# Varieties of Moral Mistake

Zoë Johnson King, University of Southern California

*Abstract*. Some philosophers think that if someone acts wrongly while falsely believing that her act is permissible, this moral mistake cannot excuse her wrongdoing. And some think that this is because it is morally blameworthy to fail to appreciate the moral significance of non-moral facts of which one is aware, such that mistakenly believing that one's act is permissible when it is in fact wrong is itself morally blameworthy. Here I challenge the view that it is blameworthy to fail to appreciate the moral significance of non-moral facts of which one is aware. This view seems okay if we focus on examples of people mistakenly believing that their wrongful acts are permissible. But it is not remotely plausible when we consider other varieties of moral mistake – such as believing that one's action is required when it is in fact supererogatory, believing that one's action is wrong when it is in fact permissible, and believing false things about the moral properties of others' actions and of merely possible actions. The upshot is that those who maintain that moral mistakes cannot excuse are sent back to the drawing board; we need a new explanation of why this would be the case.

## 1.

The question at issue in this paper is: Which moral mistakes are morally blameworthy?

A *mistake* is when someone believes a proposition that is in fact false (mis-taking it to be true). And a *moral* mistake is a mistake about a proposition with moral content. This includes mistakes about actions' deontic statuses, about the goodness or badness or relative betterness and worseness of states of affairs, about the presence and strength of moral reasons or duties and the relationships of defeat between them, and about the extension of "thick" terms – i.e. about what is honest, fair, generous, demeaning, callous, sanctimonious, pusillanimous, and so on. So, for example, if someone grabs another person's coffee and runs out of the café while believing that this is *permissible* or *fair* or *polite* or *generous* or *in accordance with duty* or *supported by the overall balance of moral reasons* or *supported by a reason that is not defeated in the circumstances* or *conducive to a better state of affairs than all of the available alternatives* or *praiseworthy*, then she is morally mistaken. But to believe that the stuff she has stolen is *horchata* would not make her morally mistaken, since propositions about what is coffee and what is horchata are not moral propositions. Lastly, *blameworthiness* is a matter of something's deserving blame or being fittingly blamable, rather than of anyone's having all-things-considered most reason to form or express a blaming attitude toward the thing. (So we are not talking about punishment here, and we are not concerned with forward-looking reasons to blame or with considerations of standing.)

The question at issue in this paper has not yet received much direct philosophical attention. But a related question has received a great deal of attention: philosophers working on moral responsibility have

extensively discussed the question of whether it excuses wrongdoing, in whole or in part, for the wrongdoer to mistakenly believe that her action is permissible.

There is already a broad consensus that we should divide that sort of moral mistake – believing that one's action is permissible, when it is in fact wrong – into two sub-varieties. Some people are mistaken about whether their actions are wrong because they are mistaken about their actions' morally relevant features, while others are mistaken about whether their actions are wrong despite being well-aware of all of the morally relevant features. The latter agents are mistaken because they incorrectly appraise the moral significance of certain facts about their actions of which they are well-aware. As far as I know, everyone agrees that mistakes of the first variety sometimes at least partially excuse. This is particularly plausible with respect to actions' far-off consequences; if someone is innocently mistaken about whether her action has a certain far-off terrible consequence, when in fact it is wrong because of this terrible consequence, then she at least seems less blameworthy than she would have been had she known about it and steamrolled ahead anyway. What is more controversial, and more interesting, is whether the same courtesy extends to moral mistakes of the second variety. For example, if our agent had been aware of the terrible consequence but had mistakenly believed that it was not really all *that* terrible, and thus that her act was still permissible, then would her mistake still have served as an excuse?

Some authors in this literature find the following claim intuitively compelling:

> No Excuse: When someone acts wrongly, and they are aware of all of the facts that explain why their action is wrong, but they mistakenly believe that what they are doing is permissible because they incorrectly appraise the moral significance of those facts, this moral mistake is *no excuse*.

And, in the course of defending No Excuse, some authors also endorse this further claim:

> Blameworthy Mistake: If someone is aware of all of the facts that explain why their action is wrong, but they mistakenly believe that what they are doing is permissible because they incorrectly appraise the moral significance of those facts, this moral mistake is *itself morally blameworthy*.

I can feel the intuitive pull of No Excuse and Blameworthy Mistake, and I think that they might be true. But, if these claims are true, then the fact that they are true is surely not a brute fact. There must be some explanation of *why* it is always blameworthy to incorrectly appraise the moral significance of facts that explain why your action is wrong, such that your moral mistake cannot serve as even a partial excuse. This paper goes in search of a plausible explanation. My conclusion will be negative: I can't find one. But that is not the only point of the paper. I also want to emphasize a point of philosophical methodology. For there is a much broader view about the moral status of incorrect moral appraisals – a view that I will introduce in the next section – to which I think some defenders of No Excuse and Blameworthy Mistake are attracted. And the appeal of this broad view is understandable if we focus, as those in the literature on moral responsibility are wont to do, on agents who mistakenly believe that what they are doing is permissible when it is in fact wrong. But I think that it becomes obvious that the broad view is not remotely plausible when we begin to consider other varieties of moral mistake. So, I would like to encourage philosophers who are attracted to No Excuse and Blameworthy Mistake to address the question at issue in this paper. My point will be that, since incorrect moral appraisals are not in general morally blameworthy, we need to pay more attention to the question of *which* moral mistakes are morally

blameworthy in order to determine in a principled way whether No Excuse and Blameworthy Mistake could be true.

Here's a roadmap. In the next section, I survey some things philosophers have said in defense of No Excuse and Blameworthy Mistake, and I introduce the broad view to which I suspect that some defenders of these views are attracted, noting that the remarks offered in support of No Excuse and Blameworthy Mistake look as though they generalize to this broad view. In section 3, I discuss a smorgasbord of counterexamples to the broad view; taken together, these counterexamples strongly suggest that incorrect moral appraisals are not in general morally blameworthy. Section 4 discusses and rejects some strategies for mounting a principled defense of the broad view by invoking epistemic theses about when we have sufficient justification to believe moral truths. Section 5 considers four natural ways of attempting to weaken the broad view such that it avoids my suite of counterexamples but still supports No Excuse and Blameworthy Mistake, arguing that none of these weakened theses does the trick. Section 6 concludes. Along the way, I also clarify the relationship between No Excuse and Blameworthy Mistake and the popular "quality of will" approach to moral responsibility, criticize the too-quick distinction between moral and non-moral mistakes that is often drawn in the literature, identify two important choice-points for anyone who thinks that blameworthiness is a matter of how one relates to "what is in fact morally significant", argue that there is a puzzling asymmetry between mistakes about one's own actions and mistakes about others' actions, and call attention to a wide range of currently-overlooked varieties of moral mistake.

## 2.

Let's start by looking at what people say in defense of No Excuse and Blameworthy Mistake.

Some philosophers have focused primarily on No Excuse, but, in passing, made remarks that indicate that they are also sympathetic to Blameworthy Mistake. For example, here is Matthew Talbert (2013):

> Even if a wrongdoer is ignorant of the fact that her behavior is wrong, and even if this ignorance is not her fault, her actions may still express the contemptuous judgment that certain others do not merit consideration, that their interests do not matter, and that their objections can be overlooked… certain features of an unwitting wrongdoer's behavior can qualify her for blame regardless of whether she is culpable for her ignorance (p.234).

This passage is about blameworthy actions rather than blameworthy beliefs. The view is that actions expressing "contemptuous" judgments are blameworthy regardless of whether the agent realizes that the action is wrong (or that her judgment is contemptuous, presumably), and regardless of whether the agent is blameworthy for not realizing this. Talbert's view thus denies what is known as a "tracing" account of blame for unwitting wrongdoing, according to which someone can be blameworthy for acting wrongly when they do not realize that their action is wrong only if that epistemic state is itself blameworthy.[1] Since he is not a tracing theorist, Talbert is not primarily concerned with the question of when moral mistakes are themselves blameworthy. His remarks in this passage support No Excuse without bearing directly on Blameworthy Mistake. But Talbert does say that a mistaken moral judgment can be *contemptuous*. And later in the same paper he says the following:

[1] The most famous tracing views in the literature are Smith (1983), Zimmerman (1997), Rosen (2003, 2004), and Levy (2009). The now-widely-used phrase "unwitting wrongdoing" comes from Smith.

> [I]t is… appropriate for [a victim] to insist that his welfare is valuable and to see [the wrongdoer]'s rejection of this claim as a manifestation of ill will… If we agree with [the victim] that his welfare is valuable, and that a person of good will would see his welfare as a source of reasons, we should conclude that [the wrongdoer]'s judgment about the significance of [the victim]'s welfare reasonably elicits blaming responses on [the victim]'s part (p.240).

To say that something reasonably elicits blaming responses is to say that it is blameworthy. So, here Talbert is saying that a wrongdoer's moral judgment can be blameworthy if it is appropriate for us to see it as "a manifestation of ill will". Moreover, Talbert says that we should take moral judgments to manifest ill will when we disagree with their underlying assumptions about what is morally valuable, about what is a source of moral reasons, or in some other way about the moral significance of a certain set of considerations (e.g. considerations pertaining to someone's welfare). On this view, then, not only is an action blameworthy when it expresses a contemptuous judgment, but the contemptuous judgment is also itself blameworthy *qua* manifestation of ill will. This supports Blameworthy Mistake. And, more generally, Talbert's approach suggests that if we think that someone has incorrectly appraised the moral significance of certain facts about their actions of which they are well-aware, then we should regard these moral mistakes as manifestations of ill will and therefore blameworthy. (I say a little more on this below.)

