Ethical Vagueness and Practical Reasoning*

Billy Dunaway
University of Oxford

Abstract

This paper looks at the phenomenon of ethical vagueness by asking the question, how ought one to reason about what to do when confronted with a case of ethical vagueness? I begin by arguing that we must confront this question, since ethical vagueness is inescapable. I then outline one attractive answer to the question: we ought to maximize expected moral value when confronted with ethical vagueness. And I close by developing one interesting application of this attractive answer. There is a literature criticizing the idea that one ought to maximize expected moral value when one is uncertain as to which moral theory is true. The position developed here suggests that some of these arguments overgenerate, as the approach to reasoning under ethical uncertainty that they aim to refute shares important structural features with the very plausible approach to reasoning in the presence of ethical vagueness.

Vagueness—or indeterminacy, or indefiniteness—is pervasive, and the practical domain is not exempt from it. Cases where it is vague what we ethically ought to do are widespread. One response to this situation is to acquiesce in ethical vagueness and allow it to infect related notions of practical normativity, blameworthiness, rationality, and the like. On this approach, when there is vagueness in the ethical status of an action, it is also the case that it is vague whether one rationally ought to perform that action, it is vague whether one is blameworthy if one performs it, and so on. But this isn’t the only approach to the issue, and I will set it aside for present purposes. After giving some structure to the issue of ethical vagueness, I outline an approach to achieving definite answers to the question of what one rationally ought to do, when it is vague what one ethically ought to do.

*Special thanks to Elizabeth Barnes, Ross Cameron, Jennifer Carr, Matti Eklund, John Hawthorne, Miriam Schoenfield, Mark Schroeder, Alex Silk, Julia Staffel, David Manley, Robbie Williams, and participants at the Rocky Mountain Ethics Congress and the Indeterminacy in What We Care About conference at the University of Leeds for discussion of the issues in this paper.
1 Cases of ethical vagueness

We cannot avoid the questions for practical reasoning posed by ethical vagueness by denying that ethical vagueness exists. One hallmark of vague terms is susceptibility to a certain kind of sorites series. Our use of ethical terms will be susceptible to a sorites series.

1.1 The general argument

It is permissible to kill an amoeba to save a modern, healthy, and fully functioning adult member of the species *homo sapiens*. And it is wrong to kill another equally healthy human to save the first human. But there will be some creatures from along the line of human evolutionary history for which it is implausible to say that it is wrong to kill the creature in such a case, and also implausible to say that it is not wrong to kill the creature in such a case. These will be cases where it is vague whether killing the creature is wrong.¹

In more detail: let \( c_1 \ldots c_n \) be a series of cases where in each there is a creature in front of you, and also a button you could press which would kill the creature to save the healthy adult human being. If you don’t press the button and kill the creature, the human dies. The creatures in each case are all evolutionary ancestors of human beings. They are arranged in order of ascending complexity: in \( c_1 \), it is a pre-historic amoeba in front of you, which you can kill to save the human, and in \( c_n \), another human is in front of you. And for each case between \( c_1 \) and \( c_n \), another creature from human evolutionary history is in front of you, slightly more developed than the creature in the case before it. (That is, in \( c_2 \) a creature from our evolutionary history slightly more developed than the amoeba is there; in \( c_3 \) it is a slightly more developed creature still, and so on, until in \( c_{n-1} \) an immediate ancestor of modern humans is in front of you.) Thus, for each \( c_i \) and \( c_{i+1} \) (\( 0 < i \leq n \)), the creature in front of you in \( c_{i+1} \) is slightly more developed than the creature in front of you in \( c_i \).

This constitutes a sorites series for ‘wrong’, since the following three claims hold. First, \( c_1 \) is a case where it is clearly permissible to press the button. Second, \( c_n \) is a case where it is clearly wrong to press the button. And third, for any case \( c_i \) in between \( c_1 \) and \( c_n \), there is significant pressure not to assert the conjunction of the following claims:

Wrong, \( c_i \) It is not wrong to press the button in \( c_i \);

Wrong, \( c_{i+1} \) It is wrong to press the button in \( c_{i+1} \).

Thus ‘wrong’ is similar to paradigmatically vague terms like ‘bald’, where similar claims apply. First, someone with no hairs on their head is bald. Second,

¹Thus I am treating a kind of susceptibility to a sorites series as a good indication of vagueness. For more on sorites-series as a theory-neutral characterization of vagueness, see Greenough (2003).
someone with 1,000,000 hairs on their head is not bald. And third, for any number \(i\) between 0 and 1,000,000, there is significant pressure not to assert the conjunction of the following claims:

**Bald**\(_i\) A person with \(i\) hairs is bald;

**Bald**\(_{i+1}\) A person with \(i + 1\) hairs is not bald.

This is strong evidence that ‘wrong’ is vague.\(^2\)

It is useful to think about how series of cases with this structure arise in matters close to everyday life. One can think of the development of a sperm and egg immediately prior to fertilisation, which then becomes a human child, in similar terms. It is permissible to destroy the sperm and egg prior to fertilisation; it is impermissible to destroy the child. But there will be some points along the development of the organism where it is vague whether it is wrong to destroy the thing that will become the child. Similarly with enhancement: drinking a cup of coffee is permissible; taking a drug that gives one the same cognitive boost but then kills you in a month isn’t. Euthanasia provides another case: not undertaking a small task to prolong the life of a person with many years of life ahead of them is wrong; not undertaking the same task to prolong for one second the life of someone with a painful disease is permissible. Examples of this kind abound; the purpose of the present paper isn’t to enumerate them, but it bears keeping in mind that the conclusions of this paper have straightforward application to many everyday examples of this kind.

1.2 Precise measurement and absolute vs. comparative ethical predicates

One might be tempted to think that the above sorites series for ethical predicates depends on some specific assumptions about ethical predicates that are in principle dispensable. In particular, one might suspect that it depends on the assumption that ethically relevant properties cannot be precisely measured and compared. Or, one might suspect that it depends on the assumption that ethical predicates are not comparative in structure. I will briefly sketch below why these suspicions would be unwarranted.

First, the existence of ethical vagueness has little to do with the unavailability of precise measurements of ethically relevant properties, or of incomparability among such properties. Assume for the moment that a simple Utilitarian theory in the style of Bentham (1781) is true, on which the only ethically relevant property of an action is how much net utility it produces; the right action is, on this theory, the one that produces the greatest amount of net utility. Assume that only happiness contributes to positive utility, and pain to negative utility. Moreover, these quantities can be precisely measured (so that if an act produces some utility, it can be assigned a real number corresponding to the amount of positive or

\(^2\)See Shafer-Landau (1995) for a different argument for a similar conclusion.
negative utility produced) and compared (so that one act produces more net utility than another iff the real number assigned to the former is greater than the latter). Even given all of these assumptions, there will be ethical vagueness. This is because it can be vague whether an act produces any positive or negative utility at all.

Here is a sorites series illustrating this: Imagine a series of cases, in which, much like before, one can press a button to save a human. But let the button in each case operate as follows: in \( c_1 \), it destroys 1,000,000 amoebas, in \( c_2 \), it destroys the same number of slightly more complex creatures, and in \( c_n \) it destroys 1,000,000 humans who are intrinsically the same as the human to be saved. On the version of Utilitarianism spelled out above, it is permissible to press the button in \( c_1 \) since no negative utility is produced by killing an amoeba, and significant positive utility is gained by saving a human. And it is wrong to press the button in \( c_n \). But there will be cases in between \( c_1 \) and \( c_n \) where it is vague whether the creatures that will be destroyed by the button can experience happiness or pain at all; in these cases it is vague how much net utility pressing the button produces (and vague how much net utility refraining from pressing the button produces) hence it is vague, according to the view under consideration, whether it is wrong to press the button in these cases.\(^3\)

Nor does the vagueness in ethics depend on the assumption that ethical predicates serve to pick out a threshold on a scale of ethically relevant factors. To be sure, this is one way in which ethical vagueness might arise: if ‘wrong’ applies to those actions that meet some threshold (for instance, if some amount of autonomy violation constitutes wrongness), then it will be very natural to think that there is no precise threshold at which ‘wrong’ begins to apply.\(^4\) This picture suggests that it is only the absolute, non-graded ethical expressions like ‘wrong’ that admit of vagueness, but that the underlying gradable structure might be perfectly precise. We might then suspect that vagueness could be avoided by only countenancing comparative ethical predicates which feature in sentences of the form ‘\( A \) is more wrong than \( B \)’.

