

UNPRINCIPLED VIRTUE



AN INQUIRY INTO MORAL AGENCY

Nomy Arpaly

OXFORD
UNIVERSITY PRESS

2003

THREE

MORAL WORTH

The previous chapter demonstrated that reasons for action need not enter into consciousness to be excellent reasons or to justify actions, and similarly that those reasons entering consciousness may be the worst of one's reasons for acting, failing to give substantial justification to any action at all. Thus, awareness of one's reasons need play no role in an evaluation others might correctly give of those reasons. These conclusions will serve us well as we turn from rationality to morality and ask about the moral significance of those reasons, conscious and unconscious, for which agents act.

Moral Worth

Sometimes a person does the right thing, but we are not particularly impressed. The reader of Gustave Flaubert's *Madame Bovary* is told of the protagonist's good works:

She sewed clothes for the poor, she sent wood to women in childbirth; and on coming home one day, Charles found three tramps eating soup in the kitchen. Her little girl, whom her husband had sent back to the nurse during her illness, returned home. She wanted to teach her to read; even Berthe's crying no longer irritated her. She was resigned,

universally tolerant. Her speech was full of elevated expressions. She would say: "Is your stomach-ache any better, my angel?" (Flaubert 1965, 155)

The average reader does not admire Madame Bovary for her charitable actions. Neither does Flaubert, who refers to her as "indulging" in "excessive charity." But why, really? Madame Bovary's charitable actions are clearly not "excessive" in any sense that implies that it is somehow morally wrong of her to perform so many of them. Another character in another novel who sewed clothes for the poor and sent wood for women in childbirth might simply be deemed charitable, her feeding of the tramps humane, and her use of sentimental platitudes unimportant. But Madame Bovary's good works, we are told, are something she takes up after a long illness brought about by the traumatic end of her first extramarital affair and are part of a conversion that, for all its remarkable intensity ("She saw a state of purity floating above the earth, mingling with heaven. She wanted to become a saint" [154]), is relatively short. Madame Bovary loses her interest in orphans as soon as her next lover comes along.

Perhaps Madame Bovary's charitable acts are excessive in the sense that it is excessive for a teenager to spend all of her savings on a hobby she is to abandon next week or to spend every waking moment with a certain boy due to a passing infatuation. Madame Bovary's actions appear to be motivated by a mere infatuation with morality, as opposed to a serious love. Perhaps her actions are indulgent in that even though she desires to be moral, and performs her good works because they are moral, one suspects that she has an ulterior motive of sorts. Like the romantic who is in love, not with her lover but with love itself, Madame Bovary seems to be in love not so much with morality but with the romance of morality—with an image of herself as one of the grand ladies who retired into solitude to shed at the feet of Christ the tears of hearts that life has wounded. The desire to become the kind of person who cares about morality may be a worthy motive, but something about Madame Bovary's dwelling upon the long-trimmed trains of these ladies' long gowns, something about the quasi-sexual passion with which she kisses the image of Christ, seems suspicious—as if there were more than one way in which one might aspire to be devoted to morality, and this is not the right one. Why the same actions prompt us to morally

praise (or condemn) some agents more or less than others is what I call the question of moral worth.

A Few Clarifications

The moral worth of an action is the extent to which the agent deserves moral praise or blame for performing the action, the extent to which the action speaks well of the agent. I will speak interchangeably of "a morally praiseworthy action" and "an action with positive moral worth." Deviating from the Kantian use of the term 'moral worth', I will also speak interchangeably of an action with negative moral worth as an action for which the agent is blameworthy and as a "morally blameworthy action." Obviously, the extent to which an agent deserves praise or blame for her action depends in part on the action's *moral desirability*. We talk about an action's moral desirability when we ask whether it is right or wrong, or how grave a wrong it is, or whether it is the best possible action. To send money to Oxfam is morally desirable; to speed, while intoxicated, through a residential neighborhood is not, and so on.

Two actions that are equal in moral desirability may be of different moral worth. To give a simple example, two people may donate equal amounts of money to Oxfam, but one of them may do so out of concern for improving the state of the world, while the other does so purely at the urging of her accountant. Even if the two agents' charitable actions are equally morally desirable—both of them have done the right thing—it is not true that both agents deserve the same degree of praise. Similarly, a person who is rude to her colleagues due to the stress caused by grave news may merit less blame than a person who is rude to her colleagues due to her belief that her status as a mathematical genius allows her to dismiss the feelings of mere mortals.

While this distinction between moral desirability and moral worth appears to be trivial, it is often ignored in casual (and sometimes serious) philosophical discussion. For example, one often hears that "Kant is concerned with the motives for our actions, while utilitarians only care about the results." If one pays attention to the distinction between moral desirability and moral worth, one finds this classroom cliché to be quite false. When it comes to moral desirability, Kantians are not interested in motives for actions but in whether the action, under a certain

description, is permitted by universal law. Recall Kant's Prudent Grocer, who prices his merchandise fairly because a reputation for honesty tends to increase his profits. Despite the Prudent Grocer's unimpressive motive, Kant never denies that the grocer does *the right thing* or that he performs the action required of him by duty. In this sense, Kantians clearly care about results. On the other hand, when it comes to moral worth, a number of utilitarians are just as concerned about motives as the Kantians. In fact, just as Mill (1979, 17–18) repeats the claim that the rightness or wrongfulness of actions has nothing to do with the agent's motives, he also repeats the claim that motives are relevant to our moral evaluation of the agent:

The motive has nothing to do with the morality of the action, though much with the worth of the agent. . . . The motive, that is, the feeling which makes him will so to do, if it makes no difference in the act, makes none in the morality [of the act]; though it makes a *great difference in our moral estimation of the agent*. (Italics added)

As the moral worth of an action involves that kind of "estimation" merited by the agent for performing the action, it is not implausible to think of Mill as allowing motives to be relevant to the moral worth of actions, even if they have nothing to do with the rightness or wrongness of the action. And quite apart from the right interpretation of Mill's text, it is perfectly consistent to view the moral desirability of actions as depending entirely on their expected consequences, and the moral worth of individual actions as depending to some degree on the agent's motives. One can believe that giving to charity is desirable because it promotes happiness, but that an agent giving to charity out of a desire to promote happiness deserves more praise for her action than does her counterpart who is merely concerned with her tax situation or with the impression she makes on her peers; and one can believe so even while holding that we are morally required, for the sake of utility, to lavish indiscriminate praise on all who help the poor.

The last point made—that the moral worth of an act need not correspond to the moral desirability of treating an agent as if she were praiseworthy or blameworthy—is worth further elaboration. When I say that an agent is praiseworthy (or blameworthy) for an action, I do not mean to imply that the agent should necessarily be praised (or blamed) for this

action: what people deserve is not always what they should be given. If, for example, an armed criminal enters a crowded room and shouts, "Give me some moral praise, or I shall kill everyone," it may be morally imperative to praise her, but that alone does not make her praiseworthy for her action. The praise that we may be required to give her is *unwarranted*, or undeserved. The moral worth of an action is the extent to which the agent deserves praise or blame for the action, not the extent to which the agent should be morally praised or blamed for it. The purpose of this chapter is to capture the conditions under which such praise or blame is *warranted*, not those under which it is *required*.¹

Responsiveness to Moral Reasons

Consider again the case of the prudent grocer. One does not need to know the details of Kant's discussion to agree that there is a sense in which the grocer, who is motivated only by a desire for profit, is not particularly praiseworthy for his policy of fair pricing. One is happy, of course, that the grocer does the right thing. But one cannot shake off a sense that this is a mere accident. The grocer aims at increasing his profits. By a lucky accident, it so happens that the action that would most increase his profits is also a morally right action. While this is all well and good, one is not inclined to give the grocer moral credit for this accident. But what, exactly, is this "accidental" quality that we perceive in the grocer's doing of the right thing? It is *not* simply the fact—if it is a fact—that the profit motive does not reliably produce moral actions. We can, with some difficulty, imagine a world in which some invisible hand or other makes it true that the profit motive reliably produces morally right actions, and we can place Kant's grocer in that world, and still we shall not free ourselves from the sense that there is something accidental in the fact that he does the right thing. It is accidental in the same way as it is accidental that a person who reads *Lolita* for the love of scandal reads an aesthetically superior book, or the fact that a person who buys cheap beer because he likes it accidentally makes a money-saving choice. The former is attracted to the novel for reasons that are of no interest to the aesthetician who pronounces it beautiful, the latter is attracted to cheap beer for reasons that are of no interest to the thrifty, and Kant's grocer is attracted to fair pricing for reasons that are of no interest to the ethicist.

The salient feature of Kant's case, I would like to suggest, is that the grocer's morally right action does not stem from any responsiveness on his part to moral reasons. In pricing fairly, the grocer acts for a reason that has nothing to do with morality or with the features of his action that make it morally right. The reasons for which he acts have to do only with his own welfare; and whatever it is that makes his action morally right, the fact that his action increases his welfare is certainly not what makes it morally right. His reasons for action do not correspond to the action's *right-making features*. An important truth about moral worth seems to be the following:

Praiseworthiness as Responsiveness to Moral Reasons (PRMR):

For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons—that is, the reasons for which she acts are identical to the reasons for which the action is right.

To put the point another way, what we praise in the morally praiseworthy agent is her responsiveness to moral reasons. Which reasons exactly are moral reasons is not a question I can deal with here, as the moral reasons to perform an action are the same reasons that make the action right, and what exactly makes an action right is a question that Kantians, utilitarians, Aristotelians, and others are still debating. Here I will make do with intuitive assumptions: what makes the grocer's action right is surely not the fact that it increases his profits and surely has to do with fairness. There are also action-theoretic concerns that will need to be left for another occasion, such as those having to do with hierarchies of motives and intentions. But I trust that the notion of the reasons for an action's rightness being the same as the reasons for which an agent acts (i.e., an action is right because it alleviates the suffering of a person, and an agent performs it in order to alleviate the suffering of a person) makes sense. PRMR is the intuition that Kant invokes when he argues that only the good will is necessarily good, and Aristotle appeals to it when he makes clear his view that defending one's city would not be virtuous if motivated by the desire for fame or a fee. Thus, it may look tempting, or even trivial, to move from PRMR to the more elegant Kantian claim that all and only morally praiseworthy actions are right actions performed *from duty*, or the venerable Aristotelian view that to act virtuously one needs

to perform fine actions *for the sake of the fine*. But as these claims are often understood in contemporary discussion, such a move would be an error. Usually, "acting from duty" and "acting for the sake of the fine" are taken to indicate not simply acting for moral reasons, but rather acting for reasons *believed or known* to be moral reasons—acting out of a desire to do that which is right. But moral worth is fundamentally about acting for moral reasons, not about acting for reasons believed or known to be such, and distinguishing the two is important in evaluating moral agents.

Moral Responsiveness *De Re*

Consider the Kantian claim that all and only morally praiseworthy right actions—or, in his language, all and only dutiful actions that "have moral worth"—are performed from duty. This claim is often understood as stating that an agent who does the right thing deserves moral praise for her action if and only if she does the right thing out of concern for morality: an agent is morally praiseworthy for her action if and only if she did the right thing because she was concerned with doing the right thing. In "On the Value of Acting from the Motive of Duty" (1993), Barbara Herman phrases the doctrine of the motive of duty in the following way:

For a motive to be a moral motive, it must provide the agent with an interest in the general rightness of his actions. 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 non-accidental effect of the agent's concern. (1993, 6)

Herman is mistaken if "an interest in the rightness of his action" is interpreted in the most obvious way—that is, as an interest in doing the right thing or the moral thing *under this description*, in a *de dicto* sense: a concern for doing what one feels or believes, even as a background belief, that one morally ought to do. For a right action to have (positive) moral worth, it is neither sufficient nor necessary that it stem from the agent's interest in the rightness of his action.

To see that it is not *sufficient*, one can look at cases in which an agent does the right thing out of concern for doing the right thing, and

still the fact that he did the right thing appears to be an accident. This is often the case when the agent has a mistaken view of moral rightness, and thus no matter how earnestly he tries to do the right thing, he still does not act for moral reasons. Here is such a case:

The Extremist: After the assassination of Yitzhak Rabin, some Jewish extremists expressed the opinion that the murder was a horrible thing simply because it involved a Jew killing a Jew. Imagine for a moment that Ron is such an extremist, believing deeply that killing a person is not generally immoral but that killing a fellow Jew is a grave sin. Ron would very much like to kill Tamara, but he refrains from doing so because he wants to do the right thing and he believes the right thing to do is to refrain from killing Jews like Tamara.