Similarly, here are Clayton Littlejohn and Maria Alvarez (2017):

> Our opponents think that a subject might form a reasonable or rational false belief about her obligations on the basis of moral ignorance or mistaken belief. [They think that i]f the agent were to act on this belief or from this ignorance and act impermissibly, it would be inappropriate to blame her for her actions even if she failed to see what her obligations were only because of a mistaken moral belief or because of moral ignorance… We think the opposition is missing something important. In acting from moral ignorance, the subject's actions can manifest *de re* unresponsiveness. We think that it is entirely appropriate to hold people accountable for actions that manifest this kind of unresponsiveness. (pp.65-66)

This passage, too, is about blameworthy actions. The view is that if actions manifest *de re* unresponsiveness — that is, if they "manifest [a] failure to respond to the reasons that determine what an appropriate response would be" (Littlejohn forthcoming) or "show that you're willing to injure the interests that morality tells us to protect" (Littlejohn 2013, p.143)[2] — then these actions are blameworthy, regardless of what the agent thinks about the permissibility of what she is doing and regardless of the reasonableness or rationality of her moral mistake. It is "entirely appropriate" to blame people for failing to respond to what are in fact moral reasons, say Littlejohn and Alvarez, even if they not only *falsely* but *reasonably* or *rationally* believe that their action is morally permissible. This passage supports No Excuse,

[2] These are two glosses that Littlejohn gives on the phrase "*de re* unresponsiveness" in other work, which I include here just because Littlejohn and Alvarez do not explain their understanding of this phrase in the article block-quoted in the main text. The two glosses are not logically equivalent and I find the second one a bit obscure. But the underlying picture is clear enough: it is one according to which the true first-order moral theory identifies certain things as morally significant — i.e. as "the interests… to protect" and "the reasons that determine what a [morally] appropriate response would be" — and that actions are then blameworthy if they manifest insufficient concern for these morally significant things, the agent's moral beliefs notwithstanding.

then. But it does not address Blameworthy Mistake. However, later in the same paper Littlejohn and Alvarez offer further remarks that support both claims:

> Any argument for the conditional claim that we cannot blame people for acting on blameless moral beliefs is, inter alia, an argument that we can blame people for forming moral beliefs that they would be blamed for acting on… If [it is really true] that anything that can excuse a mistaken moral belief is something that would excuse actions that manifest *de re* unresponsiveness, perhaps this is an indication that sometimes nothing excuses the action or belief (p.74).

This second passage contends that moral mistakes can be blameworthy if acting on them would be blameworthy – that is to say (on the authors' view), if acting on them would manifest *de re* unresponsiveness. In such a case, Littlejohn and Alvarez say, perhaps nothing excuses the action *or the belief*. They do not spell this out, but I take it that the authors' view is that an action manifests *de re* unresponsiveness when the agent is aware of all of the considerations that in fact explain why her action is wrong and yet she performs it anyway — thus being unresponsive to the in-fact-morally-relevant considerations of which she is well-aware. And the authors' remarks in this second passage indicate that they would hold such an agent blameworthy not only for her action but also for her moral mistake (if she makes one). So this view supports Blameworthy Mistake.

Most remarks bearing on Blameworthy Mistake have been made in passing in the context of defenses of No Excuse, like the ones above. But one philosopher who has explicitly defended Blameworthy Mistake is Elizabeth Harman (2011, 2015, 2019, *ms*). Here is how she states her view (2015, pp.67-68, emphases original, paragraph breaks omitted):

> I hold that people who do morally wrong things while caught in the grip of false moral views are blameworthy for their actions and are *also* blameworthy for their beliefs. But they are not blameworthy for their actions *merely because* they are blameworthy for their beliefs; and they are not blameworthy merely for having *caused* themselves to behave in this way. Rather, they are blameworthy for both their actions and their beliefs for related reasons — because both their actions and their beliefs involve their failing to care adequately about what matters morally: Believing that one's wrong action is morally required involves caring inadequately about the features of one's action that make it morally wrong, because believing that an action is morally wrong on the basis of the features that make it wrong is a way of caring about those features. False moral belief is blameworthy.

Harman, like Talbert, endorses a non-tracing account of blameworthiness for unwitting wrongdoing, denying that people who act wrongly while believing that what they are doing is permissible are blameworthy merely because they are blameworthy for their beliefs. But Harman emphasizes that she holds that the mistaken moral beliefs are also themselves morally blameworthy — thereby endorsing not only No Excuse but also Blameworthy Mistake. And her account of why the beliefs are blameworthy is similar to Littlejohn and Alvarez's: it is that the beliefs "involve" the agent's failing to care adequately about what in fact matters morally. Harman thinks that correctly appraising an action's moral wrongness "is a way of caring" about the features of the action that make it wrong. And she thinks that if someone is aware of all of the features of her action that in fact make it wrong, but she incorrectly appraises those

features' moral significance and thus mistakenly believes that the action is permissible (or required[3]), this moral mistake "involves caring inadequately" about the things that in fact make the action wrong. It is the inadequate caring involved in this variety of moral mistake, according to Harman, that renders the mistakes blameworthy.

These three defenses of Blameworthy Mistake all look as though they will generalize quite broadly. Let's use the phrase *informed incorrect appraisals* to refer to moral mistakes that an agent makes when they believe a moral proposition that is in fact false, but they are aware of all of the considerations that in fact explain why it is false (and why its negation or contrapositive is true — e.g., why the action is wrong, when the agent mistakenly believes it to be permissible), and they believe the false moral proposition only because they incorrectly appraise those facts' moral significance. Talbert, Littlejohn and Alvarez, and Harman focus on informed incorrect appraisals *of the permissibility of an action that the agent is currently performing*. But their proposed explanations of why these informed incorrect appraisals are blameworthy look as though they generalize to *all* informed incorrect appraisals. If an informed incorrect appraisal of the considerations that make your action wrong involves caring inadequately about those considerations, then one would think that, in general, informed incorrect appraisals of morally significant considerations' moral significance involve caring inadequately about those considerations. Likewise, if it manifests *de re* unresponsiveness to fail to recognize moral reasons that in fact make your current action wrong, then one would think that, in general, it manifests *de re* unresponsiveness to fail to recognize moral reasons. (Indeed, Littlejohn and Alvarez suggest in a footnote that mistaken moral beliefs can themselves manifest *de re* unresponsiveness, attributing this view to Harman — see p.66, fn.5). And Talbert's formulation is already quite general: he holds that we should see appraisals of considerations' moral significance with which we disagree as manifestations of ill will.[4] These ill-will-manifesting informed incorrect appraisals might lead an agent to act wrongly, or they might not. But, if it is informed incorrect appraisals *themselves* that manifest ill will, then one would think that they do so whether or not the considerations incorrectly appraised are considerations that make the agent's action wrong.

This suggests a general view, to which I suspect that some defenders of No Excuse and Blameworthy Mistake are attracted:

> Universalism: All informed incorrect appraisals of facts' moral significance are themselves morally blameworthy.

According to Universalism, you relate to the things that are in fact morally significant in a blameworthy manner — you manifest ill will toward them, are *de re* unresponsive to them, fail to care adequately about

---

[3] N.B. In the quoted passage Harman says "required", but elsewhere she focuses on cases in which the agent believes that her action is merely permissible (e.g. 2011 pp. 452-54, 456-57). Note also that Harman has written a paper entitled "Morally Permissible Moral Mistakes" (2016), but her use of the term "moral mistake" in that paper is not my usage here: Harman is in that paper talking about *actions*, rather than mistaken moral *beliefs*.

[4] I am not sure exactly how general Talbert intends to be. He focuses on cases in which there is a direct victim, and in which the victim's well-being, rights, or interests are what the wrongdoer incorrectly appraises. So he may not intend for his view to generalize to all informed incorrect appraisals of morally significant things' moral significance, rather than just those having to do with human well-being, rights, and interests. But Talbert certainly does *not* think that only victims can appropriately judge that incorrect appraisals of their moral status are blameworthy (as the second quotation above — about the reactions of third parties — illustrates). And his emphasis is on the fact that blamers *disagree* with the moral appraisals of those they blame. He sees our taking the incorrect appraisals to manifest ill will as expressions of this first-order moral disagreement. So I see no reason why his view would not generalize to all informed moral appraisals with which we disagree.

them, etc. — when you incorrectly appraise their moral significance. This moral significance could amount to their making your current action wrong, or they could have a different sort of moral significance. Either way, you relate to them in a blameworthy manner when you are aware of them and yet you fail to draw the correct conclusion about whatever sort of moral significance they do in fact have. One way to express this view would be to say that it is *being out of touch with moral reality* that is fundamentally blameworthy, and that people can be out of touch with moral reality in either a conative or a cognitive way: we are out of touch with moral reality in a conative way if we are indifferent to what is in fact morally significant, we care about what is morally insignificant, or our relative degrees of concern do not align with the facts about their objects' relative moral significance, and we can also be out of touch with moral reality in a cognitive way – according to Universalism – if we incorrectly appraise the moral significance of facts of which we are well-aware. (Indeed, these moral mistakes may *constitute* inadequate caring, as Harman suggests.) On this view, it is as if the things that are in fact morally significant call out to us to intellectually appreciate their significance whenever they are present.

