Varieties of Moral Worth and Moral Credit*

Holly M. Smith

In an earlier article I proposed the following Humean account of blameworthiness:

S is blameworthy for performing act A if, and only if:
1. Act A is objectively wrong,
2. S had a reprehensible configuration of desires and aversions, and
3. This configuration gave rise to the performance of A.¹

A parallel account of praiseworthiness would stipulate that the agent had a commendable configuration of desires and aversions that gave rise to an objectively right act.

An account of blameworthiness or praiseworthiness should answer two questions: it should tell us what makes a person responsible for what she does, and it should tell us what makes a person good or bad for what she does. I now think that the accounts just described do not provide fully satisfactory answers to the question of what makes a person responsible for what she does. For example, these accounts incorporate no conditions on the genesis of the agent’s desires and aversions. Thus they permit an agent to count as blameworthy even though her reprehensible desires might have been instilled in her by a hypnotist or mad neuroscientist. Many theorists, otherwise sympathetic to these accounts, would find this result unacceptable, because they would believe such an agent not to be responsible for what she does. This worry suggests that the accounts

* A distant ancestor of this paper was read at the Society for Women and Philosophy Session in honor of Elizabeth Beardsley, eastern division meetings, American Philosophical Association, Winter 1984. For their helpful discussion, either on the earlier or the more recent paper, I would like to thank Elizabeth Beardsley, Gerald Dworkin, Alvin Goldman, Joshua Hoffman, Geoffrey Sayre-McCord, Ferdinand Shoeman, Susan Wolf, and Michael Zimmerman.

1. Holly M. Smith, “Culpable Ignorance,” Philosophical Review 92 (1983): 556. In this article the definiendum was the concept of being to blame; here I have changed it to the concept of being blameworthy. There is a sense of “blameworthy” (and “praiseworthy”) for which it is not necessary that the agent have performed a wrong (right) act but has merely acted from reprehensible (commendable) motives. I shall not discuss this sense in this article.

Ethics 101 (January 1991): 279–303
© 1991 by The University of Chicago. All rights reserved. 0014-1704/91/0102-0112$01.00

279
cannot provide a full answer to the question of what makes an agent responsible until they are augmented to include some suitable condition concerning the genesis of the relevant desires.

However, the accounts in question are more promising as answers to the question of what makes a person good or bad for what he does. In essence they answer this question by saying it is the configuration of the person’s desires that makes him good or bad for his action. In this article I want to examine this answer. In particular I want to investigate what kind of configuration of desires and aversions counts as reprehensible or commendable, and so renders the person blameworthy or praiseworthy. In the earlier article I made a brief stab at describing the kind of configuration that counts as reprehensible:

Such a configuration might consist for example in a stronger desire to advance one’s own interest than an aversion to harming other people. Note that it is the configuration that is crucial here. Certain desires contribute to the undesirability of the configuration of which they form a part, even though neither their existence per se nor their absolute strength is objectionable. Thus a concern for one’s own welfare is not bad in itself, nor bad even if very powerful, so long as it is counterbalanced by sufficiently strong aversions to harming others. It is the co-existence, or the relative strength, of such desires that is significant. Note also that a configuration may be undesirable because certain desires, such as a concern for other people’s welfare, are missing from it.

This passage leaves a number of questions unanswered. The question which I wish to explore in this article can be introduced by considering the following story. Suppose a certain agent, let us call her Mother Teresa, spends her entire life in the performance of good and even supererogatory deeds. Not only does she consistently perform morally worthy acts, but she is also universally understood as a person who is motivated solely by morally good desires. Not only does she perform admirable acts which most people would be tempted not to perform, or to which they would have the most severe aversion, but in addition she appears completely unmoved by the aversive qualities of the acts. She sees, and is moved by, only the moral good or the moral duty to be done. In order to do what

2. Ibid.

3. The Mother Teresa story was suggested to me by Keith Lehrer in a discussion of these topics. There is some question whether or not any human agent could be motivated solely by morally good desires. Given the conditions of human existence, a human being who was not motivated by normal physiologically based desires, such as the desire for food and warmth, would not long survive. Yet we cannot simply “add” the physiologically based desires on top of the morally good desires, since there are many cases in which the two kinds of desires conflict. Perhaps the kind of agent we have in mind is one in whom the physiologically based desires would somehow disappear whenever they conflict with duty. Alternatively, perhaps the morally good desires include desires for food and warmth, at least when these do not conflict with higher demands of duty. Or perhaps morality requires one to seek food and warmth for oneself except in certain kinds of cases.
is morally right she needs to overcome none of the normal human pulls of self-interest, or greed, or disgust, or overly strong concern for the welfare of friends and family. She is universally admired as a moral paragon. Now let us suppose Mother Teresa dies, and her private diaries are discovered. To everyone's great surprise the diaries reveal that contrary to former belief, Mother Teresa was actually ridden by the same human temptations as the rest of us: she, too, wanted physical comfort, good health, a normal family life, luxuries, and diversions. These desires often dictated the neglect of duty, the performance of morally objectionable acts, and the avoidance of supererogatory conduct. Only continual, powerful exercises of moral will enabled her to resist these temptations and persist in her performance of duty and compassionate acts.

The information revealed in the diaries leads us to reconceptualize Mother Teresa, not as a moral paragon but rather as a superlative example of what we might call a moral resistance fighter. But do we think she is a better or a worse person than we thought before the discovery of the diaries? Do we think she is more or less praiseworthy for being a moral resistance fighter than for being a moral paragon?

Different writers have dealt with this issue in different ways. One strategy which has received a good deal of attention among recent moral philosophers is to say that there are several different concepts or models of praiseworthiness, each exemplifying a different kind of configuration of desires and aversions. Consideration of the Mother Teresa case might lead us to identify three such models. In describing these models, and for the remainder of the discussion, I shall simplify discussion by assuming that the agents in question are fully knowledgeable about the circumstances and consequences of their acts, so that uncertainty or ignorance never lead them to perform an act they did not intend.

1. The moral purity model, according to which an agent is laudable for performing a right act if and only if the agent performs that act out of a morally good desire and in the absence of any desire not to perform the right act.

2. The moral fitness model, according to which an agent is praiseworthy for performing a right act if and only if a morally good desire was present and sufficed for the performance of the right act.

3. The battle citation model, according to which an agent is creditable for performing a right act if and only if a morally good

4. Aristotle, e.g., thought that the moral paragon was more admirable. One way to deal with the question would be to say that Mother Theresa as a moral paragon exemplifies one virtue while Mother Teresa as a moral resistance fighter exemplifies a different virtue. These two virtues are independent of each other, and even inconsistent in the sense that a single individual can only exercise one of the virtues at a time (although someone who is a paragon might also possess, latently, the virtue exemplified by the resistance fighter, in the sense that if he were to suffer countermoral temptations, he would be able to resist them). This analysis presents an interesting problem case for proponents of the "unity of the virtues" thesis.
desire won a hard battle in the war against temptation to perform the wrong act.5

The terms “moral fitness” and “battle citation” I have borrowed from Richard Henson, although he uses somewhat different terminology to describe these models. The contrast between the second and third model has been examined most closely by Elizabeth Beardsley, who labels them, respectively, concepts of moral worth and moral credit.6 I will use Henson’s labels for the models, but have employed Beardsley’s terms in the different adjectives (praiseworthy, creditable) used to describe an agent whose configuration of desires satisfies the model in question. Note that the moral fitness model can be seen as a superordinate concept, of which the moral purity and battle citation models can be seen as subordinate concepts, distinguished by the presence or absence of a desire not to perform the right act.

