

*forthcoming* in *Ethics* in 2016

### **Morally Permissible Moral Mistakes**\*

Elizabeth Harman

Abstract: I argue for a moral category which has been ignored or underappreciated by moral theorists: morally permissible moral mistakes. A moral mistake is something that an agent should not do, all things considered, for moral reasons. I argue against the common thought that every moral mistake is morally wrong. Sometimes one has a supererogatory option that one should take, all things considered. In such a case, if one fails to take that option, then one makes a morally permissible moral mistake. I argue that recognizing this moral category is useful to our moral thinking in a variety of ways.

---

\* Thank you to Tyler Doggett, Jamie Dreier, Julia Driver, Adam Elga, Alex Guerrero, Peter Graham, Christopher Heathwood, Carrie Ichikawa Jenkins, Hallie Liberto, Barry Maguire, Sarah McGrath, Jennifer Morton, Alastair Norcross, Michael Otsuka, Tamar Schapiro, Laurie Shrage, Holly Smith, Sharon Street, Jonathan Quong, two anonymous referees, and the Associate Editors of *Ethics*, for comments on drafts of this paper or related papers. Thank you to audiences at the Analytic Legal Philosophy Colloquium, the Arizona Normative Ethics Workshop, Australian National University, the Bellingham Summer Philosophy Conference, Cornell University, Corridor, the Eastern APA, Massachusetts Institute of Technology, the Mohonk Philosophy Conference, New York University, Oxford University, the Pacific APA, Princeton University, Rutgers University, Syracuse University, University of California, Los Angeles, University of Colorado at Boulder, University of Sydney, University of Texas at Austin, University of Vermont, and Yale University, where this paper or related papers were discussed.

## **Introduction**

In this paper, I argue for a moral category of actions (and omissions) which has been ignored or underappreciated by moral theorists: morally permissible moral mistakes. A moral mistake is something that an agent should not do, all things considered, such that the considerations that tell against doing it – that make it something she should not do – are moral considerations. It is commonly thought that everything we should not do *for moral reasons* – that is, every moral mistake – is morally wrong. But, I argue, this is not true.

I argue that there are morally permissible moral mistakes. I argue that morally permissible moral mistakes can have some features that may be surprising: they do not just arise out of reasons of beneficence; they can be morally good things to do; and they can be praiseworthy. I distinguish the category of morally permissible moral mistakes from several other moral categories with which it might be confused. I also argue for another category which is the flipside of morally permissible moral mistakes: behavior that one should engage in, all things considered, for moral reasons; this category includes but is not exhausted by morally required behavior; some morally supererogatory behavior falls into this category as well.

Finally, I argue that recognizing these two categories is useful in a number of ways. It helps us to see some features of supererogatory behavior that otherwise are ignored. It dissolves a puzzle about supererogation. It makes new moral views available to us. It makes new interpretations of our own and others' commitments possible. And it enables us to make sense of combinations of views that otherwise appear inconsistent.

### **I. Some moral mistakes are morally permissible**

It is intuitively natural to characterize morally wrong actions as follows: morally wrong actions are those actions which (a) one ought not to perform, all things considered, and (b) one ought not to perform *for moral reasons*—that is, the reasons against those actions, which explain why one should not perform them, are moral reasons.<sup>1</sup> More generally, the following two claims appear to be true:

Morality is overriding: If it would be morally wrong for S to  $\phi$ , then S ought not to  $\phi$ , all things considered.<sup>2</sup>

All moral mistakes are morally wrong: If S ought not to  $\phi$ , all things considered, and the reasons against  $\phi$ ing that win out to make it the case that S ought not to  $\phi$  are *moral reasons*, then it would be morally wrong for S to  $\phi$ .

I believe that morality is overriding, and I will not take issue with that claim.<sup>3</sup> In this paper, I will argue that it is false that all moral mistakes are morally wrong. I will explore what follows when we recognize that there are morally permissible moral mistakes.

First, some terminological clarification. Throughout the paper, I use “moral reason” to refer to *other-regarding* moral reasons. It is a hard question whether one is ever morally required to treat oneself well; I will not address this question. If we counted self-regarding reasons among moral reasons, the claim that all moral mistakes are morally wrong might appear obviously false. Often I ought not to do something, all things considered, because it would be bad for me, but it is not morally wrong to do it; if in this case the reasons against the action—that it would be bad for me—count as moral reasons, then we would have too easy a counterexample to the claim that all moral mistakes are morally wrong.<sup>4</sup>

I will use the terms “mistake” and “moral mistake” as follows:

S’s  $\phi$ ing would be a mistake iff (def)

S ought not to  $\phi$ , all things considered

= (equivalently)

S ought not to  $\phi$ , in light of all of S's reasons<sup>5</sup>

S's  $\phi$ ing would be a moral mistake iff (def)

S ought not to  $\phi$ , all things considered

and

the reasons against S's  $\phi$ ing that win out to make it the case that S ought not to  $\phi$  are moral reasons

Given these terminological stipulations, the claim that I called "All moral mistakes are morally wrong" does indeed assert that all moral mistakes are morally wrong.

I want to clarify the definition of moral mistake as follows. For a particular behavior, how do we tell whether it is a moral mistake? Do we confine our attention to a subset of the agent's reasons, the moral reasons, and ask what the agent ought to do in light of those reasons? No, we do not. Rather, we consider what the agent ought to do in light of *all* of her reasons; this includes both reasons *for* the behavior and reasons *against* the behavior; and it includes both *moral* and *non-moral* reasons. If the agent ought not to engage in the behavior, all things considered, we then turn our attention to the reasons *against* engaging in the behavior that win out to make it the case that she ought not to engage in the behavior. If *these reasons* are moral reasons, then the behavior in question is a moral mistake.<sup>6</sup>

I will now offer a counterexample to the claim that all moral mistakes are morally wrong:

Amanda is a philosophy professor who has a two-year-old daughter. It is 11pm.

Amanda receives an email from her undergraduate student Joe, with a third draft of the paper that is due tomorrow at 12noon. She has already commented on his first two drafts.

Joe is struggling in the class, but she can tell that he is on the verge of some kind of breakthrough. If Joe fails the class, he will lose his scholarship and have to drop out of school. It would take half an hour to read the draft and write the comments, and Amanda is tired. Her daughter will wake her up early. Amanda realizes that she is not morally obligated to spend the thirty minutes to give Joe comments, but nevertheless she deliberates about whether to do it. Upon reflection, Amanda thinks, “I should do it!” She’s right. She gives him the comments.

Here’s my claim about the Amanda case: this could be a true story.<sup>7</sup> First, Amanda is not morally obligated to give Joe comments. Some professors never give comments to undergraduates on drafts of their papers before the deadline; they are not failing to fulfill any moral obligations. Some give comments on up to one draft of each paper. It is certainly not morally obligatory to give comments on three drafts of a paper, and certainly not when the third draft comes in so close to the deadline. Second, while it is not morally required for Amanda to give Joe comments, it may nevertheless be true that all things considered, she should give the comments. Finally, when we ask what reasons there are in favor of Amanda’s giving Joe comments, we see that they are *moral reasons*: that Joe would benefit from the comments, and that they might enable him to stay in school. In this case, Amanda’s giving Joe comments is *supererogatory*, but still it is the thing she should do.

If we plug in *Amanda* for *S* and *failing to give Joe comments* for  $\phi$ , then we get a counterexample to the claim that all moral mistakes are morally wrong: Amanda ought not to fail to give Joe comments; the reasons against failing to give Joe comments (that is, the reasons in favor of giving Joe comments) that win out to make it the case that she shouldn’t fail to give him comments are moral reasons; and yet failing to give him comments is not morally wrong.<sup>8</sup>

I also have a more general argument against that claim:

1. Sometimes a morally supererogatory action is the action that an agent ought to perform, all things considered.
2. In some of those cases, all the reasons in favor of the supererogatory action are moral reasons.