If Universalism were true, it would explain why Blameworthy Mistake is true; the narrower claim about failures to realize that one's current action is wrong would be true because it follows from a more general principle. Since No Excuse and Blameworthy Mistake stand in need of explanation, this gives Universalism some initial appeal. And, indeed, I suspect that quite a few philosophers who defend No Excuse and Blameworthy Mistake are indeed sympathetic to something like Universalism. I suspect, for instance, that it is this stronger principle that Harman has in mind at the end of the passage quoted above, when she makes the generic claim that "[all?] False moral belief is blameworthy".

Two quick points of clarification about Universalism before we move on. First, I have drawn the conative/cognitive comparison in order to emphasize that this view fits well with the "quality of will" tradition in philosophical thinking about moral responsibility. But quality-of-will theorists need not be Universalists. Some quality of will theorists (like Talbert and Harman) take an agent's evaluative judgments to be at least part of what the quality of her will consists in. But others focus on conative states and do not take cognitive states to have aretaic significance in and of themselves. Indeed, though Littlejohn and Alvarez (p.66, fn.5) and Harman (2011, p.460; forthcoming p.13, fn. 6) both cite Nomy Arpaly's work as an inspiration, Arpaly herself holds that the quality of our wills is determined by our intrinsic desires and that our moral beliefs are an epiphenomenon without direct aretaic relevance (see especially Arpaly and Schroeder 2013, sections 7.2 and 7.3, and compare Arpaly 2015, pp.151-155). So, we might accept that praiseworthiness and blameworthiness depend on what people care about without accepting Harman's further claims that forming correct moral beliefs *is* a way of caring and that forming mistaken moral beliefs *involves* caring inadequately – a point to which I will return in section 4. What I have argued is just that, if informed incorrect appraisals of facts' moral significance do indeed constitute inadequate caring, then it is hard to see why this would hold only of informed incorrect appraisals of facts whose particular variety of moral significance is to make one's current action wrong.

Second, some philosophers sympathetic to Universalism (or Blameworthy Mistake or No Excuse) might be tempted to speak of informed incorrect appraisals of *non-moral facts'* moral significance, rather than simply in terms of informed incorrect appraisals of *facts'* moral significance. That is because it is common in the literature to distinguish moral from non-moral mistakes, saying that everyone agrees that non-moral mistakes sometimes excuse but that there is disagreement over whether the same courtesy extends to moral mistakes. This is not how I have put things. I avoid this way of putting things because I think that it is confused. Not all mistakes about facts' moral significance are mistakes about *non-moral* facts' moral significance; people can be mistaken about the *further* moral significance of facts that are themselves moral. For example, just as someone might be mistaken about whether a white lie *constitutes*

*dishonesty* or *is bad (at all)*, similarly she might be mistaken about *how bad dishonesty is* or *whether the dishonesty involved in this white lie matters more than the value of the feelings spared* or *whether this dishonesty constitutes a rights-violation* or *whether this dishonesty still matters given that the person being lied to is dishonest all the time*. And so on. The first two of these mistakes are mistakes about the moral significance of some non-moral facts. But the others are all mistakes about the further moral significance of the fact that the act is dishonest, which is itself a moral fact. I assume that philosophers sympathetic to Universalism, Blameworthy Mistake, and/or No Excuse would want to include incorrect appraisals of moral facts' moral significance among their targets, since making mistakes about moral facts' moral significance is a way of being out of touch with moral reality just as much as making mistakes about non-moral facts' moral significance. What is important, then, is that the proposition about which the agent is mistaken is a moral proposition, rather than that the considerations whose moral significance she incorrectly appraises are non-moral.[5]

That was my best attempt at explaining why some philosophers might be attracted to Universalism. To be clear: it does not much matter if none of the above authors are in fact committed to Universalism precisely as I have stated it. What matters is that anyone who accepts Blameworthy Mistake must think that there is *some* explanation of why that thesis is true (since it cannot simply be a brute fact) and the above authors all offer remarks suggesting that they think that something in the vicinity of Universalism provides the explanation. If they reject Universalism, then it behooves them to defend Blameworthy Mistake on other grounds. But I am about to argue that Universalism is false and that no obvious alternative to it both (a) is intuitively plausible and (b) supports Blameworthy Mistake — and, by extension, No Excuse. It may be that some authors in the literature have simply assumed that something roughly along the lines of Universalism explains Blameworthy Mistake, without giving the matter much thought and without formulating the relevant principle precisely. To those authors, I offer the argument of this paper as a reason to think that they cannot simply help themselves to this assumption and must actually formulate a viable principle if they are to adequately defend Blameworthy Mistake (and No Excuse), as well as some reason to worry that doing so will be trickier than they have supposed.

Universalism is false. And it becomes clear that it is false once we turn our attention away from incorrect appraisals of one's own action's wrongness and consider other varieties of moral mistake. Showing this is my task for the next section.

## 3.

To repeat: a moral mistake is a mistake about a proposition with moral content. This includes mistakes about actions' deontic statuses, about the goodness or badness or relative betterness and worseness of states of affairs, about the presence and strength of moral reasons or duties and the relationships of defeat among them, and about the extension of "thick" terms. Clearly, there are a *lot* of moral propositions. So there are a lot of moral mistakes – and a lot of informed incorrect appraisals – that somebody could make. You might make a different sort of mistake about your action's deontic status besides believing it to be permissible when it is in fact wrong, such as believing it to be wrong when it is in fact permissible,

[5] Someone might be tempted by the view that all incorrect appraisals of facts' moral significance are *ultimately* incorrect appraisals of non-moral facts' moral significance, since any intermediate moral facts are themselves explained by non-moral facts. On this view, if X explains Y and Y explains Z, and if someone is mistaken about Z because she misunderstands the significance of Y, then she misunderstands the significance of X. I am not sure whether this is true, so I have phrased things in a manner that is consistent with it but does not require it.

believing it to be required when it is in fact *merely* permissible, or believing it to be merely permissible when it is in fact supererogatory. Or you might be mistaken about a moral feature of your action besides its deontic status, such as whether it is honest, fair, generous, demeaning, callous, sanctimonious, pusillanimous, and so on. Or you might be mistaken about whether someone else's action has any of these features, rather than your own action. Or you might be mistaken about whether an action that someone could perform (but doesn't) has these features, rather than an action that anyone actually performs. Universalism entails that moral mistakes of these other varieties are all blameworthy so long as the agent is aware of the morally relevant non-moral facts. And that, I will now argue, is highly implausible.

Let's focus initially on mistakes about your own action's deontic status. One way to be mistaken about your action's deontic status is to take it to be permissible when in fact it is wrong. But this is far from the only way to be mistaken about your action's deontic status. An action can be wrong, merely permissible, required, or supererogatory. And you can believe it to have any of these statuses. This yields twelve ways of being mistaken about your action's deontic status:

| | *You believe the act is…* | | | | |
|---|---|---|---|---|---|
| *Really, it is…* | | Wrong | Permissible | Required | Supererogatory |
| | Wrong | Correct | 1 | 2 | 3 |
| | Permissible | 4 | Correct | 5 | 6 |
| | Required | 7 | 8 | Correct | 9 |
| | Supererogatory | 10 | 11 | 12 | Correct |

The existing literature focuses on variety 1, and to a lesser extent on variety 2. But that leaves ten other ways of being mistaken about your action's deontic status that have gone more-or-less completely ignored. And it is highly implausible that these varieties of moral mistake are all blameworthy. Consider:

> *Modest*: During a genocide, Heidi goes to great lengths to hide members of the targeted racial group in her house and then smuggle them to safety. She saves hundreds of lives, incurring massive personal risks in doing so (since if her activities are discovered then she and her family will be killed). Later, when she is celebrated as a hero, Heidi demurs, saying that she just did what anybody in her position would have done.
>
> *Akratic:* Adam is a highly conscientious hedonist actualist act-utilitarian. He thinks that most of the time what he really *should* be doing is selling his belongings and using the money to buy malaria nets. But he can't bring himself to do that. He is thus convinced that almost everything he does is morally wrong. Nonetheless, Adam regularly goes far out of his way to benefit others around him in minor respects, and in everyday interactions he is extremely careful to always be polite and kind. He berates himself for not having the guts to do more, but comforts himself with the thought that he at least tries to bring about small net improvements in the world each day and takes care to minimize the harm that he causes.
>
> *Collegial:* Michaela goes to the first meeting of a salsa class at her local community college, enjoys it, gets along with the instructors, and appreciates their decision to volunteer their time. She then feels obligated to keep attending the class for the rest of the

> semester, despite the fact that it will continue to run even if she drops out. Michaela feels this way because she knows that these sorts of classes go best when they have a cohort of regular attendees, and she feels committed to being a regular attendee by dint of having attended and enjoyed the first class.