We can see that Mother Teresa as a moral paragon is laudatory under the moral purity model and praiseworthy under the moral fitness model, but not creditable under the battle citation model, since her (supposed) configuration of desires does not include any temptations against which war must be waged. By contrast, Mother Teresa as the resistance fighter is praiseworthy under the moral fitness model and creditable under the battle citation model. We must, I think, to say that in some sense Sam deserves more honor for his act than Margaret does—because he actually suffers more for doing the morally good act than Margaret does. This is true even if both Sam and Margaret would have been willing to suffer the intense pain for this cause, and indeed would have been equally willing. Nonetheless, Sam actually suffers while Margaret does not. Usually willingness to suffer is closely connected with actual suffering, but not always, as this case shows. The sense in which Sam’s act deserves more honor than Margaret’s reflects the kind of appraisal recognized by the Purple Heart, which is given by the United States military for wounds or death in combat. Heroism in combat is another matter, closer to what I have called moral credit; it is recognized by the U.S. military in such awards as the Distinguished Service Cross, given for exceptional heroism in combat, heroism which may be demonstrated even in the absence of physical injury. We need, then, to distinguish the honor a person deserves for his act—honor bestowed in virtue of the actual personal cost of performing that act—from the moral worth or credit he deserves in virtue of the goodness of his motive or his difficulty in overcoming temptation.

5. Reflection suggests there may still be a fourth kind of positive appraisal of persons for their acts. Consider the following case. Out of benevolence, Sam volunteers to serve as a research subject for an experimental drug that may cure a fatal disease (which Sam does not have). He is assured that the drug will have only minimal side effects. Margaret volunteers for the same experiment for the same reasons. Margaret only experiences minimal side effects, but the drug causes Sam to suffer prolonged intense pain. On the models I have just described, Sam and Margaret would be appraised equally. Yet we want, I think, to say that in some sense Sam deserves more honor for his act than Margaret does—because he actually suffers more for doing the morally good act than Margaret does. This is true even if both Sam and Margaret would have been willing to suffer the intense pain for this cause, and indeed would have been equally willing. Nonetheless, Sam actually suffers while Margaret does not. Usually willingness to suffer is closely connected with actual suffering, but not always, as this case shows. The sense in which Sam’s act deserves more honor than Margaret’s reflects the kind of appraisal recognized by the Purple Heart, which is given by the United States military for wounds or death in combat. Heroism in combat is another matter, closer to what I have called moral credit; it is recognized by the U.S. military in such awards as the Distinguished Service Cross, given for exceptional heroism in combat, heroism which may be demonstrated even in the absence of physical injury. We need, then, to distinguish the honor a person deserves for his act—honor bestowed in virtue of the actual personal cost of performing that act—from the moral worth or credit he deserves in virtue of the goodness of his motive or his difficulty in overcoming temptation.

under the battle citation models, but not laudatory under the moral purity model, since her configuration of desires (as we know them after reading the diaries) does include temptations which must be fought.

Parallel models of blameworthiness can be described as follows:

1. The moral corruption model, according to which an agent is censurable for performing a wrong act if and only if the agent performs that act in the absence of any morally good desire.

2. The moral unfitness model, according to which an agent is blameworthy for performing a wrong act if and only if a nonmoral desire was present and sufficed for the performance of the wrong act.

3. The battle discredit model, according to which an agent is discreditable for the performance of a wrong act if and only if a morally good desire lost the battle against temptation to perform the wrong act.

A sociopath possessing no moral desires and performing a wrong act would be criticizable under the moral corruption and moral unfitness models but not under the battle discredit model. A more normal individual who performs the wrong act because her morally good desires are simply inadequate to the task would be criticizable under the moral unfitness model and the battle discredit model but not under the moral corruption model.

Each of these models seems to express a distinct kind of assessment we make when judging agents’ praise- and blameworthiness. Examination of the models indicates that they ascribe praise- or blameworthiness to an agent in virtue of two kinds of features: (a) the presence or absence of certain kinds of desires (e.g., morally good desires) that led to her act, and (b) the strength of certain kinds of desires that led to her act. The moral purity model, for example, assesses the laudability of an agent in terms of the presence of morally good desires, and the absence of any desire not to perform the right act. A full account of this notion of laudability would clarify which kinds of desires count as “morally good.” For example, it would indicate whether Kant is correct in claiming that the only morally worthy “desire” (or motivational state, more broadly speaking) is respect for duty, or whether there are some intrinsically nonmoral desires, such as an aversion to harming people, which are morally worthy as well, and provide the basis for assessing a person as morally pure in her action. This venerable question is a difficult one that I shall not attempt to answer here. Instead I shall focus on the question raised by the other models, namely, how the strength of a person’s desires

7. In her comments on this paper, Susan Wolf described several examples that suggest the moral quality of a desire cannot simply be “read off” its propositional content. In one of her cases she points out that a desire to tear a live human being to bits and eat him would be morally evil in a person, but hardly so in a tiger. Other cases suggest that the relation of a desire to other desires, or the genesis of a desire, may affect the moral assessment we would make of it.
determines his moral status for the acts he performs. This question has not received adequate attention. In pursuing it I shall leave behind the moral purity and moral corruption models (in which strength of desire is not an issue) in order to focus on the other models. In discussing these models I shall speak generically of morally good and bad desires, without attempting to resolve what the nature of these desires must be.

The moral fitness and unfitness, as well as the battle citation and discredit models, all embody certain ambiguities that need resolution. In investigating how strength of an agent’s desires can affect his praise- or blameworthiness, our first task must be the resolution of these ambiguities.

CLARIFYING THE FITNESS MODELS

According to the moral fitness model, an agent is praiseworthy if and only if a morally good desire is present and suffices for the performance of the right act. But different interpretations of this idea are possible, depending on what notion of “suffices” one employs. In order to lay these differences out clearly, I shall assume, here and for the remainder of this article, that it is possible to assign a numerical measure to the strength of a desire. This assumption will be examined at the end. To simplify discussion I shall also restrict my discussion to agents who have fairly simple psychologies, characterized only by “first-level” desires to perform certain acts (thus disregarding agents who possess “second-level” desires with respect to the possession of their first-level desires). Examination of agents who have more complex (and therefore more realistic) psychologies must be deferred to another occasion.