Therefore:

3. It is false that all moral mistakes are morally wrong: there are cases in which an agent ought not to  $\phi$  ( $\phi$  = fail to perform some supererogatory act), and the reasons against  $\phi$ ing (that is, the reasons in favor of the supererogatory act) that win out to make it the case that she shouldn't  $\phi$  are moral reasons, and yet  $\phi$ ing is not morally wrong.

I find premise 1 hard to doubt, but some people might question it. They might say that when it comes to the supererogatory, we are simply in the realm of the permissible—not just the morally permissible, but also the permissible regarding what one ought to do, in light of all one's reasons.

As a first pass, an objector to premise 1 might assert this:

Every supererogatory action is permissible, and merely permissible, in light of all one's reasons: that is, it is not the case that one should perform it, all things considered, nor is it the case that one should not perform it, all things considered.

This claim is far too strong to be plausible, however. Importantly, sometimes supererogatory actions are mistakes; their agents should not perform them, given all their reasons. Consider the following two cases:

Betsy suffers from severe migraines, but has learned how to avoid them. If she ever gets a mild headache, she lies down with her eyes closed for thirty minutes; otherwise, about 1/3 of the time, a mild headache leads into a migraine. Betsy feels a mild headache

coming on when her neighbor Timmy, a six-year-old boy, comes to the door and asks if he could perform a song he's been practicing for her. Betsy knows that Timmy would appreciate her attention, but also that he will understand if she says she's not feeling well. Listening to Timmy's song, and thus risking a migraine, would be supererogatory. The migraines are really awful, and Timmy won't be too upset if Betsy says "no." All things considered, Betsy should tell Timmy it's not a good time for his song, and go lie down right away.

Brenda has been working hard for years, trying to build up her skills and experience so that if her dream job ever becomes available, she might have a shot at it. Her dream job is now available, and Brenda has just applied when she learns that her acquaintance Rose is also applying. Rose has been working equally hard as Brenda toward the same dream job, but Rose is currently in a somewhat worse job than Brenda is, and Rose is supporting a family who would benefit from the increased salary that comes with the dream job, while Brenda is not supporting anyone. Rose stands to gain more from getting the job than Brenda stands to gain from getting the job. Brenda knows the field well enough to predict that she and Rose will be the strongest contenders for the job, though she cannot predict who would succeed between them. Brenda is aware that withdrawing her application would be doing something supererogatory for Rose. But all things considered, given how important this is to Brenda and how hard she has been working toward it, Brenda knows that she should not withdraw her application.

These two cases, as described, could be true stories. Cases like these show that sometimes it is the case that one should not perform a supererogatory act. (The Brenda case also shows that

sometimes one should not perform a supererogatory act, even though one would be sacrificing less than one would be providing to another person.)<sup>9</sup>

The objector to premise 1 is left with the following asymmetrical claim:

Some supererogatory actions are such that their agents should not perform them, all things considered, but no supererogatory actions are such that their agents should perform them, all things considered.<sup>10</sup>

It would be odd if this claim were true. I can certainly think of cases when I performed supererogatory actions and I think I was making *the right choice* – not just a justified choice, but a choice that was a better choice than my alternatives, and not morally better but all things considered better (sometimes the *morally best* choice is a *worse choice*, all things considered<sup>11</sup>); and, more distressingly, I can think of cases when I failed to perform a particular supererogatory action and I think I made *the wrong choice* – I should have performed that action, though it was not morally required.

Someone might resist my argument by adopting the asymmetrical claim and denying premise 1. Or someone might resist my argument by granting premise 1 but denying premise 2 – holding that it is sometimes true that an agent ought to perform a supererogatory action, all things considered, but only because of non-moral reasons in favor of the action. Those who adopt either of these views can take my paper as arguing for a weaker conclusion: that the view that some moral mistakes are not morally wrong is a *coherent* view that should be taken seriously. This weaker conclusion, as I'll argue, is important in its own right.

Let me summarize the disagreement between the view I am arguing for and the common view that I am opposing.



Both views agree that morality is overriding, so both hold that if a behavior is morally required, then all things considered one should behave that way; and if a behavior is morally wrong, then all things considered one should not behave that way.

Both views agree that behaviors that are morally permissible without being morally required can fall into all of the following three categories: something one should not do, all things considered; something one should do, all things considered; and something that is neither such that one should nor should not do it. Both views agree that morally permissible behaviors can fall into all three of these categories because of *non-moral reasons* in favor of or against these behaviors. Suppose I am playing a low-stakes game of poker, in which it is morally permissible to either call the current bet or fold; nothing morally significant hangs on what I do. It may be that I *should* call the bet, all things considered, though it is not morally required to call the bet. In this case, I *should not* fold, all things considered, though it is not morally wrong to fold my hand.

The two views disagree regarding the following category of behaviors: moral mistakes, that is, those behaviors that should not be engaged in, all things considered, *for moral reasons*. The common view holds that this category is the same category as: morally wrong behaviors. I am arguing that this is incorrect. This category really encompasses two moral categories: morally wrong behaviors and morally permissible moral mistakes.

It follows that the two pictures also disagree regarding the following category of behaviors: those behaviors that should be engaged in, all things considered, *for moral reasons*. The common view holds that this category is the same category as: morally required behaviors. I am arguing that this is incorrect. This category does not just include morally required behaviors; it also includes some supererogatory actions.

My argument that there are morally permissible moral mistakes proceeds by pointing out that some supererogatory actions have the following two features: all things considered, they should be performed; and the reasons they should be performed are the very *moral reasons* that make them morally good things to do. Nevertheless, these actions are supererogatory; they are not morally required.

## II. Morally permissible moral mistakes can arise from moral constraints

I've argued against the claim that all moral mistakes are morally wrong. Someone sympathetic to that claim might try to rescue some of its spirit by saying the following:

*Some of our moral reasons are reasons of beneficence; they can make an action morally good without making it morally required. But other moral reasons are in their nature morally constraining. They cannot make an action good without making it morally required, nor can they make an action bad without making it morally wrong. Our constraint-based moral reasons morally require whenever they require at all.*

Such a thought motivates the following claim:

All constraint-based moral mistakes are morally wrong: If S ought not to  $\phi$ , all things considered, and the reasons against  $\phi$ ing that win out to make it the case that S ought not to  $\phi$  are constraint-based moral reasons (such as reasons not to lie, not to kill, and to keep one's promises), then it would be morally wrong for S to  $\phi$ .

This claim is not true, however. Consider the following case.

Cara has promised to pick Sam up at the airport. It was important to him to get Cara to *promise* because Sam is a nervous flier who feels very anxious and strung out after traveling, and it would make him feel a lot better to see a friendly face at the other end of his travels. Unforeseeably, Cara's car breaks down. The mechanic can fix the car in two

days for \$100 or in two hours for \$200. This means it would cost an extra \$100 to be able to pick Sam up. Cara realizes that it would be morally permissible not to pick Sam up under the circumstances, and that Sam would understand that too. Nevertheless, Cara continues to deliberate. Upon reflection, she concludes, “I should still pick Sam up, because I promised.” She is right. She picks him up.

Here’s what I claim about the Cara case: this could be a true story. It could be that Cara ought to pick Sam up, all things considered, because she promised to pick him up, even though she is not morally required to pick him up, because there is an unexpected significant cost to picking him up. In this case, if Cara were to fail to pick Sam up, she would be making a *constraint-based* morally permissible moral mistake.

In saying that the Cara case could be a true story, I do not mean to suggest that it is always the case that we should keep our promises; of course that is not true. My claims are just these:

It is not unreasonable to continue to deliberate about whether to keep one’s promise, even after realizing that circumstances have changed so that one is not morally obligated to keep one’s promise.