Depending on how we fill out the details, *Modest* can be a moral mistake of the eleventh or twelfth variety in the table – Heidi does something that is in fact supererogatory while believing either that it is required or that it is merely permissible. And this case is not at all far-fetched. On the contrary, the literature on supererogation is full of cases in which people who perform what seem to us to be heroic acts are inclined to meet praise with demurrals. As far as I am aware, no-one in this literature argues that these agents are blameworthy for being morally mistaken despite being well-aware of all of the facts that in fact explain why their actions are supererogatory. Instead, their underestimation of their actions' deontic statuses is widely seen as a form of modesty that is at worst morally neutral and that might even be praiseworthy.[6]

Depending on how we fill out the details, *Akratic* can be a moral mistake of the fourth, seventh, or tenth variety – Adam does something that is either permissible, required, or supererogatory (depending on what it is that he does) while believing that it is wrong. But his moral mistake does not seem blameworthy either. On the contrary, when I think of the person I know who most closely resembles this case, I gather that most of his friends find it touching and rather sweet that he berates himself so much while being so good. When he buys his colleague a coffee and brings it to their meeting because he knows that the colleague has a full day, or when he gets tongue-tied with expressions of condolence when a friend shares some bad news, his beneficiary's reaction is rarely to think that he is blameworthy for believing that what he is doing is not good enough. And, again, this case is not particularly far-fetched (although there are not all that many conscientious hedonist actualist act-utilitarians). More broadly, when we talk to guilt-inclined friends who are worried about offending or upsetting or otherwise wronging someone, but we think that what they did was totally fine, our reaction is usually to reassure them and not usually to judge them blameworthy for their moral mistake.

*Collegial* is a moral mistake of the fifth variety – Michaela believes that what she is doing is required when in fact it is merely permissible. And, again, this is a quotidian phenomenon: many people are quick to feel obligated to contribute to collective goods or to help individuals or causes with which they have only a passing involvement, in virtue of this involvement, where what is at stake is in fact not even sufficiently important for their behavior to count as supererogatory rather than merely permissible. The fact that such people think that the importance of what is at stake and their level of involvement with it are jointly sufficient to generate a moral requirement may strike us as sweet, sentimental, or just plain silly. But it again seems more touching than blameworthy.

It is easy to come up with cases like this. Happily, people who hold themselves to high moral standards – at least some of the time – are everywhere. Many such people are frequently mistaken about their acts' deontic statuses, "seeing" requirements and prohibitions where in reality there are none. Moreover, these

[6] Driver (1989) argues that modesty *requires* underestimating the moral significance of one's acts and character traits, calling it a 'virtue of ignorance'. If this is correct, then failing to appreciate the moral significance of non-moral facts of which one is aware is sometimes praiseworthy. Driver's view has fallen out of fashion, with later authors arguing that modesty is consistent with accurately estimating one's acts' and character traits' moral significance (e.g. Flanagan 1990, Brennan 2007, Bommarito 2013). But even these authors do not say that people who mistakenly underestimate their acts' deontic status would be blameworthy – they just say that modesty does not require underestimation.

moral mistakes frequently constitute informed incorrect appraisals in the sense introduced earlier: agents like Heidi, Adam, and Michaela are usually well-aware of all of the facts that in fact explain why their actions have different deontic statuses to those that they believe them to have. Heidi is well-aware of what she is doing and of the risks involved; Adam understands full well that he is making small net improvements in the world; Michaela is aware of all of the morally relevant facts about her relationship to the salsa class and its instructors. These people are mistaken about their actions' deontic statuses only because they are mistaken about the moral significance of facts of which they are aware. And yet it does not seem, intuitively, as though their moral mistakes are blameworthy.

Cases like this put pressure on Universalism. For they indicate that, although Universalism might look appealing when we think only of agents who believe that their current action is permissible when it is in fact wrong, it does not plausibly extend even to other informed incorrect appraisals of the deontic status of an action that the agent is presently performing. If it is true that one relates to the things that are in fact morally significant in a blameworthy manner — manifests ill will toward them, is *de re* unresponsive to them, fails to care adequately about them, etc. — whenever one incorrectly appraises their moral significance, then Heidi, Adam, and Michaela all relate to what is in fact morally significant in a blameworthy manner. But that is intuitively not the case. These agents' informed incorrect appraisals just don't seem blameworthy. And the familiar features of the cases suggest that the same will intuitively hold of a wide range of other, similar moral mistakes.

This already looks bad for Universalism. But it gets worse. Consider:

> *Niche.* Ashiyr is considering the latest obscure variation on the trolley problem, involving several trolleys in a Rube Goldberg-like setup and complex relationships between the would-be switcher and various people tied to the tracks. He forms a belief about whether it would be permissible to flip the switch in this case. But he gets it wrong.
>
> *Minor*. Jūlija has promised to take care of her friend's basil plant while the friend is away. She is invited to stay at another friend's house for a couple of nights. She forms the belief that doing so would be mildly inconsiderate, in light of her promise, since it would mean that she cannot water the basil for a couple of nights. But she does not think that this inconsiderateness is sufficiently morally important to render it impermissible for her to stay at her friend's house. In fact, Jūlija has gotten one thing right and one thing wrong: going to her friend's house is indeed permissible, and is not even particularly inconsiderate, since she can water the basil right before leaving and right after returning.
>
> *Whatever*. Gyeong-Hui is watching the film *Me And You And Everyone We Know.* Observing the character Peter teaching his little brother Robby how to use chat rooms, she forms the belief that Peter is being a somewhat irresponsible older brother (but not that he is acting wrongly). Then she carries on enjoying the rest of the film.

In *Niche*, Ashiyr is morally mistaken about a merely possible action — not one that he is currently performing, nor even one that anybody will ever actually perform, but one that someone *could* perform. Cases like this are so far removed from real life that their interest is purely academic. When Ashiyr forms a false belief about the deontic status of switch-flipping in this particular iteration of the trolley problem, then, his moral mistake does not seem particularly blameworthy. And that is so even though Ashiyr is aware of all of the morally-relevant facts (by default, since the case is a thought-experiment, so Ashiyr can stipulate that the facts are as he chooses them to be), and is mistaken only because he incorrectly

appraises their moral significance. Being mistaken about a moral matter as obscure as this seems fairly innocuous.

In *Minor*, by contrast, Jūlija is morally mistaken about her actual action of going to stay at a friend's house for a couple of nights. But she is not mistaken about this action's deontic status. She is mistaken about whether it has a certain thick property: the property of being mildly inconsiderate. But this moral mistake doesn't seem blameworthy either. Granted, Jūlija is aware of all of the morally relevant facts that in fact explain why going to stay at her friend's house for a couple of nights is not particularly inconsiderate – facts about what is within the scope of her promise, how to care for basil, and so on – and she misidentifies her action as mildly inconsiderate because she incorrectly appraises these facts' moral significance. But this just doesn't seem like a big deal. Jūlija's moral mistake is so minor that it again seems perfectly innocuous.

*Whatever* combines the factors at play in *Niche* and *Minor*: Gyeong-Hui is momentarily morally mistaken about whether a merely possible action — one that occurs in a work of fiction — has a certain thick property. Universalism entails that Gyeong-Hui is blameworthy for this moral mistake. But that is a real stretch. Holding that it is always blameworthy to momentarily impute a thick property to the actions of a character in a film, however minor or inconsequential this property may be, when in fact those actions do not bear that property, just seems absurd. Of course, it is not as though Gyeong-Hui's moral mistake is positively praiseworthy. But "blameworthy" and "praiseworthy" are contraries, not contradictories. Some things are neither blameworthy nor praiseworthy – they are just *fine*. And great deal of minor moral mistakes seem intuitively to fall into the "just fine" category, with Gyeong-Hui's among them.

As well as providing further intuitive counterexamples, mistakes about merely possible actions and thick properties put some pressure on the philosophical picture underlying Universalism. This picture, recall, is one according to which we relate to what is morally significant in a blameworthy manner whenever we incorrectly appraise its moral significance. But it is unclear how this picture applies to varieties of moral mistake beyond the usual suspects, because the phrase "what is in fact morally significant" is not very precise. (This phrase is usually left unanalyzed in the literature, ostensibly in the interests of remaining neutral between first-order moral theories.) The existence conditions of "what is in fact morally significant" are unclear. So it is unclear whether, when someone makes a mistake about a thought-experiment like the one in *Niche*, she is mistaken about what is *actually* morally significant or only about what is *counterfactually* morally significant. This must be decided in order to determine how a view like Universalism applies to mistakes about merely possible actions. Furthermore, given that facts about thick properties can *reflect* the moral significance of other facts while themselves having *further* moral significance, it is unclear which of the facts in these explanatory chains count as "what is morally significant" and thus what we must appraise correctly in order to avoid blameworthiness. For illustration, compare someone who fails to recognize her action as dishonest but understands the moral significance of dishonesty with someone who recognizes her action as dishonest but fails to understand dishonesty's moral significance. Which of them relates to what is morally significant in a blameworthy manner? This, too, must be decided in order to determine how a view like Universalism applies to mistakes about anything besides non-moral facts (at the bottom of the explanatory chains) and facts about actions' deontic statuses (at the top of the chains) – in other words, to *very* many of the varieties of moral mistake.

To sum up: Universalism faces death by counterexample once we consider just a handful of the currently-underexplored varieties of moral mistake. It also becomes clear that proponents of such a view face a number of important choice-points that have not thus far been addressed.