Let us consider four individuals, each of whom performs a right action motivated by a complex of desires and aversions. Without describing the concrete details of the cases, we can divide the agents’ desires into morally good desires, cooperating desires—desires that have no intrinsic positive moral value but that favor the performance of the right action—and countervailing desires—desires that disfavor the performance of the right action. Both cooperating and countervailing desires may have negative moral value, or they may have neutral moral value; countervailing desires might have positive moral value. For example, a professor might distribute student evaluation questionnaires in her course out of (1) a morally good desire to obey a reasonable university regulation, and out of (2) a (morally neutral) cooperating desire to avoid unpleasant censure from her department chairman if she fails to distribute the questionnaires, and despite (3) a (morally neutral) countervailing desire to avoid collecting data that will make it more difficult for her to achieve tenure. The four cases to be considered are in table 1. In table 1 and subsequent tables, desires disfavoring the right act are given negative scores, while desires favoring the right act are given positive scores. The

---

8. For simplicity I shall assume throughout that there is only one right action in each case.
<table>
<thead>
<tr>
<th>Countervailing Desires</th>
<th>Cooperating Desires</th>
<th>Net CV/CP</th>
<th>Morally Good Desires</th>
<th>Net Desires</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adele .......... -100</td>
<td>+120</td>
<td>+20</td>
<td>+15</td>
<td>+35</td>
</tr>
<tr>
<td>Ben ............. -100</td>
<td>+90</td>
<td>-10</td>
<td>+15</td>
<td>+5</td>
</tr>
<tr>
<td>Carol ........... -10</td>
<td>+12</td>
<td>+2</td>
<td>+15</td>
<td>+17</td>
</tr>
<tr>
<td>David ........... -10</td>
<td>+8</td>
<td>-2</td>
<td>+15</td>
<td>+13</td>
</tr>
</tbody>
</table>

**NOTE.**—Negative scores = desires disfavoring the right act; positive scores = desires favoring the right act; net CV/CP = the net score of the countervailing and cooperative desires; and net desires = the net score of the countervailing, cooperating, and moral desires.

"CV" score represents the strength of the countervailing desires, while the "CP" score represents the strength of the cooperating desires. The "Net CV/CP" score represents the net score of the countervailing and cooperative desires, while the "Net Desires" score represents the net score of the countervailing, cooperating, and moral desires.

In Adele's case, a morally good desire was present and sufficed for the performance of the right act—at least in the minimal sense that, given her total configuration of other desires, the right act was indeed performed. However, her morally good desire was in no way necessary for the performance of that right act; had she had a weaker moral desire, or no moral desire at all, still she would have performed the right act, since the cooperating desires would have carried the day. Let us call a model in which the morally good desire suffices, in the circumstances, for the production of the right act, the "minimal sufficiency model." In such a model, the morally good desire may or may not be necessary for the performance of the right act.

In Ben's case, on the other hand, his morally good desire sufficed for the performance of the right act in a more substantial sense, which we might call "weak sufficiency." Given his total configuration of other desires, the right act was performed; and moreover, his morally good desire was necessary for the performance of that act. Had it been substantially less strong, or not present at all, the right act would not have been performed. His cooperating desires play a major role in his performance of the right act, but they do not carry the day by themselves.

In Carol's case, her morally good desire is not necessary for the performance of the right act; her cooperating desires would have carried the day by themselves. However, her moral desire sufficed for the performance of the right act in a sense that we might call "strong sufficiency": had her cooperating desires not been present, still her moral desire would have been enough to bring about the performance of the right act by itself.

Finally, in David's case, his morally good desire is necessary for the performance of the right act, since his cooperating desires would not have carried the day. Moreover, his morally good desire would have been
sufficient to overpower the countervailing desires even in the absence of any cooperating desires. We may call this “complete sufficiency.” (See table 2.)

The possibility of these different forms of sufficiency raises the question of which form is most appropriate in the context of the moral fitness model for evaluating agents. In order for an agent to be praiseworthy for her right act, must her morally good desires be minimally sufficient, weakly sufficient, strongly sufficient, or completely sufficient for the performance of that right act? Turning to the parallel model of moral unfitness, we can see that it does not lend itself to the same question. For an agent to be morally unfit, she must perform a wrong act out of a nonmoral desire that suffices for the performance of that act. Such an agent is one whose countervailing desires prevail over the combined strength of her moral desires (if any) and her cooperating desires (if any). In such a case the agent is simply morally unfit for the task that faces her; we do not distinguish among different kinds of desires that favor the performance of the wrong act, or the different contributions they make to its performance. Hence we need not ask what form of sufficiency is pertinent.

Returning to the moral fitness model, our question is this: in order for a person to count as praiseworthy for her right act, must her morally good desire be minimally sufficient, weakly sufficient, strongly sufficient, or completely sufficient for the performance of that act?

One way to answer this question would be to say that each of these agents is praiseworthy for his or her action but that there is a progression in terms of degree of praiseworthiness. Thus, it might be said that even though their moral desires have equal strength, Adele is less praiseworthy than Ben, Ben is less praiseworthy than Carol, and Carol is less praiseworthy than David. I find a certain attraction in this idea. However, I also find it hard to say exactly what the ordering should be. Perhaps Adele should.

9. In her comments, Susan Wolf suggested that praiseworthiness might require not just that the moral desire be sufficient to bring about the right act but also that it exceed some minimum level. Since we are focusing here on the strength of desires, this might be filled out as a suggestion that the moral desire must exceed some minimum level of strength. Such suggestions often tie the idea of a “minimum level” to the concept of what motivations are normal, or average, for human beings. However, it is important to avoid any concept of statistical normalcy here. Otherwise it would turn out that we can render a morally mediocre person praiseworthy for her act merely by strategically killing that segment of the population whose morally good desires are stronger than hers. The idea of a minimum level is an important one that I shall not attempt to pursue further in this article.

10. However, we could introduce an additional model of moral criticism in which we distinguish immoral from nonmoral countervailing desires, whose combined force prevails over the combined force of morally good and cooperating desires. We might then be concerned with whether the immoral desires would have sufficed by themselves to prevail over the combined force of morally good desires and the cooperating desires, whether the immoral desires were necessary for the defeat of the good desires, and so forth. Under such a model we would have to consider different forms of sufficiency. I shall not explore such a model here.
<table>
<thead>
<tr>
<th></th>
<th>Countervailing Desires</th>
<th>Cooperating Desires</th>
<th>Net CV/CP</th>
<th>Morally Good Desires</th>
<th>Minimal Sufficiency</th>
<th>Weak Sufficiency</th>
<th>Strong Sufficiency</th>
<th>Complete Sufficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adele</td>
<td>-100</td>
<td>+120</td>
<td>+20</td>
<td>+15</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ben</td>
<td>-100</td>
<td>+90</td>
<td>-10</td>
<td>+15</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Carol</td>
<td>-10</td>
<td>+12</td>
<td>+2</td>
<td>+15</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>David</td>
<td>-10</td>
<td>+8</td>
<td>-2</td>
<td>+15</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**NOTE.**—Negative scores = desires disfavoring the right act; positive scores = desires favoring the right act; and net CV/CP = the net score of the countervailing and cooperative desires.
be least praiseworthy (since her morally good desire seems to play no role at all), and perhaps David should be most praiseworthy (since his morally good desire was necessary for the performance of the right act, and sufficient even in the absence of cooperating desires to secure its performance). But what about Ben and Carol? Should Ben get more credit because his morally good desire plays a crucial role (even though it would not have sufficed by itself), or should Carol get more credit because her morally good desire would have sufficed by itself (even though it was not really necessary for the performance of the right act)? I don't find it easy to say which of these two agents is the most praiseworthy; but neither am I attracted by the idea that they are equally praiseworthy. Perhaps there are just different kinds of praiseworthiness here.