Here is a case of a person doing the right thing—refraining from killing Tamara—because he very much wants to do the right thing and he believes, rightly, that refraining from the murder would be the right thing to do. He does not, however, do the right thing for the relevant moral reasons. The morally right reasons not to kill Tamara have to do with her being a person; they have nothing to do with her ethnicity. Ron's reasons for (in)action have nothing to do with the right-making feature of his (in)action. Hence the impression that it is merely *accidental* that Ron did the right thing in this case. After all, he might just as well have done the wrong thing had Tamara been Latino instead of Jewish. Just as, in the case of the prudent grocer, it is fortunate that the prudent action also happens to be the moral one, in Ron's case it is fortunate that favoring Jews in a certain way (not killing them when one would otherwise like to) is moral. Ron is not morally praiseworthy for his action, for he does not act for the morally relevant reasons, but rather for reasons that he mistakenly believes to be morally relevant. (There will be more on the likes of Ron in the sections on misguided conscience.)

Thus, for an agent to be morally praiseworthy for her right action it is not sufficient that her action be motivated by a desire to do what is right. One might be tempted to suggest instead that one is morally praiseworthy for one's right action if and only if one acts out of a desire to do what is right and has a decent conception of the right—moral knowledge, or knowledge of the virtues. An agent has to act out of a sense of

duty, but the sense of duty ought to be combined with knowledge of what one's duty is.

But even if acting on a desire to do what is right assisted by knowledge of the right is sufficient for granting one's right action moral worth, it is not *necessary* for it. In the case of the extremist, a person acts out of concern for morality (a desire to do the right thing) and still does not act for moral reasons. Let us now look at cases where a person does the right thing for moral reasons but does not in any way act out of a desire to do the right thing. It is exactly ignorance of the virtues, or lack of moral knowledge, that can lead a person to this state. The same ignorance that led the extremist to mistake racist reasons for moral reasons can sometimes cause a person to mistake moral reasons—for which he acts—for something else. Sometimes we are smarter or act more reasonably than we think, sometimes what we take to be our lesser work is our best, and sometimes we act for moral reasons just when we think we do not.

Let us look first at a case of what I have elsewhere called 'inverse *akrasia*' (Arpaly and Schroeder 1998). I use the term 'inverse *akrasia*' to refer to cases in which an agent does the right thing but does so against her best judgment. Perhaps the most famous case of inverse *akrasia* in world literature, already briefly mentioned in chapter 1, is the case of Huckleberry Finn. Huckleberry Finn befriends Jim, a slave, and helps him escape from slavery. While Huckleberry and Jim are together on a raft used in the escape, Huckleberry is plagued by what he calls "conscience." He believes, as everyone in his society "knows," that helping a slave escape amounts to stealing, and stealing is wrong. He also believes that one should be helpful and loyal to one's friends, but loyalty to friends is outweighed by some things, such as property rights, and does Miss Watson, Jim's owner, not have property rights? Hoping against hope to find some excuse not to turn Jim in, Huckleberry deliberates. Like many children (and adults), Huckleberry is not very good at abstract deliberation, and it never occurs to him to doubt what his society considers common sense. Thus, he fails to find a loophole. "What has poor Miss Watson done to me," he berates himself, "that I can see her nigger go away and say nothing at all?" Having thus deliberated, Huckleberry resolves to turn Jim in, because it is "the right thing." But along comes a perfect opportunity for him to turn Jim in, and he finds himself psychologically unable to do it. He accuses himself of being a weak-willed boy, who has not "the spunk of a rabbit" and cannot bring him-

self to do the right thing, and eventually shrugs and decides to remain a bad boy.

Obviously, Huckleberry Finn does the right thing. His action is morally desirable. It is fortunate that he takes the course of action that he does. Does his action have positive moral worth? The answer to this question depends on our reconstruction of Huckleberry's motives, and, on the reconstruction I find most plausible, Huckleberry is morally praiseworthy for his action. Now there are many stories one can tell about Huckleberry's motives. In some of these stories, it would be rather clear that Huckleberry's helping Jim instead of turning him in would be a morally desirable action with *no* moral worth. For example, if Huckleberry's action were motivated by an unconscious desire to do things that would make the adults around him angry, then it would not have much moral worth. There would simply be no connection between the reasons for which he acts and the right-making features of his action. Similarly, if Huckleberry were to help Jim because of the operation, within himself, of some purely atavistic mechanism — akin, perhaps, to the tendency some animals have not to harm creatures with infantile features — we would also not regard his action as morally praiseworthy, because he would not be acting for reasons at all (this is, perhaps, what Kant imagines when he talks of acting out of *mere inclination*, and it may also remind us of Aristotle's idea of "natural virtue"). This is the interpretation favored by Bennett (1974), who sees Huckleberry Finn as merely squeamish, soft-hearted, unable to see a man in chains. In these two scenarios, Huckleberry is a racist boy who accidentally does something good. There is, however, a scenario in which Huckleberry is morally praiseworthy for his action, and I would guess this is the scenario Mark Twain had in mind, though whether he did is of no consequence for my argument.² On this interpretation, Huckleberry Finn is acting from neither squeamishness nor a desire to upset the adults. Rather, during the time he spends with Jim, Huckleberry undergoes a perceptual shift. Even before meeting Jim, the way Huckleberry viscerally experienced black people was inconsistent with his "official" racist views. There are people who sport liberal views but cross the road when a person of a different race appears or feel profound disbelief when that person says something intelligent. Huckleberry, from the beginning, appears to be the mirror image of this sort of person: he is a deliberative racist and viscerally more of an egalitarian. But this discrepancy between Huckle-

berry's conscious views and his unconscious, unconsidered views and actions widens during the time he spends with Jim. Talking to Jim about his hopes and fears and interacting with him extensively, Huckleberry constantly perceives data (never deliberated upon) that amount to the message that Jim is a person, just like him. Twain makes it very easy for Huckleberry to perceive the similarity between himself and Jim: the two are equally ignorant, share the same language and superstitions, and all in all it does not take the genius of John Stuart Mill to see that there is no particular reason to think of one of them as inferior to the other. While Huckleberry never reflects on these facts,³ they do prompt him to act toward Jim, more and more, in the same way he would have acted toward any other friend. That Huckleberry begins to perceive Jim as a fellow human being becomes clear when Huckleberry finds himself, to his surprise, *apologizing* to Jim — an action unthinkable in a society that treats black men as something less than human. As mentioned above, Huckleberry is not capable of bringing to consciousness his nonconscious awareness and making an inference along the lines of "Jim acts in all ways like a human being, therefore there is no reason to treat him as inferior, and thus what all the adults in my life think about blacks is wrong." He is not a very clear abstract thinker, and there but for the grace of God go all of us. But when the opportunity comes to turn Jim in and Huckleberry experiences a strong reluctance to do so, his reluctance is to a large extent the result of the fact that he has come to see Jim as a person, even if his conscious mind has not yet come to reflective awareness of this perceptual shift. To the extent that Huckleberry is reluctant to turn Jim in because of Jim's personhood, he is acting for morally significant reasons. This is so even though he does not *know* or *believe* that these are the right reasons. The belief that what he does is moral need not even appear in Huckleberry's unconscious. (*Contra* Hursthouse 1999, my point is not simply that Huckleberry Finn does not have the belief that his action is moral on his mind while he acts, but that he does not have the belief that what he does is right *anywhere* in his head — this moral insight is exactly what eludes him.) He is also unaware, or only dimly aware, of the fact that he is acting for these reasons in the first place. But he is acting for moral reasons all the same, in the *de re* sense of the expression "moral reasons." Huckleberry Finn, then, is not a bad boy who has accidentally done something good, but a good boy. No doubt he is imperfect, and one who would be better if some of his moral

convictions were changed, but as he is, he is better than many, including his counterpart who is liberal in conviction but not in deed.

Huckleberry Finn is not an isolated case. While inverse *akrasia* may be a dramatic phenomenon, it is not dissimilar to a considerably more common type of behavioral inconsistency—the person whose explicit views with regard to morality and politics are terribly wrong but who in everyday life “cannot hurt a fly.” We all have friends, family members, or acquaintances of this sort. We can all recall the likes of a student who, waving his copy of *Atlas Shrugged* in one’s face, preaches that one should be selfish and then proceeds to lose sleep generously helping his peers. If philosophers were right in believing that only those actions subjected to prior deliberation are done for reasons, or that only actions derived by deliberation from one’s moral principles are done for moral reasons, we would have to view these people as bad people who happen to have some fortunate inclinations in their makeup. More commonly, however, we treat these people as fundamentally good people who happen to be incompetent abstract thinkers. While the existence of such people may sometimes be baffling—note the fascination with the character of Oskar Schindler—such people are as commonplace their opposites: people with wonderful convictions who are immoral in their actions and emotions. The idea that we can sometimes act for moral reasons without knowing that we act for moral reasons is not strange when posed against the background of epistemology and psychology, fields in which many have maintained that we can know without knowing that we know, believe without knowing that we believe, or act for a reason without knowing that we act for a reason. It only seems strange if we are used to imagining people as divided, more or less, into the Faculty of Reason—the seat of deliberation from which all meaningful action springs—and that shady realm of emotion, inclination, and instinct, more or less atavistic forces that mechanically push and pull the agent. On this picture, an agent who is pulled into action by the Appetite can only be blamed indirectly (for failing to restrain his appetites or emotions properly) and can only be praised indirectly (for managing to train his appetites or emotions so that they appear in the right place at the right time). One may attempt, with difficulty, to force the plethora of different cases in which one condemns the viscerally racist liberal, the honesty-preaching Polonius, or the child-beating Christian into a picture in which it is only a lack of self-control that one condemns. But Huckle-

berry Finn, Oskar Schindler, and our *Atlas Shrugged*-toting friend are obviously not praiseworthy for any kind of self-training or character-building on their parts. They are praiseworthy because, despite any character-building imposed on them by their misguided selves or others, some of their moral common sense, much of their moral goodness—that is, their responsiveness to moral reasons—remains intact.⁵ (Issues surrounding the development of self-control will be dealt with in fuller detail in chapter 4.)

Blame and Moral Unresponsiveness

To recapitulate: some people do the right thing by some sort of lucky accident, while others do the right thing in response to moral reasons. Those who do the right thing in response to moral reasons are those who are morally praiseworthy for their actions. It makes sense, then, that something similar is true for people who do the wrong thing. Sometimes the fact that one did the wrong thing appears to be accidental, while at other times it seems to stem from what can be called *ill will* or from a deficiency of good will. I take good will to be the same as moral concern (to be discussed in the next section) and as responsiveness to moral reasons. I take a person to be *responsive to moral reasons* to the extent that she wants noninstrumentally to take courses of action that have those features that are (whether or not she describes them this way) right-making and not to take courses of action that have those features that are (whether or not she describes them this way) wrong-making features. If good will—the motive(s) from which praiseworthy actions stem—is responsiveness to moral reasons, deficiency in good will is insufficient responsiveness to moral reasons, obliviousness or indifference to morally relevant factors, and ill will is responsiveness to sinister reasons—reasons for which it is never moral to act, reasons that, in their essence, conflict with morality. In other words, the person who is deficient in good will acts without regard for the wrong-making features of his action, while the person who has ill will performs his action exactly because of its wrong-making features.

Imagine, for example, that Jeanne is very rude to Joseph and, unsurprisingly, hurts his feelings. Many scenarios can be imagined that fill in her motives. Perhaps she comes from an aggressive culture in which

"shut up" is a commonly used phrase and has not yet met people who find it grossly offensive when used by a peer. In this case, the fact that she did something wrong is, in a clear sense, accidental. Perhaps, on the other hand, she has no such excuse, and she acts the way she does because she desires to vent the tensions of a long day by saying exactly what comes to her mind, which happens to be offensive to Joseph in an obvious way. In this case, she is blameworthy, because her action indicates a failure to respond to morally relevant considerations; she should be motivated by the fact that Joseph is likely to be hurt by what she says, but she is unmoved by this and so acts in an inconsiderate manner. A third possibility is that she is rude to Joseph because she enjoys inflicting suffering on others and wishes to hurt Joseph. In this scenario, she is even more blameworthy than in the previous one, as her action not only expresses a deficiency of good will but also expresses ill will. There is nothing about the desire to vent one's feelings that essentially conflicts with morality. Like the more celebrated motives of love and money, it can sometimes accidentally conflict with morality while at other times it leads to morally good or neutral actions. On the other hand, whether one thinks of morality in Kantian, utilitarian, Aristotelian, or simple, commonsensical terms, a desire to inflict suffering for its own sake is essentially in conflict with morality. To do something purely because it would inflict suffering on a fellow human being is to act for sinister or "anti-moral" reasons: it is to perform a wrong action for reasons that are, to a substantial degree, the reasons for which it is wrong. While in the "venting" case the features of Jeanne's action that make it wrong seem to be ignored by her, in the sadism case she performs the action exactly because of the features that make it wrong. Other things being equal, a person is more blameworthy for a given wrong action if she acts out of ill will (for sinister reasons) than she would be if she were to act out of a lack of good will (for neutral reasons, while ignoring moral reasons to the contrary). This does not, however, imply that a sadist or a proponent of a sinister ideology is always an all-around worse person than an opportunist. A chilling, profound indifference to moral reasons, the kind compatible with, say, killing for profit, is much worse than a mild predilection for the mildly immoral, such as a taste for unnerving one's underlings.