In the course of such deliberation, one might conclude “I should still keep my promise” and one might be right about that.<sup>12</sup>

In some such cases, one should keep one’s promise *because one promised* and not for some other reason.

### **III. Some moral mistakes are morally good and praiseworthy things to do**

I've argued that not all moral mistakes are morally wrong, and I've argued that not all constraint-based moral mistakes are morally wrong. Someone might try to rescue something of the spirit of the original claim by saying the following:

*At least we can say this about moral mistakes: they are never morally good actions; and they are never praiseworthy (if the agent knows what she is doing).*<sup>13</sup>

Here is the claim under consideration:

No moral mistakes are morally good actions or praiseworthy actions: If S ought not to  $\phi$ , all things considered, and the reasons against  $\phi$ ing that win out to make it the case that S ought not to  $\phi$  are *moral reasons*, then it is not the case that S's  $\phi$ ing would be a morally good action, and it is not the case that S would be praiseworthy for  $\phi$ ing (if S knew the nonmoral facts that make  $\phi$ ing something S ought not to do).

This claim is false. We can see that it is false by considering a case in which an agent ought to perform a particular supererogatory action, all things considered, but she also has another supererogatory action available to her, which is less morally good than the one she ought to perform. Consider this case, which is similar to the Amanda case we saw earlier:

Deborah is a professor with a two-year old child. At 11pm, two undergraduates email drafts to Deborah. Jim sends his third draft; she's already given him comments on two drafts; he is struggling but nearing a breakthrough. If Jim fails the class, he will lose his scholarship and have to drop out of school. It would take her thirty minutes to give him comments. Tom also sends a draft; it's a second draft, with one new section that she hasn't seen; the first draft was already reasonably good. Tom would benefit from comments, and it would only take her ten minutes to give him comments, but he'd do fine without comments too. Deborah definitely shouldn't give them both comments. She

needs sleep. She's not morally required to give comments to either one. Deborah deliberates about what to do. In fact, she should sacrifice thirty minutes of sleep to give Jim comments. But that's not what she decides to do; she spends ten minutes giving Tom comments instead.

(Again, my claim is that this is a *possible* case.) Deborah makes a mistake: she does something that she should not do, all things considered. (She comments on Tom's paper rather than Jim's.) Her mistake is a moral mistake; the reasons against doing what she does, that make it something she should not do, are *moral reasons*—they are the reasons in favor of commenting on Jim's paper instead. But what Deborah does is a morally good thing: she gives Tom comments and helps his paper to be better even though she's already given him comments on one draft, at some real cost to herself, the cost of some sleep. What Deborah does is praiseworthy. It's not as praiseworthy as commenting on Jim would have been, but still it is praiseworthy.<sup>14</sup>

#### **IV. Some supererogatory actions should be performed, but are not morally optimal**

In this section, I want to point out that when one has an option that one ought to take, all things considered, for moral reasons, this is often not the *morally best* option one has. The question of which action is morally best is a very different question from the question of what one ought to do, all things considered. Sometimes, it will turn out that one ought to do a particular thing, because of the moral considerations to be said for doing it, without that being the morally best option one has. Consider the following case; it is similar to the cases of Amanda and Deborah above.

Ellen is a professor with a two-year old child who wakes her early every morning. At 11pm, two undergraduates email drafts to Ellen. Jeff sends his third draft; she's already given him comments on two drafts; he is struggling but nearing a breakthrough. If Jeff

fails the class, he will lose his scholarship and have to drop out of school. It would take her thirty minutes to give him comments. George sends a sixth draft; she's already given him comments on five drafts. George is struggling, and is in danger of having to drop out of school, in which case he will also be deported. To help George, Ellen would have to pull an all-nighter. She'd have to give him extensive comments, and then have a long email exchange with him that would take all night. Ellen is not morally required to help either of them tonight. She needs sleep, and for this reason, she should not help George. But she can afford to lose the sleep involved in helping Jeff. All things considered, she should help Jeff.

I claim that this is a possible case. In this case, Ellen's morally best option is to stay up all night to help George. All things considered, she should not take that option. But there is a supererogatory action that she should perform, all things considered. She should help Jeff. Helping Jeff is a supererogatory action that she should perform, all things considered, that is not morally optimal.

## **V. Other Moral Categories**

Some moral mistakes are morally wrong; I have argued that some moral mistakes are not morally wrong. Let me introduce some more terminology:

S's  $\phi$ ing is a mere moral mistake iff (def)

S's  $\phi$ ing is a morally permissible moral mistake

All morally wrong actions are moral mistakes. A mere moral mistake is *merely* a moral mistake in that it is a moral mistake but (unlike some moral mistakes) it is not also morally wrong.

Now that we have seen that there are morally permissible moral mistakes, we can recognize some interconnected distinctions. The following distinctions are *inclusive*; the first category is a subset of the second category:

morally wrong actions *versus* moral mistakes

morally required actions *versus* actions an agent ought to perform, all things considered,  
for moral reasons

The following distinctions are *exclusive*; no behavior falls into both categories:

morally wrong actions *versus* mere moral mistakes

morally required actions *versus* actions an agent ought to perform, all things considered,  
for moral reasons, but which are not morally required

To better understand the category of *mere moral mistakes* (that is, morally permissible moral mistakes) and the four distinctions above that the category gives rise to, I will now examine some well-recognized moral distinctions and discuss whether these distinctions enable us to characterize mere moral mistakes. If so, then we have already been talking about mere moral mistakes in talking about these other categories. If not, then the category of mere moral mistakes is not already captured by these distinctions.

#### **a. What is Objectively Wrong *versus* What is Subjectively Wrong**

Someone might think that we can see an instance of a mere moral mistake by considering cases in which an agent has false beliefs about her situation. A wife poisons her husband, thinking she is giving him the cure; she is blameless for her false belief. In this case, what she does is objectively wrong (she poisons him); but it is not subjectively wrong (she gives him what she has every reason to believe is a cure).

Proposal A: The wife makes a mere moral mistake: she does something she should not do, for moral reasons (poison her husband), but what she does is morally permissible (she gives him what she has every reason to believe is a cure).

The proposal fails. The wife does not make a mere moral mistake, because the normative statements above are, we might say, true only relative to different bodies of evidence – that she should not do what she does is true relative to her actual situation, ignoring her evidence; but that it is permissible to do what she does is true relative to her evidence. When an action is a mere moral mistake, that is so because *it is something the agent should not do for moral reasons* and *yet it is morally permissible* –where these normative statements are true *relative to the same body of evidence*.<sup>15</sup> There can be objective mere moral mistakes and there can be subjective mere moral mistakes. But one cannot identify something as a mere moral mistake by exploiting both objective and subjective moral claims.

If we consider the case of Amanda, who has no non-moral false beliefs about her situation, had she failed to give Joe comments, that would have been an objective moral mistake and also a subjective moral mistake. If we consider a variant of the case in which, unbeknownst to Amanda, she's going to be woken up at 4am by a prearranged test of the fire alarm system in her apartment building, that is a case in which, objectively, she ought to go to sleep right away (the impending fire alarm makes sleep much more urgent), and failing to give Joe comments is not a mistake, objectively. But if in this case, she fails to give Joe comments, she does make a subjective moral mistake.

**b. What is morally best *versus* what is morally required**

Consider the following proposal:



Proposal B: What is morally best *is the same as* what one ought to do, all things considered, for moral reasons. S's  $\phi$ ing is a mere moral mistake just in case S's refraining from  $\phi$ ing is morally best though S's refraining from  $\phi$ ing is not morally required.