In an influential article in which she investigates related questions from a Kantian standpoint, Barbara Herman argues that an act has moral worth only if it is “no accident” that the morally good desire leads to the performance of the right action. By “no accident” she appears to mean that even were the external, contingent circumstances surrounding the agent’s action different, still he would have performed the right act. Thus, for example, as external circumstances now stand, a shopkeeper has both a moral reason and a cooperating reason of prudence to give inexperienced customers the right change. However, under different social arrangements (e.g., ones in which there is little competition for business, or laws against fraud are not enforced), this would change. Prudence, which is now a cooperating reason, would become a countervailing reason and favor defrauding the inexperienced customer. In order for his act to qualify as morally worthy, according to Herman, the shopkeeper’s moral reason would have to carry the day even in those altered circumstances, since only if it survives this test can we say that it is no accident that his moral desire led to the right act.

In asking what configuration of motives would meet this no-accident test, Herman considers a psychological structure she labels the “greater strength model.” According to this model, the morally good desire not only suffices given the agent’s actual configuration of countervailing and cooperating motives, but it also would have prevailed even if the cooperating motives had “reversed direction” and combined with the countervailing motives to favor the wrong act. An example of an agent whose psychology satisfies this model is Donna (see table 3). “Donna (1)” is Donna as she actually is; “Donna (2)” is Donna as she would be if her cooperating desires reversed direction. If Donna’s cooperating desires had joined her countervailing desires, her morally good desires would have confronted

11. "And when we say that an action has moral worth, we mean to indicate (at the very least) that the agent acted dutifully from an interest in the rightness of his action: an interest that therefore makes its being a right action the nonaccidental effect of the agent's concern" (Herman, p. 366).
12. Ibid.
TABLE 3

<table>
<thead>
<tr>
<th></th>
<th>Countervailing Desires</th>
<th>Cooperating Desires</th>
<th>Net CV/CP</th>
<th>Morally Good Desires</th>
<th>Net Desires</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donna (1)</td>
<td>-20</td>
<td>+30</td>
<td>+10</td>
<td>+60</td>
<td>+70</td>
</tr>
<tr>
<td>Donna (2)</td>
<td>-20</td>
<td>-30</td>
<td>-50</td>
<td>+60</td>
<td>+10</td>
</tr>
</tbody>
</table>

NOTE.—Negative scores = desires disfavoring the right act; positive scores = desires favoring the right act; net CV/CP = the net score of the countervailing and cooperative desires; net desires = the net score of the countervailing, cooperating, and moral desires; Donna (1) = Donna as she actually is; Donna (2) = Donna as she would be if her cooperating desires reversed direction.

a combined countermoral force of $-50$. Nonetheless, they would have prevailed. Donna’s psychology satisfies the greater strength model, and meets the no-accident test.

However, none of the agents described in table 1 have configurations of desires that satisfy the greater strength model, since their morally good desires are too weak to have prevailed over an opposing combination of their other desires. For example, David's morally good desire has strength of $+15$, which would not have prevailed over the $-18$ opposing combined force of his other desires. Hence none of these agents satisfy the greater strength model or meet the no-accident test. On Herman's account, then, none of these agents is morally praiseworthy for his or her act.

These judgments seem too harsh. Some of these agents, and certainly David, are praiseworthy for the acts they perform. (One could object at this point that the notions of “moral worth” and “moral praiseworthiness” are distinct, so that cases in which we deny the presence of moral worth may not be cases in which we deny the presence of praiseworthiness. Herman is interested in the Kantian notion of moral worth, which may or may not have a clear bearing on our notion of praiseworthiness. However, since in this article I am interested in praiseworthiness, I shall take her suggestions as suggestions about this notion, whether or not they were intended that way.) Indeed, Herman herself dismisses the greater strength model, because she rejects the requirement implicit in it that a worthy moral motive must be able to prevail in altered external circumstances (let us call this the “altered external circumstances test”). She argues persuasively that it is enough that the motive is able to prevail in its actual circumstances. If we are assessing an agent’s overall moral virtue it is appropriate to inquire whether his morally worthy motives would have prevailed in altered circumstances, since to be morally virtuous is just to have the disposition to do the right kind of act in a wide variety of external circumstances. But assessing an agent’s moral worth for his performance of a particular act involves a very different judgment from assessing his overall moral virtue. To require that the morally good desire of a praiseworthy agent have the capacity to prevail in different circumstances is, Herman argues, illegitimately to conflate moral worth with
moral virtue.\textsuperscript{13} Possession of moral worth for an act does not require that one would have done the right act even if one found oneself in different circumstances.

This seems correct. Moreover, if we began to assess the worthiness of moral motives by their ability to prevail in altered circumstances, it would be difficult to stipulate what the relevant range of “altered external circumstances” is. In a counterfactual situation in which an agent’s cooperating desires switch their allegiance, so to speak, in order to combine with her countervailing desires, there is every likelihood that the strength of the desires would change as well as their alignment. Thus, in the shopkeeper’s case, if he inhabited a social environment in which laws against fraud were not enforced, so that prudence aligned against duty instead of with it, it is possible that the strength of the shopkeeper’s morally good desire would be less, while the strength of his self-regarding desires would be greater. Would we want to say that his present act is not praiseworthy just because there are possible alternative circumstances of this kind in which his morally good desire would not prevail? This is completely implausible. Yet once we allow performance in altered circumstances to test moral worth in the present circumstances, we may not be able to resist this unacceptable conclusion.\textsuperscript{14}

As Paul Benson has pointed out, Herman appears to have agreed to two inconsistent constraints on the account of moral worth. On the one hand, she requires that morally worthy acts must arise from morally good desires in a nonaccidental way, that is to say, would have arisen from these desires even if the agent’s external circumstances had been different, and therefore his desires had aligned with each other in a different fashion. On the other hand, she asserts that counterfactual failure to do what is right in circumstances other than the actual one has no bearing on an action’s worth.\textsuperscript{15} Although Herman proposes a test

\textsuperscript{13.} Ibid., pp. 368–69.

\textsuperscript{14.} Indeed, in addition to the strengths of actual desires changing, completely new desires might appear in the counterfactual situation, or existing ones disappear. There is a general problem about how to define and limit the range of cases involving “altered external circumstances.” Should we allow counterfactual situations in which the past as well as the present is different, a category in which it would be natural to interpret my case in the text as falling? Should we only allow cases in which the external circumstances of the agent, but not her internal psychological states, change? This is surely too strict, since (as in Donna’s case) we must at least allow the agent’s beliefs and most specific desires to change, in order to link up appropriately with the altered external circumstances. But can we draw a line between changes in the agent’s most specific desires and changes in her less specific, or even basic, desires? To accept an altered external circumstances test would be to open a can of philosophical worms.