I said that to act specifically in order to hurt a person is to act for a reason that essentially conflicts with morality. I should emphasize that I

assume here that it is possible to act in order to hurt a person and have this be one's reason for action *simpliciter*, without a further motive of obtaining pleasure and also without something like Kantian self-love in the background—that is, it is possible to decide to indulge one's sadistic desire. Very few of us believe that everything we do is done for pleasure or for self-interest, and the view is one against which Tim Scanlon (1998) and Stephen Darwall (1997) have argued convincingly. It is also false that all we act for is self-interest unless we act for moral reasons. I said that cruelty for its own sake should be seen as an anti-moral reason no matter whether one has a Kantian, utilitarian, or commonsense view of morality. Here, when referring to a Kantian view of morality, I am referring to a view of moral desirability or rightness that is based on the idea of respect for persons or on something resembling Kant's universal law formula. Some proponents of Kantian moral psychology could argue that 'anti-moral reasons' is an oxymoron of sorts, and that both in the case of the sadistic agent and the case of the selfish agent, the proper way to formulate the agent's reasons for action would be to say that an agent has chosen the satisfaction of an inclination over morality. In this sense, both agents act out of self-love or self-interest, even though different inclinations served as "incentives" for their decisions. I think, however, that such a view risks underestimating the moral difference between the sadist and the egoist. Intuitively, other things being equal, we are more morally outraged by a person who performs an action out of cruelty for its own sake, or out of an Iago-like desire to defile all that appears good, than we are by the person who performs the same action because he is tempted by money or fame, or even because he clearly cares about money or fame above all other things. The latter makes us feel that we are in the presence of ruthlessness and selfishness, or of weakness; the former makes us feel that we are in the presence of pure evil. Even on a small scale, a Jeanne who is rude to her colleagues because she is inconsiderate is less bad than a Jeanne who is rude to her colleagues out of sadism or out of racial prejudice. Being overly susceptible to love or money is bad; being susceptible *at all* to sadistic pleasure is bad—even if one never acts on one's sadistic motives, attempts to suppress them as much as possible, and never makes a decision to indulge them. It is these day-to-day intuitions that my account explains and that other views need to explain as well.

I shall assume, then, that there are sinister or anti-moral reasons for action: it is possible, in other words, to perform actions for reasons that are, morally speaking, reasons to avoid them. (I take the burden of proof to be on anyone who wants to show that it is not possible.) Now consider again the person who acts not for sinister reasons, but who fails to respond to relevant moral factors. What is it to lack responsiveness to moral reasons? There seems to be more than one form this lack of responsiveness can take. Sometimes, one is deficient in some sort of moral perception — blind, as it were, to some kinds of moral factors. One is almost tempted to accuse such a person of *localized amorality*. Some people, for example, do not seem to be able to grasp the idea of personal autonomy. Imagine the sort of parent who not only fails to see that something could be wrong with her extremely paternalistic treatment of her adult son, but who also does not understand the concept of paternalism even after it has been explained over and over. ("But if no one tells you what to do, how would you know?" she asks in bewilderment.) More often, however, one perceives all of the morally relevant features of a situation but is not sufficiently moved by them. One need not be an amoralist or sociopath of any kind for this to happen. It is enough that one be a more common though much less discussed figure: the all-too-human person who cares about morality, but *not so very much*.⁶ One might call such a person "the half-moralist." Here is one story about a half-moralist:

Mary realizes perfectly well that if she does not promptly mail her former roommate, now in Russia, the book he lent to her, she will break a promise, in addition to causing unnecessary pain. Even so, she does not send it, because it is ten degrees below zero outside and because it would be nice if she finished writing her novel before her birthday, and so she cannot really spare the time to make a special trip to the post office. She is not akratic but is acting on a decision. Not being an amoralist, it is not the case that she is indifferent to the institution of promising, the interests of her fellow human beings, or other relevant moral factors. She cares about these things enough to have avoided more serious breaches of trust, but not enough to go to the post office.

The first kind of morally unresponsive person is analogous to the person who does not notice when his clothes do not match. The second kind is analogous to the person who notices very well when his clothes are un-

matched but does not care about his appearance enough to make sure that he is never stuck with nothing to wear but his orange shirt and purple pants.

Deficiency of perception and deficiency of motivation are, admittedly, hard to tell apart. Many people have wondered if their spouses don't see the dust on the floor or don't care that it is there, and the diagnosis is hard because there is often a combination: a person who does not care about cleanliness tends to be less trained at spotting dust, and a person who is not good at spotting dust may find it hard to be motivated to learn to see it and add a new worry to his life. Similar things apply to the person of mediocre moral sensibilities. A person who does not care much about morality may not give much thought to some things to which a more morally concerned person would pay more attention, and she may thus be less competent in perceiving other people's feelings, putting herself in their shoes, and so on. In fact, this is almost always the case, a fact on which I will draw later in this chapter when I discuss the marks of moral concern. Still, it seems useful to distinguish, in comparing people's moral responsiveness, the motivational from the perceptual — much in the same way that, when comparing two people's health-consciousness, it may be useful to point out that one of them seems to be more aware of health factors but the other one seems to be more motivated by factors of which he is aware. By distinguishing the motivational from the perceptual side of responsiveness, I do not preclude the possibility that failing to care about morality is irrational (after all, failing to care about one's health or bank account may be irrational as well), nor even the view that moral factors when noticed are always *somehow* motivating. All that I assume here is that two agents can be motivated by the same moral reasons exactly, but one may be motivated by them much more than the other. This does imply that moral motivation is not proportional to understanding of morality and/or noticing morally relevant factors, and thus it does not imply anything incompatible with the most Platonic views of moral *akrasia*. It is simply to say that people can be aware of moral factors and yet not be maximally motivated by them. Again, I take the burden of proof to be on anyone who wishes to argue that Mary the half-moralist must have a problem *understanding* or seeing the moral reasons she has to return the book to her former roommate.

Degrees of Moral Concern

This discussion of moral unresponsiveness has brought to the foreground the idea that, given that a person is motivated by moral (or anti-moral) reasons, she can be motivated by them more powerfully or less powerfully. Let us go back to the realm of praise. Kant's prudent grocer and his ilk demonstrate to us that moral praise is warranted for the agent who acts for moral reasons, as opposed to other reasons. But suppose we know that an agent acts for moral reasons. What bearing should the strength of his moral motivation, the amount of concern that he has for moral considerations, have on our assessment of his action's moral worth? I would like to offer the following answer:

Praiseworthiness as Responsiveness to Moral Reasons (revised version): For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons—that is, for the reasons for which the action is right (the *right reasons* clause); and an agent is more praiseworthy, other things being equal, the deeper the moral concern that has led to her action (the *concern* clause). Moral concern is to be understood as concern for what is in fact morally relevant and not as concern for what the agent takes to be morally relevant.⁷

What does being *deeply* or *strongly* concerned for something, or being *more concerned* for something than someone else, amount to?⁸ I take concern to be a form of desire. To say that a person acts out of moral concern is to say that a person acts out of an intrinsic (noninstrumental) desire to follow (that which in fact is) morality, or a noninstrumental desire to take the course of action that has those features that make actions morally right. To say in general that a person has moral concern would be to say that she has an intrinsic desire (*de re*) that morality be followed or that the courses of actions that have those features that make actions right be taken (which would imply, for example, a desire that other people act morally as well). Much work has been devoted to the study of desire, and very different views have been defended by philosophers of mind.⁹ Giving a full account of concern in general or moral concern in particular would require choosing a theory of desire, which I do not wish

to do here. What I would like to do is, first, sharpen our intuitive idea of concern by making a few points about what depth of concern does *not* amount to and, second, point out what I take to be important markers of depth of concern in everyday life.

It is important to note what depth or strength of concern does not amount to. It does not amount simply to *intensity of feeling*. I may care about the well-being of my friends more than I care about drinking Coca-Cola, even though now, thirsty as I am, I experience my desire for Coca-Cola considerably more intensely than I experience my concern for my friends. You care for your loved ones even while you are so angry at them that you do not “feel” any love toward them, or you are so panicked about missing a flight that nothing is further from your mind than your loved ones. Nor does being deeply concerned with something amount to a type of reflective endorsement. Two people can reflectively endorse identical things but be very different in their level of concern for these things. Erica and I may reflectively endorse the same kind of political action, but she may be more concerned with it than I am, which may explain why she is at a demonstration while I am writing. It is also natural to say that I am less *committed* to political action than Erica is, and this may tempt some readers to think of caring in terms of *commitment*—reflective endorsement with some sort of emotional backing, perhaps. This would also be misleading, as we may deeply care about things that we do not reflectively endorse at all. Tamara may care deeply about Todd and her relationship with him even though she believes she should not do so, or even though she is utterly unaware of her deep concern, ignoring it in practical deliberation. A frustrated but perceptive Todd may say, “She does not know how much she cares about me, and that is why she is not committed to the relationship.” Ultimately, Tamara may say that she had not known how much she cared about Todd until, for example, Todd was in danger. Similarly, one may care about beauty, religion, money, or the opinions of other people even though one does not reflectively believe that any of these is important.

Let me now discuss three features *associated*, other things being equal, with strength of concern. I shall talk generally about moral concern or concern for morality (where “morality” is used in the *de re* sense). But what I say about such concern also holds true for more particular moral concerns—for example, concern for the happiness of others, concern for justice, and even more specific moral concerns. The first feature

of concern I wish to mention is a *motivational* one—what can be called the “diehard” quality of stronger concerns as motivators. It appears to be the case that the more you care about something, the more it would take to stop you from acting on your concern. Other things being equal, the person who cares more about the football team will show up for a game on a stormy day while the person who cares less will stay at home. Both go to games for the same reasons, but one of them does so from a deeper love of the team. Thus, other things being equal, the person who cares very much about morality (again, in the *de re* sense of morality), or about any specific moral consideration, would tend to be motivated to action by it in situations in which the rest of us would not. I will say more about this feature in the subsequent discussion of benevolence.

A second feature has to do with the concerned person's *emotional makeup*. Other things being equal, caring about a team makes wins pleasant and losses painful. More than this, the person who cares about a team is likely to experience shame at its bad performance, pride at its good performance, anxiety when an important game approaches, a sense of utter despair if it turns out that a key player has been involved in a serious drug fraud, and other such emotions that utterly baffle the person who does not possess such a concern. More of these phenomena exist, and the more one is concerned with the team, the more their intensity increases. The morally concerned person, other things being equal, will find the thought of doing something wrong quite painful. As a result, she will not simply think, “I have done something wrong,” but will also feel pain at the thought—that is, she will feel *guilty*. She will also feel anger when reading about atrocities in the paper, sadness when wondering, as Kant did, if “anything straight can be fashioned from the crooked timber of humanity,” as well as admiration for moral heroes and disgust at the thought of being in the company of some immoral individuals. We imagine the amoralist as someone who never experiences these emotional reactions in any form (even if, for some reason or other, he does not happen to live a life of crime). The more one cares about morality, the more it colors one's emotional world (though again, this is true other things being equal, and other things are hardly ever equal when it comes to people's emotional lives). A third and related feature of moral concern is a *cognitive* feature. A person concerned with morality is, other things being equal, “morality conscious”—he notices morally salient things that a person indifferent to morality would not notice. It is

a feature of the human mind that we learn more about things in which we are emotionally invested: other things being equal, a child who cares about playing the piano learns better how to play the piano than a child who is forced to do so. A person who cares about birds will notice a bird on the roof, while in the same situation a person who does not care about birds may not notice it at all. If one cares about morality, moral facts matter to one emotionally, and they are salient to one. As a result, other things being equal, a person of more moral concern will be more sensitive to moral features of situations—more apt to notice, for example, that a fellow human being is showing signs of distress or that a joke has the potential to offend certain people. I will say more about the emotional and cognitive markers of concern later in this chapter, in my discussion of blameworthy moral beliefs. The emotional and cognitive features of concern will also be relevant in the section on character, as the virtuous person is at times described as a person who has a collection of concerns that produce both an uncanny ability to recognize morally salient situations and a tendency to have the right emotional reactions at the right time.