In fact, this proposal isn’t so far-fetched, since many ethicists do countenance comparative ethical predicates. One might think of views which take the notion of weighted reasons for and against taking a certain course of action as primary in this way.\(^5\)

While there needn’t be a threshold on this picture, there will still be vagueness. One can devise sorites series where in the first case there are no reasons for acting either way, and hence where the weight of reasons in favor of \( \phi \)-ing will be 0. And in the last case, we might imagine that there is some reason in favor of \( \phi \)-ing, so the weight of reasons in favor of \( \phi \)-ing will be greater than 0. But for the cases in between will vary in some small respect so that it will be vague in some of

\(^3\)Compare Shafer-Landau (1995).
\(^4\)Dougherty (2013: 2)
them whether there is any reason to $\phi$. (Similar points apply to any degree on the relevant scale: for any degree of wrongness $d$, there will be a series of cases such that the first is definitely wrong to degree $d$, the last definitely is not, and it is vague whether some of the cases in the middle are vague to degree $d$.) The move to a purely comparative ethical notion removes some amount of vagueness from ethics, but it is a mistake to say that it eliminates the vagueness entirely.\textsuperscript{6}

2 What to do when it’s vague: definitions and assumptions

I will be discussing answers to the question, ‘what ought one to do when it is vague whether an available action is wrong?’. I will be supposing that we are trying to answer this question in a context where ethical considerations are the only considerations that bear on performing the action. In other words, when discussing cases where it is vague whether an action is ethically wrong, it will be a background assumption that there are no non-ethical reasons, prudential or otherwise, for or against performing it. Thus there is no possibility of answering the question of whether one should perform a vaguely wrong action in the negative, on the grounds that it will make me feel slightly cold. Clearly there will be cases where no non-ethical considerations are available at all; the question I will be asking is what to do in these circumstances.

Here then is one way to think about the question ‘what ought one to do when it is vague whether an available action is wrong?’ ‘Ought’ can take on various senses, one of which is ethical. But there are other senses as well, and vagueness in the ethical sense need not imply vagueness in other senses. Among these other senses is an ‘ought’ that picks out the actions that are best in view of what one knows.\textsuperscript{7} More specifically, I will be understanding this ‘ought’ as follows: $\llbracket$ought $\phi$$\rrbracket$ is true just in case $\phi$-ing is among the actions that are best, in view of what one knows. This ordering on actions is naturally heard as the one at issue on the true reading of the sentence ‘Sally ought to bet on heads’ when she is offered a bet with identical payouts on outcomes of a coin flip with a coin she knows has a bias $\beta$ (where $\beta > .5$) in favor of heads. Since it is more likely on what Sally knows that the coin will land heads and the payoffs of winning on a ‘heads’ bet are the same as the payoffs of winning on a ‘tails’ bet, the action of betting ‘tails’ isn’t among the best actions in view of what she knows. Let us call this the “rational ‘ought’ ”, though this may be misleading as there are likely other uses of ‘rational’ that don’t like up with the one I outline here.

We can then ask, what ought rationally an agent to do when it is vague whether an action available to her is wrong? With the assumptions outlined above, this amounts to the question: when faced with an action that is vaguely wrong, which worlds are best, in view of what one knows?

\textsuperscript{6}Perhaps cases like those Schoenfield (MS) describes under the heading of “insensitivity to sweetening” could be co-opted to illustrate this point.

\textsuperscript{7}Compare the semantics for ‘ought’ developed in Kratzer (1977).
The following additional assumptions will also be in place throughout:

The first is that one is certain about which ethical theory is correct. The existence of a sorites series of the kind described in §1 does not depend on uncertainty in this matter: even if one were certain that one ethically ought to maximize net utility, there are still possible cases where one would find it vague whether one ethically ought to press the button.8

The second is that one knows all of the physical facts about each case in a sorites series. That is, one knows the microphysical differences between each case in the series, what will result (microphysically) if one performs available action in a case, and the like. There very well could be some vagueness in the microphysical facts too, and in which microphysical fact “constitutes” a case, but I will ignore them for convenience. Instead, I will focus on the following: even someone who was fully able to comprehend the microphysical nature of a case, and how the world would unfold microphysically if a particular action were to be performed, would still find it vague in some cases whether a particular action produces happiness, violates autonomy, etc.

The third is that one knows certain necessary truths about each borderline case. For each borderline case \( b_i \) in a sorites series for ‘wrong’ (where it is vaguely wrong to press the button in \( b_i \)), there is a wrongness-like property \( \text{Wrong}_i \) which has its cut-off point exactly at case \( b_i \).9 That is, the property \( \text{Wrong}_i \) is such that pressing the button in every case in the series leading up to and including \( b_i \) lacks the property, and pressing the button in every case in the series after \( b_i \) has the property. (For simplicity assume that in other cases, \( \text{Wrong}_i \) is instantiated by everything that is wrong, and isn’t instantiated by everything that isn’t wrong.) Letting \( b_1, b_2 \ldots b_n \) be the borderline cases in a sorites series, then, the following are necessary truths:

\[
\begin{align*}
\text{Wrong}_0 & \quad b_1, b_2, b_3, b_4 \ldots b_n \text{ are } \text{Wrong}_0; \\
\text{Wrong}_1 & \quad b_1 \text{ is not } \text{Wrong}_1 \text{ and } b_2, b_3, b_4 \ldots b_n \text{ are } \text{Wrong}_1; \\
\text{Wrong}_2 & \quad b_1, b_2 \text{ are not } \text{Wrong}_2 \text{ and } b_3, b_4 \ldots b_n \text{ are } \text{Wrong}_2; \\
\text{Wrong}_3 & \quad b_1, b_2, b_3 \text{ are not } \text{Wrong}_3 \text{ and } b_4 \ldots b_n \text{ are } \text{Wrong}_3.
\end{align*}
\]

The idealising assumption, then, is that in a case of ethical vagueness one knows the propositions \( \text{Wrong}_0, \text{Wrong}_1, \text{Wrong}_2 \), etc. (After all, it is presumably a cognitive limitation only that prevents us from adequately specifying a property

---

8I address in §5 the relationship between uncertainty arising due to vagueness and the kind of uncertainty addressed in the moral uncertainty literature by authors such as Ross (2006) and Sepielli (2009).

9A terminological note: in §1, I used subscripted ‘c’s to refer to cases along a sorites series, where ‘c1’ names the first case in the series, and so on. I will continue with this convention but will also use subscripted ‘b’s to denote borderline cases along the sorites series, where ‘b1’ names the first borderline case in the series, and so on. (I will be ignoring higher-order vagueness throughout.)
that takes a stand on each borderline case, and nothing in the vagueness in ‘wrong’ itself suggests one couldn’t do this.)

The fourth is that the sorites series we are dealing with contain finitely many cases (and, hence, that there are only finitely many borderline cases in a sorites series). Any sorites series for ‘wrong’ doesn’t need infinitely many borderline cases, since the distinctive claim—that for any case there is substantial pressure not to assert that pressing the button in one borderline case is not wrong, but that pressing in the very next case is wrong—will hold so long as the cases in a finitely long series are sufficiently fine-grained. So I will make this assumption to make exposition simpler in what follows.