\textsuperscript{15.} Benson, pp. 369 ff. Herman herself does not see these constraints as contradictory and proposes a different model of psychological structure that allegedly satisfies both. In this model the moral motive is the only motive that actuates the agent’s choice, even though other motives (favoring and/or disfavoring the right action) may have been in existence at the time of the act. Since I find problems in understanding the idea of a desire that exists
(not described here) for moral worth that she believes meets both constraints, in fact these constraints are inconsistent. We cannot accept both constraints. Since the arguments in favor of rejecting the altered external circumstances test appear compelling, I shall reject the first requirement that the right act’s arising from the morally good desire be nonaccidental. An agent’s moral worth in performing an action cannot be tested by asking whether her morally good desire would have prevailed in altered external circumstances.

Agreeing with Herman, I have argued that an agent’s moral worth for an act does not depend on what she would have done in altered external circumstances. But what about altered internal circumstances? Does an agent’s moral worth depend on what she would have done in altered internal—for example, psychological—circumstances? Is David, for example, to be denied moral worth for doing the right act here and now, even though he would not have done the right act if his countervailing desires had had a strength of −50 instead of their actual strength of −10? Is he to be denied moral worth because he would have failed to perform the right act if his moral desires had had a strength of +1 rather than their actual strength of +15? This seems even more implausible than denying that David’s act had moral worth because he would not have done the right act if his external circumstances had been different. Testing an agent’s moral desire by its capacity to prevail under greater internal temptation may seem appropriate when we are assessing his overall moral virtue, but does not seem appropriate when we are assessing the moral worth of what he actually does on a concrete occasion for action. And testing a moral desire by its capacity to prevail if it were weaker than it actually is seems to have little appropriateness to any kind of judgment. The same kinds of considerations that led me to reject the altered external circumstances test lead me to reject as well the analogous altered internal circumstances test.

If we reject the idea that altered circumstances—either internal or external—provide a test for an action’s worth, we have found reason to reject the greater strength model as our explication of moral fitness, since its rationale relies explicitly on testing for performance in altered circumstances. We also have found reason to reject weak, strong, and complete sufficiency as our explication of moral fitness. Weak sufficiency requires that an agent’s morally good desire be necessary for the performance of the right act: the agent would not have qualified as praiseworthy if the act would not have been performed if the moral desires had been weaker or absent. But this requirement tests the agent’s ability to perform the right act under altered internal circumstances—diminution of the strength of her morally good desire. Both strong and complete

but has no connection with the agent’s choice, even though it is a desire to perform or to avoid performance of the actions available for choice, I shall not discuss this suggestion. See Herman, pp. 369–71.
sufficiency require an agent's morally good desire to prevail even if her cooperating desires had not been present. Here again the agent's ability to perform the right act is tested against altered internal circumstances—alteration of the presence of her cooperating desires. This is a test I reject; hence I can accept no moral difference between Adele, Ben, Carol, and David: in each case the agent’s morally good desire sufficed for the performance of the right act in the actual internal and external circumstances which he or she faced. Despite our initial reaction that these agents may have different kinds or degrees of moral worth for what they do, in fact there are no distinctions in their degrees of praiseworthiness for their present actions. Of course the differences among these agents expressed by the differences between the minimal, weak, strong, and complete sufficiency models might be relevant to judgments about the agents' moral virtue, that is, their respective likelihoods of performing the right act in other cases. But these differences are not relevant to their moral status for what they do here and now. Hence we are left with minimal sufficiency as our remaining explication of moral fitness.

CLARIFYING THE BATTLE MODELS

Let us turn our attention to the battle citation model of praiseworthiness. According to this model, an agent is praiseworthy if and only if a morally good desire won a hard battle in the war against temptation. This model differs from the moral fitness model in requiring the presence of "temptation"—a desire disfavoring the performance of the right act. Within this model it is possible to distinguish four variants, just as it was in the case of the moral fitness model: minimal sufficiency, weak sufficiency, strong sufficiency, and complete sufficiency. According to the minimal sufficiency variant, for example, a morally good desire is present and suffices for the performance of the right act, even in the face of temptation to do wrong. However, the morally good desire may or may not be necessary for the performance of the right act. Even if it is true that the agent would still have performed the right act if she had had a weaker moral desire, or none at all, she would still have performed the right act; she is creditable under the minimal sufficiency model.

Given our previous decision that the altered external and internal circumstances tests should be rejected—that is, that evaluation of an agent should not depend on what the agent’s morally good desires would have accomplished in counterfactual external or internal circumstances—we cannot accept the weak, strong, or complete sufficiency variants of the battle citation model. The distinction between these three variants and the minimal sufficiency model turns precisely on the question of what the agent would have done if her cooperating desires had been different, or if her moral desires had been different. We have rejected the relevance of these questions. Hence we must adopt the minimal sufficiency variant of the battle citation model. The other variants express factors that may be important to judgments of moral virtue but are
irrelevant to judgments of moral credit for what an agent does here and now to resist temptation.  

Turning to the parallel model of battle discredit, we can see that it does not lend itself to parallel variants. According to this model, an agent is discreditable if and only if a morally good desire lost the battle against opposed countervailing desires. By hypothesis, the agent's morally good desire was insufficient. His countervailing desires prevailed over the combined strength of his moral desires and his cooperating desires (if any). Since this model does not distinguish among different kinds of desires that favored the performance of the wrong act, there is no occasion to ask of some of these desires whether they, in the absence of others, would have been strong enough to prevail. Hence we need not ask what form of sufficiency is pertinent. And in any event the answer to this question would not be morally significant, since it would involve testing the agent by reference to his performance in alternative hypothetical circumstances.

COMPARATIVE ASSESSMENTS UNDER THE FITNESS MODELS

In the remainder of this article I shall focus on the minimal sufficiency variants of the moral fitness, moral unfitness, battle citation, and battle discredit models. I shall lay out several different ways in which comparative assessments of praise and blame might be made under these models.

Within these models, how might one compare different persons (or the same person on different occasions) with respect to their degrees of praise- and blameworthiness, or credit- and discreditworthiness? It is