In this section, I will focus on the first feature—“diehard” motivation. Imagine a person who cares so much for her fellow human beings, or for what she takes rightly to be her moral duty to them, that she would act benevolently even if severe depression came upon her and made it hard for her to pay attention to others. Now imagine benevolence's fair-weather friend, who acts benevolently as long as no serious problems cloud her mind but whose benevolent deeds would cease, the way some people drop their exercise programs, if there were a serious crisis in her marriage or her job. Last, imagine the person who acts benevolently on a whim. It is Sunday morning and she is awakened by a call from a charity asking for a donation. Our agent thinks, “Why not do something right?” and is moved to do something right so long as her credit card happens to be close enough to the bed.

The first agent is more praiseworthy for her actions than the second agent, because to act benevolently for moral reasons while one is depressed takes more concern for those moral reasons than to do so in happy times. Kant would have offered a different explanation. He would say that the first agent is more praiseworthy than the second agent because she acts out of one motive, duty, while the second one acts out of a motive called ‘inclination’, a basically hedonistic motive

that is as different from duty as water is from oil. This, however, need not be the case. It is quite unlikely that my first agent changes her motive from inclination to duty the moment her life becomes harder. My second agent, unlike Kant's own happy philanthropist, is not someone who "gets her kicks" out of doing good deeds (like Jane Austen's Emma), and her disposition need not be particularly sunny. She is just an ordinary person who does good for moral reasons but whose moral concern is not deep enough to override some other concerns when they appear. Certainly, she sometimes feels a sense of satisfaction with herself, but one needs to be quite a psychological hedonist to think that any action which results in pleasure is motivated solely by it. There is no reason to say that her motives are different from the motives of the first agent, any more than there is a reason to say that the person who is truly devoted to her exercise program goes to the gym for different reasons than the person whose devotion is somewhat weaker. Both may go to the gym for health reasons, but one cares *more* about her health. Similarly, my first and second philanthropists both act benevolently for moral reasons, but one of them cares more. The contrast is *not* that happy philanthropists are less praiseworthy for their actions than sad philanthropists, but rather that *fair-weather*, frivolous philanthropists deserve less praise than those whose concern for morality or for the well-being of others is more *serious* or *deep*. The third agent—the person whose moral concern is skin deep—may be called the *capricious* philanthropist and would be very presumptuous to expect much praise for an action that almost seems accidental, attributable to the charity's call and the location of the credit card more than to her depth of concern for her fellow human beings. Still, there is no reason to doubt that she has acted for moral reasons. When a person whimsically asks for milk instead of cream in the coffee she has with her chocolate cake, one does not doubt that she does it for health reasons but doubts merely the seriousness of her concern.

My account is un-Kantian in two ways: I accept the view that it is possible for people who understand morality equally well to care about morality to different degrees, and I also reject the view that the fair-weather philanthropist acts on inclination—that is, on a hedonistic motive (which qualifies her as a "happy" philanthropist)—and grant that it is possible for her to act on moral motives. If we were to believe that only foul-weather, die-hard philanthropists act for moral motives, we would

have to believe that only *very* morally virtuous people ever act for moral motives and ordinary people never do.

A natural question concerns Kant's cold-hearted philanthropist—the one who, at ordinary times rather than in sorrow, has to force himself to help people despite having no inclination to do so. Neither the cold-hearted philanthropist nor the sorrowing philanthropist just discussed enjoys his actions very much, which makes it deceptively easy to look at them as rather similar instances of "acting out of duty." However, each of them has a different story behind the fact that he drags his feet on his way to charity. The sorrowing philanthropist drags himself to action because he cares about the welfare of others so much that his concern lasts even when sorrow tempts him to stay home. The cold-hearted philanthropist has to drag himself to his good work because his philanthropy, even at the happiest times, is *half-hearted*. This could mean that he cares about the good of others just enough to wish he cared more, or it could mean that his concern for the good of humans is offset by essentially conflicting attitudes,¹⁰ such as disdain for his fellow humans. The fact that one's concern, whether for humanity or for one's wife or one's art, is enough to motivate one—albeit barely—even though one is grief-stricken is a testimony to the strength of one's concern. The fact that one's concern, in the best of times, is only enough to motivate one barely, shows a deficiency of concern or a half-heartedness. Thus, under many descriptions, the cold-hearted philanthropist is less praiseworthy than he would be if it were not for his cold-heartedness. The picture becomes complicated, however, when one remembers how underdescribed Kant's case is. I have followed Hursthouse (1997) in taking the coldness of this philanthropist's heart to be a sign of half-heartedness, but truly, many different things can cause a person to appear cold or to experience himself as indifferent, and these things differ in morally significant ways (recall that what one cares about most is not always that about which one feels most strongly and warmly). Thus, a cold-looking or even cold-feeling philanthropist may in fact care about morality much more deeply than a philanthropist known for a "Mediterranean temper" or a "sweet temper." As the concept of hysteria teaches us, concerns that produce songs and tears are not always more serious than those that do not. Also, even if my act of philanthropy is itself half-hearted, it may be a part of a *serious* effort (as opposed to a whimsical one) to habituate myself into being a better person, and such an effort is a praiseworthy action all by itself.¹¹

Having mentioned "half-heartedness," I should say a few more words about it. One way to be half-hearted is simply not to have enough motivation; so much is straightforward. Another way to be effectively, or functionally, half-hearted—half-hearted in a sense opposite to that of Frankfurt's wholeheartedness—is to have concern for something that essentially conflicts with the concern in question. A person may have attitudes that conflict with each other accidentally, such as a desire to be a great philosopher and a desire to excel at basketball. These desires conflict accidentally because there are simply not enough hours in each day or days in each week to pull off both projects in the same life. On the other hand, a person who values puritan Christianity but who also loves sex may have two *essentially* conflicting attitudes, and thus we may say that her love of Christianity, as well as her lust, are effectively rendered half-hearted. If her love of Christianity is much greater than her lust, or if her lust stands alone while her love of Christianity is intrinsically supported by her other attitudes, her lustful acts may even be, so to speak, quarter-hearted or less—in other words, the deeper her essentially antilustful attitudes, the less it would be true, other things being equal, that acts performed out of such lust reflect who she is. Thus, a person who lives a life of crime, sensual pleasure, and extravagance, and who goes to church and prays for forgiveness every so often, will usually be considered a person whose Christianity is effectively superficial (and not very praiseworthy, if Christianity is praiseworthy generally), even if her guilt is resilient enough to move her to go to church her entire life.

The difference between essential and accidental conflict of attitudes is implicit in my discussion of the philanthropist. There is no essential conflict between sadness and philanthropic concern, so the fact that the philanthropist helps in spite of depression only serves to underscore the depth of her philanthropy. On the other hand, if the philanthropist helps despite a contempt for her fellow human beings or an indifference to their suffering (I am ignoring here the possibility that this help is part of a self-improvement project, as above), the presence of this contempt or indifference reduces the wholeheartedness of her philanthropic acts to a level that would be expressed by someone who was not conflicted but less concerned. Naturally, it is not always so easy to tell an essential conflict from an accidental conflict. If I love the person Albert Einstein and hate physics, does my hatred of physics merely test my love

of Einstein, or does it diminish my wholeheartedness in loving Einstein? It is hard to tell. Note, however, that those who take the second attitude would probably say something along the lines of "Physics is part of *who Einstein is*":—thus conceptualizing the conflict as essential rather than accidental. Very few people would say that loving them requires liking the job they hold at a fast food restaurant.

What, then, of blameworthy actions? There are two types of blameworthy actions: those done for sinister reasons and those done for morally neutral reasons, as a result of some indifference to moral considerations. It makes sense that a person who acts for sinister reasons is more blameworthy the stronger the ill will indicated by his action. Consider, however, the wrong action motivated by moral indifference—perhaps the more common sort of misdeed. If my various philanthropists are judged differently by the depth of concern for moral considerations their actions express, various wrongdoers can be judged by the depth of indifference to moral reasons that their actions reveal. The more moral concern it requires to take the right course of action in a certain situation, the more praiseworthy an agent is for taking it and the less blameworthy an agent is for not doing so. Imagine that in a certain situation it is right to act charitably and wrong to act uncharitably. Imagine also that in this situation it is hard, and therefore requires deep moral concern, to be charitable if one is in great sorrow, but it is easy to be charitable if one is happy (perhaps the situation involves being energetic and making phone calls on behalf of a charity for a few hours). Thus, the person who is charitable in sorrow shows herself to be more praiseworthy for her action than the person who is charitable generally, but not in sorrow. Accordingly, it does not take a great deal of moral indifference to be uncharitable while sorrowing, but it does take a lot of such indifference to be uncharitable while happy. Thus, if the sorrowing philanthropist is more praiseworthy than the happy philanthropist, the sorrowing failed philanthropist is less blameworthy than the happy failed philanthropist. Hence, for example, the use of "She was under stress" to excuse minor wrongdoing. In a more dramatic vein, if one wonders if actions such as hiding Jews from the Nazis are required by morality or are supererogatory, one compromise suggestion can be that such actions *are* required, but performing them requires a degree of moral concern so rare that the person who fails to perform them shows hardly any moral indifference and thus is not particularly blameworthy.

If the depth of one's concern for the right-making features of one's action changes an action's moral worth, how does Huckleberry Finn measure up? Huckleberry, we said, cannot be denied praise for helping Jim, because he is acting for the right reasons—he performs his action because of its right-making features. But can he be said to be acting out of deep concern for the right-making features of his action? Yes—at least on the interpretation described above—because concern does not amount to reflective endorsement and does not have to be conscious. Let us compare Huckleberry to another inverse akratic, whose action I do not take to be morally praiseworthy. Joseph Göbbels, as is evident from his diaries and some of his public speeches, suffered from surprisingly frequent attacks of what he called weakness of will, attacks that he attributed to fatigue and stress. These attacks consisted in feeling compassion for the victims of the Nazi regime, and Göbbels repressed and overcame them with relative ease, the way a civil person may overcome a desire to be rude at a family dinner. Let us suppose that on a certain occasion, Göbbels, tired at the end of a long day, perhaps having had a drink, finds himself momentarily unable to resist his compassion. Against his best judgment, he does something which he will regret the next day and for which he will compensate with a new hardening of his policy. Perhaps he signs an order that will enable a Jewish acquaintance to leave Germany—an action that is easy and not particularly costly for a man of his position. Let us assume for a moment that Göbbels acts for the same reasons as Huckleberry. Of all the things that go by the name of compassion, we shall assume that Göbbels experiences the kind that involves being moved to action by the humanity of the person. Still, Göbbels's action does not impress us in the same way that Huckleberry's does. The reasons seem to be the following. Göbbels's inverse akratic compassion strikes us as shallow or capricious—not so much because of the short time his attacks last but because of the ease with which they are overcome. His compassion is also offset by the strength of his essentially conflicting attitudes—his hatred of Jews, which normally causes him to overlook their personhood, and his deep concern for the Nazi cause, for which he appears to be willing to die. Thus, his action seems to be the product of alcohol, fatigue, and stress more than anything else. We can almost imagine his friends regarding this as a condition that excuses him from blame. For us it “excuses” him from praise.

Things appear to be different with Huckleberry Finn. Huckleberry does not need fatigue or alcohol to help Jim. As he is described in the novel, one gets the impression that he helps Jim because his concern for his fellow human beings in general, and for Jim as a human being in particular, is strong enough to make him act despite any reservations he may have. The strength of his concern for Jim as a person is a testimony to the shallow nature of his racist attitudes, not the other way around. To call Huckleberry's compassion capricious simply because it conflicts with these attitudes is like calling a Mafia boss's actions capricious because he sometimes goes to church. Thus, even though both Huckleberry and Göbbels act—let us assume—for the right reasons, and even though both of them act against their best judgments, we view them differently. One of them appears to be a good (though imperfect) boy who performs a morally praiseworthy action, while the other appears to be a bad man who, thanks to fatigue, alcohol, and a changeable mood, performed a morally desirable but negligibly praiseworthy action.

Character

It is tempting to rephrase what I have said about Huckleberry Finn in terms of character. Huckleberry Finn's concern with all that is human seems to be a deep feature of his character, while his racist convictions do not seem to be so. He is basically a good person; Göbbels is basically a bad person. It is not an accident that talk of character comes naturally here. Just as the idea of doing the right things for the right reasons, or out of concern for the right-making features of one's action, accounts for the appeal of the doctrine of the motive of duty, the idea that deep concern for these features is worth more than shallow concern for these features accounts for the appeal of the Aristotelian idea that right (or fine) actions are only praiseworthy (or virtuous) if they follow from the agent's character.