With these assumptions in place, the next section begins by outlining one way of answering the question, what ought one rationally to do when it is vague whether pressing the button is wrong? I begin by explicitly answering the this question with the assumption of an Epistemicist view of vagueness in the background. I will be making this assumption not out of a commitment to the plausibility of the Epistemicist view, but rather because it affords an especially simple and workable approach to vagueness which allows for straightforward talk about vagueness which retains classical logic and standard attitudes of belief, knowledge, and credence toward vague cases. My working assumption will be that we needn’t go in for the Epistemicist view in order to accommodate these desirable features. (Exactly which alternative approaches to vagueness will mimic the Epistemicist on these matters is a question in the philosophy of language that I won’t try to answer here.) I will argue that on the Epistemicist view, it is very natural to adopt a picture on which what one rationally ought to do when confronted with ethical vagueness is to maximize expected moral value. It is like the coin flip case, where one rationally ought to maximize expected monetary value in the face of uncertainty. Insofar as other approaches to vagueness mimic the Epistemicist’s talk about classical logic and various propositional attitudes, the picture I outline here will be plausible on these alternative views of vagueness as well.

3 What to do when it’s vague: maximize expected moral value

3.1 The framework

The Epistemicist view in Williamson (1994) is the standard version of the epistemic view of vagueness, and as indicated before, I will for reasons of simplicity begin by approaching the question of what we rationally ought to do in a case of ethical vagueness from within the Epistemicist framework. Applied to our sorites series for ‘wrong’ from §1, the Epistemicist view holds that for any case \( c_i \) from \( c_1 \ldots c_n \), there is a fact of the matter whether pressing the button in \( c_i \) is wrong or not. The vagueness in ‘wrong’ consists in our inability to know, for some of the cases in \( c_i \) from \( c_1 \ldots c_n \), whether pressing the button is wrong in them or not. These are the borderline cases. Vagueness is, in short, not to be found in language or in the
world, but in what we are able to know.

This picture requires some qualification in light of the simplifying assumptions we made above. (We might alternatively try to modify the assumptions, but I won’t explore this route here.) In particular, we assumed that one can know necessary truths which are expressed by sentences that are stipulated to have a particular cut-off point in a sorites series. But if ‘wrong’ itself has a precise cut-off point then one of the propositions is equivalent to the proposition specifying the cut-off point for wrongness. That is: if the third borderline case, \(b_3\) is the cut-off point for wrongness and \(b_1, b_2,\) and \(b_3\) are not wrong and \(b_4 \ldots\) are wrong, then the proposition \(\text{Wrong}_3\) is equivalent to the proposition that \(b_3\) is the cut-off point for wrongness. Our assumption that one can know the former is inconsistent with the Epistemicist claim that one can’t know the latter.

The inconsistency isn’t unresolvable, although there are multiple ways to resolve it. One could deny that the proposition \(\text{Wrong}_3\) is the same proposition as the proposition that the cut-off point for wrongness is \(b_3\). It then the claims that one knows \(\text{Wrong}_3\) and that it is not the case that one knows that the cut-off point for wrongness is \(b_3\) are not inconsistent. Or, one could hold that these are the same proposition, but entertained under difference guises, and that the relevant knowledge attributes are guise-relative. Both have some plausibility, as the property \(\text{Wrong}_3\) is specified using enumeration of borderline cases, whereas this plausibly isn’t that are and are not it its extension, and this plausibly isn’t how one ordinarily grasps wrongness. While I have some sympathy with the guise-theoretic approach, I will simply assume it in what follows for concreteness but won’t argue for it here.\(^{10}\)

One important consequence of Epistemicism for practical reasoning in cases of ethical vagueness is the following: while each borderline case is such that we can’t know whether pressing the button in it is wrong, each borderline case is not identical to the others in every epistemic respect. Instead, the likelihood on what one knows that pressing the button in the borderline case \(b_i\) is wrong is lower than the likelihood on what one knows that pressing the button in the borderline case further from \(b_i\) is wrong, where \(i < 1\). Picturesquely: as one traverses through more borderline cases of wrongness, one should grow more confident that one has passed the cut-off point for wrongness.

This is captured by the following thesis, which we can call INCREASING CREDENCES. \(Cr\), as I use it below, is a credence function such that \(Cr(p)\) is the likelihood of \(p\) on what one knows. (These are the credences one ought to have, in the sense of the rational ‘ought’ outlined above.) Then, if \(b_i\) and \(b_j\) are borderline cases in a sorites series for ‘wrong’ where \(j > i\), and \(b_x\) is the proposition that pressing the button in case \(b_x\) is wrong (under the guise of ‘wrong’), the following is true:

\(^{10}\)[Acknowledgement blinded for review]
**Increasing Credences** \( Cr(b_j) > Cr(b_i). \)

It is very natural to supplement **Increasing Credences** with a principle concerning *how much* one’s credence in each case should increase. For instance: assuming we have a genuine sorites series for ‘wrong’, it would be highly implausible to say given what one knows, it is .3 likely that pressing the button in the borderline case \( b_i \) is wrong, but that in the very next borderline case, \( b_{i+1} \), it is .7 likely that pressing the button is wrong. (If this were rational, it wouldn’t feel so implausible to select \( b_i \) as the point where the cut-off point for wrongness lies, and hence the series wouldn’t be a sorites series.) This amounts to the requirement that one’s confidence that the cut-off point for wrongness lies at a particular borderline case be roughly the same as one’s confidence that the cut-off point is at a distinct borderline case. We can call this the **Uniformity** thesis:

**Uniformity** \( Cr(\neg b_{j-1} \land b_j) \approx Cr(\neg b_{i-1} \land b_i). \)

It is intuitively very plausible that **Increasing Credences** and **Uniformity** are constraints on one’s credences about borderline cases, given the Epistemicist framework. On this view, such cases are structurally similar to other cases where vagueness is not at issue. Suppose I have drawn a winning ticket from a 10-ticket lottery, where the tickets are numbered sequentially from ‘1’ to ‘10’. Suppose moreover that I have looked at the winning ticket but not told you. You cannot know which ticket won. I then line the tickets up on a table, with the ticket labelled ‘1’ on the left, ‘2’ next to it, and so on, with ‘10’ at the other end. I then start pointing at tickets in order, asking whether a ticket I have pointed at at some time was the winner. You cannot know the answer to this question until I point to ticket 10. But there are some constraints on your credences in this case: Let the proposition \( \text{Ticket}_1 \) be the proposition that ticket 1 won, \( \text{Ticket}_2 \) be the proposition that either ticket 1 or ticket 2 won, and so on. You cannot know any of \( \text{Ticket}_1 \text{-Ticket}_9 \). But you should be *more confident* in \( \text{Ticket}_2 \) than you are in \( \text{Ticket}_1 \), as the former is more likely on your evidence than the latter, even if you cannot know either claim. And similarly for any two propositions \( \text{Ticket}_m \) and \( \text{Ticket}_n \) where \( m > n \). Moreover for any \( n \) where \( 0 < n < 10 \), one’s evidence does not make \( \text{Ticket}_{n+1} \) substantially more likely than \( \text{Ticket}_n \). Hence one’s

---

11Note that **Increasing Credences** doesn’t imply that there is a single unique rational credence one ought to have towards the proposition that pressing is wrong in a particular borderline case. (It thus doesn’t presuppose the “rational uniqueness thesis” described in Christensen (2007) and elsewhere.) Rather, the only constraint it imposes is that any rational credence distribution must assign to cases that are closer to cases of determinate wrongness credences that are higher than the credences it assigns to cases that are further from cases of determinate wrongness. There might be other equally rational credence distributions that assign different credences to each case, but also satisfy **Increasing Credences**.

12This only requires that these credences in the cut-off points must be *roughly* the same; I do not want to rule out rationally thinking that it is somewhat more likely that, say, the cut-off point lies in the middle of the series of borderline cases rather than at the extreme edges.

13*Cf.* Hawthorne (2004)
credences in the ticket case should obey analogues of INCREASING CREDENCES and UNIFORMITY. And it is quite plausible that one’s epistemic position with respect to the borderline cases of wrongness resembles one’s epistemic position with respect to the lottery tickets in this case.