This proposal fails because often the morally best option available to an agent is not the one that she should take, all things considered. The morally best thing I could do right now might be this: go to the local hospital and offer up a kidney and some of my liver, to save two lives. But that's not what I should do right now, all things considered. So I am not making a mere moral mistake in failing to do so.

Several cases I've already mentioned also provide counterexamples. Betsy's morally best option is to listen to Timmy's song; but she should lie down instead. Brenda's morally best option is to withdraw her application, but she shouldn't withdraw. Ellen's morally best option is to help George, but that's not what she should do. Finally, consider the variant of Amanda's case in which in fact a fire alarm will go off in the night: in this situation, she ought not to give Joe comments (because of her need for sleep), but giving Joe comments is her morally best option.<sup>16</sup>

Note that while an agent always has a morally best option (or at least one option such that no other option is morally better), it is not true that an agent always has an option that she ought to take, all things considered, for moral reasons. Often one finds oneself in a situation in which one ought to do something, all things considered, but the reasons to do it are not moral reasons. Betsy and Brenda are both in that kind of situation.

**c. The morally right thing to do *versus* the morally required thing to do**

One might make the following proposal:

Proposal C: S's  $\phi$ ing is a mere moral mistake just in case S's refraining from  $\phi$ ing is the morally right thing to do though S's refraining from  $\phi$ ing is not morally required.

This proposal uses the expression "the morally right thing to do" to pick out behavior one should perform, all things considered, for moral reasons, including both morally required behavior and supererogatory behavior that one should engage in, all things considered. This strikes me as a good way to use the expression "the morally right thing to do," but I don't think that talk of "the morally right thing to do" is generally used in this way. Often people use "the morally right thing to do" to pick out only actions that are morally required. I think it would be good if talk of "the morally right thing to do" were reformed in line with Proposal C.<sup>17</sup>

**d. What one ought to do *versus* what one must do**

Proposal D: S's  $\phi$ ing is a mere moral mistake just in case S ought not to  $\phi$  but it's not the case that S must not  $\phi$ .

This proposal fails because the ought/must distinction applies even to non-moral cases. If you are coaching me while I play online chess, there may be certain moves that I *must* make; it would be a serious mistake not to make them. Other moves are such that I ought to make them, but I don't have to make them: it's a mistake not to make them, but not a serious mistake. Suppose I ought to move my rook, but it's not the case that I must move my rook. In this case, failing to move my rook is not a *moral* mistake. So the proposal fails.<sup>18</sup>

**e. The suberogatory**

Suberogatory actions are actions that it is bad to do, but not wrong to do. They are an inverse of the supererogatory, if the supererogatory is what is good to do, but not morally required.<sup>19</sup> Here are some purported examples: refusing to give a kidney to one's dying brother; taking the best available seat on the train although it prevents the couple behind one from getting

to sit together; and demanding immediate repayment of money from a friend, though it would be very easy for one to get along without the money for now, and hard for the friend to do so. We might disagree about the kidney case, holding that one is typically morally obligated to save one's sibling's life by giving him a kidney.<sup>20</sup> But the other two cases are plausible: taking the best seat and demanding the repayment of money owed are both morally *permissible*, yet they are morally bad to do. And absent unusual details in these cases, these are instances of morally permissible moral mistakes: the agents should not do these things, for moral reasons, but it is not morally wrong to do them.

In this paper, I need not take a stand on whether any behavior is *supererogatory*. The category is controversial, and it need not be embraced by those who accept the existence of the supererogatory. One could accept all of my arguments in this paper while denying the existence of the supererogatory. But I do believe in the supererogatory; so in what follows, I will talk as though some behavior is supererogatory.

Proposal E: All and only mere moral mistakes are supererogatory.

The proposal is false because some mere moral mistakes are not supererogatory. If Amanda fails to give Joe comments, she's failing to do something good; but she's not doing anything bad.

An objector might hold that since Amanda ought to give Joe comments, all things considered, for moral reasons, she does something *morally bad* and *blameworthy* if she fails to give him comments. Here it is useful to distinguish the question of whether an agent is *criticizable*, including whether she is morally criticizable, for having made a *mistake* and whether she does something *morally bad* or *blameworthy* (if she knows what she's doing). If Amanda fails to give Joe comments, then she has failed to act as she ought to act, all things considered. Amanda is thus liable to criticism: if we say "she didn't do what she ought to have

done,” we speak truly. Amanda is even liable to *moral criticism*: she failed to react appropriately to her moral reasons. But once we recognize the phenomenon of *morally permissible moral mistakes* that are *failures to perform supererogatory actions*, we should reject the following claim:

Whenever a person is liable to moral criticism, she has done something morally bad and is blameworthy.

Rather, a person may be liable to moral criticism because she failed to do something morally good that she should have done, all things considered, but that was supererogatory. Many supererogatory actions are such that failures to perform them are not suberogatory: these failures are neither morally bad nor blameworthy.

The proposal may fail in another way as well: it may be that some suberogatory actions are not moral mistakes, because all things considered, one *should* perform them. I am undecided on this point. For the supererogatory, there are cases in which something would be *good* to do, but all things considered, one *shouldn't* do it, for example because doing it would be too burdensome, or because doing it would involve forgoing something beneficial for oneself; one should do something morally worse instead. Perhaps, similarly, there are cases of suberogatory actions that would be *bad* to do, but all things considered one *should* perform them, because refraining would be too burdensome, or because going ahead would be beneficial to oneself; these would also be cases in which one should do something morally worse rather than something morally better.

While this seems in principle possible, it is interesting that there may be an asymmetry here between the supererogatory and the suberogatory.<sup>21</sup> When an otherwise morally good action would also be *burdensome*, its burdensomeness does not threaten the moral goodness of

the action (indeed, it might be seen to enhance it); by contrast, when refraining from an otherwise morally bad action would be *burdensome*, this burdensomeness *does* threaten the moral badness of the action: the fact of the burdensomeness may mitigate or eliminate the moral badness. We can see this with moral wrongness; for example, breaking a promise is generally morally wrong, and thus morally bad, but there are many cases in which it is neither morally wrong nor morally bad to break a particular promise because it has turned out to be unexpectedly very burdensome to keep the promise. Perhaps the following is true:

Whenever refraining from an otherwise morally bad action would be sufficiently burdensome (or the action would be sufficiently beneficial) to make the action *what the agent should* do, all things considered, then this burdensomeness (or potential benefit) is also sufficient to make the action *not morally bad*.

If this is true, then every suberogatory action is also a mere moral mistake. But perhaps the following is true instead:

Some actions are morally bad to do, but not morally wrong to do, and some of these are such that all things considered, given the non-moral considerations in favor of doing them, one should do them.

This also has some plausibility. Imagine one friend saying to another, “Look, given what’s at stake for you, you should do it. It’s a lousy thing to do to him, but you’re allowed to do it,” where by “you’re allowed to do it”, the speaker means that the action is not morally wrong. We can imagine someone making such a speech. Perhaps such a speech could be true. If so, then there are some suberogatory actions that are not mistakes, and so are not moral mistakes. I leave this issue open for further thought.

**f. What is justified by reasons versus what is required by reasons**

It has been argued that philosophers overlook a certain feature of reasons, which is that reasons have two dimensions of strength, not one. On the one hand, reasons have a certain requiring strength; on the other hand, they have a certain justifying strength.<sup>22</sup> A reason's requiring strength is sometimes—indeed, often—weaker than its justifying strength. Consider the fact that Mary's daughter would be thrilled if Mary baked cupcakes today. This fact has no requiring strength, I claim. Yet it has some justifying strength; it could make it reasonable for Mary to bake cupcakes today, even if Mary dislikes baking.

Proposal F: S's  $\phi$ ing is a mere moral mistake just in case S's refraining from  $\phi$ ing is justified but not required by moral reasons.