So much the worse for Universalism.

4.

Before moving forward, let me briefly discuss some epistemological claims that Universalists might offer in defense of their view.[7] I have argued that Universalism faces a lot of counterexamples. But one might think that a principled argument can be offered for Universalism on epistemological grounds, in which case its proponents could shrug and bite the counterintuitive bullets. To be clear: I don't think that such an argument can be offered. But it will be instructive to see why not.

First, Universalists might invoke the idea that facts about moral significance are *a priori*.[8] While not all of ethics can be *a priori*, since a great many moral facts depend in large part on *a posteriori* empirical facts (about such things as which bottle contains poison and whether one's salsa class will continue to run even if one stops attending), lots of metaethicists have thought that facts *about the moral significance of other facts* are the sorts of facts that we discover through reflection alone. The a priority of facts about moral significance might be thought to provide a principled argument for Universalism independently of our intuitions about cases.

But spelling this argument out precisely will be a tricky business. Some epistemologists do hold that everyone always has sufficient justification to believe all *a priori* truths, which Universalists might think helps their case. But this epistemological claim is a claim about what we have justification for, whereas Universalism is a claim about blameworthiness. There is a gap here, between has-sufficient-justification-to-do to blameworthy-for-omitting, that would need to be bridged in any successful argument from the a priority of morality to Universalism. And it is not always blameworthy to fail to do what one has sufficient justification to do; for instance, if I have sufficient justification to go walk in the park and sufficient justification to continue to work on my paper, then I am not blameworthy for working on the paper just because I had sufficient justification to walk in the park but I didn't. Moreover, the epistemological claim is a claim about what we have *epistemic* justification for, whereas Universalism is a claim about *moral* blameworthiness. So there is another gap here, between the epistemic and the moral, which would again need to be bridged for the apriority of morality to be used in an argument for Universalism. And it is not remotely plausible that we are always *morally* blameworthy for failing to believe what we have sufficient *epistemic* justification to believe, or even for failing to believe *a priori* truths – witness our blameless failures to believe huge swathes of mathematical truths, for instance. So the a priority of a fact about the moral significance of some other facts, by itself, cannot secure agents' blameworthiness for being mistaken about it.

Instead of saying that facts about moral significance are *a priori*, Universalists might try saying that everyone – or at least most people, or normal people – gains enough evidence over the course of their lives to figure out all of the facts about moral significance – or at least the most important ones, or enough of them to correctly appraise the moral significance of everything that we think about. For example, they could say something like the following:

[7] Thanks to an anonymous referee for pushing me to explore this way of defending Universalism.
[8] See DePaul and Hicks (2021) for an overview of arguments for this view and a survey of related issues.

> [O]rdinary life provides enough evidence for the moral truth that people are never in a position to justifiably believe false moral claims. Note that I do not deny that testimony can provide *some reason* to believe false moral claims. I simply deny that people are ever in a situation in which their *total evidence* supports a false moral claim. The evidence that people possess for the moral truth comes in many forms. Some of that evidence comes from one's own experience… Some of one's evidence for the moral truth comes from one's own emotional reactions to how others treat oneself and how one treats other people. Some of one's evidence for the moral truth comes from explicit moral reasoning. The moral evidence that people possess is diverse; and the claim in my proposal is that it is a rich body of evidence. (Harman 2019, pp.179-180)

> In general, most adult human beings have the capacity and opportunity for moral reflection to achieve the required moral knowledge to guide them in most of their actions… we must admit that some people are very poorly equipped to [figure out moral truths], if at all, perhaps because they were exposed to extreme forms of abuse and brutality that destroyed or thoroughly distorted their capacity for moral thinking, or because they have limited intellectual capacities… Our response is that those people are not exculpated from their moral obligations by their moral ignorance; rather, they are not subjects of the obligations in the first place. (Littlejohn and Alvarez 2017, pp.79-80)

Notice that whether epistemological claims like these can support Universalism depends on their scope. The Harman quote has the right scope to support Universalism, since it is very strong: Harman's view is that incorrect moral appraisals are *never* epistemically justified. The scope of Littlejohn and Alvarez's position is harder to discern. They start by saying that "most adult human beings" are capable of achieving "the required moral knowledge to guide them in most of their actions", which is fairly weak. But then the only agents that Littlejohn and Alvarez concede are incapable of achieving this knowledge are those who are so "thoroughly distorted" or intellectually "limited" as to not be subject to moral obligations in the first place. This looks like the very strong view that the only people who lack sufficient capacity and opportunity for moral reflection to figure out the moral truths relevant to their actions are those who are incapable of moral agency *tout court* (and thus permanently exempted from blame). Such a strong epistemological claim does have the right scope to support Universalism.

The trouble for Universalists is that, as the strength of these epistemological claims increases, so too does their implausibility. Consider Harman's suggestion that ordinary life provides sufficient evidence for us to figure out all of the moral truths. Recall *Niche*. What in his ordinary life is supposed to provide Ashiyr with sufficient evidence to recognize his mistake? His experience and emotional reactions are of limited relevance, since the case is so far removed from ordinary life. And his explicit moral reasoning is what led him astray. It is hard to see why we should think that the School of Life provides Ashiyr with sufficient evidence to see the error of his ways – unless we simply insist that everyone always has sufficient justification to believe all facts about moral significance, which is just the a priority claim again. Similarly, consider the strong interpretation of Littlejohn and Alvarez's view, according to which all moral agents have sufficient capacity and opportunity for moral reflection to figure out the moral truths relevant to their actions. Recall *Collegial*, and consider the true proposition that Michaela is not required to continue attending the salsa class. Why should we insist that she has sufficient capacity and opportunity for moral reflection to figure this out? Perhaps Michaela is a mother-of-three and carer for ailing parents who works two jobs, as a cleaner and a support worker for students with intellectual disabilities, all of which leaves her particularly sensitive to others' interests and inclined to prioritize them over her own – which then explains why she feels obligated to keep going to the salsa class. If Michaela has all of these caring

responsibilities, then she will have precious little time to ruminate on the aptness of her high degree of moral sensitivity. She may have the *capacity* for moral reflection, then, but to say that she has the *opportunity* would be to misunderstand the demands that life places on people in precarious financial positions.

In short, Universalists who want to use epistemological claims to defend their views face a dilemma: these epistemological claims must be strong enough to apply to all informed incorrect appraisals but not so strong as to be manifestly implausible.

Furthermore, regardless of scope, claims about how often people have sufficient evidence and/or opportunity to figure out moral truths cannot support Universalism all by themselves – just as the a priority of facts about moral significance cannot support Universalism all by itself. The most that these other epistemological claims can do is to show that some number of people *could* figure out some number of moral truths – or perhaps that they could figure them out *justifiedly*.[9] And, to repeat, we are not usually morally blameworthy for failing to do anything and everything that we could do and would be epistemically justified in doing. Even for weaker epistemological claims, then, bridges from the epistemic to the moral and from would-be-justified-in-doing to blameworthy-for-omitting are needed to get to Universalism. Without these bridges, Universalism lacks a principled defense, leaving it vulnerable to counterexamples. And I just gave some counterexamples.

5.

Universalism looks too strong. But Universalism can be false while No Excuse and Blameworthy Mistake are still true; there may be some other explanation of why No Excuse and Blameworthy Mistake are true. (They are surely not brute facts, as we noted earlier.) In order to determine whether they are true, then, we must look for a plausible general principle about which moral mistakes are blameworthy that entails Blameworthy Mistake, and thereby supports No Excuse, but is not as strong as Universalism.

Here is one option:

> Know Thyself: All informed incorrect appraisals of the moral significance of facts *relevant to one's own actions* are blameworthy.

Harman sometimes endorses Know Thyself rather than Universalism. For instance, her earliest paper on this topic claims that "[w]e are morally obligated to believe the moral truths relevant to our actions (and

[9] I do think that the *a priority* of morality, or perhaps one of these weaker epistemological claims, helps Universalists to resist a possible objection to their view: the objection that "ought" implies "can" and that we therefore must not hold people blameworthy for moral mistakes that they could not possibly have failed to make. (See the related discussions of possibility and difficulty in Bradford 2017 and Guerrero 2017). If facts about other facts' moral significance are *a priori*, or if we always have sufficient evidence for them and sufficient opportunity to figure them out, then it seems unduly pessimistic to say that someone who is morally mistaken could not possibly have avoided her moral mistake – even if her evidence was misleading or her reasoning poor, there is still a sense in which she had everything she needed to figure out the moral truth and thus *could* have figured it out. But it is important to remember that this argument is a reply to a possible objection to Universalism, rather than a positive argument for Universalism. It may be true that "ought" implies "can", but nobody thinks that "can" implies "ought". So even if there is a sense in which any of us at any time *can* figure out all of the moral truths, this does nothing to show that we are blameworthy for being mistaken about them.

thus not to believe false moral claims relevant to our actions)" (2011, p.459). The "relevant to our actions" qualifier commits Harman to Know Thyself rather than Universalism, at least in this paper. And Know Thyself does entail Blameworthy Mistake, since facts that make one's current action wrong are surely relevant to one's actions. Moreover, if we are not only *epistemically justified* in believing all true propositions about moral significance that are relevant to our actions (as Harman also thinks, and as we discussed in the previous section) but also *morally obligated* to believe all such propositions and *morally prohibited* from believing any false ones, then we can certainly be blameworthy for mistakenly believing that what we are doing is morally permissible. On such a view, moral mistakes of this variety are themselves directly morally prohibited.