3.2 Moral value and expected value maximization

On this picture, a case of ethical vagueness induces uncertainty about the ethical facts. In cases of uncertainty like this, it is very natural to take the rational ‘ought’ as requiring one to perform the action that maximizes expected value. But the credences from the previous subsection provide only half of the resources needed by an expected value maximization framework. This subsection sketches how the other half might be filled out, and the recommendations from the rational ‘ought’ that result.

Let’s confine ourselves to a specific ethical theory for the sake of illustration, and a simple explanation of how vagueness arises for this theory. Suppose one is certain that a simple Utilitarianism is correct, and that there is a sorites series for ‘wrong’ on this theory because it is vague, in some cases, whether the amount of net utility produced by pressing the button and killing the creature in front of you is greater than the amount of net utility produced by refraining from pressing the button and letting the human die. According to the Epistemicist, then, one cannot know whether in these cases pressing the button produces more net utility than refraining from pressing it. There are then several aspects to the decision problem of what one rationally ought to do in such a borderline case.

First, in each borderline case $b_x$, there are two available actions: pressing the button and not pressing the button. There are also two possibilities given what one knows in each borderline case: that pressing the button would be wrong, and that pressing the button would not be wrong (these are represented as $b_x$ and $\neg b_x$, respectively). This gives a total of four possible outcomes, one for each quadrant below:

<table>
<thead>
<tr>
<th></th>
<th>$b_x$</th>
<th>$\neg b_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>press</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no press</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Second, given our assumption of a simple Utilitarianism, the amount of moral value contained in each possible outcome where pressing the button is wrong is easy to calculate.

(i) **Press and Wrong**: If one presses the button and it is wrong to press the button, then pressing the button produces less net happiness than the alternative; one’s action has significant negative moral value.
(ii) **No press and Wrong**: If one refrains from pressing the button when pressing the button would be wrong, then one’s action produces more net happiness than the alternative; one’s action has significant positive moral value.

Turn next to the two outcomes where pressing the button is not wrong. There are two ways for pressing the button not to be wrong: on one, it is because pressing the button produces the same amount of net happiness as not pressing (and hence both actions are permissible); on the other, it is because pressing the button produces more net happiness (and hence is obligatory). On the simple Utilitarianism we have assumed there will definitely some cases of the first kind: if it is a borderline case where pressing is not wrong, then there must be some nearby cases where neither available action produces more net happiness than the other; otherwise all of the nearby cases would be cases in which pressing is not wrong, and it wouldn’t be vague whether pressing is wrong. Perhaps there are borderline cases of the second kind too; for simplicity, I factor these into the value produced by taking an action when pressing the button is not wrong by combining the net utility of an outcome of an action when pressing the button is not wrong. This brings us to the second two possible outcomes:

(iii) **Press and Not wrong**: If one presses the button when it is not wrong to press the button, then on average one’s action has a small amount of positive value. In at least some of the cases where pressing isn’t wrong, one’s action is of neutral moral value; this is because in these cases pressing and not pressing produce the same net happiness. There might also be cases where pressing the button is not wrong because it produces more net happiness; in these cases, pressing has significant positive value. Averaging out the moral value in these cases, one’s action has small positive moral value.

(iv) **No press and Not wrong**: If one refrains from pressing the button when it is not wrong to press the button, then on average one’s action has a small amount of negative value. In at least some of the cases where pressing isn’t wrong, one’s action is of neutral moral value; this is because in these cases pressing and not pressing produce the same net happiness. There might also be cases where pressing the button is not wrong because pressing produces more net happiness; in these cases, refraining from pressing has significant negative value. Averaging out the moral value in these cases, one’s action has small negative moral value.

All of this can be summarised in a table like the following for each borderline case $b_x$ (as before, $b_x$ is the proposition that pressing the button in $b_x$ is wrong; $\neg b_x$ is its negation):

<table>
<thead>
<tr>
<th></th>
<th>$b_x$</th>
<th>$\neg b_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>press</td>
<td>large -</td>
<td>small +</td>
</tr>
<tr>
<td>no press</td>
<td>large +</td>
<td>small -</td>
</tr>
</tbody>
</table>
With this decision table in hand, we have a rough heuristic for answering the question of what one rationally ought to do in a case of ethical vagueness. One ought to maximize expected moral value, where the expected moral value of pressing the button and not pressing the button in a borderline case \( b_x \) are calculated by multiplying the values of the possible outcomes of an action by their likelihood.\(^1\)

Even without being any more specific than ‘large positive value’, ‘small negative value’, etc. in specifying the values contained in the various possible outcomes, we can note an interesting feature of the decision problem. Given INCREASING CREDENCES and UNIFORMITY, there will be some cases \( b_x \) where the likelihood on what one knows that \( b_x \) is true is around .5.\(^2\) In these cases, because of the values of (not) pressing the button in the various possible outcomes, one rationally ought to refrain from pressing the button. Thus in the cases where the likelihood on what one knows that pressing the button is wrong is .5 or greater, there is a determinate fact about what one rationally ought to do: one ought to refrain from pressing the button.

To illustrate, take the case where the likelihood on what one knows that pressing the button is wrong is .5. Why one ought to refrain from pressing the button is straightforward: in these cases, there is an equal chance that one produces a large negative outcome or a small positive outcome if one presses the button. Conversely, there is an equal chance that one produces a large positive outcome or a small negative outcome if one doesn’t press the button. Thus refraining from pressing has higher expected moral value.\(^3\)

\(^1\)More formally, this goes as follows:

\[
\begin{align*}
\text{EMV(press)} &= Cr(b_x)V(\text{press} \land b_x) + Cr(\neg b_x)V(\text{press} \land \neg b_x) \\
&= Cr(c_x)(\text{Large} -) + Cr(\neg b_x)(\text{Small} +) \\
\text{EMV(no press)} &= Cr(b_x)V(\text{no press} \land b_x) + Cr(\neg b_x)V(\text{no press} \land \neg b_x) \\
&= Cr(b_x)(\text{Large} +) + Cr(\neg b_x)(\text{Small} -)
\end{align*}
\]

\(^2\)Given INCREASING CREDENCES one’s credence that one has crossed the cut-off point for wrongness should increase as cases get closer to cases where it is determinately wrong to press the button. Presumably, for the first borderline case \( b_1 \), one should have a very low credence (say, \( \approx .01 \)) that \( b_1 \) is the cut-off point for wrongness. For the last borderline case \( b_n \), one should have a very high credence (say, \( \approx .99 \)) that it is past the cut-off point for wrongness. Given UNIFORMITY, there must be some case in between where one’s credence that it is past the cut-off point for wrongness should be roughly .5.

\(^3\)For concreteness: suppose the value of pressing if pressing is wrong is -10, and the value of pressing if pressing isn’t wrong averages out to +2. And conversely, suppose the upside to not pressing if pressing is wrong is +10, while the disvalue to not pressing if pressing isn’t wrong averages out to -2. The the expected moral values (assuming a case where credence .5 that pressing is wrong is appropriate) are as follows:

\[
\begin{align*}
\text{EMV(pressing)} &= .5(-10) + .5(2) \\
&= -4 \\
\text{EMV(not pressing)} &= .5(10) + .5(-2) \\
&= 4
\end{align*}
\]
It is also straightforward to work out how to the rational ‘ought’ applies to other borderline cases where one’s credence that pressing the button is less than .5, though the verdict of the rational ‘ought’ in any particular case will depend on the details of each case, which we haven’t supplied here. In some cases, the expected value of refraining from pressing is higher even if one is somewhat confident that pressing isn’t wrong. This doesn’t apply to every borderline case, however. If one is sufficiently confident that pressing isn’t wrong, then pressing will have higher expected moral value. One rationally ought to press the button in these cases—though such cases will be in the minority, given increasing credences and uniformity, and the assumption that the total number of borderline cases is finite. Let’s summarise this result by saying that one ought rationally to be somewhat conservative in cases of vague ethical wrongness.\textsuperscript{17}

4 Interlude: complications and connections

Before exploring the upshot of the expected moral value maximization approach to ethical vagueness, a few points are in order. Each of these should be explored in much greater detail, but owing to considerations of space I will simply note them here and then move on.

