is a theory of objective rightness. But there is also a utilitarian account of subjective rightness, which tells us to maximise *expected* utility. For instance, the expected utility of an action that has a 20% chance of producing 5 utils and an 80% chance of producing 10 utils is:  $(20\%)(5) + (80\%)(10) = 1 + 8 = 9$ . Of course, expected utility calculations are apt to be intractable in real life, and so utilitarians since John Stuart Mill have suggested that we instead employ simple "rules of thumb", the prescriptions of which tend to approximate those of subjective utilitarianism.

This seems to give me what I want - an action-guiding moral code. But alas, my satisfaction is short-lived. For not only am I uncertain about how much utility my actions will produce; I am also uncertain about whether utilitarianism is right in the first place. So now what am I to do?

In other work, I've tried to develop an answer to this question. Just as there is a version of utilitarianism that tells you what to do when you're uncertain about utility, there may be a kind of meta-rule that tells you what to do when you're uncertain about utilitarianism, or more generally, when you're uncertain about morality. If subjective utilitarianism directs us to maximise expected *utility*, then a plausible meta-rule might direct us to maximise expected *rightness* or *moral* value, or to do what's most likely to be morally acceptable. Someone who is uncertain about morality can then guide her actions by a meta-rule like this (or perhaps by a rule of

thumb that approximates it - the University of Pennsylvania philosopher Alex Guerrero has suggested the basically self-explanatory "Don't Know, Don't Kill" principle).

But if we can be uncertain about moral rules, then surely we can also be uncertain about meta-rules.

This actually presents two very different problems. Let's call one of them the "Inconsistency Problem": Suppose you are uncertain which view of morality is right. A philosopher, Tim, suggests that you guide your action by meta-rule A. But you find that you are uncertain whether A or another meta-rule, B, is right, and return to Tim for advice. What should he tell you *now*? He might simply say, "I already told you what to do. Use meta-rule A. End of conversation." But if Tim is so willing to refuse this request for guidance, to respond with such insensitivity to your expressed uncertainty, why would he have even offered meta-rule A in the first place? Why wouldn't he have simply said, "Morality already tells you what to do. End of conversation"? More likely, Tim would offer you a *meta*-meta-rule - "AA", let's call it. In that case, though, you might well ask of Tim: "First you said to follow meta-rule A; now you say to follow meta-meta-rule AA. But their prescriptions might be inconsistent. So which is it?" I think Tim courts trouble if he says "A", but also if he says "AA", or "both", or "neither". I try to resolve this problem in a paper called "What to Do When You Don't Know What to Do When You Don't Know What to Do...".

The problem I want to work through here is one I call the "Guidance Regress Problem" Like many regress problems, it makes especially vivid the possibility that we may not be able to get what we want in philosophy. I wanted morality to be action-guiding. But we've been presuming that it can't be action-guiding unless it tells me what to do in the face of the relevant uncertainty. But now it seems as though there's no upper limit to my uncertainty. I can be uncertain about utility, uncertain about morality, uncertain about meta-rules, meta-meta-rules, meta-meta-meta-rules, and so on. This seems to imply that my actions will inevitably be unguided, and threatens to deprive the project of developing meta-rules of any chance at fulfilling my aims in pursuing it.

Let's slow down a bit, though. The idea that all of my actions are unguided leaps of faith is surely at odds with my everyday experience. I mostly just *do things*; occasionally, I employ a moral rule; every once in a blue moon, I employ a meta-rule or a meta-meta-rule. But despite

basically making my living wallowing in uncertainty, I cannot report having ever felt condemned to take a leap of faith on account of the regress I've described. So what's going on here?

I think we need to distinguish between two types of uncertainty. The first is dispositional, not necessarily conscious, the sort of attitude I have towards any claim I wouldn't bet my life on. The second is conscious, involving a feeling of directionlessness, the kind that appears when I deliberate, and disappears when I'm "in the zone". I am uncertain in only the first sense about what the strings on a guitar are; I am uncertain in both senses about what the strings on a banjo are. That is why I can simply play an A7 on a guitar, but can play an A7 on a banjo only *by trying*.

This distinction helps us see how this "guidance regress" can and does end: Although I am in some sense uncertain about nearly everything, I only need a rule to navigate that tiny fraction of my uncertainty that is conscious. So long as I'm not consciously uncertain about moral rules, I have no need for meta-rules. And so long as I'm not consciously certain about meta-rules, I have no need for meta-meta-rules. And so on.

Some may find this solution unsatisfying because it seems to rely on a contingent feature of human psychology, rather than on a necessary fact about thought-as-such. But I question whether this dissatisfaction is warranted. I might be bothered if I thought the ability to halt the regress in this manner could be *lost*, but that's not the same as it's being *contingent*. Nor is it even obvious to me that it's contingent. It might be contingent that my conscious uncertainty tends to give out *here* rather than *there*, but nonetheless necessary that it gives out *somewhere*.

Rather, I find this solution unsatisfying for another reason: It isn't clear to me that halting the regress by dissolving conscious uncertainty is really any better than simply taking an unguided leap of faith. Of course it *feels* better to act from a unified perspective on the world than *despite* a *disunified* one. But the real problem with leaps is not that it feels bad to take them. This is at most an indicator of the real problem - namely, that leaps of faith are morally risky.

But is dissolving your conscious uncertainty any less risky? Suppose a doctor is consciously unsure whether it's okay for your child to take a certain kind of medicine. Now consider two scenarios: In one, the doctor simply writes a prescription despite her nagging doubts. In the other, she experiences a sudden rush of confidence, perhaps after taking a huge swig of coffee, or hearing "Eye of the Tiger" on the hospital radio, and then writes the prescription. I doubt that you would find the second possibility more comforting than the first. Both involve the doc-

tor being reckless with regard to your child's health. So why should you find it more comforting just because you're the one experiencing surges in confidence rather than taking leaps of faith?