Consider Aristotle's treatment of bravery. For an instance of defending one's city in war to be praiseworthy, it is not enough for Aristotle that it is performed for the right reasons (“for the sake of the fine”). However pure the defender's motive is, if he is only capable of defending his city because he is pathologically fearless, or because he is such a skilled

soldier that war does not frighten him, or because he has used drink to calm his nerves, his action is not morally praiseworthy. If, like the half-hearted philanthropist, he needs only the aid of soldierly self-control to drag himself to action, he is only partially virtuous. The only agent who is fully morally praiseworthy for defending his city is the agent whose ability to overcome the fear of death is the result not of wine or even of mechanical self-control, but of the fact that he genuinely *fears moral disgrace more than he fears death*—in other words, his concern for the right-making features of his action is so strong that it defeats the fear of death. Such deep concern for the right-making features of defending one's city is to be found in the *brave* person—the person in whom risking his life for the sake of doing the right thing expresses a virtue of character. The virtuous person is different from the rest of us in his possession of the markers of moral concern. He has a die-hard disposition to do what is fine (in the case of the brave person, fight for his city); he has a different emotional life from that of the nonvirtuous person (in the case of the brave person, the thought of disgrace frightens him; in general, Aristotle makes clear that virtuous people are pleased and pained by things that would not please or pain the rest of us), and he has an ability to recognize the right occasions for certain kinds of action that a mere theoretical knowledge of ethics does not give us. That the mental life of the virtuous person is to a large extent colored by her concerns has been pointed out by John McDowell (1979).

For the Aristotelian, the moral worth of every fine action depends to a large extent on the character of the agent performing it. This claim should not be confused with the common saying that for the Aristotelian, the reason an action is right is that the virtuous person would perform it. There are some reasons to doubt the latter claim: Aristotle never says that defending one's city in war is a fine action *because* that is what the virtuous person would do. However, he does take pains to explain that of all of the many people who would perform the fine action of defending their cities in war, only those of virtuous character deserve genuine praise for their actions.

Why should Aristotle, or anyone else, believe that the praiseworthiness of an individual action depends on the character from which it stems? If one thinks of character as a stable disposition of some sort, the idea may seem fairly strange. Why should the moral worth of the right action that Steve performs on November 13, 1999, be any different if we

know that Steve has performed similar actions for twenty-six years and will most likely continue to do so for the rest of his life?

The answer is that the mere frequency or predictability of an action should not matter at all to its moral worth—unless frequency or predictability are taken to be signs of *deep moral concern* in the agent. After all, the pathologically fearless man or the merely well trained soldier may have just as stable a disposition as the brave person to defend the city in war, and they, too, may act for the right reason among other reasons—but their actions do not express virtue.¹² Or consider Steve. Steve, a professor, has not missed class in twenty-six years. When I mention this fact about Steve, I am complimenting him: it says something good about Steve that he has never missed a class in twenty-six years. But why? Because, I would like to claim, we normally take the fact that a professor never misses class as a sign of his devotion—a sign that he cares about his students and his role as a teacher very much. Steve's stable disposition in matters of teaching is *evidence* that his teaching and his students matter to him very much. Such evidence, however, can be overridden, in which case Steve's stable disposition to come to class would not be perceived as virtuous at all. Imagine, for example, that Steve's incredibly stable disposition to come to class is due to the fact that he, unlike other teachers, experiences pain or discomfort in very few circumstances, so that coming to class while injured or sick is not onerous for him. Or imagine that Steve, having only an average amount of concern for teaching, still cannot imagine avoiding a class, because he deeply desires to be perceived as a perfect worker, is deeply afraid of getting fired, or is blindly devoted to a work ethic regardless of whether he works at the university or at a chemical weapons factory. If any of these were the case, we would no longer be inclined to think that Steve's clockwork predictability in coming to class says anything particularly good about him. Steve's devotion to his students is his virtue: his actions over the last twenty-six years are mere evidence—as we have seen, incomplete evidence that can be overridden—of this virtue. If we wish to be neo-Aristotelian, we might want to say that Steve's coming to class on a particular day despite having a severe migraine, say, is especially praiseworthy because it stems from a virtue of character. This would not be tenable if “stems from a virtue of character” meant “stems from a stable disposition,” but it makes perfect sense if it means “stems from a markedly deep morally relevant concern”—in this case, deep devotion to one's students. Note that “deep

concern" does not mean "long-lived concern." It happens to be empirically true that deep concerns usually do not change overnight, and so, if Steve's behavior toward his students changed from day to day, we would be inclined to assume that his benevolent actions toward them are at least somewhat *capricious*—that is, they do not reflect a particularly deep concern for them. Such an assumption would seem more plausible than thinking that Steve is a sort of Dr. Jekyll, whose genuinely deep concerns change—whose character fluctuates—with alarming frequency. But on the rare occasions when our assumption that deep concerns do not change frequently or rapidly proves mistaken, we do not tend to think of the changed person as capricious. For example, if it were known to us that a teacher, previously indifferent to his students, having read *Up the Down Staircase*, underwent a conversion of sorts and developed a genuinely deep concern for his students, we would have no reason to regard his devotion to his students as any shallower than Steve's—even if the conversion happened only a year ago and thus did not have a chance to manifest itself for twenty-six years. We might say naturally that the person's character changed. Of course, in real life, nobody knows for sure if such a person's new concern is deep or not, and so we tend to postpone our judgment until we see if it lasts, but the "test of time" is only that—a mere test for the depth of a concern, and an imperfect one at that. Brevity does not make a concern shallow any more than longevity makes a concern deep—in a comfortable climate, one may be a fair-weather friend for many years.

Thus, the idea that character matters for the moral worth of individual actions makes sense if one takes character to be about depth of concern and not about predictability or frequency. Many traditional objections to character-oriented views lose their force in this way. To give only one example, suppose that as the Nazis come to power, your longtime fair-weather friend, who has never done you wrong before, cheerfully informs on you. A traditional objection states that virtue ethics is committed to excusing his action as "out of character," but if "in character" does not mean "predictable" or "in keeping with historical trends," then "unpredictable" and "out of keeping with trends" does not always mean "out of character," either.

In defense of the critics of virtue ethics, it can be said that virtue ethicists themselves have rarely if ever questioned the connection between character and predictability or stability—and it is easy to see how

this mistake came to be. As I have previously indicated that deep concerns die hard, it is plausible that the person who is deeply concerned, say, with his students or with his city, would be disposed, *other things being equal*, to act for the sake of his students or his city more than most. If one assumes that other things are usually equal, one can conclude that concern for a certain morally relevant thing does imply considerable predictability. If one adds some version of the view that concern for some morally relevant things always goes together with concern for other morally relevant things, one has to conclude that depth of moral concern implies even more predictability.

Unfortunately for traditional virtue ethicists, both of these assumptions are dubious. This has been shown by various psychological experiments (see Doris 1998; Doris forthcoming; Harman 1999), but it can also be deduced from more intuitive evidence. The falsity of the second assumption can be seen from the fact that a corrupt lawyer may be more likely to return your book on time than an activist whose life is devoted to global justice, or that soldiers who are brave in battle are not any less likely to fear expressing unpopular opinions in public. It may seem harder, at first glance, to show the falsity of the first assumption—that there is a noticeable correlation between the extent to which one cares about *x* and the frequency of one's acting for the sake of *x*, but this too can be shown. Imagine that one works in an academic department in which the workload is typically very high. If one passes through the corridor, greeting each and every member of the department as one walks, some members are more likely than others to greet one back and to engage one in conversation. One may be tempted to think of those who do not do so as less friendly or less talkative. This, however, is likely to be a mistake—an instance of the *internal attribution error* (Harman 1999). The people least likely to greet one in the corridor are those whose schedules are the busiest, the grouchiest-looking ones are nearing their sixth year as assistant professors, and so on; conversely, the people who are most likely to engage one in conversation are those in the more comfortable stages of their careers. Having realized this, one may still be tempted to believe that at least the few individuals whose pattern of behavior deviates from this correlation are showing their true colors in terms of their deep concern, or deep lack thereof. The man who stops to engage in conversation despite nearing his tenure decision is surely one who deeply desires to engage in conversation, and the woman who tends

to pass you in the corridor with almost no greeting despite being a full professor on leave is truly unconcerned with making conversation. But this, of course, need not be true. It might be that the man, though misanthropic, has a tendency to react to extreme stress with a compulsion to talk, and the woman, though deeply desirous of her colleague's company, fears making a fool of herself or has been habituated by some sexist norm not to say too much. In a world in which people are rarely allowed to act on all their hearts' desires, it is little wonder that one's actions do not always reflect one's concerns very well.

While these tales show serious problems for the traditional notion of a character trait, they do not show that there is no fact of the matter about moral concern. The person who loves more is not always the person who says "I love you" more often, but some people do love more than others. The person who is more devoted to her students is not always the person who comes to class more often, but some people are devoted to their students more than others. And the connection between depth of concern and moral worth is an important truth aimed at by talk of character.

No Such Thing as a Good Nazi

This account of moral worth is bound to be incomplete. In fact, it is such almost by definition, for if moral worth has to do with responding to the right-making features of one's action, the full theory of moral worth depends on a full theory of the right, which would tell us what these right-making features are. Are they utility-maximizing features, person-respecting features, or eudaimonia-conducive features? Or are they something else? It is worth noting, though, that intuitions about moral worth also enlighten us about the right. One reason most people believe that ethical egoism is wrong is the fact that we do not tend to hold praiseworthy people who do good for the sake of doing well. If egoism were true, such people would be acting for morally relevant reasons and thus should be perceived as praiseworthy for their actions. Unfortunately, this sort of insight (or any other kind) is not likely to end the debates about right and good actions anytime in the near future, and so there is inevitably going to be some incompleteness in a theory of moral worth.

But while this account is not meant to be complete, two nagging questions need to be addressed. A relatively simple question, concerning

the role of moral knowledge, shall be addressed at the end of this chapter. This section will deal with the more complex question of the *misguided conscience*. The world is full of people who are devoted to morally wrongheaded causes. Some of these causes are terribly wrongheaded, like Gobbles'. Others are less wrongheaded but still wrong: perhaps one is devoted to promoting chastity, convinced that chastity is an important moral cause. The chastity promoter, while she is harmless compared to the Nazi, still causes moral harm in the world. She may, for example, choose to donate her money to the Chastity Society instead of the Cancer Society, and she may make a variety of morally wrong choices due to being blind to the virtues of some unchaste individuals or the vices of some chaste ones (e.g., she may vote for the morally inferior candidate or send her child to a private school inculcating false moral views). Whatever the degree of one's wrongheadedness, one must ask, Is there any merit in the devotion of a person to a wrong cause? If a person *does* conquer his desire to help a slave escape, does it always speak only badly of him? Is there ever anything to say in favor of the idealistic Nazi soldier who *does* conquer his inverse akratic urges in service of what he takes, ever so mistakenly, to be a good cause?

Some would suggest that there may be something good about having self-control or the ability to "stick to one's guns"—an ability on which many loyal Nazis prided themselves and which Huckleberry Finn accuses himself of lacking. Note, however, that the loyal Nazi does not necessarily have great self-control—his hatred of Jews and love of power may simply be stronger than any "temptation" for him, a state that does not speak well of him. Similarly, Huckleberry Finn need not be particularly lacking in self-control—it may simply be the case that his reluctance to treat a human being like escaped livestock is stronger than any inhibitions he may have, which does speak well of him. But to the extent that there is a property called "self-control" that is independent of the content of the desires to be controlled, it is hard to see it as anything but morally neutral, a tool that we want the moral agent to have and we hope the immoral agent lacks. Let me explain.

Recall the difference between Aristotle's brave person and Aristotle's self-controlled (enkritic) person. Aristotle's brave person performs as he does on the battlefield *not* because he has great self-control, but because his concern for honor exceeds his fear of death—thus showing not self-control, but virtue. Many cases in which we tend to think that a person

has self-control are in fact cases analogous to Aristotelian bravery—cases in which a person is simply more concerned with one thing than another. For example, if I am acting kindly toward Russell even though I am very angry at him, it may show that I have great self-control—that is, an ability to suppress whatever emotions I wish—or it may be the case that I have no self-control at all, but I am so concerned with Russell's well-being that even my anger cannot override my concern. In the latter case, the fact that I act kindly despite my anger shows something good about me. In the former, all it shows is that I have a sort of mental discipline—a morally neutral skill that, like a soldier's training, can be used for bad ends just as easily as for good ones. We want the moral person to have self-control, but only for the same reason we want the moral person to have money—namely, because self-control helps the moral person attain her goals (just as it helps the immoral person attain hers). To come back to Huckleberry Finn, the fact that he is motivated to help Jim despite his convictions may similarly show one of two things: it *may* show that he has less than an ordinary person's self-control, or it may show that his concern for Jim's humanity is so strong that it defeats his convictions and inhibitions. I was assuming that the latter is the case, which would make Huckleberry paradoxically similar to Aristotle's brave person. The brave person's desire to avoid disgrace is stronger than his fear of death. Huckleberry Finn's well-motivated desire to help Jim overrides his attachment to his misguided convictions. The brave person's desire to defend his city is strong enough that it does not need the aid of self-control to win his inner struggle; similarly, Huckleberry Finn's desire to help Jim is strong enough that it does not need the aid of an unusual lapse of self-control or lapse of consciousness (produced by alcohol, stress, etc.) to win in his own inner struggle. In the cases of both Huckleberry and the brave person, it is a virtuous concern that wins the struggle, and that is all that matters to the moral worth of their actions.

