My Virtue Situation

Gilbert Harman

December 4, 2005

Here are some more or less autobiographical remarks describing my education in virtue ethics.

1 Background

At Swarthmore College in the late 1950s, Michael Scriven taught us about double-blind testing and how such testing indicated that certain sorts of psychotherapy are no more effective than simply doing nothing, a result that raised the question why psychotherapists were so certain their psychotherapy was effective. What psychological principles might explain their misplaced confidence? Much later I became interested in whether such psychological principles would also lead to the conviction that people had character traits even if they didn’t.

Meanwhile, outside of class, we were reading Sartre’s Being and Nothingness\(^1\) which describes how people present themselves to others as having a certain identity as a certain sort of person, an idea that Erving Goffman develops in

\(^1\) Translated by Hazel E. Barnes. New York: Philosophical Library (1956).
The Presentation of Self in Everyday Life. These works raised such questions as whether people “really are” the sorts of people they act as if they are.

We had been led to read Sartre through selections in Walter Kaufmann’s Existentialism from Dostoevsky to Sartre. We also read Kaufmann’s translations of Nietzsche and his book on Nietzsche. Contrasting conceptions of character play important roles in Nietzsche’s writings, in particular in his contrast between master and slave morality.

One of my Swarthmore classmates, John Darley, went on to do ground-breaking psychological research on surprising situational effects on behavior. He and I have been colleagues at Princeton for many years.

David Baltimore, another Swarthmore classmate, went on to win a Nobel Prize in Biology and has more recently been President of the California Institute of Technology. After Swarthmore, he and I shared an apartment in Cambridge for a year while I started graduate school in philosophy at Harvard and he did biology at MIT. In 1961 we discussed Hannah Arendt’s discussion of “the banality of evil” in her reports in The New Yorker about the trial in Israel of Adolf Eichmann.

After graduate school, I began teaching at Princeton University, doing sections in Walter Kaufmann’s course on “Hegel, Nietzsche, and Existentialism,” and in Stuart Hampshire’s historically oriented “Introduction to Moral Philosophy.”

Kaufmann wrote and lectured about the dangers of what he called exegetical thinking—treating texts as sacred truths to be interpreted “charitably,” that is, as not conflicting with one’s own ideas, allowing one to read one’s ideas into the sacred text so that one can read them back out endowed with authority.

---

Hampshire and I talked about writing a paper together about psychological determinism. Our idea was that the strength of desire is not fixed ahead of time, but only after one decides to do something, that decision making it the case that that’s what one most wanted to do. Before one’s decision, it is not determined what one will do; after one’s decision, it becomes true that one’s decision was determined by one’s desires. We thought this had some connection with Kant’s distinction between practical and theoretical standpoints and with Sartre’s ideas about freedom of the will and his distinction between what something is in itself and what it is for itself. Alas, we never actually wrote the paper.


2 Ethics and Human Flourishing

As time went by I became increasingly interested in such functionalist approaches to ethics, especially those that seek to derive everything from a conception of human flourishing. Contemporary writers included not only Foot and Katz, but also Ayn Rand, Elizabeth Anscombe, Abraham Maslow, Henry Veatch, Tibor

Machan, David Norton, John Finnis, Alasdair MacIntyre, and Robert Nozick, among others.9

A standard way to introduce the ideas behind this approach notes that artifacts which have a particular purpose or function can be evaluated in terms of how well they serve their functions. A clock ought to keep time accurately, a good clock is one that does so and something is wrong with a clock if it fails to do so.

A related account applies to bodily organs which have specific functions. A heart ought to pump blood regularly. A good heart is one that does so and something is wrong with a heart that fails to do so. The function of a bodily organ is often related to the flourishing of the animal with that organ. Similarly for functional parts of plants, such as roots. Good roots are roots that are adequate for the flourishing of a plant.

Plants and animals can themselves be evaluated in terms of what counts as their flourishing. A flourishing oak tree is a good example of an oak tree—a “real oak tree,” one might say, whereas a sickly tree might be a “poor oak tree.”

People with certain functions or roles can be evaluated in terms of those functions or roles. A teacher is a good teacher to the extent that he or she teaches well, which perhaps means that she does well at helping students to learn the subject being taught. A teacher ought to help students learn well and there is something

wrong with a teacher whose students do not learn the subject.