\textsuperscript{17}The same result applies on some assumptions where it can be known that in the vague cases there is no gap between actions that are ethically wrong and ethically required (that is, it is known that there are no “merely permissible” actions). This kind of situation arises if we assume Utilitarianism but take the source of ethical vagueness to be vagueness in whether certain creatures can experience happiness. There will still be borderline cases of wrongness, for instance those from the modified sorites series in §1: in each, one can kill 1,000,000 creatures in order to save an adult human, but it is vague whether the 1,000,000 creatures in each can experience happiness. Given Epistemicism, each borderline case is a case where one can’t know whether the 1,000,000 creatures can experience happiness, and hence can’t know whether pressing the button and killing the 1,000,000 creatures is wrong. For the same reasons as before, though, for each borderline case there is a certain likelihood that the creatures in that case can experience happiness, given what one knows.

In addition, decision table will have the same structure as before, and hence the “somewhat conservative” recommendation of the rational ‘ought’ will still apply. To see this, we can again divide a borderline case \( b_x \) into four possible outcomes. In the outcome where one presses the button and pressing is wrong, one saves the life of a human (say, a gain of +1,000 units of happiness) but kills 1,000,000 happiness-experiencing creatures (say, a loss of -1,000,000 units of happiness). When one presses when pressing isn’t wrong, one gains the life of the human (+1,000) and loses nothing. When one refrains from pressing when pressing isn’t wrong, one loses the life of the human (-1,000) and gains nothing. And when one refrains from pressing when pressing is wrong, one loses the human (-1,000) but saves the 1,000,000 creatures (+1,000,000). The decision table then looks as follows:

\[
\begin{array}{c|cc}
\text{press} & b_x & \neg b_x \\
\hline
\text{press} & -999,000 & 1,000 \\
\text{no press} & 999,000 & -1,000 \\
\end{array}
\]

Thus the decision table has the same structure as before, and the somewhat conservative conclusion follows for the same reason. Thanks to Elizabeth Barnes for discussion here.
4.1 Rankings of value

One thing that could be asked is how non-Utilitarian ethical theories yield a ranking of the moral value contained in outcomes. Obviously the ranking won’t be as straightforward once we move away from the assumption of a simple Utilitarianism. There is however some precedent for determining such a ranking; for starters one might look to the work in “consequentializers” such as Dreier (1993), Louise (2004), and Portmore (2007) as they derive something like the “moral value” contained in each outcome (relative to a person, time, etc.) for standard deontological moral theories.

4.2 Vagueness in the rational ‘ought’

Another question one might ask is whether there is vagueness in what one rationally ought to do in borderline cases. If it is vague what one knows, or if the likelihoods given what one knows are vague, or if the moral value contained in various possible outcomes is vague, then there will be some vagueness in the notion of expected moral value. This is compatible with there being some vagueness in the rational ‘ought’ itself; the only claim here is that vagueness in the rational ‘ought’ doesn’t extend to every case of ethical vagueness. The rational ‘ought’ gives guidance in some (but not necessarily all) cases where the ethical ‘ought’ is silent owing to vagueness.

4.3 Non-Epistemicist approaches to vagueness

The expected value maximization approach to the rational ‘ought’ was implemented here under the assumption of an Epistemicist approach to vagueness. I will exploit this aspect of the implementation in §5 below in evaluating the relationship between a rational ‘ought’ and normative uncertainty. One issue worth keeping in the background, however, is an expected moral value approach can be sustained once the assumption of Epistemicism is given up and replaced with an alternative approach. Some authors assume that questions of decision-making under vagueness will be independent of questions about the nature of vagueness; see in particular Wasserman (2013) and Williams (2013).

One worry is that certain linguistic approaches to vagueness inspired by views on which vagueness is semantic underdetermination (Fine 1975) or semantic ambiguity (Lewis 1982). It is not clear that the underlying account of ethical vagueness as constituted by indecision with ‘wrong’, in particular, can support Uniformity and Increasing Credences. (Note that ‘the likelihood that pressing the button is wrong in b_n is i’ is false an all precisifications or disambiguations of ‘wrong’ for any borderline case b_n and any integer 0 < i < 1.) Nevertheless the attractive answer to the practical question may well be capturable on many approaches to vagueness, such as views which require that it be vague whether you believe that pressing that button is wrong (Dorr (2003)), views on which these cases require a special kind of attitude toward the proposition that pressing the
button is wrong (Field (2000), Schiffer (2000)), ontic views of vagueness (Barnes (2010)), and views which assign degrees of truth to vague sentences (Edgington (1997)). I do not wish to rule out these alternative approaches here, but will only note that alternatives may suffer a serious deficiency if they cannot mimic the Epistemicist on this matter.

5 An application: reasoning under broad ethical uncertainty

Recently there has been a burgeoning debate over rational responses to ethical uncertainty. The target question is the question of what one rationally ought to do when one is uncertain as to which ethical theory is true. The paradigm cases of interest in this literature are cases where, say, credence .9 that an ethical theory on which it is permissible to eat meat is true, but .1 credence that a theory on which eating meat is impermissible is true. In such cases Ross (2006), Sepielli (2009) have claimed that one rationally ought to do the action which maximizes expected moral value, where the expectation of the moral value produced by an action is determined by (i) one’s credences in competing moral theories, and (ii) the moral value of the action according to each theory one has some credence in. According to this rational ‘ought’, one is required (given appropriate value assignments) that one not eat meat when uncertain whether one is morally required to be a vegetarian. I will call this general picture about the relationship between moral uncertainty and rational action the EMV-maximization approach.

The kind of moral uncertainty in these paradigm cases is not identical to the ignorance in cases of ethical vagueness. The uncertainty in the paradigm cases is a product of one’s not knowing which moral theory is correct—i.e., one might not know whether hedonistic utilitarianism or some form of deontology is correct. In such cases it is prima facie compelling to follow Sepielli and Ross and accept a rational ‘ought’ which requires one to hedge one’s bets in response to the fact that, for all one knows, a moral theory which prohibits or requires certain actions is true. (For instance, since for all one knows a moral theory which prohibits eating meat is true, one rationally ought to hedge one’s bets and not eat meat.) Uncertainty owing to vagueness is not identical to this moral uncertainty—unlike in the paradigm cases from Sepielli and Ross, in cases of vagueness one can know which moral theory is true and yet still not know what one ethically ought to do. Sepielli and Ross are in the first instance concerned to address the distinct question of how to act only when one doesn’t know which moral theory is correct. Call this ignorance broad moral uncertainty.

Broad moral uncertainty is to be contrasted with narrow moral uncertainty, the kind of ignorance one can be faced with even when one knows the correct moral theory. As we have seen in §1, any moral theory will admit of cases of vagueness. In these cases, even subjects who know which ethical theory is correct, and who are not faced with any broad moral uncertainty, are still subject to some kind of uncertainty about what to do. The previous sections have outlined an attractive
approach to the rational ‘ought’ in cases of narrow moral uncertainty.

While broad and narrow uncertainty are not strictly speaking the same phenomenon, they nonetheless share certain structural features. One might think of each of the claims that wrongness is \textit{Wrong}_1, that it is \textit{Wrong}_2, that it is \textit{Wrong}_3, etc. as each constituting a distinct ‘theory’ of wrongness under a certain guise. Of course if one is not subject to broad moral uncertainty, one will know which theory is correct under some guise, and hence will know (under some guise) which of these theories is the right one. But it is natural to think that, under the \textit{Wrong}_1 guise, \textit{Wrong}_2 guise, etc., one cannot know which of these theories is correct. The existence of ethical vagueness means that a certain kind of ignorance is unavoidable, even if broad moral uncertainty can be eliminated.