But Know Thyself is subject to most of the counterexamples that I gave in section 3. For the fact that one's action is wrong (when it is wrong) is far from the only moral truth that is "relevant" to one's action. On the contrary, the fact that one's action is supererogatory, merely permissible, or required (when it has one of these other deontic statuses) is equally relevant. And that's just for starters. Facts about our actions' many and varied thick properties, our reasons or duties and their relative strength and relationships of defeat, the value our actions could promote and the relative degrees of value of different states we could realize, and so on, are all relevant to our actions, whether they explain why the actions are wrong or have some other sort of moral significance. If Know Thyself is true, then, the moral mistakes of agents like Heidi, Adam, Michaela, and Jūlija are all blameworthy, since their moral mistakes are about facts relevant to their actions. But, as we have already seen, these varieties of moral mistake just don't seem blameworthy. So this view is a non-starter.

Here is another option:

> Quality Of Will Spade-Turn: All informed incorrect appraisals of facts' moral significance *that manifest ill will* are blameworthy.[10]

Since Universalism fits naturally with the quality-of-will approach to moral responsibility, those attracted to Universalism might want to just build this right in to the view. They might do this by explicitly restricting their view to informed incorrect appraisals that manifest ill will. And they might then respond to intuitive counterexamples of the sort that I have given by insisting that Heidi, Adam, Michaela, Ashiyr, Jūlija, and Gyeong-Hui's informed incorrect appraisals do not manifest ill will.

I don't think that this strategy can work. To see why, consider these examples:

> *Mean*: Héctor learns that, during a genocide, some people went to great lengths to hide members of the targeted racial group and then smuggle them to safety, incurring massive personal risks and saving hundreds of lives in so doing. When these people are celebrated as heroes, Héctor frowns, shrugs, and says that they just did what anybody in their position would have done.
>
> *Ungrateful*: Anushka is a highly diligent hedonist actualist act-utilitarian. She thinks that almost everything that the people around her do is morally wrong; most of the time, she thinks, they really should be selling all their belongings and using the money to buy malaria nets. Sometimes Anushka's friends go out of their way to benefit her in minor respects, such as by bringing her a coffee when she has a full day or being polite and

[10] Thanks to an anonymous referee for encouraging me to consider this view.

> kind in their interactions with her. She is polite in return, but privately reminds herself that what they are doing is wrong.
>
> *Judgey*: Judy learns that her neighbor Michaela went to the first meeting of a salsa class at their local community college, enjoyed it, got along with the instructors, and appreciated their decision to volunteer their time. Judy forms the belief that Michaela is now obligated to keep attending class for the rest of the semester — despite the fact that the class will continue to run even if she drops out. Judy knows that these sorts of classes only run well if they have regular attendees and feels that Michaela has committed to being one of those attendees by dint of having attended and enjoyed the first class.

*Mean, Ungrateful*, and *Judgey* are analogous to *Modest*, *Akratic*, and *Collegial* (respectively), except that in these cases the agents are mistaken about other people's actions rather than their own. But these moral mistakes seem much worse when they are mistakes about other people's actions than when they are mistakes about the agents' own actions. Heidi's mistake seems like a form of modesty, but Héctor's mistake cannot be seen in this way, since we cannot be modest about other people's actions. Instead, Héctor just seems mean-spirited. And something similar seems true of Anushka; while Adam's tendency to be harsh on himself can be touching, Anushka's tendency to be harsh on others – like Héctor's dismissive reaction to heroism – seems insufficiently appreciative of the good that the others are doing. Moreover, since some of the good acts that Anushka dismisses are favors done for her by her friends, her harshness sometimes seems like a form of ingratitude that might even be blameworthy. Lastly, while it seems okay for Michaela to take herself to be obligated to keep attending the salsa class when she is in fact under no such obligation, Judy's mistake seems excessively judgmental. So Michaela's over-sensitivity to the morally significant considerations at stake in her own circumstances looks much better than Judy's over-sensitivity to the morally significant considerations at stake in Michaela's circumstances. Yet Judy and Michaela's moral mistakes are about the exact same considerations and circumstances.

If Heidi, Adam, and Michaela's mistakes do not manifest ill will and are therefore blameless, then the same goes for Héctor, Anushka, and Judy's mistakes. That is because Heidi and Héctor, Adam and Anushka, and Michaela and Judy (respectively) each make *the exact same moral mistakes*. Heidi and Héctor believe the same false propositions, namely that Heidi just did what anybody in her position would have done and that her action is therefore either morally required or merely permissible (depending on how we earlier filled out the details of the case), but not supererogatory. Likewise for Adam and Anushka and for Michaela and Judy. So, if it is the making of a moral mistake *per se* that manifests ill will, then the agents in each of these pairs stand or fall together. Either their mistakes manifest ill will, in which case both agents are blameworthy, or their mistakes do not manifest ill will, in which case neither is blameworthy. But this leaves the proponent of Quality Of Will Spade-Turn in an uncomfortable position. For it does intuitively seem as though Héctor, Anushka, and Judy's mistakes manifest ill will. So it will not do to respond to apparent counterexamples to Universalism by simply insisting, on an *ad hoc* basis, that the moral mistakes in each apparent counterexample do not manifest ill will. For the very same mistakes in at least three of these counterexamples *do* seem to manifest ill will when made by other people. This contrast needs explaining; we must unturn the spade and keep digging.[11]

[11] One might say that Héctor, Anushka, and Judy's mistakes are blameworthy because they *underestimate* Heidi, Anushka's friends, and Michaela. But that cannot explain our differing intuitions between the first three cases and the second three cases, because it does not point to a difference between them. Recall: these agents make *the exact same mistakes*. If Héctor underestimates Heidi, then Heidi underestimates Heidi. If Héctor's mistakes involve some sort of ill will towards Heidi, then, so do Heidi's mistakes. And ditto for the other pairs of agents.

The contrast between *Mean, Ungrateful*, and *Judgey* and *Modest*, *Akratic*, and *Collegial* also hints at a further challenge for the idea that we are blameworthy for failures to appreciate facts' moral significance. This idea, coupled with the intuitive contrast between these six cases, suggests that Héctor, Anushka, and Judy's mistakes constitute failures to appreciate the moral significance of the facts about Heidi, Adam, and Michaela's circumstances, whereas Heidi, Adam, and Michaela's own mistakes do not constitute failures to appreciate these facts' moral significance. And that is perplexing, given that the agents in these pairs make exactly the same mistakes about exactly the same things. The only way that I can see for it to be true is for the morally significant things in Heidi, Adam, and Michaela's circumstances to be *more morally significant* for Héctor, Anushka, and Judy than they are for Heidi, Adam, and Michaela. But that is bizarre. On this view, the morally significant things do call out to us to intellectually appreciate their significance, but they issue this call somewhat selectively – and, in particular, they issue the call *less* strongly to the agents who are performing the very actions for which they are significant and *more* strongly to third-parties who are considering those actions from the sidelines. It is very unclear why the morally significant things would do that. So this contrast makes the sort of philosophical picture that underlies No Excuse, Blameworthy Mistake, and Universalism even more perplexing.

Here a proponent of Quality Of Will Spade-Turn might suggest that informed incorrect appraisals of the deontic status of Heidi, Adam, and Michaela's actions are *underlain by different attitudes* when they are made by Héctor, Anushka, and Judy than when they are made by Heidi, Adam, and Michaela themselves, and that this contrast explains why the mistakes in my first three examples involve no ill will whereas the mistakes in the second three do involve ill will. After all, beliefs with the same content can be underlain by different attitudes when they are held by different people; the belief that New York is superior to Los Angeles can be self-aggrandizing when held by New Yorkers but self-abasing when held by Angelenos, for instance.[12] And, as I've said, Heidi's mistake seems to amount to a laudable form of modesty and Adam and Michaela's mistakes seem sweetly conscientious, whereas Héctor's mistake seems mean-spirited, Anushka's seems ungrateful, and Judy's seems judgey. It is intuitively plausible that modest and touching mistakes manifest good will, whereas mean-spirited, ungrateful, and judgey mistakes manifest ill will. Moreover, a proponent of Quality Of Will Spade-Turn might emphasize that this is all entirely consistent with their view; their view is that all informed incorrect appraisals that manifest ill will are blameworthy, and, with the proviso that what determines whether any particular informed incorrect appraisal manifests good or ill will can be something other than its propositional content, such a view can readily accommodate our intuitions about all six examples.