Of course, people can be evaluated not just as teachers or plumbers but as people. It is sometimes suggested that one’s life has a certain purpose and that one can be evaluated in terms of that purpose just as a plumber can be evaluated in terms of the purpose of plumbing. Sartre agrees to this with the proviso that the relevant purpose cannot be set by God or society but has to be a purpose that one freely chooses, a purpose that can change from moment to moment as the result of one’s choices.\(^\text{10}\)

It is more commonly suggested that a conception of a good person derives from a conception of what it is for a person to flourish in a sense of *flourish* that goes beyond what is involved in a plant or animal flourishing, because a good person is not just a healthy happy person.

The idea seems to be that people flourish in the relevant sense to the extent that they lead lives that it is good to lead, lives of the sort parents most want their children and themselves to lead. One *ought* to lead such a life. A *good* life is one of that sort. Something is *wrong* with a life that is not of that sort.

Matters are complex, however, since we cannot simply *equate* being a *good person* with having a good life in this sense. Good people can be subject to misfortunes that lead them not to flourish. It is not a straightforward matter to say how human flourishing might be related to what a person ought to do, what it is to be a good person, and what it is for there to be something wrong with a person.

In my paper, “Human Flourishing, Ethics, and Liberty,”\(^\text{11}\). I argue that any reasonable ethics based on human flourishing has to involve moral relativism, because what counts as flourishing is itself relative to one or another set of values.

(An extreme example is represented in Nietzsche's contrast between master and slave moralities.)

One way to base ethics on the basic value of flourishing, is to adopt a consequentialist theory in which the goal of ethics is to maximize human (or animal) flourishing. There are, of course, a variety of such theories, each with counterintuitive consequences, except for those too vague to yield results that might be assessed for intuitiveness.

Another way to get from human flourishing to ethics is to take as an ideal the life of someone who flourishes. Then the right thing to do in any situation is the thing that a person who flourishes would do in that situation. But three problems arise. First, the situation may be one that a person who flourishes would never be in. Second, a person who flourishes may do so because of being able to rely on an excellent character of a sort that the person who needs advice does not have—the flourishing person may be able to trust himself to carry out a promise to do something requiring a kind of bravery that the person needing advice does not have, in which case the person needing advice should not make the promise even though the flourishing person would make the promise.

Third, and perhaps most important, it is completely implausible that our conception of what it is to lead a good noble life is more basic than our conception of what someone ought morally to do on a particular occasion. Surely we figure out what we think a person of good character would do by figuring out what we think is the right thing to do. But then we are not deriving the right thing to do from a conception of human flourishing, we are going in the other direction.
3 Social Psychology

In 1980, I attended a faculty seminar at Princeton University given by the psychologist Nancy Cantor discussing Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*. This seminar had a great impact on me, most immediately on my subsequent thinking and writing about human reasoning and rationality, later in thinking about virtue.

Nisbett and Ross mention research indicating that people are subject to a *fundamental attribution error*, a “tendency to attribute behavior exclusively to the actor’s dispositions and to ignore powerful situational determinants of the behavior” (31). This idea intrigued me very much and in the decade following Nancy Cantor’s seminar I tried to find out more about it, without much success for many years.

I regularly discussed this issue with the psychologist Ziva Kunda’s, asking her whether research in social psychology showed that there were no character traits as we ordinarily think of them. She always responded that the issue was “complicated.” Later she took it up in her textbook.

there is surprisingly little consistency in people’s friendliness, honesty, or any other personality trait from one situation to other, different situations. ...[W]e often fail to realize this, and tend to assume that behavior is far more consistent and predictable than it really is. As a result, when we observe people’s behavior, we jump to conclusions about their underlying personality far too readily and have much more confidence than we should in our ability to

predict their behavior in other settings (page 395). . . .