Proponents of a rational ‘ought’ which requires EMV-maximization in response to broad moral uncertainty have faced a number of criticisms. Typically, however, these critics do not distinguish between broad uncertainty and narrow uncertainty, and as a result some of their criticisms are ipso facto criticisms of a rational ‘ought’ which requires maximizing expected moral value in response to narrow moral uncertainty. But the latter is a very attractive as an answer to the practical question of what to do in cases of ethical vagueness. Any criticism entailing that there is no rational ‘ought’ which requires EMV-maximization in cases of narrow uncertainty might plausibly be accused of overreaching.

I will focus on a few examples of overreaching criticisms below. But it should be emphasized here that not every criticism of expected moral value maximization in response to broad uncertainty risks overreaching. For instance what I will say below does nothing to defend the EMV-maximization accounts of Sepielli and Ross from the objection that they cannot account for the distinction between obligatory and supererogatory actions (see Lockhart (2000) and Hedden (forthcoming) for more elaboration). Nor will I talk about the problem of intertheoretic value comparisons (see Sepielli (2014) and Hedden (forthcoming) for more on this issue). Instead I will focus only on one broad family of criticisms, namely those which traffic in the distinction between ‘de dicto’ and ‘de re’ concern for morality. Several authors in the literature claim that in cases of broad uncertainty, a rational ‘ought’ requiring EMV-maximization thereby requires agents to act out of de dicto concern for morality. And they have concluded that such approaches are on these grounds objectionable. I will claim below that these objections are much less compelling when a rational ‘ought’ requiring EMV-maximization in cases of narrow moral uncertainty is in the picture.

5.1 Concern de dicto and de re: preliminary observations

Several authors have alleged that acting so as to EMV-maximization in cases of broad uncertainty betrays a kind of vice on the part of the actor; roughly this is the vice of caring about avoiding doing things which have the property of being morally wrong—whichever property that might turn out to be—rather than caring about avoiding actions in virtue of their wrong-making features. According to
these authors, the expected moral value maximizer can be faulted for, in the words of Smith (1994), caring about morality de dicto, and not caring about morality de re.

I will focus on two different versions of this objection below. Brian Weatherson (2014) uses an analogy with a decision-making procedure one may employ when deliberating about one’s own welfare. Take someone who makes decisions out of concern for their own welfare as such, rather than out of concern for the things that constitute welfare: for instance, suppose this person is someone who is unsure as to whether aesthetic appreciation was among the things that promoted their welfare, and then treats the potential for aesthetic appreciation as a reason to go to art galleries. (We might further imagine that this person weights this reason in accordance with their credence that aesthetic appreciation contributes to welfare, and hence begins to look like an expected welfare-maximizer.) This person is someone who is concerned for their own welfare as such. They don’t care for aesthetic appreciation itself, but only care about it as a potential contributor to their own welfare.

There is a similar distinction for moral obligation: someone who thought there was a small chance that vegetarianism is morally required would refrain from eating a steak only if she cared for avoiding moral wrongness as such, in the same way as someone who cared only about their welfare as such. And according to Weatherson, each of these motivations would be deficient:

The rational person values their health, happiness and friendships (and whatever goes into the actual list of things that constitute welfare). They don’t simply value their welfare, and desire to increase it. That’s why it would be perverse for Bruce to go to the gallery. He would only go if he had a strange motivation. And it is why it would be perverse for Martha to turn down the steak. To do so she would have to care about morality, whatever it is, not about the list of things that Smith rightly says a good person will care about. (Weatherson 2014: 152)

Elizabeth Harman (2014) raises the objection from de dicto concern without recourse to an analogy:

In my view, people who act wrongfully are blameworthy not in virtue of what their moral beliefs and credences were, but in virtue of what their non-moral beliefs and credences were, and how these influenced their choices. Someone who knows she is killing an innocent person, and does so anyway, does not care adequately to avoid killing the innocent. A view of blameworthiness that can undergird the claim that false moral views do not exculpate is this:

A person is blameworthy for her wrongful behavior just in case it resulted from her failure to care de re about what
is morally important—that is, from her failure to care adequately about the non-moral features of the world that in fact matter morally.

A person cares *de dicto* about morality if she wants *to be moral*. A person cares *de re* about morality if she wants *to keep her promises, to help the needy*, etc., and if keeping one’s promises, helping the needy, etc. are in fact morally important.

Thus on Harman’s view, as I will read it, a rational ‘ought’ licenses action out of *de dicto* concern if it is sensitive to both moral and non-moral beliefs and credences. Likewise a rational ‘ought’ licenses action out of *de re* concern for morality if it is sensitive to non-moral beliefs and credences only.\(^\text{18}\)

In what follows I will grant that the EMV-maximization approach to the rational ‘ought’ does require agents to be *motivated* or *desire* things in the way Weatherson and Harman describe. One might question this assumption, but I will not pursue this question further here.\(^\text{19}\) Instead I will grant for the sake of argument that *de dicto* motivations must be present, and suggest that the existence of narrow moral uncertainty shows that not to be as objectionable as Weatherson and Harman suppose. I explore the case for this claim in more detail in the following subsections.

5.2 Lessons from narrow uncertainty: Weatherson on the analogy with welfare

The argument against and EMV-maximizing rational ‘ought’ in Weatherson (2014) is an argument from analogy. We begin with the plausible claim that certain kinds of actions out of concern for one’s welfare, which he calls actions out of a “*de dicto*” concern for one’s own welfare, are irrational. The relevant kind of irrational action can be grasped by comparing two claims about how the rational ‘ought’ interacts with an agent’s (call him ‘Bob’) uncertainty about how an available action will impact his welfare:

**B1** Bob’s welfare is such that it is irrational for him to do something that might undermine it for no compensating gain.

---

\(^{18}\)This reading treats the rational ‘ought’ as tied to considerations of blameworthiess, so that one rationally ought to \(\phi\) (in the relevant sense) iff one would not be blameworthy if one \(\phi\)-ed. Many philosophers are sympathetic to this kind of connection between rationality and blameworthiness, though see Hawthorne and Srinivasan (2013) for dissent.

\(^{19}\)For instance one might take a lesson from Railton (1984) to heart and treat the rational ‘ought’ as analogous to a “criterion of rightness”—for example, as analogous to a moral principle which requires maximizing utility. One can, with Railton, deny that such a criterion also functions as a “decision procedure”—that is, one can hold that hedonistic utilitarianism does not require agents to made decisions by calculating the total amount of pleasure produced by each available action. Similarly one might do what an EMV-maximization ‘ought’ requires, without deciding on the action because it is the action that maximizes expected moral value. Thus applying the Railtonian distinction between a criterion of rightness and a decision procedure to the rational ‘ought’ suggests considerations of appropriate wants or motivations are irrelevant to the viability of an EMV-maximization approach.
It is irrational for Bob to do something that might undermine his welfare, whatever that turns out to be, for no compensating gain.\(^{20}\)

Weatherson then glosses the differences between these two claims and evaluates them in the following passage:

If welfare turns out to be health, happiness and learning, then the first claim \([B1]\) says that it is irrational to risk undermining your health, happiness and learning for no compensating gain. And that is, I think, right. But the second claim \([B2]\) says that for any thing, if that thing might be welfare, and an action might undermine it, it is irrational to perform the action without a compensating gain. That’s a much stronger, and a much less plausible, claim.\(^{21}\)

The analogy with broad moral uncertainty is fairly straightforward: there are two related (though importantly different) claims involving the rational ‘ought’ in cases of moral uncertainty for an agent (call her ‘Sue’). Just as Bob can perform actions which undermine his welfare by bringing about states of affairs where his total welfare is lower than what it could have been, likewise Sue can undermine the satisfaction of her moral obligations by bringing about a state of affairs where her moral obligations have not been satisfied. The analogues for \([B1]\) and \([B2]\) when undermining of obligation-satisfaction is in play are as follows:

\[\text{S1}\] Sue’s moral obligations are such that it is irrational for her to do something which might undermine her satisfaction of them for nor compensating gain.

\[\text{S2}\] It is irrational for Sue to do something that might undermine her satisfaction of her moral obligations, whatever those turn out to be, for no compensating gain.