Now, in fact I think that this is more-or-less the right thing to say about these examples. But notice that it gets us no closer to a defense of Blameworthy Mistake. Indeed, if anything, it gives us some reason to doubt that Blameworthy Mistake will turn out to be true. For Blameworthy Mistake is the thesis that a certain variety of moral mistake *always* manifests ill will, where that variety is picked out by the mistake's content: believing that one's current action is permissible when it is in fact wrong (and one is aware of all of the facts that explain why it is wrong). And the strategy that we are considering for a proponent of Quality Of Will Spade-Turn to accommodate my six examples is for them to offer the proviso that what determines whether a moral mistake manifests good or ill will can be something *other* than its content. Once they have added this proviso, it is unsatisfactory for them to insist without further argument that, nonetheless, there are certain contents such that moral mistakes with these contents *always* manifest ill will, regardless of what else is going on. In particular, once they have made explicit that whether a moral mistake manifests ill will usually depends on the attitudes underlying the mistake and not on its content,

[12] Thanks to an anonymous referee for suggesting this reply and offering a version of this example to illustrate it.

it is deeply unsatisfactory for them to then insist that, nonetheless, mistakenly believing that one's current action is permissible when it is in fact wrong *always* manifests ill will, regardless of the attitudes underlying the mistake (as must be the case for Blameworthy Mistake to turn out true). Here the spade-turning aspect of the view is particularly unhelpful. It would be good to have a positive argument as to why mistakes with *this* content always manifest ill will, while mistakes with other contents can vary in the quality of will that they manifest. But such an argument is precisely what the spade-turning view refuses to provide. The view is logically consistent, then, but by its very nature it is not a view for which a principled defense can be mounted. So it is not much use a *defense* of Blameworthy Mistake.

The lesson to draw here is that, in the present context, Quality Of Will Spade-Turn just pushes the bump under the rug. To defend Blameworthy Mistake, if one is a quality-of-will theorist, one must defend a specific account of *what it takes for a moral mistake to manifest ill will* that entails that it always manifests ill will to think that one's action is permissible when it is in fact wrong and one is aware of all of the facts that explain why it is wrong. Without this supplement, Quality Of Will Spade-Turn leaves open the possibility that some such moral mistakes do not manifest ill will and are blameless. And, in that case, Quality Of Will Spade-Turn does not support Blameworthy Mistake and No Excuse; it is consistent with these theses, but also with their negations, which does not provide any support for the theses.

Now, to her credit, Harman does propose a specific account of what it takes for a moral mistake to manifest ill will. Her account is as follows (forthcoming p.22):

> Inadequately caring about what is morally significant occurs if one forms a specific belief about the issue (even if one merely has an implicit belief) or if the issue is relevant to one's behavior. But the requirement to care adequately about what is morally significant does not require believing all moral truths (not even implicitly).

Here Harman is fending off the objection that, on her view, adequate caring requires moral omniscience. Her view, recall, is that beliefs and failures to believe are blameworthy when they involve inadequately caring about what is morally significant.[13] And Harman says that it does not constitute inadequate caring to simply fail to believe a true moral proposition, but it *does* constitute inadequate caring to form a mistaken belief about a moral matter — even a merely implicit belief — or to fail to believe a true moral proposition that is relevant to one's behavior. So this view of which moral mistakes manifest ill will does entail that all informed incorrect appraisals of the wrongness of one's current action involve inadequate caring, just as Blameworthy Mistake says. Indeed, this view is much stronger than Blameworthy Mistake. For it ascribes inadequate caring in all cases in which someone fails to believe a true moral proposition that is relevant to their behavior, whether they are morally mistaken about it or they simply lack any beliefs about it. And it ascribes inadequate caring in all cases in which someone "forms a specific belief", implicit or explicit, about a moral issue. That is, any moral issue whatsoever. (Notice that Harman's account is disjunctive; forming a "specific belief" – that is false, I presume – about a moral issue is on this account sufficient for inadequate caring, whether or not the issue is relevant to one's behavior.) So this view about which moral mistakes manifest ill will entails Universalism.

We already rejected Universalism, though. And the earlier counterexamples to Universalism tell equally against this view about when inadequate caring occurs. It just does not seem, intuitively, to be the case

[13] For present purposes I take "involves inadequately caring about what is morally significant" and "manifests ill will" to be equivalent, so that Harman's proposal is a way of developing Quality Of Will Spade-Turn (i.e. of turning the spade).

that agents like Heidi, Adam, Michaela, Ashiyr, Jūlija, and Gyeong-Hui care inadequately about what is in fact morally significant, though they do form specific beliefs about moral issues (most of which are relevant to their behavior) and they get the issues wrong. So Harman's view about inadequate caring looks too strong.

Something like this would be good if it could be defended:

> Must Identify Wrongness of Own Actions: All informed incorrect appraisals of the moral significance of facts *whose moral significance is to make one's own action wrong* are themselves morally blameworthy.

This is a weakening of Harman's thesis that we are morally obligated to believe the moral truths relevant to our actions. That claim is too strong, as we saw earlier. But perhaps philosophers who are attracted to Harman's thesis could retreat to the view that we are morally obligated to believe that actions we might perform *are wrong*, when in fact they are (or would be) wrong and we are aware of all of the facts that explain why they are (would be) wrong. This would avoid the counterexamples to Harman's thesis that I raised earlier. And it would perfectly capture the central cases with which proponents of No Excuse and Blameworthy Mistake are chiefly concerned, all of which involve agents' failures to recognize the wrongness of their own current actions.

However, the fact that Must Identify Wrongness of Own Actions so perfectly captures these cases should raise a suspicious eyebrow. This perfect mesh occurs because Must Identify Wrongness of Own Actions and Blameworthy Mistake are logically equivalent; the former is just a restatement of the latter, using the term "informed incorrect appraisals" that I introduced in this paper. But we are looking for a distinct principle or idea that can *explain* why Blameworthy Mistake would be true, not a restatement of that very claim using specialist terminology coined by its critics. So simply invoking Must Identify Wrongness of Own Actions to explain Blameworthy Mistake won't do.

Here is one idea.[14] We might say that, if it is wrong for somebody to perform a certain action, then she must have *decisive* moral reasons against performing that action. And, if she is aware of all of the facts that in fact explain why her action is wrong, then she must be aware of those decisive moral reasons – though not necessarily of their decisiveness. So, perhaps it is *decisiveness* that is the special sort of moral significance involved in making an action wrong, and perhaps it is moral decisiveness that it is blameworthy to fail to recognize when one is well-aware of all of the relevant facts. This proposal has some intuitive appeal; however exactly we are supposed to understand the metaphor of morally significant things' calling out to us to intellectually appreciate their significance (such that we care inadequately about them, are *de re* unresponsive to them, etc., when we fail to do so), it doesn't seem outlandish to think that this could apply just to the considerations whose moral significance is in fact decisive. Indeed, one might think that moral considerations that count decisively against performing an action *just are* considerations that count decisively against believing the action to be morally permissible.[15] If the considerations count decisively against both the action and the belief, then it is easy to see why both the action and the belief would be blameworthy.

[14] Thanks to an anonymous referee for suggesting this idea to me.

[15] Ideas along these lines have appeared in Littlejohn (2012, pp.206-209) and Kiesewetter (2016). See also the extensive discussion of decisiveness in Lord (2017).

This idea ties the fate of No Excuse and Blameworthy Mistake to the coherence of the concept of decisiveness, the plausibility of the idea that wrong actions are wrong because of the decisiveness of the reasons against them, and the plausibility of the idea that it is blameworthy to fail to intellectually recognize the decisiveness of one's reasons. The first two of these are moderately contentious metaethical notions that it would take us too far afield to explore in this paper, so for present purposes I simply grant that something in their vicinity might be correct. But I am much less sure that there could be something distinctively blameworthy about failures to recognize decisiveness. There is something odd about this idea, which we can grasp either through examples or by considering the nature of decisiveness itself. I will close by describing this oddness.

Many of the considerations that can make an action wrong, when they are part of a set of considerations that count decisively against performing it, still count against performing actions even when these actions are *not* wrong all-things-considered. Just as dishonesty can make an action wrong, for instance, it can also count against performing an action that is not wrong. Even an action that is morally required might be dishonest, in which case the dishonesty will still count against performing it. (This is what we sometimes call a moral "residue" or "remainder".) But if there is something uniquely blameworthy about failures to recognize moral decisiveness, then it is blameworthy to incorrectly appraise considerations' moral significance when they are part of a set that makes an action wrong, and not otherwise. That seems odd. I would have expected defenders of No Excuse and Blameworthy Mistake to say instead that one relates to each consideration that is in fact morally significant in a blameworthy manner – one cares inadequately about it, is *de re* unresponsive to it, et cetera – whenever one incorrectly appraises *its* moral significance, regardless of what else is going on. Consider:

> *Saved By The Bell.* Sascha's advisee is giving her first ever conference talk, via Zoom, and she asks Sascha to attend. Sascha believes that the facts about their relationship to their advisee, her stage of the program, and her request that they attend, provide no moral reason at all for them to attend the talk. They plan to read a book instead. However, on the morning of the student's talk, Sascha gets a phone call from their sister who is having a panic attack. So Sascha spends the morning helping their sister.

Here are my first-order assumptions about this case. First, Sascha's belief is false; the facts about their relationship to their advisee, her stage of the program, and her request that they attend, together provide a strong — though not indefeasible — moral reason for them to attend the talk, especially given that it's just on Zoom. Second, since there is nothing more pressing for Sascha to do and they are just going to read a book, this strong moral reason to attend the talk is initially decisive. So, at the time when they initially plan to skip the talk and read a book, they are planning to do something wrong. Third, Sascha's strong moral reason to attend the talk *ceases* to be decisive when they get the phone call from their sister. At this point, given their relationship to their sister and her urgent need, they are permitted to skip the talk in order to help her (as they do).