This proposal fails because an action might be justified in the above sense (it might be a reasonable thing to do) without being the action that the agent ought to perform, all things considered. The notion of justification cannot be substituted for all-things-considered requirement. The proposal is working with only one kind of requirement, but to understand the phenomena I have been describing, we need two kinds of requirement: there is what is morally required; and there is what an agent ought to do, all things considered.<sup>23</sup>

#### **g. Perfect Duties versus Imperfect Duties**

Consider the following proposal:

Proposal G: S's  $\phi$ ing is a mere moral mistake just in case S has an imperfect duty to refrain from  $\phi$ ing but S lacks a perfect duty to refrain from  $\phi$ ing.

This proposal fails because there are many actions that would satisfy an imperfect duty, and so it is true that one has an imperfect duty to perform them, but that are not such that one should perform them, all things considered. Consider the case of Betsy, who should go lie down to

stave off a possible migraine rather than listen to the boy at her door. Listening to him would fulfill an imperfect duty; but it is not the case that she should listen.<sup>24</sup>

## **VI. What one ought to do for moral reasons**

In this section, I will clarify and characterize the category that is the flipside of moral mistakes: the category of behavior that one ought to engage in, all things considered, for moral reasons. We can understand this category more narrowly or more broadly; I will argue that the category should be construed more broadly.

Let's begin by noting that the following four kinds of behaviors do seem to form a unified group; these behaviors have in common that their agents should engage in them, all things considered, and that moral considerations in favor of them play a crucial role in making this the case:

- all morally required options
- some supererogatory options: those that an agent should perform, all things considered, such that *all* the reasons in favor of performing them that win out to make it the case that they should be performed are moral reasons
- some supererogatory options: those that an agent should perform, all things considered, such that moral considerations in favor of performing them crucially explain why they should be performed; these include some supererogatory options that are not morally optimal (as we saw in considering the case of Ellen)
- all refrainings from suberogatory options such that one should not take those suberogatory options, all things considered<sup>25</sup>

Here are two possible characterizations of the category, one narrow and one broad:

S's  $\phi$ ing is something that S ought to do, all things considered, for moral reasons,

*narrowly construed* iff (def)

S ought to  $\phi$ , all things considered, and *all* the reasons in favor of  $\phi$ ing that win out to make  $\phi$ ing what S ought to do, are moral reasons.

S's  $\phi$ ing is something that S ought to do, all things considered, for moral reasons, *broadly*

*construed* iff (def)

S ought to  $\phi$ , all things considered, and there are some moral reasons in favor of  $\phi$ ing that play a crucial role in explaining why S ought to  $\phi$ , all things considered.

Two considerations show that we should construe this category broadly rather than narrowly.

First, we want this category to include all morally required actions. But some morally required actions have moral reasons in favor of them but also have other considerations, such as substantial prudential considerations, in favor of them. When we ask what reasons in favor of such an action *win out* to make it the case that one should perform the action, the answer includes both some moral reasons and some prudential considerations. The narrow construal of the category will not include these morally required actions.

Second, I claim that the four types of behavior listed above do form a natural group, and the narrower construal leaves out the third type listed above: it leaves out supererogatory actions that ought to be performed, all things considered, but are not morally optimal. Consider Ellen's helping Jeff. One reason in favor of helping Jeff is that Ellen would still be able to get some sleep, whereas by helping George she would not; this is not a moral consideration. The broader construal allows the category to include all four types of behaviors listed above.

## **VII. A Broader Category of Morally Permissible Moral Mistakes**



One issue raised in the last section applies as much to *morally permissible moral mistakes* as to *what one ought to do, all things considered, for moral reasons*. I have commented a few times above that every morally wrong action is a moral mistake. This is strictly speaking false, as I have defined “moral mistake.” Some morally wrong actions are also prudentially disastrous, so if we ask why one should not perform them, the answer yields both moral and self-interested considerations. These are mistakes for moral reasons, but not only for moral reasons; so they do not satisfy my definition of “moral mistake.” We can distinguish “moral mistake” as I use it in this paper from the following broader notion:

S’s  $\phi$ ing would be a moral mistake, *broadly construed*, iff (def)

S ought not to  $\phi$ , all things considered

and

there are moral reasons against S’s  $\phi$ ing that play a crucial role in explaining why

S ought not to  $\phi$

The broader category is more natural than the narrower category because the broader category succeeds in including all morally wrong behavior.

I defined “moral mistake” more narrowly because the narrow definition gives us what we might think of as *pure moral mistakes*; I see an implicit denial of the existence of pure moral mistakes in much of contemporary moral philosophy. What is at issue in this paper is whether there are any pure moral mistakes. I have argued that there are. Once we recognize pure moral mistakes, we can see that they are part of a broader category, *moral mistakes, broadly construed*. In this paper, the term “moral mistake” refers to the narrower category, because this paper seeks to introduce the concept of a “morally permissible moral mistake” by talking about these pure

cases and arguing for their existence. It is only once we recognize the pure cases that we can see the interest and importance of the broader category.

My claim that all morally wrong actions are moral mistakes is true only for *moral mistakes, broadly construed*. But all other claims in this paper use “moral mistake” as I have defined it – the narrow notion.

### **VIII. New Moral Understandings**

Recognizing that there are such things as mere moral mistakes can serve at least four purposes in our moral theorizing.

First, it can dissolve the following “puzzle of supererogation.”<sup>26</sup> On the one hand, supererogatory actions are not required—by definition, they are above and beyond the call of duty. On the other hand, people sometimes do perform supererogatory actions. What can be going on when someone chooses to perform a supererogatory action? Since she is performing the action, she must believe it is *the thing to do*, and so she must believe it is required. Thus, it appears that it is impossible for anyone to perform a supererogatory action *while understanding that it is supererogatory*. That is the puzzle. We might add on: Could there really be a category of action that is above and beyond the call of duty but such that no one who performs it knows that it is above and beyond the call of duty? This may suggest that there are no supererogatory actions. Also: It’s natural to think that there is a special kind of praiseworthiness that attaches to supererogatory actions.<sup>27</sup> But can someone be praiseworthy in that special way if she thinks that she is just doing what is required? This may suggest that no one is ever praiseworthy in the special way that seems to attach to the supererogatory.

The statement of the puzzle gets things wrong in at least two ways. First, one *could* choose to do something without thinking it is *the thing to do*, for example when one chooses one

of several identical soda cans from the refrigerator. I think there is a hidden assumption in the puzzle: that when one chooses to perform a supererogatory action, one is surely not taking that action to be *on a par with* one's non-supererogatory options, but rather one must take one's action to be *the thing to do*. Let's grant that assumption (for now). The more significant flaw in the reasoning in the development of the puzzle is this: there are two kinds of *requirement* that are relevant. One is *moral requirement*; the other is *requirement in light of all one's reasons*. To put this another way: an agent who performs a supererogatory action could take what she is doing to be the thing she should do, in light of all her reasons, without taking it to be morally required.

(We can see what I have said as offering a solution to the puzzle on the puzzle's own terms – granting the assumption that an agent who performs a supererogatory action always takes her action to be *the thing to do*. But that assumption is surely too strong. Some people who perform supererogatory actions do not take those actions to be *the thing to do*. There must be cases in which a supererogatory action is on a par with one or more non-supererogatory options; an agent who performs one of these actions may well recognize this, and choose the action without taking it to be the thing to do.)

Recognizing the existence of mere moral mistakes can also provide a second, third, and fourth benefit in our moral reasoning. It makes new views available to us. It makes it possible for us to interpret our own views or others' existing views in new ways. And it may help us to make sense of combinations of views that otherwise seem inconsistent.