16. Notice this implies that an agent is morally praiseworthy for what he does even if his morally good desire is too weak to have prevailed by itself, and his cooperating desire, which really carries the day against his countervailing desire, is itself a morally evil desire. For example, a judge might (let us assume) perform the right act in imposing a sentence of life imprisonment on a convicted criminal. However, the judge's motives include a weak countervailing desire to show leniency by imposing a lighter sentence (-10), a strong immoral desire to impose the life sentence in order to retaliate against the criminal for alienating the affections of the judge's wife (+100), and a weak morally good desire to see justice done (+5). The judge qualifies as praiseworthy under the moral fitness model, even though none of us may feel much respect for him as a person. But this lack of respect stems from our judgment that such an individual is highly likely to impose unjust punishments in other cases. (The existence of this kind of case was pointed out in discussion by Ferdinand Schoeman [April 1990].) As Michael Zimmerman and Geoffrey Sayre-McCord pointed out in discussion (April 1990) the minimal sufficiency model (like many accounts of praise- and blameworthiness) implies that two agents can differ in praiseworthiness because their external circumstances differ (and therefore their acts differ), even though their basic motivational structures are identical. Thus "luck in one's circumstances" makes a difference to one's moral status. For example, both Charlie and Chuck might have moral desires of strength +15, and desires to avoid physical pain of strength -20. In his circumstances Charlie can do the right act without risking physical pain, so he performs the right act and is praiseworthy; by contrast in his circumstances Chuck can only do the right act by risking physical pain, so he performs the wrong act and is blameworthy. We assess them differently for their acts, even though their basic psychologies are identical. Again, an appraisal of their moral vice or virtue might assess them as equal.
clear that we do make such comparisons in everyday discourse. The following grounds for making such comparisons might be suggested:\textsuperscript{17} (1) the degree of "wrongness" or "rightness" of the act; (2) the degree of evil or good of the motives that led to the act; (3) the likelihood of the act's being right or wrong (in the estimation of the agent); and (4) the strength of the motives that led to the act. Since I am assuming throughout this article that the agents are fully knowledgeable about the circumstances and consequences of their acts, so that uncertainty or ignorance never affect their decision making, we can set aside factor number 3, the likelihood of the act's being right or wrong. Factor number 1 may appear strange; typically we use "right" and "wrong" as all-or-none concepts—either an act is wrong or it isn't. However, it is clear that some acts (taken in themselves, and leaving aside the agent's motives) are morally worse than others. Other things being equal, it is worse to kill someone than to commit theft. This is what I mean by the degree of "wrongness" or "rightness" of an act. However, I believe that degree of wrongness and rightness does not itself affect the agent's degree of blameworthiness or praiseworthiness for an act. Of course the agent's beliefs about the nature of her act affect her degree of blame or praise for it: an agent who does what she believes to be killing a person is more blameworthy than someone who commits what she believes to be theft.\textsuperscript{18} However, this difference arises from differences in the motives that led to these acts (given the agents' beliefs), not differences in the acts themselves. For suppose Rachel and Jane both send ten dollars to CARE. Rachel's money saves the lives of ten children, while Jane's money goes to a country where food is more costly, and so saves only one life. Rachel's act is then a better act in itself than Jane's act, but we cannot conclude that Rachel is more praiseworthy than Jane. Or suppose that John and Peter each steal a piece of jewelry, believing it to be highly valuable, but the "diamond" in John's stolen jewelry turns out to be fake while the diamond in Peter's is real. Peter's act is worse in itself than John's. But we should not conclude that Peter is more blameworthy than John. Hence the degree of rightness or wrongness of the act does not seem directly to affect the degree of praiseworthiness or blameworthiness of the agent. A full discussion of this issue would entangle us in the topic of moral luck, which is beyond our current scope. However, I will provisionally reject factor number 1 as irrelevant.


\textsuperscript{18} Actually, matters seem more complicated than this. Suppose two agents risk their lives to swim out to save drowning creatures: one believes (correctly) her creature to be a dog, while the other believes (correctly) her creature to be a child. Since the moral imperative to save the life of a dog is presumably less stringent than the moral imperative to save the life of a child, the dog-saving agent does something more highly supererogatory than the child-saving agent, and might well be judged more praiseworthy on that account, even though her planned act is less obligatory.
TABLE 4

<table>
<thead>
<tr>
<th>Countervailing Desires</th>
<th>Morally Good Desires</th>
<th>Battle Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fred</td>
<td>-100</td>
<td>+150</td>
</tr>
<tr>
<td>Gail</td>
<td>-1</td>
<td>+3</td>
</tr>
</tbody>
</table>

**NOTE.**—Negative scores = desires disfavoring the right act; positive scores = desires favoring the right act.

Clearly the degree of evil or good of the motives that led to the act (factor number 2) is relevant to the agent’s degree of praise or blame, credit or discredit. An agent who does wrong out of cruelty is more criticizable than one who does the same wrong out of cowardice. Indeed, difference in good and evil of motives probably explains our intuition that factor number 3 (the likelihood of the act’s being right or wrong) is relevant to the degree of criticism the agent deserves. An agent who tolerates the strong possibility of causing great harm is more to blame than one who merely tolerates the weak possibility of causing great harm.19

But what about factor number 4—the strength of the motives that led to the action? Can different degrees of strength of motive affect how criticizable the agent is?

I think it is clear that degrees of strength are relevant to our judgments here. Consider the following pair of cases (from now on I will simplify the discussion by only considering cases in which there are no cooperating desires, since they play no direct role in judgments under the models we are considering). Let us assume that Fred and Gail performed their right acts out of the same kind of good motives (e.g., respect for duty), and despite resistance from the same kind of countervailing desire (e.g., self-interest). (See table 4.) Here both Fred and Gail are creditable in the battle citation sense; in each case, the agent’s morally good desire prevailed over the countervailing desires. However, Fred faced a harder trial than did Gail, and his morally good desire was equal to the harder task set for him. We may judge that Fred is more creditable than Gail, and that judgment must be based on the comparative strength of the agents’ desires, since we make the judgment even though there is no difference in the intrinsic goodness of their morally good desires, or the intrinsic evil of their countervailing desires.

But how precisely does strength of motive make a difference in comparing the degree of criticism or praise two agents deserve? Let us focus first on evaluations under the moral fitness and moral unfitness models.

According to the weak moral fitness model, the agent is morally praiseworthy when her morally good desire is sufficient, given her other desires, to lead to performance of the right act; her morally good desire

may or may not be necessary for the performance of the right act. In comparing the degrees of praiseworthiness of two agents, clearly one possible ground for comparison would be the agents' absolute strengths of moral commitment. Consider the two cases presented in table 5. Both Ophelia and Peter are praiseworthy, under the moral fitness model, for their acts; however, Peter, who has the greater commitment to morality, appears more praiseworthy than Ophelia.

Similarly, an agent who has a stronger commitment to morality appears less blameworthy under the moral unfitness test when he does wrong than an agent with a weaker commitment. Consider the two cases in table 6. Here Steven appears less morally fit than Rachel, and so more blameworthy. Let us say that such judgments issue from an "absolute model" for evaluating degrees of moral worth.

What underlies these judgments? Why does Peter appear more morally fit than Ophelia? Each of these agents had sufficient commitment to morality to bring about performance of the right act. From one point of view, Peter's greater commitment to morality has no particular function or value in the actual circumstances in which he acted. Similarly, neither Rachel nor Steven had a sufficiently strong commitment to morality to meet the test each of them faced. The mere fact that Rachel had a greater commitment than did Steven did not enable her to do better than he in the actual circumstances. Of course, the greater commitments of Peter and Rachel would have enabled them to fare better than Ophelia and Steven in alternative circumstances—but I have argued that how an agent would do in alternative circumstances is no test for how praise- or blameworthy the agent is in the actual circumstances. For this reason we cannot accept greater absolute strength of moral commitment as a comparative measure of praise- and blameworthiness if we are covertly using it as an indication that the agent would do better in alternative circumstances. However, we may care about absolute strength of moral commitment for other reasons, not having to do with potential performance in counterfactual situations. We may simply think that a stronger moral commitment is morally superior. If we judge that one agent is better than another for what he does in the present circumstances because he does it from a stronger—and therefore better—moral commitment, then the absolute model can be accepted as one model for assessing comparative degrees of praise- and blameworthiness.
Let us consider a second possible model for comparing degrees of blame and praise by reference to strengths of desires. On what we can call the “adequacy model” for measuring degrees of moral worth, what matters is not just the absolute strength of the agent’s commitment to morality but, rather, how adequate her commitment is to the task at hand. Thus an agent who does wrong, but whose moral commitment was almost strong enough to overcome her countervailing desires, would be less blameworthy than one who does wrong from a weak moral commitment that falls far short of the strength necessary to overcome temptation. The latter agent is more criticizable because she was less fit for the task that actually faced her. An agent who does what is morally required, and whose commitment to morality far exceeds what was necessary to overcome opposition, would be more praiseworthy than one who does right from a moral commitment that was barely sufficient to accomplish its job. The former agent was far more fit for the task that actually faced her. These judgments can be seen in the cases shown in table 7 and table 8. On the adequacy model, Teresa’s act is more praiseworthy than Van’s, since her commitment to morality exceeds the necessary level by a greater amount than does Van’s. On the adequacy model of moral unfitness (see table 8), Wendy’s act is more blameworthy than Zeb’s, since her commitment falls short of what is required by a greater amount.\(^{20}\)