Again the same distinction applies: if one’s moral obligations are simply to avoid violations of autonomy, then the \([S1]\) says that if an action risks violating someone’s autonomy, then it irrational to perform it for no compensating gain. \([S2]\), by contrast, says that if it might be that Sue is obligated to maximize happiness, then an action that undermines her maximization of happiness is an action that it is irrational to perform without a compensating gain. \([S2]\), like \([B2]\), before, is much stronger. Is it also less plausible?

The analogy with Weathersonian de dicto concern for welfare certainly suggests that it is. However the existence of narrow moral uncertainty makes an alternative analogy available.

Take Zoe, who is faced with a case of ethical vagueness: there is an action (pressing the button) which is such that it is vague whether that action is wrong, and hence vague whether Zoe is morally obligated not to do it. As before, there are two related claims about Zoe which involve the rational ‘ought’:

\(^{20}\) Weatherson (2014: 150)
\(^{21}\) Weatherson (2014: 150)
Zoe’s moral obligations are such that it is irrational for her to do something which might undermine her satisfaction of them for no compensating gain.

It is irrational for Zoe to do something that might undermine her satisfaction of her moral obligations, whatever those turn out to be, for no compensating gain.

Suppose the actual cut-off point for wrongness corresponds to the cut-off point for the property $Wrong_2$. Then, pressing the button in the borderline case $b_2$ is not wrong but pressing in the borderline cases $b_3 \ldots b_n$ is wrong. \textbf{Z1} then amounts to: it is irrational for her to press the button, when pressing might instantiate $Wrong_2$, for no compensating gain. This seems plausible: if Zoe is uncertain whether pressing the button instantiates $Wrong_2$, then undermining an attempt to press the button would be irrational. But cases of narrow uncertainty are, as we noted earlier, compatible with the absence of this kind of uncertainty: Zoe’s uncertainty is perfectly compatible with her knowing whether pressing the button instantiates $Wrong_2$.

Owing to the vagueness in ‘wrong’, Zoe might know whether pressing the button in a case instantiates $Wrong_2$, but still not be in a position to know that wrongness is the property $Wrong_2$. In such a case, for all she knows, pressing in $b_2$ is wrong. So, owing to vagueness, pressing the button in $b_2$ might turn out to undermine her satisfaction of her moral obligations, whatever those turn out to be; this is because wrongness might turn out to be $Wrong_1$. Since pressing in $b_2$ is merely permissible, there is no compensating gain in not pressing in $b_2$. Thus if Zoe is in $b_2$, she is in a case where (i) wrongness might turn out to be $Wrong_1$, (ii) if it did, then pressing the button would undermine her satisfaction of her moral obligations, and (iii) there is no compensating gain to pressing the button.

The claim \textbf{Z2} above says it is irrational for Zoe to press the button in a case like this where (i)-(iii) obtain. And in this case, that seems right: if she were to press the button, then she would be acting as if she did know that wrongness is $Wrong_2$ and not $Wrong_1$, and ex hypothesi she is not in a position to know this. All she can know is that wrongness is one of $Wrong_1$, $Wrong_2$, $Wrong_3$, and so on. (Note the situation is not one where she simply does not know this because she fails to believe it, or it would be hard for her to come to know this.) It is very natural to think that rationality cannot require one to act is one has knowledge that one cannot have; and if this is the case then \textbf{Z2} is true.

The upshot is that sometimes it is irrational to do something that might undermine performance of your moral obligations, whatever those might be. I haven’t argued that the same must be true for \textit{broad} moral uncertainty. But the Weathersonian argument against EMV-maximization in response to broad moral uncertainty relies on an analogy with plausible claims about welfare maximization under uncertainty. What narrow moral uncertainty does is show that there are other analogies which are more friendly to the EMV-maximizer. The case against S2 should remain open.
5.2.1 Harman on factual uncertainty

Harman (2014) presents a version of the argument that action in accordance with a rational ‘ought’ that requires EMV-maximization results in objectionable behavior. At the center of her argument is a thesis called “Actualism”. The thesis is as follows (in what follows I will replace Harman’s use of ‘subjectively should’ with the expression ‘rationally ought’ in order to maintain continuity with previous sections):

A person’s moral beliefs and moral credences are usually irrelevant to how she (subjectively) should act. How a person (subjectively) should act usually depends solely on her non-moral beliefs and credences [...] according to Actualism, what a person (subjectively) should do depends crucially on what’s actually the true moral theory, and not on what people believe is the true moral theory.\(^{22}\)

The intended contrast is with the “Uncertaintist” view, which allows that, contra Actualism, moral credences are relevant to what one rationally ought to do.

We can think of the difference between Actualism and Uncertaintism by looking at the prescriptions they make for a pair of agents are alike in their non-moral credences and beliefs (and preferences), but who have wildly different credences in moral propositions. Take for instance the agents Carol and Daniel, who are non-moral cognitive duplicates: for every non-moral belief or credence Carol has, Daniel has it as well, and vice versa. According to Actualism, since what an agent rationally ought to do is determined only by (i) their credences in non-moral propositions, and (ii) the facts about what is morally required, Carol and Daniel rationally ought to do the same thing in the same situations.\(^{23}\) These agents are, by hypothesis, identical in every respect which the rational ‘ought’ is sensitive to. So their rational obligations, according to Actualism, are the same.

Uncertaintism, on the other hand, holds that there might be further differences between Carol and Daniel which crucial to what they rationally ought to do in a given situation. If, while remaining non-moral cognitive duplicates, they adopt distinct attitudes toward moral propositions, the rational ‘ought’ may tell the

---

\(^{22}\)Harman (2014: 5). I have omitted for the sake of simplicity the qualification that requires the occurrences of ‘usually’; Harman adds: “her moral beliefs and credences are relevant only insofar as they provide warrant for beliefs and credences about what her non-moral situation may be”. Since cases where moral credences support a certain credence in a non-moral proposition will not be at issue in the following discussion, I will ignore this component of the official statement of Actualism.

\(^{23}\)Note that Actualism does not require the same things of Carol and Daniel simply because they are non-moral cognitive duplicates. If their beliefs are silent on certain details of their external environment, it is consistent with their being duplicates in the relevant sense that Carol ought to press a certain button while Daniel ought not to. For if neither Carol nor Daniel have beliefs about the consequences of pressing the button, then their external environments might differ by being such that in Carol’s pressing the button would cause instant death to millions of people while in Daniel’s pressing would have more benign consequences. Whether this is a desirable feature of a rational ‘ought’ to have is a question I will leave unanswered here.
agents to act in different ways.\footnote{Some details in a full statement would need to be worked out for this characterization of the Uncertaintist view to be fully satisfactory. For instance typically agents who differ in credences in a moral proposition $m$ will also differ with respect to their non-moral credences about what their respective credences in $m$ are; agents are more likely to know what their own credence is than they are to know what their counterpart’s credence is. So difference in moral beliefs will typically make agents fail to be non-moral cognitive duplicates. Moreover agents who have different moral views will also typically have preferences over outcomes which in at least some cases where the outcomes in question are assessed differently by the relevant moral views. Presumably these differences in preference are irrelevant to the rational ‘ought’ when moral considerations are present, but preferences cannot in general be irrelevant to the rational ‘ought’ since it should be sensitive to an agent’s preferences in cases where no moral considerations are present. I will not attempt to iron out these details here, as the main thrust of Actualism is clear enough, and what I have to say below will not depend on how these issues are resolved.} For instance if Daniel is confident that eating meat is impermissible, he rationally ought to act as if the likelihood that eating meat is impermissible is high. He is being irrational if he goes to a restaurant and orders a steak for dinner. By contrast if Carol has no such beliefs, she will not be rationally required to act in the same way in Daniel’s circumstance where steak is on the menu.