In *Saved By The Bell,* the considerations to which Sascha is unresponsive go from being decisive to being non-decisive. But does Sascha's incorrect appraisal of the moral significance of their relationship to their student, her stage of the program, and her request, get *better* after this shift in decisiveness has occurred? Intuitively, it seems not. For Sascha incorrectly appraises these considerations' moral significance consistently throughout the case — they believe throughout the case that they have no moral reason at all to attend the talk, but this is false throughout, even after the phone call. (At this point they still have a strong reason to attend, but it is outweighed by their sister's urgent need.) Intuitively, if Sascha's informed incorrect appraisal of the moral significance of this suite of facts about their student is a way of

relating blameworthily to the student — caring inadequately about her, being *de re* unresponsive to her, etc. — then Sascha relates blameworthily to their student in these ways just so long as they underestimate the moral significance of that suite of facts, regardless of what else is going on. So, it does not intuitively seem as though the blameworthiness of Sascha's informed incorrect appraisal varies with the decisiveness of the reasons whose moral significance they incorrectly appraise. Decisiveness, or the lack thereof, doesn't seem to make a difference.

Here's another way of putting the same point. Wrong actions are wrong in virtue of the *balance* of reasons working out against them. So, for a typical wrong action, there are a bunch of considerations counting against performing it (and perhaps some counting in favor of performing it) as well as further considerations counting in favor of and/or against performing each of the agent's alternatives. The decisiveness of the reasons against performing the action is determined by a complex interplay between all of the reasons at play in the agent's circumstances, which depends on their relative strengths and also on the normative relationships (such as defeat or lexical priority) that obtain between them. Decisiveness, then, is an *extrinsic property* of any given set of reasons. And it seems odd for the blameworthiness of someone's informed incorrect appraisal of a certain set of considerations to depend on an extrinsic property of those considerations. The blameworthiness seems to be between her and the considerations, so to speak, and to be none of these other facts' business. That is what cases like *Saved By The Bell* illustrate. Contrary to what one might expect from the literature on No Excuse and Blameworthy Mistake, then, these cases suggest that there is nothing particularly special about failures to recognize *wrongness*. The decisiveness of the incorrectly-appraised reasons does not seem to be doing explanatory work.

To sum up: cases like *Saved By The Bell* are interesting because they suggest that the blameworthiness of informed incorrect appraisals does not vary with the decisiveness of the considerations incorrectly appraised. Rather, it seems, the blameworthiness remains constant throughout the informed incorrect appraisal. But this intuition is worrying in the present context. That is because it looks as though this intuition will take us right back to Universalism; it suggests that all informed incorrect appraisals are blameworthy, regardless of the variety of moral significance that the incorrectly-appraised considerations happen to have all-things-considered at any particular point in time. And Universalism, as we have seen, looks far too strong.

I think that there is a genuine puzzle here. There *may* be a stable and intuitively plausible general principle about which moral mistakes are themselves morally blameworthy – that is, an answer to the question at issue in this paper. But it remains to be seen what this principle is. We have considered a wide range of natural candidates and found them all wanting. Without such a principle, though, it remains unclear whether No Excuse and Blameworthy Mistake are true.

6.

Where does this leave us?

In brief: if informed incorrect appraisals cannot excuse wrongdoing then there must be some explanation of why this is so. And if the explanation is that moral mistakes of this variety are always blameworthy then there must be some explanation of why *that* is so. Universalism offers a putative explanation – to which I suspect that many in the literature are attracted, though they are not always explicit about it. And Know Thyself, Quality Of Will Spade-Turn, and Must Identify Wrongness Of Own Actions offer refinements of this putative explanation. But none of these views are plausible. So, if we wish to say that

informed incorrect appraisals cannot excuse wrongdoing and because they are always themselves blameworthy, then we are sent back to the drawing board.

There may be some other explanation of why informed incorrect appraisals are always blameworthy. It remains to be seen what this explanation is. Alternatively, these varieties of moral mistake may *not* always be blameworthy. (That's what I think.) And, even if informed incorrect appraisals are sometimes blameless, it remains an open question whether they suffice to excuse wrongdoing. If a non-tracing view of blameworthiness for actions is true, then even blameless moral mistakes may not excuse wrongdoing. Alternatively, it might be that informed incorrect appraisals do sometimes excuse. I have not said enough in this paper to tell us which way to go — other than away from an assortment of overly-simplistic views and back to the drawing board. To determine whether No Excuse and Blameworthy Mistake could be true, we must roll up our sleeves and do the hard work of figuring out precisely which varieties of moral mistake are blameworthy.[16]

[16] Here's my gratitude footnote! I presented previous versions of this paper at the KCL Workshop on Mistakes, Ignorance and Blameworthiness in Fall 2020; at the LSE Choice Group in Spring 2021; at the UT Austin/St Andrews Workshop on Blame and Responsibility in Spring 2021; and at the APA Eastern Division meeting in Spring 2022. I am grateful to the organizers and participants of all of these sessions for the formative feedback that I received, and especially grateful to my APA commentators Emily McRae and Keshav Singh. I am also grateful to Rima Basu, Tom Dougherty, Amy Flowerree, Georgi Gardiner, Liz Jackson, Cat Saint-Croix, and Ralph Wedgwood for insightful discussion of earlier drafts and/or the ideas contained therein.

REFERENCES

Arpaly, Nomy & Timothy Schroeder (2013). *In Praise of Desire*. New York, USA: Oxford University Press.

Arpaly, Nomy (2015). "Huckleberry Finn Revisited: Inverse Akrasia and Moral Ignorance". In R. Clarke, M. McKenna, and A. Smith, eds., *The Nature of Moral Responsibility: New Essays*. New York, USA: Oxford University Press, pp.143-156.

Bommarito, Nicolas (2013). 'Modesty as a Virtue of Attention'. *Philosophical Review* 122 (1): 93-117.

Brennan, Jason (2007). 'Modesty Without Illusion'. *Philosophy and Phenomenological Research* 75 (1): 111-128.

Driver, Julia (1989). 'The Virtues of Ignorance'. *Journal of Philosophy* 86 (7): 373-384.

DePaul, Michael and Amelia Hicks, "A Priorism in Moral Epistemology", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/moral-epistemology-a-priori/>.

Flanagan, Owen (1990). 'Virtue and Ignorance'. *Journal of Philosophy* 87 (8): 420-428.

Harman, Elizabeth (2011). 'Does Moral Ignorance Exculpate?' *Ratio* 24 (4): 443-468.

Harman, Elizabeth (2015). "The Irrelevance of Moral Uncertainty". In R. Shafer-Landau, ed., *Oxford Studies in Metaethics*, Vol. 10. Oxford: Oxford University Press, pp.53-79.

Harman, Elizabeth (2016). "Morally Permissible Moral Mistakes". *Ethics* 126 (2): 366-393.

Harman, Elizabeth (2019). "Moral Testimony Goes Only So Far". In D. Shoemaker, ed., *Oxford Studies in Agency and Responsibility*, Vol. 6. Oxford: Oxford University Press, pp.165-185.

-- (ms). "Ethics is Hard! What follows?" Draft available online at: http://www.princeton.edu/~eharman/Ethics%20Is%20Hard%20020714%20For%20Web.pdf

Kiesewetter, Benjamin (2016). "You Ought to $\phi$ Only If You May Believe That You Ought To $\phi$". *The Philosophical Quarterly* 66, 760-782.

Levy, Neil (2009). "Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick." *Ethics* 119: 729–41.

Littlejohn, Clayton & Maria Alvarez (2017). "When Ignorance Is No Excuse". In P. Robichaud and J. Wieland, eds., *Responsibility: The Epistemic Condition.* Oxford: Oxford University Press, pp. 64-81.

Littlejohn, Clayton (2012). *Justification and the Truth-Connection*. Cambridge: Cambridge University Press.

Littlejohn, Clayton (2013). "The Unity of Reason". In C. Littlejohn and J. Turri, eds., *Epistemic Norms: New Essays on Action, Belief, and Assertion.* Oxford: Oxford University Press, pp. 135-152.

Littlejohn, Clayton (forthcoming). "A Plea for Epistemic Excuses". Forthcoming in J. Dutant and F. Dorsch, eds., *The New Evil Demon: New Essays on Knowledge, Justification and Rationality.*
Lord, Errol (2017). "What You're Rationally Required to Do and What You Ought to Do (Are the Same Thing!)". *Mind* 126: 1109-1165.

Rosen, Gideon (2003). "Culpability and Ignorance". *Proceedings of the Aristotelian Society* 103: 61–84.

Rosen, Gideon (2004). "Skepticism about Moral Responsibility". *Philosophical Perspectives* 18: 295–313.

Smith, Holly (1983). "Culpable Ignorance". *Philosophical Review* 92(4): 543-571.

Talbert, Matthew (2013). "Unwitting Wrongdoers and the Role of Moral Disagreement in Blame," in D. Shoemaker, ed., *Oxford Studies in Agency and Responsibility*, Vol. 1. Oxford: Oxford University Press, pp. 225–45.

Zimmerman, Michael (1997). "Moral Responsibility and Ignorance." *Ethics* 107: 410–26.