Much more will be said about self-control in chapter 4 (where I shall also discuss crimes of passion and Freudian slips). But what has been said so far suffices at least to show one important point: if self-control is morally neutral, it is hard to see how any morally positive value can ever be attached to morally misguided loyalties, for how can anything positive come from a neutral power being exercised for a bad end? To say that there is never any value to a misguided conscience, however, is to do injustice to the complexity of many misguided conscience cases. To see

how a misguided conscience can say *something* good about its owner, though, we must first take a closer look at the reasons it usually shows one to be bad.

Why Racists, Sexists, and Megalomaniacs Are (Usually) Bad, Not Just Mistaken

There is a lively argument between historians as to whether Hitler was "sincere"—that is, whether Hitler in fact believed his professed views. The question of Hitler's sincerity is perceived as morally disturbing in a special way. What if it turned out that Hitler sincerely believed that Jews amounted to horribly harmful creatures who are out to destroy humanity, and that therefore he was attempting to do the world a lot of good by helping to get rid of them? Would we have to admit that Hitler was not evil, after all? But if Hitler was not evil, who is?

We may know too little about Adolf Hitler's private life to make a study of his moral psychology, but the little that we know is quite consistent with the possibility that whatever the sincerity of his beliefs about Jews, Hitler despised *morality* and only used its language to capture the heart of the masses. If Hitler in fact despised morality and wished to promote something he knew to be evil, there would be no problem in explaining why he was evil; it is easy to show that he acted for sinister reasons. We can also imagine a Hitler who thinks he has a "new morality"—perhaps a master morality as opposed to a slave morality, or an Aryan morality as opposed to a Jewish morality—according to which killing innocent human beings is often just fine. It would not be a problem to explain why such a Hitler would be bad either: like the extremist, above, he is a person who mistakenly believes that he is acting for moral reasons, but who in fact acts for sinister reasons or out of indifference to (what in fact is) morality, no matter how he happens to name it. But the person whose badness the historians fear they would have trouble accommodating is the person whose anti-Jewish actions were based not on false moral beliefs but on false *factual* beliefs: he mistakenly, or as a result of a damaged brain, believed it to be a fact that Jews were trying to take over the world and subjugate gentiles. Given this false belief, what else could a decent fellow do?

Fortunately for the uncomfortable historians, Hitler does not fit this description very well, but the same question could be raised of many other bad people and evildoers of the more ordinary sort. We can well imagine a more run-of-the-mill anti-Semite justifying her ill treatment of Jews by citing her belief that Jews are subhumans who are trying to conquer the world and that treating them badly may improve the state of the world by letting them know that they are not likely to conquer it in the near future. This person, too, is blameworthy, and, if her racism is rabid enough, she is obviously evil. Yet it seems as if all this person gets wrong is *the facts*: she does not have false beliefs about morality but simply false beliefs about Jews. If she acts badly toward Jews, it seems as if her false beliefs about them—sometimes referred to as her “ignorance”—should excuse her action. Why, then, is she blameworthy?

Racists, sexists, and their ilk are not the only sort of people who are regarded as bad or vicious despite the fact that their behaviors seem attributable to a false belief. Imagine a person who snubs his coworkers because he believes, falsely, that he is much superior to all of them in intelligence, virtue, skill, and practically everything else. Or consider a parent who beats his children quite severely because he believes that beating builds character (one may imagine Klaus Mann's Professor Urat believing similarly that his cruelty helps his students). These people also seem to have false factual beliefs: the belief that one's coworkers are annoying idiots (assuming that the believer has reasonable definitions of his adjectives) is a factual belief—we can easily imagine what it would be like for it to be true. The belief that children benefit from “good beatings” is similarly a factual belief about child psychology, though a false one. Yet we often take people like the megalomaniac and the child-beater to be blameworthy.

Usually, a false factual belief does seem to be a good excuse, and often it provides an exemption from moral blameworthiness. A paradigmatic case would be the following: Boko Fittleworth (a character in a P. G. Wodehouse novel) overpowers and traps a man whom he spots hiding in his would-be father-in-law's garden shed at midnight, because he believes this man to be a burglar. In fact, the man is not a burglar but a business tycoon whose presence in the shed is part of a secret, unlikely, and harmless plot in which the future in-law is a willing participant.

Though Boko, of course, gets himself in serious trouble with his future father-in-law, and though he owes the businessman an apology, he is

still not blameworthy for his action. It is wrong to treat a fellow human being violently, but Boko has a perfect excuse—“But I thought you were a burglar.” We need to understand why the anti-Semite and her friends do not have a similar excuse—something like “But I thought you were a member of a worldwide conspiracy.”

The beginning of the answer has to do with the epistemic rationality of Boko's belief, when contrasted with the epistemic irrationality of anyone's belief, in this day and age, that all Jews are all monstrosity greedy people engaged in a conspiracy to control the world. It is eminently epistemically rational to believe that an unknown man hiding in a known garden shed at midnight and refusing to identify himself is probably up to no good. Thus, Boko's belief is what I would call a false, epistemically rational belief—what we would usually call an *honest mistake*. On the other hand, unless one has just arrived from another planet, armed with a seriously flawed travel guide to Earth, it is difficult for one to reach the belief that all Jews are greedy people engaged in a conspiracy to control the world by way of an honest mistake. Many people who have held such beliefs about Jews met enough Jewish people, and knew enough of history, to be able to see the unlikely nature of that sort of conspiracy. Their beliefs are not honest mistakes, nor the result of mental retardation, but the result of motivated irrationality, sometimes expressed in words such as “I was ready to believe anything about Jews, as long as it was bad.”

We do not take people to be bad if they hold bigoted beliefs out of pure ignorance, or mostly out of ignorance. This is hard to show, of course, because such cases of pure ignorance are extremely rare, but I would imagine that if an alien arrived on Earth with a used copy of a generally reliable Guide to the Solar System telling her that all black Earthlings are foolish and all pale Earthlings clever, she would be misled, not irrational, and she would not be bad simply for having made this mistake, even if she acted on it. For a more mundane tale, consider the case of Solomon, a boy who lives in a small, isolated farming community in a poor country. Solomon believes that women are not half as competent as men when it comes to abstract thinking, or at least are not inclined toward such thinking. Solomon's evidence for his belief is the fact that all the women in his community, despite his attempts to engage them in learned conversation, seem to discuss nothing but gossip, family, and manual work, that the few people in his community who are interested in abstract thinking are all men, that no one he knows of has ever

doubted that women are worse abstract thinkers, and that the community's small, outdated library contains abstract work written by men only. Solomon's belief is false, but it is not *particularly* or *markedly* irrational. It is not particularly irrational because Solomon is not exposed to striking counterevidence to it, and he is exposed to a consensus and "expert" opinion in its favor (just think how many of our own everyday beliefs are grounded simply on expert opinion and lack of clear counterexamples). True, if Solomon were to think more carefully, he might find reasons to change his mind, but in his case "thinking more carefully" would involve having the intelligence of John Stuart Mill without the advantage of having known at least one woman who is inclined toward abstract thinking. While no person is perfectly rational, it is unfair to accuse a person of being exceptionally irrational just for not being exceptionally intelligent.¹³ The boy Solomon is more ignorant than irrational, and he is also, intuitively, more ignorant than morally vicious—he seems much less deserving of the title "sexist pig" than a contemporary American who holds the same view.

Imagine, on the other hand, that Solomon gains a scholarship and finds himself a student in an excellent academic institution, where he proceeds to study his favorite abstract topic. In college, Solomon sits shoulder to shoulder with brilliant female students and is taught by brilliant female professors. At the end of his first year as a college student, if Solomon were rational, he would have changed his mind about the aptitude of women for abstract thinking. If at the school year's end Solomon still believed that all women are bad abstract thinkers, his belief would now be not only false but also irrational. He would no longer be simply mistaken, but *prejudiced*. At the point when we regard him as irrational, we also regard him as suffering from a serious moral flaw.

So an honest mistake is an excuse and often exempts one from blame; prejudice, which is an irrational belief, does not. But why, exactly? There is nothing morally vicious about irrationality *per se*. After all, there seems to be nothing morally vicious in believing—as irrationally as can be—that Elvis Presley is still alive or that a lottery ticket is less likely to win if it carries the same number as the ticket that won last month.¹⁴ Yet even if there is *nothing* morally vicious in epistemic irrationality *per se*, there still seems to be something vicious about the anti-Semite, the educated yet sexist Solomon, the megalomaniac, and the

child-beater, a vice that does not exist in the person who commits the Gambler's Fallacy.

Consider the sentence "I hate him because he is disgusting." Often, this sentence evokes smiles. It does not seem to provide a true explanation of why "I hate him," because unless there are obvious reasons to think of the object of hate as disgusting in some way or another, the hearer is likely to infer that the speaker puts the cart before the horse: the truth would be more likely expressed in the words "I see him as disgusting because I hate him." Suppose for a moment that all of this is true of Joshua and me. I mistreat Joshua, and I tell you that I do it "because he is disgusting." You, a rational observer, see that no one would see Joshua as disgusting unless her belief formation were colored by hate or some other emotion or desire. You call my bluff: "No," you say, "you treat Joshua like that because you hate his guts. The fact that, with all he has done for you, you still believe that he's disgusting only shows how much you hate him. You really are unfair."

This is a story in which my attribution of an action (mistreating Joshua) to a believed state of affairs (he is disgusting) is not to be taken at face value. Perhaps my mistreatment of Joshua is motivated simply by hatred, the belief "he is disgusting" being a mere epiphenomenon of the hatred. Perhaps my mistreatment of Joshua is simply the result of my wanting to hurt someone and his being a readily available victim, and my belief that "he is disgusting" is an excuse that I develop to rationalize my action, however inadequately. Perhaps I believe that Joshua is disgusting because for some reason or other it is comfortable for me to believe that he is, and my action is my perverse way of confirming and strengthening my belief. In all these cases, the motive for my action seems to be quite unsavory, and all these cases are different from the case in which I mistreat Joshua because I heard from a very reliable (but, in this case, inaccurate) source that Joshua is a serial killer and my actions toward him are understandable or even justified given that assumption. Only the latter case would be analogous to the case of Boko Fittleworth.

I would like to suggest that the anti-Semite and her friends are also generally presumed to act out of unsavory motives, and their claim that they do X because they believe that Y is as suspicious as my "because he is disgusting." Thus, my point is *not* that, à la Robert Adams (1985), the anti-Semite's *belief* is itself blameworthy, but that the anti-Semite's belief

plays a rather superficial role in a drama in which a person is motivated by sinister desires. When looking at the blameworthiness of my actions when I hurt Joshua, what matters is the fact that I hurt him because I hated him, and my "because he's disgusting" does not matter much. The anti-Semite hates the Jew "because he's disgusting" or rather believes that he's disgusting because she hates him, or even worse, because she wants to hate *someone*. Sartre has already said that if the Jew did not exist, the anti-Semite would have invented him. Presumably, Sartre meant that if the Jew did not exist, the anti-Semite would have found someone else to scapegoat, and if the anti-Semite did not come across the world-conspiracy idea, she would have found a well-poisoning or a Tzar-killing story to believe instead. The anti-Semite mistreats the Jew because she wants to mistreat someone and is therefore ready to believe anything about the Jews, as long as it is bad. Thus the anti-Semite is not doing what she does because of an honest mistake, but because of a sinister motive; and the more far-fetched her beliefs about the Jews, the more the evidence shows that her sinister motivation is strong. If unjustified hatred is a vice, a hatred so strong that it produces near delusions is an even stronger vice. That is why, if Hitler's strange beliefs about the Jews were the result of hatred's distortion of his perception of the world, it would justify seeing him as *both* evil *and* crazy—the way we see him most of the time.