In fact, all four of these benefits are available even to those who are not convinced by my arguments for *the existence* of mere moral mistakes. My paper can also be seen to argue for the weaker claim that it is *conceptually coherent* to suppose that there are some mere moral

mistakes, and thus to suppose that particular actions or types of action are mere moral mistakes. By seeing that this is conceptually coherent, we can recognize the availability of certain interpretations of existing views, and the possible coherence of certain combinations of views.

New views are made available:

I have recently developed the following view about the ethics of being an anonymous donor of sperm or eggs:<sup>28</sup> this is a morally good thing to do, and is praiseworthy, but all things considered, one should not do it, because of moral reasons one has regarding one's children (including one's genetic children); being a gamete donor is a morally good, morally praiseworthy, morally permissible moral mistake. Our reasons regarding our own children include reasons to be available to them when they need our help, and reasons to have real relationships with them; anonymous donors set themselves up not to do the things these reasons support. I will not defend this view here. But the view captures a number of different things that each have some intuitive appeal. My arguments in this paper—for the existence and coherence of mere moral mistakes—do some work to show that these views can actually be held together.<sup>29</sup>

More generally, the category of mere moral mistakes includes those actions that have serious moral things to be said against them, but are not morally wrong. We must be careful to consider, when we find a behavior problematic for moral reasons, that it might be merely a moral mistake rather than morally wrong. When moral objections to a practice strike us as weighty, we fail to recognize that there are two different ways moral objections to a practice can win out: they may make the practice morally wrong, or they may make it merely a moral mistake.

This brings me to the third benefit of taking seriously the category of mere moral mistakes: we can interpret our own views or others as implicitly committed to the view that certain kinds of action are mere moral mistakes. I will give two examples.

Consider the view that one should not buy or eat meat, and yet one should not criticize or impede the meat-purchasing and meat-eating practices of others. It is striking that vegetarians often seem to have this view: they keep a strict practice of avoiding meat themselves, for moral reasons; and yet they are willing to support restaurants that serve meat, and are willing to eat alongside others eating meat, saying nothing at all in criticism. It is not generally true that it is morally fine to fail to criticize or impede in any way the morally wrong behavior of others. The practices of these vegetarians makes perfect sense, however, if we attribute to them the following implicit belief: that buying and eating meat is a mere moral mistake. They might even think that buying and eating meat is *suberogatory* – that it is bad to do, though not morally wrong. Yet still appropriate responses to suberogatory actions (which are, after all, morally permissible) vary greatly from appropriate response to morally wrong actions.<sup>30</sup>

Consider next the view that one should not have an abortion, for moral reasons, and yet that abortion should not be illegal.<sup>31</sup> One way to make sense of this view is that those who hold it are implicitly committed to the view that abortion is a mere moral mistake.

## **IX. Further Considerations in Favor of the Existence of Mere Moral Mistakes**

Our moral lives include a substantial realm of moral permissibility. Each of us has her own projects and her own interests, as well as her own loved ones, and we have agent-relative permissions to privilege our projects and our interests in substantial ways. Nevertheless, we often have reasons regarding others which tell in the other direction; these include reasons of beneficence (but not only those). We regularly face questions of whether to privilege ourselves over others, even in cases in which it is permissible to privilege ourselves. It makes sense to deliberate about what to do in these cases, even though we know that we are not required to follow our other-regarding reasons: our other-regarding reasons continue to exert a force on us

even though they do not in these cases render the behavior they favor morally required. It would be wrong to think that in these cases, all one's options are on a par. Sometimes doing a supererogatory thing is a mistake; the cases of Betsy and Brenda from section I are two examples. Sometimes failing to do a particular supererogatory thing is a mistake: just as one's self-regarding reasons may win out in some of these cases, similarly one's other-regarding reasons may win out in some of these cases.<sup>32</sup>

#### **X. The Significance of Mere Moral Mistakes**

One might ask: what is the significance, or upshot, of recognizing that a particular action (or each typical action of a particular kind) is a mere moral mistake?

The examples of mere moral mistakes that I have given in this paper, of Amanda's failing to give Joe comments, Cara's failing to pick Sam up, Deborah's giving comments to Tom, and Ellen's going to sleep without helping anyone, are all cases in which an agent fails to perform a supererogatory action that she ought to perform, all things considered, but the failure does not involve doing anything *morally bad*; it is simply a failure to do something *morally good*. These mere moral mistakes are not blameworthy, but the agent had an alternative that would have been praiseworthy (if she knew what she was doing).<sup>33</sup> Such a mere moral mistake is criticizable – the agent has not responded appropriately to her reasons – but not morally blameworthy, even though the reasons to which she failed to respond adequately are *moral reasons*. (Indeed, as we have seen, some of these mere moral mistakes are praiseworthy.) So we see here a kind of criticism that falls short of blameworthiness. It would simply be inappropriate for the potential recipients of supererogatory help in these cases to *resent* the agent who does not help him, even if he sees that all things considered, the agent should have helped him.

If the suberogatory exists, then there is another kind of mere moral mistake: suberogatory behavior, which involves doing something *morally bad* though not *morally wrong*. A suberogatory mere moral mistake is typically blameworthy. (It is blameworthy if the agent knows what she is doing, and if there is no other factor that undermines blame).<sup>34</sup> But suberogatory actions are not blameworthy in the same way that morally wrong actions are blameworthy: the content of appropriate blame and resentment is different. The victim of a suberogatory action can appropriately resent the agent *for treating her badly*, but she cannot appropriately resent the agent *for treating her wrongfully*, as that has not happened. So this is a way that suberogatory actions are more morally innocent than morally wrong actions.<sup>35</sup> But these actions are not as morally innocent as other morally permissible actions. For example, they are not as innocent as non-suberogatory failures to perform supererogatory actions.

All mere moral mistakes are *mistakes*, so they are less innocent than *some* failures to perform supererogatory actions. Many failures to perform supererogatory actions are merely failures to do something that would have been morally good, but these failures are not *mistakes*: it is often not the case that, all things considered, one should have performed the supererogatory action in question. Mere moral mistakes are *mistakes*: they are failures to act as one all things considered should act, and they are failures to respond to one's *moral reasons* appropriately. These are moral failings. But only some of these moral failings (the suberogatory ones) are blameworthy.

Let's suppose that each instance of an individual's eating meat is a mere moral mistake. Some people think eating meat is morally wrong; given our supposition, these people are *overstating* the moral failing involved in eating meat. Some people think that eating meat is

morally permissible, but that it would be *morally better* to refrain from eating meat; if that is the whole of their position, these people are *understating* the moral failing involved in eating meat.



---

<sup>1</sup> In this paper, I use “ought to” and “should” interchangeably.

<sup>2</sup> This claim and the next one are universal generalizations, though I leave the statement of the universal generalization implicit. “S” ranges over agents, and “ $\phi$ ” ranges over behaviors, which include actions and omissions.

<sup>3</sup> A number of different claims have been called the claim that “morality is overriding”. My indented statement specifies what I mean by that terminology.

<sup>4</sup> I think it is an open question whether all moral reasons are other-regarding. As I understand it, a moral reason is any consideration that tells in favor of a behavior being morally required, morally good to do, morally wrong, or morally bad to do. Note that it need not be the case that all reasons that are *morally relevant* are *moral reasons*. For example, the consideration that  $\phi$ ing would inconvenience me, or would be bad for me, may be morally relevant to whether  $\phi$ ing is morally required, without that consideration’s being a moral reason (that consideration may not tell in favor of  $\phi$ ing’s being morally wrong or morally bad, nor tell in favor of refraining from  $\phi$ ing’s being morally required or morally good). Douglas Portmore’s “Are Moral Reasons Morally Overriding?” (*Ethical Theory and Moral Practice* 11 [2008]: 369-388) points out that we should distinguish *reasons that are morally relevant* from *moral reasons*.