The absolute and adequacy models deliver different assessments in the same case. For example, in table 8, the absolute model appraises Wendy and Zeb as having the same degree of blameworthiness, while

\(^{20}\) Agents may be judged morally fit or unfit even in cases where there is no countervailing desire. In such cases, the absolute and adequacy models would deliver the same assessments of comparative blame and praise. Here again we can ask whether the degree to which an agent’s commitment to morality exceeds, or falls short of, the necessary strength should really affect his praise- or blameworthiness. One could say that what counts is whether the strength of this commitment is sufficient to do the job. How much it exceeds, or falls short of, this necessary strength is not really relevant to how fit the agent is for this particular moral task. It may tell us something about how well the agent would do in alternative circumstances (where the challenge might be harder), but we have rejected this factor as irrelevant to the agent’s moral status for the present act. If this is our reason for being attracted to the adequacy model, we should reject it. However, we may retain it if our underlying thought is simply a judgment that the person more morally fit for the task facing him is morally better than the less morally fit person.
the adequacy model assesses Wendy as being more blameworthy. These models capture different, and sometimes incompatible, ways of judging people's moral character for what they do. A full account of our implicit moral theory must contain them both.

COMPARATIVE ASSESSMENTS UNDER THE BATTLE MODELS

Let us now turn our attention to the battle citation and battle discredit models. The central idea of an act that is creditable or discreditable under these models is the idea of an act that is morally difficult for the agent to perform, difficult because the agent must resist the temptation to neglect her duty. But how might we measure, or conceive of, “moral difficulty” in performing an act? Reflection on the concepts of difficulty that are used in other contexts suggests we possess two rather different models for assessing degrees of difficulty. Consider how we might assess the difficulty of leaping a given hurdle by a given athlete. One assessment would turn simply on the height of the hurdle: how high the athlete must jump in order to clear it. On this kind of assessment, an athlete who must jump a three-foot hurdle faces a more difficult task than an athlete who must only jump a two-foot hurdle. A contrasting type of assessment would take into account not just the height of the hurdle but also the athlete's capacity to clear it. On this view, a seasoned athlete facing a three-foot hurdle would have a less difficult task, than, say, a six-year-old child facing a two-foot hurdle.

Both these models for assessing degrees of difficulty can be extended into the sphere of moral evaluations under the battle citation and battle discredit models. On the first model, which we might call the “simple model,” the level of difficulty of performing an act is simply a function of the “hurdle”: the strength of the agent's countervailing desires not to perform the act. An agent whose countervailing desires are stronger has

---

**TABLE 7**

<table>
<thead>
<tr>
<th>Countervailing Desires</th>
<th>Morally Good Desires</th>
<th>Gap</th>
<th>Moral Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teresa . . . . . . . . .</td>
<td>−1</td>
<td>+20</td>
<td>+19</td>
</tr>
<tr>
<td>Van . . . . . . . . .</td>
<td>−18</td>
<td>+20</td>
<td>+2</td>
</tr>
</tbody>
</table>

_Note._—Negative scores = desires disfavoring the right act; positive scores = desires favoring the right act.

---

**TABLE 8**

<table>
<thead>
<tr>
<th>Countervailing Desires</th>
<th>Morally Good Desires</th>
<th>Gap</th>
<th>Moral Unfitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wendy . . . . . . . . .</td>
<td>−100</td>
<td>+20</td>
<td>−80</td>
</tr>
<tr>
<td>Zeb . . . . . . . . .</td>
<td>−25</td>
<td>+20</td>
<td>−5</td>
</tr>
</tbody>
</table>

_Note._—Negative scores = desires disfavoring the right act; positive scores = desires favoring the right act.
TABLE 8

<table>
<thead>
<tr>
<th></th>
<th>Countervailing Desires</th>
<th>Morally Good Desires</th>
<th>Gap</th>
<th>Moral Unfitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wendy</td>
<td>-100</td>
<td>+20</td>
<td>-80</td>
<td>Yes</td>
</tr>
<tr>
<td>Zeb</td>
<td>-25</td>
<td>+20</td>
<td>-5</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**NOTE.** — Negative scores = desires disfavoring the right act; positive scores = desires favoring the right act.

a harder task than one whose countervailing desires are weaker because the former has more resistance to overcome. The psychological “hurdle” she must clear is higher. Thus an agent who fails to do her duty out of severe temptation is less discreditable than one who fails to do her duty out of weak temptation. And an agent who fulfills her duty in the face of strong temptation is more creditable than one who fulfills her duty in the face of weak temptation. The former managed to clear a higher hurdle. This model helps to explain why we think Fred is more creditable than Gail in table 4. It also explains why we think Ivy is less discreditable than Jack in table 9. Neither Ivy nor Jack succeeds in doing the morally right act; but Ivy faces more severe temptation than Jack, and so is less discredited by her failure.21

On the second model, which we might call the “capacity model,” the level of difficulty is a function not only of the strength of the agent’s countervailing desires but also of his capacity to overcome them—a capacity measured by the strength of his opposing morally good desires. It is, so to speak, the gap between his countervailing and morally good desires that counts, not the absolute strength of the countervailing desires, or the absolute strength of the morally good desires. Thus an agent who fails to do his duty, and whose morally good desire falls far short of his countervailing desire, is less discreditable than one who fails but whose moral desire falls short by only a small amount from being powerful enough to overcome his temptation. It would have been harder for the former agent to overcome his temptation, since a greater increase in the strength of his moral desire would have been necessary. By the same token, an agent who fulfills her duty, and whose morally good desire greatly surpasses the strength of her temptation to do wrong, receives less credit than one who surmounts temptation but just barely. The former agent underwent a less challenging trial, just as the seasoned athlete leaping the three-foot hurdle undergoes a less challenging trial than the six-year-old child leaping the two-foot hurdle.22

21. Note that the countervailing desires of these agents might be further subdivided into morally neutral and morally evil desires; how much strength their desires in these categories possess might also affect how discredited they are.