It is interesting to compare the behavior of those whose motivation resembles the anti-Semite's to the behavior of those who are honestly mistaken—or those who are honestly correct. One characteristic difference between them is that the motivatedly irrational person tends to make more of her belief than the honest believer ever does. The Bus Drivers' Union in England believed that women are weaker than men—and in the 1970s used this belief as a premise in arguing that women should be barred from driving buses. Compare this approach with that of Plato, who also believed that women are weaker than men but pointed out that as this is true only in general terms, any woman who demonstrates that she is strong enough to be a guard should be recruited as a guard. Plato's approach is more characteristic of the person who "simply" believes that women are weaker than men. The bus drivers' demand is so unwarranted by the belief that they claim as its reason that one suspects that the "weakness" argument is a mere rationalization of a desire to keep

women "in their place," and if the excuse of "weakness" did not exist, they would have had to invent another one.

Another common difference between the innocent and the motivatedly irrational person is that the latter is likely to show signs of bad faith. The learned officers of the Holy Inquisition based their actions on the belief that they were not punishing their victims, but rather were rewarding and curing them by saving their souls from Hell. Yet, it seems, the inquisitors often treated their personal enemies to their "cure," even if they believed them to be relatively without "sin," and never offered it to their friends, presumably even if they believed them, privately, to have a lot of sins weighing on their souls. Their behavior is significantly different from the behavior of those who, after a good experience with a certain method of therapy, even quite a dreadful one, begin to urge everyone to try it. Thus it seems likely not only that some inquisitors winked inwardly when they repeated their claim that they were not administering punishment, but also that even some of those who never did were motivated in their actions not by the desire to save souls, but by the desire to hurt people they disliked or to exercise indecent amounts of power. The belief that they were saving souls was just a particularly comfortable, sinister rationalization.

Let us take an even closer look at why the anti-Semite and her friends are blameworthy. There are some cases in which the answer is easy: they are acting from morally bad motives. The anti-Semite may be the sort who desires to mistreat others, and so whatever she says about conspiracies, her actual motive is the simple desire to mistreat others. The man who mistreats his children may have a similarly sinister motive, whatever his tales of character. The person who takes herself to be superior to all her coworkers may simply want to have unfair power over people and humiliate them, her real or imaginary superiority being a convenient inner excuse for action. But in other cases, one's motive for acting (and even one's motive for believing) is not by itself evil. For example, perhaps the man who beats his children was beaten himself, and his need to love his own parents, and rationalize to himself the fact that he has allowed them to mistreat him, attracts him to the belief that there is nothing wrong with the way his parents treated him, as they were only trying to build his character; in order to prevent cognitive dissonance, he beats his own children himself. The person who mistreats her coworkers may be deeply insecure and evade this insecurity by telling herself that

she is such a genius that her peers' criticisms mean nothing to her. The anti-Semite may be anti-Semitic because all her friends are also anti-Semitic, and she deeply wishes to fit in. Solomon, if he keeps his youthful belief about the intellectual inferiority of women despite the counter-evidence offered by his new circumstances, may simply be tired of the onerous job of constantly changing and adjusting his attitude to a new, perplexingly sophisticated world and be clinging nostalgically to the epistemic comforts of home.

All of these motivational stories look quite innocent. Wishing to love one's parents, to avoid insecurity, to fit in with one's friends, or to have a simple view of the world are all perennial motives that influence our actions and beliefs, rationally or otherwise, often enough, whether we are morally bad or good or average. What is wrong, then, with the child-beater whose motives are nothing but the desire to love his parents and improve his children's characters or the anti-Semite who only wants to fit in with her friends? Blameworthiness involves either the presence of a sinister motive for action or the marked absence of concern for morally relevant factors. If no sinister motive is present, the guilt of these agents must have to do with the *absence of moral concern*. Perhaps the child-beater is only trying to love his parents, perhaps the anti-Semite is only trying to fit in with her friends, perhaps the megalomaniac is only trying to evade insecurity—but they are trying to do this *at the expense of others*, at the expense of morality. There is nothing wrong with wanting money, but a person who kills for money is blameworthy because he fails to respond to pertinent moral reasons. The same is true for the person who beats his children in order to better love his parents. He, too, is criminally selfish.

This point needs refinement, because the analogy between the cases seems forced. Consider the following three cases:

(1) A person beats his children quite severely because he wishes to please his parents, who reward him for delivering such beatings.

(2) A person beats his children quite severely because he wishes to please his parents, as above. He is, however, self-deceived about his motives. To make himself feel better, he rationalizes that he beats his children because "it builds character."

(3) A person beats his children quite severely because he believes that beating one's children builds character. His belief is the result of his strong wish to believe that his parents, abusers themselves, are decent people; this wish, in turn, is the result of his wish to love them.

(1) is the typical case of the person who, for love or money, defies morality. (2) is similar to (1), except that a rationalization is added. The person's belief in "character" is, in a sense, irrelevant to his blameworthiness—only one fact matters, namely the fact that he values the approval of his family more than his children's well-being. Things are quite different in (3). In (3), the child-beater's belief that beating builds character, however irrationally formed, plays a real role in his behavior. His desire to love his parents led him to form the belief. Now, however, he beats his children not, as in (1) or (2), out of a desire to love his parents *simpliciter*, but rather because he believes that his beatings will improve their characters. Why, then, is he still blameworthy for his action? The answer is that if he cared appropriately about morality or about the welfare of his children, he would not have formed the belief that severe beatings build character, or he would not have been motivated by it to beat his children.

This is a very delicate point, as it can easily be confused with two false claims. One is the claim that since the beater's belief that beating builds character is self-deception, he is blameworthy for this very act of belief formation. After all, he knew at some level that his belief was false and that it was going to cause him to behave badly, and if he cared more about his children he would not have caused himself to have that belief. This claim rests on an overly Cartesian notion of belief formation or self-deception, and I have explained in chapter 1 my reasons for rejecting the idea that self-deception is an intentional, voluntary action. Belief formation, as far as I am concerned, is not intentional, as is shown by the fact that one cannot be forced to change her beliefs at gunpoint. Another claim I wish to avoid is the claim that we are responsible for our irrational beliefs in the way that one is responsible for "culpable ignorance" or criminal negligence—the view that we have a duty to "check" on our beliefs the way that we have a duty to check the brakes on our cars. This may be true in some limited contexts, but as we do not even know which of our beliefs need checking more than others, and as it is in the nature

of some kinds of irrationality that one cannot detect it in oneself, it makes no sense to say that we always have such a duty.

On the other hand, it is true, as I said earlier, that what we care about influences our deliberation, perception, and other cognitive processes. Freud's love of antiquities, mentioned in chapter 1, may have caused him to imagine antiquities shops when they did not exist, but he never failed to notice a real antiquities shop, even those that most people would have passed without noticing. If you care passionately about being at work on time, you are less likely to underestimate the amount of time you need to get to work—which is why you, unlike the rest of your group, are never late. If you are a lover of birds and care very little for buildings, you might well remember the bird that you saw on top of Buckingham Palace and not the building itself. If feminism is your life, you may have trouble reading nineteenth-century novels because you feel sorry for the female characters, and if your loved one is a vegetarian, a movie about Argentina is bound to make you wonder how on earth vegetarians get along there. If truth matters to you a lot, it will take longer for you to accept a piece of gossip than it would take for a person to whom truth does not matter much—not because you would feel like you have a duty to doubt it but because you "naturally" will doubt it more.

Similarly, those who care about treating people fairly are not quick to believe a very unlikely, morally incriminating story about a third party, even if this third party is someone of whom they would love to think badly. If a woman is told by an unreliable source that the husband who left her is in fact a criminal, she will probably be disposed to believe the story—except if she cares very much about treating people fairly, in which case she is likely to come to no firm conclusion; the high stakes involved in treating a person as a criminal when he is not one would be too salient for her to accept a cock-and-bull story so quickly. If you care about fairness, you will not easily delude yourself, in the face of counterevidence, that your coworkers are stupid, however soothing of insecurity such a belief might be (of course, if you respect your coworkers' personality, you would not even find the *actual* stupidity of coworkers a reason to mistreat them). If respect for persons is a deep concern for you—deeper than the desire to be popular with your anti-Semitic friends—you will think more than twice before accepting the view that Jews are not persons in the face of powerful evidence to the contrary. If you care deeply about your children, you will not be quick to believe, in

the face of glaring counterevidence, any view that implies that it is helpful to beat them severely. That is why, while some people who were beaten by their parents beat their own children, many do not. In fact, many people who made excuses for their abusive parents for years snap out of their "denial" the moment they have children whose own interests are at stake. Thus, a person who was severely beaten as a child is justified in being angry at his parents, accusing them of selfishly caring more about maintaining their illusions about their parents or whatnot than about the suffering of their child. Thus, even if the anti-Semite, the megalomaniac, or the child-beater is not motivated by sinister motives, she is still guilty of chilling indifference to moral reasons.

Good Motives, Bad Causes

To recapitulate: believing something such as "Jews are germs" is not an excuse for a bad action if only a mind that is maddened with hatred, or that is chillingly indifferent to fairness and compassion, could accept that belief in the first place. Thus, it seems that if a belief in Nazi doctrine speaks ill of a person, loyal devotion to Nazism would speak even worse of her. It could only speak well of her to the extent that the devotion is fueled by morally good motives. Above, we said that the sheer desire to "stick to one's guns" is neutral—not by itself a good motive. What good motives *could* there be in one's attachment to a bad cause?

There could be none, if it were true that being an advocate of an ideology always implied having the beliefs required by the ideology, and if being an advocate of a person, theory, or country always involved being an advocate of what the person, theory, or country really is. However, things are much more complicated than this. To say that Jennifer believes, for example, in the truth of the Bible is not to imply that it is in general possible to take a text from the Bible, put "Jennifer believes" in front of it, and produce a true statement. Jennifer's love of the Bible may be largely due to a strong sense of the rightness of the "do unto others" slogan and the idea that all human beings somehow deserve love. When Jennifer feels the need to defend the Bible, she feels that she is protecting these things—and to the extent that these are things that move her, her devotion to the Bible does speak well of her, regardless of what the Bible might say about homosexuals or the sins of the fathers being visited

upon the sons. On the other hand, Dana may be attracted to the Bible largely because it provides her with a good excuse to look down at people she dislikes or to hate homosexuals, in which case her devotion speaks ill of her. The two people may be very different morally, and they would be wrong if they assume, as people often do, that they are likely to get along because "they are both Christians." While they may almost always produce identical statements about their beliefs, we feel that there is a sense in which Christianity "means different things to them."

Because any cause—including a bad cause—may "mean" very different things to different people, and thus a bad cause can mean good things to some people, there will be cases in which a person's attraction to a bad cause would speak well of him. It would rarely speak *nothing* but well of him, because it is rare that the reasons which attract a person to an unwholesome cause are all wholesome, or that a bad cause means nothing but good things to someone, but there may still be something good about it. Klaus Mann provides us with the (surely real) character of Hans Miklas, a young, uneducated, and very naive actor who becomes a member of the Nazi Party while it is still a fringe group. Hoefgen, a fellow actor and a socialist, refuses to share the same stage with him out of dramatized disgust. Yet the reader gets the impression that Miklas is not really a bad man. His attraction to the Nazi ideology appears to rely a lot on his desire for social justice and on his disgust with the (real) corruption around him. In his mind, he associates "German honor" not with conquering other countries, but rather with a corruption-free public life. The meshing of "honor" and social justice in his mind is the result of his believing the propaganda according to which many of the rich who are getting unjustly richer are Jewish. After Hitler comes to power, Miklas, despite being promoted and respected due to his long-standing party membership, quickly discovers that not only has social injustice remained ubiquitous, but it has in fact increased, and so has corruption. He is so devastated that he consciously destroys himself by publicly denouncing the regime. This action authenticates the suspicion that, from the start, his attraction to Nazism was a morally mixed bag, containing an element of true attraction to justice. One wonders if he is not in fact better than his colleague Hoefgen, whose socialism does not stop him from choosing to cooperate with the regime when he can easily leave the country instead. The reason one wonders is not that devotion to a cause, regardless of content, is morally better than fickleness, but because, as

his later actions testify, Miklas's devotion was partially a result of responsiveness to truly moral reasons.