<sup>5</sup> Note that I am using the word “mistake” to apply to *actions (or failures to act)* rather than *thought processes*. When we say that someone has made a mistake, very often we mean that she has reasoned poorly. But we also sometimes refer to actions as mistakes. We say “it was a mistake to call that poker hand” or “don’t make the mistake of paying the carpenter before she finishes the job.” A person can perform an action that is a mistake without making any mistake of reasoning, as when someone correctly realizes what she should do but then out of weakness of

---

will fail to do it; one might truly say, “She made a mistake. How did it happen, you ask? She knew what she should do but she was weak-willed.”

<sup>6</sup> Note that the expression “win out” should not be taken to imply that there are always reasons on the other side. Some reasons might *win out* to settle that one should do something because there are *no reasons* against doing it.

<sup>7</sup> My claim about this case is that it *could be a true story*: that the normative claims that are part of the case *could be true* of a case with these non-normative features. I don’t want to claim (about this case or the other cases in this paper) that the non-normative features of the case that I lay out make it clear that the normative claims hold in the case. Rather, I just need the weaker claim (about this case and my other cases) that the normative claims *could be true* in *some* case like this.

Thus, my use of described cases in this paper is a bit different from the common use in which a philosopher describes a case and then claims that a certain normative claim *must* be true in the case as described.

<sup>8</sup> The claim that all moral mistakes are morally wrong is implicitly held by most ethicists, including both those who believe that some actions are supererogatory and those who deny that any actions are supererogatory. This paper does not engage the debate regarding whether any actions are supererogatory, though it does take a side in that debate. There clearly are some supererogatory actions; the cases I offer in this paper provide some examples.

Some philosophers who recognize the supererogatory also recognize the suberogatory: actions that are morally bad to do but not morally wrong to do. (I will remain neutral on whether any behavior is suberogatory.) If the suberogatory exists, then (as I discuss below), either suberogatory behavior is a subclass of morally permissible moral mistakes, or the two categories

---

overlap. While philosophers who believe in the suberogatory do recognize that some moral mistakes are not morally wrong, no one (to my knowledge) has recognized that there is a bigger category, morally permissible moral mistakes, that encompasses both suberogatory behavior (if it exists) and some failures to engage in supererogatory behavior (where these failures are not suberogatory). And no one to my knowledge has recognized that there can be morally permissible moral mistakes that are morally good to do and praiseworthy.

I discuss the suberogatory in part (e) of section V.

While most *moral philosophy* implicitly assumes that all moral mistakes are morally wrong, *ordinary moral thinking* involves implicit commitment to the existence of morally permissible moral mistakes. It is common for one person to say to another, “you ought to do it, though you don’t have to do it” in a context in which both know that only moral considerations favor doing it, and in which by “you don’t have to do it” she means “it is not morally required.”

<sup>9</sup> Joseph Raz (“Permissions and Supererogation,” *American Philosophical Quarterly* 12 [1975]: 161-168) claims that no supererogatory actions are such that their agents should not perform them. Barry Curtis (“The Supererogatory, the Foolish and the Morally Required,” *Journal of Value Inquiry* 15 [1981]: 311-318) and Gregory Mellema (*Beyond the Call of Duty: Supererogation, Obligation, and Offence* [Albany: State University of New York Press, 1991]) hold that no supererogatory actions are foolish; they hold that a foolish helping action is not supererogatory. But Mellema claims, contra Curtis, that some actions are supererogatory even though “the moral reasons in favor of acting are clearly overridden by reasons of self-interest in favor of refraining” (176). Thus, Mellema appears to endorse the view that some supererogatory actions should not be performed, but are not foolish. Perhaps Brenda’s withdrawing her application is an example.

---

<sup>10</sup> An objector to premise 1 could hold that *all* supererogatory actions are *mistakes*, that whenever a person has a supererogatory option, all things considered, she should not take that option. It is not true that whenever anyone performs a supererogatory action, she is making a mistake. Nevertheless, some people have committed themselves to this view. For example, Douglas Portmore's "Position-relative consequentialism, agent-centered options, and supererogation" (*Ethics* 113 [2003]: 303–332) has this implication, pointed out by Betsy Postow ("Supererogation again," *Journal of Value Inquiry* 39 [2005]: 245–253) and repudiated in Portmore, "Are Moral Reasons Morally Overriding?"

<sup>11</sup> See part b of section V.

<sup>12</sup> In Elizabeth Harman, "Morality Within the Realm of the Morally Permissible" (in *Oxford Studies in Normative Ethics*, ed. Mark Timmons [Oxford: Oxford University Press, 2015]), I argue that looking closely at deliberation about whether to perform some supererogatory actions shows that there are some morally permissible moral mistakes.

<sup>13</sup> An uninteresting case of a moral mistake that is praiseworthy occurs when someone does something objectively wrong (so it is a moral mistake) but she doesn't know what she's doing: a woman gives her husband what she has every reason to believe is the cure to his illness but it's actually poison. She poisons her husband (that's a moral mistake) but she is praiseworthy (she tries to save his life). An interesting case would be one in which an action is a moral mistake due to non-moral features of the case that the agent knows about, but nevertheless the agent is praiseworthy for acting. That is the kind of case I claim is possible.

<sup>14</sup> The existence of these "second-best" supererogatory actions, such as Deborah's commenting on Tom's paper, is underappreciated in discussions of the supererogatory. For example, in Horgan and Timmons's "Untying a Knot from the Inside Out: Reflections on the "Paradox" of

---

Supererogation” (*Social Philosophy and Policy* 27 [2010]: 29-63), they gloss the paradox of supererogation as arising because supererogatory actions are “morally best” (29). Similarly, Jamie Dreier’s “Why Ethical Satisficing Makes Sense and Rational Satisficing Doesn’t” (in *Satisficing and Maximizing*, ed. Michael Byron [Cambridge: Cambridge University Press, 2004], 131-154) proposes that supererogatory actions are those that are required from the perspective of beneficence, and Portmore’s “Are Moral Reasons Morally Overriding?” says “for there is a sense in which supererogatory acts are acts that agents morally ought to perform” (379). Paul McNamara’s “Supererogation, Inside and Out: Toward an Adequate Scheme for Common Sense Morality” (*Oxford Studies in Normative Ethics*, ed. Mark Timmons [Oxford: Oxford University Press, 2011] 202-235) points out that sometimes a supererogatory action is not an agent’s morally best option, that such an action can nevertheless be praiseworthy, and that this phenomenon is underappreciated.

<sup>15</sup> By speaking of moral claims being true relative to bodies of evidence, I don’t intend to make any controversial claims about how we should understand objective and subjective uses of normative expressions. My intention is to grant to the Proposal that all the claims it makes have *true readings*, while explaining that the Proposal fails. (There is a rich literature on objective and subjective uses of normative expressions, including the following: Janice Dowell, “Flexible Contextualism about Deontic Modals: A Puzzle about Information-Sensitivity,” *Inquiry* 56 [2013]: 149-178; Frank Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection,” *Ethics* 101 [1991]: 461-482; Niko Kolodny and John MacFarlane, “Ought: Between Subjective and Objective” [manuscript]; John MacFarlane, *Assessment Sensitivity: Relative Truth and its Applications* [Oxford: Oxford University Press, 2014]; and Holly Smith, *Making Morality Work* [manuscript].)

---

<sup>16</sup> Paul McNamara has some work on the supererogatory in which it might seem that he makes one of my central claims. Both McNamara and I say things like this: “Sometimes one ought to do something that is not morally required, and the thing one ought to do is a supererogatory action.” But McNamara and I use “ought” in quite different ways, so he is not making the same claim that I am making. He uses “ought” to pick out an agent’s morally best actions. So when he says “sometimes an agent ought to do something that is not morally required” he simply means that sometimes an agent’s morally best option is not morally required. And when he says “sometimes an agent ought to perform a supererogatory action” he simply means that sometimes an agent’s morally best option is a supererogatory action. McNamara makes these kinds of claims in his “Supererogation: Inside and Out,” and in other papers. McNamara does not address the question of whether it is ever true that one ought to perform a supererogatory action, all things considered.