The lesson to be drawn is that the waning of conscious uncertainty is only a solution to the *psychological* problem of how we can act without leaping. It's not a solution to the *normative* problem of how we can manage moral risks non-recklessly. For it seems like, however many orders of meta-rule you apply, you will eventually have to take the sort of risky action that we originally developed meta-rules to avoid. Maybe it will come sooner, maybe later; maybe it will *feel* risky at the time, maybe it won't. But it will happen, unless we somehow reach dispositional certainty at some level.

What, then, is the value of developing meta-rules? They cannot completely solve the normative problem; and while they may help us with the psychological problem, so can a pep talk or a Good Stiff Drink.

I think the right thing to say is that meta-rules offer us a normative advantage by *forestalling* moral recklessness, rather than by eliminating it entirely. More precisely, there is a sense in which it is better to leap in the face of uncertainty about meta-rules than to leap in the face of uncertainty about ordinary moral rules, better still to leap in the face of uncertainty about meta-meta-rules, and so on. But why?

To answer this question, we'll need a better understanding of how leaps of faith work. On one view, they are utterly arbitrary; there is no explanation at the level of desires, affections, and values that would rule out any conceivable action. For example, the doctor who leaps in the case above might just as well give your child turpentine instead of medicine, or suddenly take the day off to go porcupine-hunting. It may seem obvious that if this is what a leap is, it's always better to ascend to another lever of reflection than to take one at the current level.

But this is a very implausible view. It fails to recognise an important feature of unguided actions - that they are only unguided *relative to some other actions*. For example, prescribing the medicine would be unguided relative to treating the child without prescribing it. But each action (and their disjunction) would be guided relative to giving the child turpentine; whatever the doctor is uncertain about, she's certain that she shouldn't do *that*. Just as importantly, the theoretical help apparently afforded by this view of leaps is entirely illusory. For if leaps are truly random, there is no reason to think that the doctor uncertain among *meta-rules* would not randomly select a crazy meta-rule like "*minimise expected moral value*" or some such.

On a more plausible account, to leap in the face of uncertainty among moral theories is simply to implicitly accept and act on some plausible meta-rule without considering the possibility of other meta-rules. For example, one might respond to moral uncertainty by doing what one regards as most likely to be morally right, without ever holding this response up for explicit scrutiny. Against this background, the practice of developing meta-rules for moral uncertainty would seem to have three important effects. First, it makes the would-be leaper explicitly aware of competing meta-rules - alternatives to the one he had previously taken for granted. Second, it may prompt him to evaluate the merits of the various meta-rules, which may make him more sure of some rules and less sure of others. Third, it will cause him to act on his implicit acceptance of a *meta*-meta-rule - that is, unless he explicitly considers the various meta-meta-rules, in which case the his confidence in these may shift around and he may act on implicit acceptance of a *meta*-meta-meta-rule. And so on.

This helps us restate our question more helpfully: Why is it worse to act on implicit acceptance of a meta-rule now than to consider and evaluate other possible meta-rules and act on implicit acceptance of a *meta*-meta-rule later? (Or substitute “*meta*-meta-meta-rule” for “*meta*-meta-rule” and the latter for “*meta*-rule”. Again: *and so on*.) Well, let’s consider the three steps I’ve just mentioned, but in reverse order. Suppose first that you have already considered the possibility that there may be alternative meta-rules, and have gone through the process of evaluating them. Now you must decide whether to implicitly accept and act on some meta-meta-rule, or to do what you were going to do in the first place - namely, just act on the meta-rule that you had been taking for granted.

From this perspective, it must surely seem better to implicitly accept and act on a meta-meta-rule that, as it were, aggregates the opinions of the various meta-rules you find plausible, than to simply act on one of those meta-rules as though there were no alternatives. To do the latter would be to grant no weight whatsoever to rules that, according to your own (now more panoramic) perspective, have some chance of being correct. Drawing on an earlier observation, we might say that while acting on this meta-meta-rule is unguided relative to acting on some competing meta-meta-rule (that’s why it’s a leap), it is not unguided relative to acting simply on a meta-rule.

Now, given that the meta-meta-rule upon which you will act is an aggregator of the verdicts of various meta-rules, it will make sense, once you become aware of these meta-rules, to

subject them to evaluation. For it is preferable to aggregate the verdicts of meta-rules that are better supported by argument rather than those that are not as well supported. I should say that it's not entirely obvious why this is so — that is, why there's a *practical* advantage to thinking deeply about the rules upon which you will act as opposed to foregoing such reflection. My preferred explanation is basically that there is no relevant difference between moral reasoning and evidence-gathering, and there is almost always some value from the agent's perspective to gathering additional evidence, as I.J. Good demonstrated in his classic 1967 paper "On the Principle of Total Evidence".

It is not clear what to say about the rationality of the first step - i.e. becoming explicitly aware of the meta-rule you'd been taking for granted, and its many alternatives. The matter rests on quite delicate questions about the nature of mental states and about the rational significance of conscious awareness. But this is a moot point for readers of this paper, for you are all now incurably aware that your heretofore implicit theories about how to deal with moral uncertainty are not the only possible theories. The only open questions for you now are whether to think more about the relative merits of these theories, and whether to take into account all those theories you find plausible rather than adopt the blinkered perspective of considering only one. And I have argued both that you should think more about meta-rules, and that, ideally, you should in some way, shape, or form take into account the recommendations of all meta-rules that you find plausible.