Harman’s argument for Actualism rests on some subtle and interesting reflections on what is required for moral blameworthiness. (See in particular Harman (2014: §3).) I won’t go into these arguments here, as the question I am interested in here is one which Harman does not explicitly address: the question of what Actualism implies for cases of narrow moral uncertainty. It is clear that on the above gloss Actualism is inconsistent with an EMV-maximization approach to broad moral uncertainty. What someone who has broad moral uncertainty rationally ought to do is the same as what someone who has no broad moral uncertainty (but otherwise has the same preferences and non-moral credences) rationally ought to do.

As we might expect, Actualism has the same consequence for cases of narrow moral uncertainty: take an agent who, owing to vagueness, is uncertain which of $Wrong_1$, $Wrong_2$, $Wrong_3$, etc. is the property wrongness. Consequently she is uncertain as to whether it would be wrong for her to press the button in the borderline case $b_2$. She nonetheless is certain of all of the relevant non-moral facts in some sense: in our simplifying assumptions in §2, we restricted our attention to agents who know everything about the distribution of microphysical properties in their world. So she might be certain as to the non-moral character of the case she is in. Such certainty does not eliminate the possibility of ethical vagueness, and consequently our agent will not be able to know whether pressing the button in the borderline circumstance $b_2$ is wrong. According to Actualism, in such a case, what she rationally ought to do depends only on what in fact she is morally required to do. (Ex hypothesi there is no factual uncertainty, so the only relevant factor is what the agent morally ought to do.) If wrongness is the property $Wrong_2$, then in case $b_2$ it is not wrong for her to press the button. Hence, according to Actualism she is rationally permitted to press the button.
The (rational) high credence that wrongness is not the property \( \text{Wrong}_2 \) is irrelevant to her rational obligations according to Actualism. Someone who, through some kind of epistemic irresponsibility, was certain that wrongness is the property \( \text{Wrong}_2 \), would have all of the same rational obligations in \( b_2 \). Actualism requires someone who adopts the proper response to vagueness, by being uncertain about the wrongness of pressing the button in \( b_2 \), to act in the same way as the person who is irrationally certain about the wrongness of pressing the button.

This makes Actualism highly problematic. First, as point out above, the person who has adopted the true belief that wrongness is \( \text{Wrong}_2 \) is being epistemically irrational; they believe something they cannot know. They are plausibly \textit{acting} irrationally if they act in accordance with Actualism’s requirements. For not only does actualism allow them to press the button in \( b_2 \), it requires them to refrain from pressing the button in \( b_3 \), since pressing the button is wrong in \( b_3 \) on present assumptions. So Actualism implies that an agent is rationally required to do very different things in situations that are virtually indistinguishable from each other. It is implausible that rationality can require agents to act differently in these situations.

Second, according to Actualism, someone who adopts the attitudes toward borderline cases that are rational is rationally obligated to do the same thing as the first agent. (Since the agent can be a non-moral cognitive duplicate with the first, she will have all of the same non-moral beliefs and credences. If there are no differences in the situations in which they are placed, then the actual moral requirements on them are exactly the same. So Actualism says that they are rationally required to act in the same way.) But this person is doing better epistemically by adopting rational moral credences. Actualism requires for her to act in the same way someone who had different and irrational moral beliefs would act. This seems implausible.

Note that while these are implausible consequences of Actualism in cases of narrow moral uncertainty, analogous consequences are not a problem for the Actualist treatment of broad moral uncertainty. One might have false moral views, and Actualism says that these are irrelevant to what one rationally ought to do. But in a case of broad moral uncertainty, it is presumably at least possible for one to know the moral truths one is at present uncertain of. Even if one wrongly believes

---

25Such an agent would of course, on our present assumptions, be \textit{right} about which property wrongness is. But since she has no way of knowing this, she would violate some epistemic norms on belief-formation in coming to this belief.

26Of course on our assumptions the agent knows other non-moral facts which distinguish \( b_2 \) from \( b_3 \), since she knows facts about the microphysical differences that constitute them. But this knowledge will not support rationally treating the cases very differently. If she does not know that wrongness is \( \text{Wrong}_2 \) and not \( \text{Wrong}_3 \) (and ex hypothesi she does not), then it would be irrational for her to act differently in these cases simply in virtue of knowing that there are some microphysical differences in their makeup. This is so even if the agent has adopted through irrational means the true belief that wrongness is \( \text{Wrong}_2 \); what rationality requires is that she give up the belief and not that she treat nearly indiscriminable situations very differently.
that eating meat is permissible, one in principle could come to know that it is not. So Actualism does not obviously require one to act as if one is in an epistemic position which it is not possible for one to be in when dealing with cases of broad moral uncertainty.

The problems with narrow moral uncertainty for Weatherson and Harman are not unrelated. Both are are arguing for principles which rule out action out of de dicto concern for morality only. For Weatherson this is because a principle analogous to B1 but not B2 captures genuine requirements on rational action. For Harman it is because Actualism is true. But it is quite plausible that one’s actions in cases of narrow moral uncertainty should display a de dicto concern for morality—it is not possible for one to do any better epistemically and, as it were, to come to know morality de re. So any principles which constitute a blanket rejection of action out of de dicto concern will ipso facto overreach.

6 Conclusion: further issues and the way forward

The arguments in §5 do not establish that the EMV-maximizing approach to moral uncertainty, which is favored by Sepielli and Ross, is unobjectionable. Rather they only suggest that one way of arguing against the EMV-maximizing picture is incorrect. The rational ‘ought’ does not ignore moral uncertainty without exception. But this does not mean that the rational ‘ought’ is always sensitive to moral uncertainty. Just because one should hedge one’s bets when one is in a situation where one cannot know for vagueness-related reasons what one morally ought to do, it does not follow that a small credence in a radical and false moral theory has the same effect on one’s rational obligations.

Articulating a positive theory of the rational ‘ought’ is a further task beyond what I have attempted here. Some natural pictures suggest themselves: the rational ‘ought’ is sensitive only to unavoidable moral uncertainty, or perhaps only epistemically rational moral uncertainty. However these are just pictures and a full theory will need to give more precision about what the privileged kind of uncertainty amounts to.

Regardless of how these positive proposals are worked out, it is clear that the results will be of immense interest to some important practical questions. As §1 emphasized, vagueness that arises around the beginning and end of life, or in cases involving borderline humans (or borderline happiness-experiencers, borderline autonomous agents, etc.) will be structurally analogous to the idealized sorites series we have been interested in here. These are paradigms of cases where interesting practical questions can be asked about how to act. And they are extremely important questions, not merely questions about wild thought experiments, as the borderline cases are bound to be present in some real-life scenarios. It would be highly disappointing if we were forced to give the Actualist’s answer to these questions, or to dismiss countenancing the vagueness as fetishistic concern for morality as such. A rational ‘ought’ that recommends
maximizing expected moral value is a live and natural answer to these questions, 
regardless of whether it is applicable to the full range of moral uncertainty, as 
some enthusiasts have hoped.

References

627, 2010.

Citations to the 1996 edition published by Oxford University Press.

David Christensen. Epistemology of Disagreement: The Good News. The 

Cian Dorr. Vagueness without Ignorance. Philosophical Perspectives, 17:83–113, 
2003.

Tom Dougherty. Vague Value. Philosophy and Phenomenological Research, pages 
1–21, 2013.


Dorothy Edgington. Vagueness by Degrees. In Rosanna Keefe and Peter Smith, 

Hartry Field. Indeterminacy, Degree of Belief, and Excluded Middle. Noûs, 34(1): 


Elizabeth Harman. The Irrelevance of Moral Uncertainty. In Russ Shafer-Landau, 


John Hawthorne and Amia Srinivasan. Disagreement without transparency: 
Some bleak thoughts. In David Christensen and Jennifer Lackey, editors, The 

forthcoming.

Angelika Kratzer. What “Must” and “Can” Must and Can Mean. Linguistics and 


Miriam Schoenfield. A puzzle about vague value. MS.