<sup>17</sup> Usage of “morally right” varies. Often people use “morally right” to mean “morally permissible”: when they say “that is the morally right thing to do” they pick out a morally required action by using the definite article (“that is the (only) morally permissible thing to do”) and when they say “that isn’t morally right” they mean “that isn’t morally permissible.” (Thanks to Jamie Dreier.) Another usage allows “morally right” to cover any morally good action, including both all morally required and all supererogatory actions (excluding merely morally permissible but morally neutral actions). No one to my knowledge has used “morally right” in the way I suggest in the main text (to apply to all and only behavior that one should perform, all things considered, for moral reasons, including all morally required behavior and also some supererogatory behavior).

---

<sup>18</sup> We might consider a similar proposal regarding “morally ought” and “morally must”. My response to this proposal is similar to my response to Proposal C. It might be a good idea to use “what one morally ought to do” to pick out exactly those ways that one should behave, all things considered, for moral reasons; but that is not how that expression is currently used.

<sup>19</sup> The supererogatory is discussed in Roderick Chisholm, “Supererogation and Offence: A Conceptual Scheme for Ethics,” *Ratio* 5 (1963): 1-14; David Heyd, *Supererogation* (Cambridge: Cambridge University Press, 1982); Mellema, *Beyond the Call of Duty*; Julia Driver, “The supererogatory,” *Australasian Journal of Philosophy* 70 (1992): 286-295; and McNamara, “Supererogation: Inside and Out.” Hallie Liberto (“Denying the Supererogatory,” *Philosophia* 40 [2012]: 395–402) argues that there are no supererogatory actions. Chisholm and Heyd use the word “offence” for supererogatory behavior; Mellema distinguishes two types of supererogatory behavior, “offence” and “quasi-offence.”

<sup>20</sup> Indeed, we might hold, as Heyd does (*Supererogation*), that there are no very seriously bad supererogatory behaviors, but that supererogatory behavior is always only somewhat bad. On this view, behavior that is very seriously bad is always morally wrong.

<sup>21</sup> Other asymmetries between the supererogatory and the supererogatory are argued for by Heyd (*Supererogation*, Chapter 6) and Mellema (*Beyond the Call of Duty*, Chapter 8).

<sup>22</sup> Joshua Gert, “Moral Worth, Supererogation, and the Justifying/Requiring Distinction,” *Philosophical Review* 121 (2012): 4: 611-618.

<sup>23</sup> Horgan and Timmons (“Untying a Knot”) argue that reasons can function in three ways: they can require, justify, or “confer merit” on an action. We might consider the proposal that failing to  $\phi$  is a moral mistake just in case  $\phi$ ing would have merit. This proposal fails because an action might have merit without being the thing that an agent should do.

---

<sup>24</sup> Horgan and Timmons (“Untying a Knot”) argue that a supererogatory act involving great self-sacrifice cannot fulfill an imperfect duty, so they might deny that Betsy’s listening to the boy would fulfill an imperfect duty. They say that the duty of beneficence does not “call for” great self-sacrifice, *therefore* instances of great self-sacrifice do not “fulfill” this duty. But they are wrong to think that a duty cannot be fulfilled by an action that it does not call for. If I promise to give you some food for helping me move, and if buying you some pizza would have fulfilled my promise, but instead I cook you a gourmet meal, then I do fulfill my duty to you. Doing more than is called for can be a way of fulfilling a duty.

<sup>25</sup> As I’ve already said, I am not taking a stand on whether there is any supererogatory behavior, nor on whether, if there is any, all supererogatory options are moral mistakes. Note that behavior of the fourth kind on the list – refraining from supererogatory behavior one should not engage in – may fall entirely within the two kinds listed above it, because it may be that all refrainings from supererogatory behavior are supererogatory; I will not take a stand on this question either. As I read them, Chisholm’s “Supererogation and Offence,” Heyd’s *Supererogation*, and McNamara’s “Supererogation: Inside and Out” do not countenance any supererogatory behavior that it is supererogatory to refrain from, while Driver’s “Suberogation” and Mellema’s *Beyond the Call of Duty* argue that at least some supererogatory behavior is supererogatory to refrain from (though Mellema uses the word “quasi-supererogation” in such cases).

<sup>26</sup> This “puzzle of supererogation” is not the same as what is sometimes called “the paradox of supererogation”, though they are related. The paradox asks how there can be actions that are favored by moral reasons but not morally required. (The paradox is discussed in Raz, “Permissions and Supererogation”; Dreier, “Why Ethical Satisficing”; Horgan and Timmons, “Untying a Knot”; and others.) The puzzle I am discussing asks how someone can take her own



---

action to be favored by moral reasons without also taking it to be morally required; this puzzle is raised in J. O. Urmson, “Saints and Heroes,” in *Essays in Moral Philosophy*, ed. A. Melden (Seattle: University of Washington Press, 1958): 198-216.

<sup>27</sup> Heyd’s *Supererogation* develops the view that supererogatory actions have a special kind of value.

<sup>28</sup> Elizabeth Harman, “Gamete Donation as a Laudable Moral Mistake” (manuscript).

<sup>29</sup> My development of this view also involves my making a further surprising claim that I have not argued for here: that one can have a supererogatory option that is morally worse than a non-supererogatory option. In the gamete donation case, I hold that helping people to conceive is supererogatory, but refraining from donating, so as to avoid creating genetic children to whom one would be unavailable, is a morally better choice, though it does not involve doing anything morally good and is not itself supererogatory.

<sup>30</sup> I develop this view about vegetarianism and accommodation of meat eating in Elizabeth Harman, “Eating Meat as a Morally Permissible Moral Mistake,” in *Philosophy Comes to Dinner*, ed. Andrew Chignell, Terence Cuneo, and Matt Halteman (Routledge, forthcoming). I do not mean to suggest that once stated, this explanation will strike us as the best explanation of vegetarians’ accommodation behavior. My claim is that this explanation becomes available once we recognize the concept of mere moral mistakes.

<sup>31</sup> Thanks to Kirsten Egerstrom for asking me about this.

<sup>32</sup> In Elizabeth Harman, “Morality Within the Realm of the Morally Permissible,” I offer further defense of the existence of morally permissible moral mistakes. I consider four ways that objectors might resist my claim that in some cases, agents should perform particular supererogatory actions, and would otherwise be making morally permissible moral mistakes.

---

One objector claims that such actions are really *morally required*; another objector claims that such actions are *made eligible* by the agent's reasons but are not such that they *should be performed*; a third objector acknowledges that such actions *should* be performed, but claims that *non-moral reasons* always play a crucial role in explaining why this is so; finally, a fourth objector claims that such actions really *should not* be performed. (And these objections could be combined to handle different cases differently.) I argue that while each objector may be right that *some* apparently supererogatory actions are really as described by the objector, we have no reason to believe that *every* apparent case of a supererogatory action that should be performed must actually have the form described by one of these objectors.

<sup>33</sup> That is, if she knew the non-moral features of her situation.

<sup>34</sup> Michael McKenna (*Conversation and Responsibility*, [Oxford: Oxford University Press, 2012]) argues that suberogatory actions show that an action can be blameworthy without being morally wrong, contrary to some standard views of blameworthiness.

<sup>35</sup> Note that particular suberogatory actions may be morally worse, and more blameworthy, than particular morally wrong actions. (See Driver, "Suberogation.") This is similar to the phenomenon that particular supererogatory actions may be less morally good, and less praiseworthy, than particular morally required actions. (See Chisholm, "Supererogation and Offence.")