Can one really act for a mix of morally good and bad reasons? It seems Miklas does, because, though some frustration and hatred have obviously affected his infatuation with the Nazi movement instead of some other workers' movement and have made him all too willing to believe anti-Semitic propaganda, a true concern for justice surely contributes to his great disappointment with the Nazi regime. (To complicate the character even further, Miklas's Nazism is also the result of his ignorance and cognitive limitations, which makes it hard, in his case, to tell innocent false beliefs from irrational, ill-motivated prejudice.) It may seem, from a Kantian¹⁵ point of view, that one either is motivated by morally relevant reasons or is not, without the possibility of mixed motives, but this is not the way we think when we judge people's motivations. Granted, in some contexts it may seem as if it is. To borrow an observation from Korsgaard, if a student tells us that he has decided to study calculus for the sake of the intellectual challenges involved, but then we learn that it is also a required class, we tend to take quite a cynical attitude to his claim that he is motivated by the intellectual challenge. But consider a student who has always wanted to take calculus since her first contact with high school physics, but has been inhibited about doing so because she fears that her cognitive abilities are not up to the task. Upon entering college, she discovers that a certain degree requirement can be met by one of a few classes, including Introductory Calculus. She decides to take calculus, thankful for the requirement at the same time as she fears it. While this student would not have taken calculus if it had not been required, it would be unjust on the part of her math professor to treat her on par with her peer who is dragged, kicking and screaming, into the required class, trying to avoid intellectual challenge to the extent that is compatible with earning a good grade. Analogously, issues of moral worth become complicated when we consider those whose motives are mixed. To fail to kill a person because she is a person is to respond to moral reasons; to fail to kill a person only because she is a fellow Jew is not to respond to moral reasons; but what does one say of those who would not kill someone because she is "both a human being and a fellow Jew"? As with the tale of calculus, the answer depends on the details of the case and on the amount of concern for Jewishness, as opposed to concern for personhood, involved in the person's (in)action. Sorting

this out is especially difficult because some people use the word "person" to mean "a fellow Jew," and some people would talk about helping a fellow Jew but mean in this something as close to helping a person as it gets.

Thus we may vindicate Kant's view that, as moral worth depends on the complex matter of people's motives, diagnosing moral worth in real life can be very difficult—perhaps even more difficult than Kant himself believed.

Knowledge of the Virtues

In addition to the view that there must be something morally good in self-control, there is a related, even more venerable view, according to which there is something especially good in having knowledge of the virtues, correct moral convictions, and/or the ability to deliberate well on moral matters. If acting for moral reasons is not the same as acting for what we take to be moral reasons, and depth of moral concern is not the same as conscious commitment to moral principles, how can this be true? I have conceded earlier that Huckleberry Finn would be a better person if his conscious moral convictions were not appalling. However, some people may want to argue that in addition to this fact, it is also true that there is something especially regrettable in the fact that Huckleberry Finn's character deficiency is located specifically in his ability for moral *deliberation*, as opposed to somewhere else. I believe that there is something to this claim, but that the goodness of having the ability to deliberate well about morality can be maintained without denying that Huckleberry Finn's action is morally praiseworthy.

Intellectual knowledge of morality and the ability to deliberate well about all matters moral have obvious things to recommend them. To begin with, an agent in Huckleberry Finn's position, who does not know his virtues from his vices, is likely to try to make himself a worse person, whereas the person who knows his virtues from his vices is likely to try to make himself a better person. More important, there are certain types of morally desirable actions that are very hard to perform if one does not have the right moral principles or at least the ability to deliberate well on moral matters. For example, it is very hard to vote for the right political candidate if one cannot deliberate well on moral matters (and, to some extent, on other matters such as economics and human behavior).

While some of those who voted for Hitler were obviously bad people, others strike us as morally average people (or, in the case of Miklas, somewhat better) whose unreflectiveness, simplicity, and incompetence at deliberation made them easy prey for the candidate who promised jobs, more national pride, and order in the streets. Given the complex moral decisions that even the most mundane life has to offer, it seems that we have a moral duty not to allow ourselves or our children to be too stupid, unreflective, or uninformed, especially with regard to morally relevant issues. The fact that a person—even a good person—cannot deliberate very well or is ignorant when it comes to morally relevant matters has a special sadness to it.

Conclusion

It has been my goal in this chapter to sketch a quality-of-will-based theory of moral worth and show that such a view is potentially useful in understanding some facts of moral life. On my view, people are praiseworthy for acts of good will and blameworthy for acts of ill will or the absence of good will, and the amount of praise or blame they deserve varies with the depth of their motivation or the extent of their indifference. But since, I hold, good will is wanting, noninstrumentally, to perform actions that have whatever property it is that makes actions right, a full account of moral worth is impossible until we know what property it is that makes actions right. I shall not try to fill this gap in this book, nor shall I attempt a full account of acting for reasons. Other gaps that remain in my account concern the significance of my view for moral responsibility and autonomy. I shall discuss this topic in the next chapter.

11. Smith later points out that it would be more precise to say that the person who believes that she should ϕ in *C* is committed to believing that if she were fully rational, she would desire that, were she her actual, imperfectly rational self, she would ϕ in *C*. Smith makes this correction, for example, in his reply to Swanton (1996) in order to answer objections based on the idea that sometimes one's hypothetical, fully rational self is so different from the actual self that it makes little sense for the two selves to desire the same things. I take his response to be generally effective, and in any case the difference between the versions of Smith's claim does not matter to my subsequent discussion.

12. It is especially not my point to ascribe to Smith any particular theory about the genesis of akratic action.

13. If we thought of Sam not as developing or failing to develop a desire to become a hermit, but as developing or failing to develop the absolute judgment "I must become a hermit" from his all-relativized judgment that he should become a hermit, the big picture would not change: it would still make Sam's psyche more coherent not to develop that absolute judgment, which would accommodate one belief but contradict many more.

14. A similar position to the coherence view of rationality is the view that judgments of practical rationality, like judgments of probability, are always judgments of rationality given a certain set of beliefs or conditions. Smith may be alluding to this position in arguing that Sam would be irrational in not becoming a hermit given that he believes that he should become a hermit. I would, however, argue that given the whole story about Sam's beliefs and desires, he would be less rational in becoming a hermit than in not becoming one, and that since the latter involves more that is relevant to what Sam has a reason to do, it is a better basis for determining whether, all things considered, Sam acted rationally or not.

15. A judgment recalling the remarks of Frankfurt in "Rationality and the Unthinkable" (1988).

16. My point regarding Paul's case and Sam's case does not depend on the claim that emotions have cognitive content. Even if they do not, it is still true that Sam is not irrational simply because he felt anxiety while he was deliberating, but because anxiety, or something else, caused him to make a poor inference. In the case of Paul, even if one does not assume that his anxiety was by itself somehow reasonable, it is obvious that it did not prevent him from reaching a practical conclusion in a reasonable manner.

17. In some contexts, the view that *akrasia* is a form of irrationality is contrasted with the view that it is the result of failure of self-control or volition. I find it hard to see how failure of self-control or volition can be a rational process—if it is not irrational, it is nonrational.

18. Again, this is evident in nearly all of the extensive literature about weakness of will. From "How Is Weakness of the Will Possible?" (Davidson 1980) to

"Where Does the Akratic Break Take Place?" (Rorty 1980) and many more, the question always appears to be taken to mean "What kind of unfortunate mental glitch causes one to act against one's best judgment?"

19. Thus, he is not acting for what Davidson calls "renegade reasons"—he is not acting for reasons that he already considered and dismissed but for reasons that he overlooked.

20. Once again, it does not matter whether we take Sam's failure to abide by his decision to be a failure to develop a desire or failure to develop a Davidsonian "absolute" judgment from a Davidsonian "relativized" judgment. In either case, he fails to form the desire or absolute judgment because of good evidence to the effect that the desire or absolute judgment is inappropriate or false, and his only irrationality is in failing to revise the (relativized) judgment he reached by deliberation.

21. Or perhaps, despite wondering what made her stay in the program for so long, she never bothers to examine her reasons for quitting or is unable to cite them due to her lack of introspective skill. Whether Emily or Sam comes to a reflective understanding of the irrationality of her or his former best judgment, or whether they eventually appreciate the reasons for their failures to follow them, does not affect my argument for the rationality of this failure.

22. Similar results are discussed by Churchland (1995).

23. One account of rationality that would seem to fit would be the one offered by Scanlon in his 1998 book or in his forthcoming article, "Reasons and Passions." I do not wish to defend this specific account of desire and reason, though I agree with his arguments in "Reasons and Passions" against the idea that there exists a Faculty of Reason.

24. I clarify the specific relation of moral knowledge to reflection and deliberation in chapter 4.

25. As you grow experienced, I am told, educated "instinct" and "intuition" will play a larger and larger role.

Chapter 3

1. I elaborate on this point in "Hamlet and the Utilitarians" (Araply 2000a); a similar point, with a fascinating discussion of the various ways in which emotions can be appropriate or inappropriate, is found in D'Arms (2000).

2. For an alternative, sophisticated treatment of Huckleberry Finn, see Hill (1998). Driver (1996) also takes for granted that Huckleberry Finn's action is meant to be understood as praiseworthy, but her account of this praiseworthiness is completely independent of Huckleberry's motives and reasons. An interesting discussion of one kind of inverse *akrasia* is in Frankfurt's "Rationality and the Unthinkable" (1988).

3. There is nothing very unusual about perceiving fairly sophisticated truths without perceiving that one is perceiving them—as people discover retrospectively when they realize that they can find their way around a city they had thought was still foreign to them, or when the confession of a cheating spouse is surprisingly unsurprising, as was argued in chapter 2.

4. Hursthouse (1999, 195) addresses a previous version of my argument.

5. For a Huck-friendly view of moral reasons, see Scanlon (forthcoming). Railton (unpublished) also defends a Huck-friendly account of moral reasons.

6. The person who cares about morality, but not very much, is implicitly discussed in Susan Wolf's "Moral Saints" (1982) and more recently, and explicitly, in work by Sigrun Svavarsdóttir (1999).

7. One can, if one is so inclined, apply Michael Smith's (1994) discussion of the distinction between concern for morality *de re* and *de dicto* here.

8. A *fortiori*, I shall not try to develop a fully fledged positive view of what it means to be concerned about something in the first place. I am not committed here to Frankfurt's view of caring, but I am not sure that I say anything in this paper with which it is incompatible—at least given the assumption that wanting to have a desire does not imply thinking that one *should* have it.

9. For a sense of the diversity of prevailing views, see, e.g., Dretske (1988), Stalnaker (1984), Strawson (1994).

10. By "essentially conflicting attitudes" I mean attitudes that conflict because of their content and not merely because of some accidental factor making it hard or unrealistic to act in accordance with both of them. For example, love of sports and love of philosophy may conflict accidentally if it so happens that one has to choose between a philosophy talk and a basketball game, or between a sports-related career and an academic one. The conflict here is a contingent one—it has to do not with the nature of sports or philosophy but rather with the fact that one cannot be in two places at the same time, or with having merely "one life to live." On the other hand, there is an essential conflict between one's love of philosophy and a craving to live the life of a "simple soul," between (certain forms of) Christianity and lust, etc.

11. My discussion of the philanthropists owes a lot to Hursthouse (1997) and Foot (1978).

12. The connection between virtue and depth of concern is evident, though never quite explicit, in Foot's *Virtues and Vices* (1978).

13. Some authors, such as Moody-Adams (1994, 291–309), would argue that since Solomon's beliefs were influenced by cultural bias, he is still irrational. I agree that some cultural bias is present in everyone's early education. However, since we all have to form beliefs under conditions of ignorance, Solomon's belief is only about as irrational as most of our everyday beliefs, and thus it can be compared favorably to the belief of a twentieth-century American

sexist, which is formed in circumstances where many more human beings would "know better." To use Russell Hardin's expression, Solomon is more "epistemically crippled" than anyone the reader is likely to know (Hardin, forthcoming).

14. For a discussion of authors who believe that there is something morally bad in irrationality itself, see, for example, Richard Feldman (2000) or Guy Axtell (1997).

15. See Herman (1993).

Chapter 4

1. After all, a person who has made a decision on the basis of nonculpable ignorance is still praiseworthy if he does what seems morally imperative, or blameworthy if she neglects an apparent moral imperative.

2. In Frankfurt (1988).

3. See Wolf (1993).

4. Strawson (1962).

5. Described in Frankfurt (1976).

6. Frankfurt (1976).

7. Frankfurt (1976).

8. It may be that it is common in North American culture to viscerally expect self-control from oneself and experience any failure of self-control as disturbing, baffling, and surprising. On the other hand, Russian novels are full of people who perceive themselves and humans in general as weak and passive, and are pleasantly surprised, and grateful to God, for any exception to this (for example, Lebedev in Dostoevsky's *The Idiot*).

9. Timothy Schroeder and I presented an extended version of these arguments in "Alienation and Externality" (1999).

10. David Copp (forthcoming), however, argues for autonomy being identical to rationality.

11. Though I pick on Mele, similar examples could have been drawn from John Martin Fischer and Mark Ravizza (1998).

12. For a further discussion of visceral self-images, see Schroeder and Apaly (1999).

13. Michael Bratman may be such an author. When claiming that planning is essential to autonomy, Bratman does not attempt to show that we are not responsible for unplanned actions but rather that creatures who never plan are never morally responsible and are not full-fledged persons. See Bratman (1999).

14. The claim that we have little control over the development of our character is made and defended by Sher (2001).

15. Thanks to David Velleman for bringing this objection to my attention.