22. Note that the capacity model might be understood as covertly invoking either the altered external or altered internal circumstances tests. But it need not be so understood. We can think of it as measuring something else we care about: simply how “far” the agent's
model explains why we think Karen is more discreditable than Len (see table 10). Although both Karen and Len faced the same challenge in the absolute sense (in each case the strength of the temptation was −20), still it would have been harder for Len to triumph over his temptation, given the strength of his morally good desires, than it would have been for Karen. Since the gap between his morally good desires and his countervailing desires is greater, he would have had to strengthen his moral desire more in order to prevail. Karen had only a small alteration to make in order to succeed. Hence Karen is more discreditable for her failure than Len is for his. The capacity model also explains why we think Mary is more creditable than Ned in the cases shown in table 11. Here Mary and Ned face similar “hurdles” in the sense that their degrees of temptation to do wrong are equally strong (−8). On the simple model their degrees of credit would be the same. However, Mary undergoes a more challenging trial than does Ned in the sense that the strength of her commitment of morality is just barely sufficient to overcome temptation, whereas Ned’s commitment is hardly tested by this choice. According to the capacity model, she is more creditable than he.

Another, and perhaps more intuitive way, we might understand the capacity model, as it applies to moral judgments, is as follows. Suppose we view people as having two levels of desire-systems: first-level desires to perform or not perform the acts in question, and second-level desires to have or not to have certain first-level desires. On this conception, an agent who does wrong might have had a second-level desire to have first-level desires of different strengths than his actual ones—first-level desires that would have led him to do his duty, instead of leading him into wrong-doing. Let us assume that second-level desires can affect the strength of first-level desires. Then a wrong-doing agent whose first-level commitment to morality is almost, but not quite, strong enough to produce the right act, would only have needed a weak second-level desire (an “effort of will”?) to improve his first-level desires sufficiently to have brought about the desired change. But a wrong-doing agent whose first-level commitment to morality was far from being strong enough to overcome temptation would have needed a very powerful second-level desire for improvement in order to have brought about the desired change. Hence it was more difficult for the second agent to alter his first-level desires sufficiently, and we should give him less discredit for failing. Somewhat parallel remarks hold for cases of positive moral credit. A right-doing agent whose (first-level) commitment to morality far surpasses her countervailing desire receives less credit for her success because it took less powerful desires on the second level to maintain the requisite first-level motives. She could have let her first-level desires “slip” a bit without endangering performance of the right act.
In the last two sections we have used rejection of the altered circumstances tests to argue that the fitness and battle models of moral evaluation should be understood according to the "minimal sufficiency" interpretation. However, when we asked how comparative judgments of different agents should be made according to these models, we discovered that each model sustains two quite different types of comparisons: the absolute and adequacy models (in the case of judgments about moral fitness and unfitness), and the simple and capacity models (in the case of judgments about battle credit and discredit). These ways of comparing agents give us quite different judgments about the relative worth of agents for their actions. In some cases the judgment delivered by one model may not only differ from that delivered by another model, but even seem positively wrong-headed from the latter perspective. For example, on the adequacy model, Karen is less blameworthy than is Len (since her level of moral commitment brings her closer to overcoming temptation), while on the capacity model Karen is more discreditable than is Len (since it was less of a challenge for her, given her level of commitment, to overcome temptation). The judgments issued by these two models always contradict each other. Nonetheless all these models seem necessary to capture the full range of our assessments of agents for their acts. Our moral thought is too complex to be adequately represented by any single model of assessment.23

CONCLUSION
This article began by delineating familiar accounts of blameworthiness and praiseworthiness. I have suggested that while these accounts might be deficient as accounts of what makes agents responsible for their acts, they hold more promise as accounts of the grounds for which agents are

23. An unanswered question here is how creditworthiness/discreditworthiness are affected by the joint factors of (1) strength of desire and (2) degree of evil or good of the desire. Is a person with a weak cruel streak, who resists weak temptation, more or less praiseworthy than a person with a strong degree of cowardice, who resists strong temptation? There is some temptation to say that moral quality of desire does not affect moral credit and discredit at all (although it clearly affects praiseworthiness and blameworthiness). It is no harder to resist a strong evil desire (say, a desire to cause pain) than it is to resist an equally strong but morally neutral desire (say, a desire to avoid pain for oneself). Thus it is plausible to say that the person who successfully resists the former is no more creditable than the person who successfully resists the latter.
criticizable or laudable for what they do. According to these accounts, the character of an agent's configuration of desires explains the assessment we make of her for what she does. The fitness and battle models were described as representing different kinds of judgments we make about a person, and I argued that these models must be understood as involving a notion of the agent's morally good desire as being "minimally sufficient" for the performance of the right act. I then turned to the question of how comparative judgments of degree of praise and blame, credit and discredit could be made under these models. Focusing on the role that an agent's strength of desire plays in such judgments, I articulated a number of different models for measuring degrees of worth and credit. These models deliver divergent assessments when applied to a given case but seem necessary to capture the full range of our complex thought about how to appraise a person for what she does.

The existence of all these models for evaluation raises a number of questions. In particular, we need to know how they link up to other features of moral thought. For example, is degree of moral credit (as measured by the simple model) linked to one kind of reward or sanction, while degree of moral credit (as measured by the capacity model) linked to another? If an agent is seriously discreditable on the simple model, does that justify others in imposing one kind of rehabilitative effort, while serious discredit on the capacity model justifies another sort of rehabilitation? These are difficult issues that must be deferred to another occasion.

In closing let me point out one final implication of this discussion. Throughout I have assumed that it is possible to impose numerical measures on strengths of desire. This assumption was not really essential to the project of differentiating the moral purity and corruption models, the moral fitness and unfitness models, and the battle citation and discredit models, or their variations, since these models could have been differentiated by using the language of necessary and sufficient conditions alone. However, in characterizing the absolute, adequacy, simple, and capacity models for comparing degrees of moral worth and credit, the assumption about numerical measures was critical: these models require not only that we be able to impose cardinal measures on the strength of desires but also that we be able to make interpersonal comparisons of strength of desire. For example, we could not have said that Ivy is less discredited than Jack on the simple model of battle discredit (in table 9) if we could not have compared the strengths of their respective temptations to perform the wrong act.24

24. In a personal communication (March 1990), William Harper has suggested to me that all that may be required for the requisite comparative judgments is measurement of the strength of the desire relative to the strongest and weakest desires in the agent's repertoire of desires. (I need to resist a countervailing desire that is halfway between my strongest and weakest desires, while you need only resist a countervailing desire that is a third of the way between your strongest and weakest desires.) Such measurements can be made without involving interpersonal comparisons. I am not inclined to think this strategy captures the judgments we want to make, but the suggestion is worth exploring.
Of course such measures and interpersonal comparisons are suspect. We could reject them, and in doing so reject these four methods of comparing worth and credit. However, my own strong feeling is that we do in fact compare moral worth and credit on these grounds, and so do assume at some level that strengths of desires can be measured. As we all know, one of the primary objections to utilitarianism is the criticism that it assumes interpersonal comparability of utility, and in particular interpersonal comparability of preference strength. What we have just seen is that any moral theory that includes the concepts of moral worth and credit may be committed to the idea that interpersonal comparability of preferences is possible. If so, then utilitarianism is no worse off in this regard than any deontological theory with a sufficiently rich account of personal responsibility. This is no small side-effect of our investigation—I will resist the temptation to say “no small benefit.”