Our notion of traits as broad and stable dispositions that manifest themselves to the same extent in a variety of situations cannot hold water. However, this does not mean that there are no enduring and systematic differences among individuals. My intuitions that I am a very different person from my brother or that my children have predictably different patterns of behavior need not be wrong. Such intuitions may be based on meaningful and stable differences among individuals but not the kind of differences implied by the traditional understanding of traits. . . . [For example,] Carol is extremely extroverted in one-on-one situations, is only moderately extroverted when in small groups, and is not at all extroverted in large groups. She will appear very comfortable and outgoing if you meet with her alone, but will clam up and appear very shy and awkward if you encounter her in a large group setting. Linda has a very different profile. She is extremely extroverted in large groups but not at all extroverted in one-on-one situations. She may appear composed and comfortable when lecturing to a large audience but withdrawn and aloof if you approach her alone (pp. 443-4).

In conclusion, it appears that we are truly quite consistent in our behavior within each situation, and it is quite appropriate to expect such consistency in others. But we run into trouble when we expect this consistency to extend to other situations as well. Even slight variations in the features of a situation can lead to dramatic shifts in people’s behavior (p. 499).

At the beginning of the 1990s, two important books were published. One was
Owen Flanagan’s *Varieties of Moral Personality*\textsuperscript{15}, which argues that moral philosophy should pay more attention to psychology and philosophy of mind. Among other things, Flanagan discusses some of the social psychological research that might be interpreted as casting doubt on the existence of character traits in anything like the usual sense. Although he argues that the challenge from this research does not succeed in undermining ordinary views about character, it seemed to me that he had underestimated the challenge.

The other book was Ross and Nisbett’s textbook, *The Person and the Situation: Perspectives of Social Psychology*\textsuperscript{16}, which describes a great deal of the research in social psychology that casts doubt on many ordinary attributions of character.

Three years later Russell Hardin published *One For All: The Logic of Group Conflict*\textsuperscript{17}. He observes that, although many terrible actions by groups, such as those in the former Yugoslavia are often attributed to historical “ethnic hatreds,” it is often possible to explain these events in rational terms. Suppose there are limited resources and a successful coalition will benefit its members more than those excluded from the coalition. Such a coalition is possible only if insiders can be distinguished from excluded outsiders and only if it is possible to keep members from defecting to other groups. Coalitions formed around ethnic or religious lines might succeed. The threat that one such coalition may form can lead other groups to form competing coalitions and to struggle against each other. If stakes are high enough, such struggles can become violent. If we attribute the resulting violence to ethnic hatred, we may very well doubt that there is anything we can do. If we understand the way the violence arises from the situation, we may see more opportunities to end the conflict.

---

\textsuperscript{15.} Cambridge, MA: Harvard University Press, 1991
\textsuperscript{17.} Princeton, NJ: Princeton University Press, 1995
In 1997, I gave a talk at Stanford urging philosophers to pay attention to research in social psychology, arguing that developments in social psychology make it unclear that we have any reason to believe people have character traits in anything like the ordinary sense. Lee Ross, who was in the audience, interrupted occasionally to correct my accounts of some of the research. Afterwards he told me my conclusions were more extreme than any he would draw.

On this occasion I met Maria Merritt who was then writing what proved to be a terrific dissertation at Berkeley, *Virtue Ethics and the Social Psychology of Character*,18. I had been thinking of character traits as ordinarily conceived and as conceived in virtue ethics as having a counterfactual aspect, so that whether one has a certain trait depends not only on how one reacts in actual situations but also in how one would react in certain other possible situations. In this view, a virtuous person would not have tried to give powerful, possibly lethal electric shocks to the “learner” in the Milgram experiment, so the results of that experiment show that hardly anyone is really virtuous. Merritt’s work indicates that there is another, Humean, conception of character traits involving only how someone actually behaves. In this other conception, it may be that the actual subjects of the Milgram experiment became unvirtuous, because of the way they acted in the experiment, without this affecting the character traits of those who did not participate. Similarly, perhaps the citizens of Nazi Germany were horrible people for not objecting to and resisting the Nazi’s treatment of Jews, although citizens in England or Colorado were not horrible people in the same way, even though they would have acted as German citizens did if they had been living in Germany. Merritt uses this idea to argue for a conception of character traits as existing in particular social contexts, without counterfactual implications.

Judith Jarvis Thomson defends a possibly related view of virtue ethics, which

18. University of California, Berkeley, 1999
I have described as “Virtue Ethics without Character Traits.” In this sort of view, terms for virtues are applied in the first instance to acts: a kind act, a courageous act, etc. and there need be no commitment to character traits of any sort.

Maybe that is the way to think of the subjects in the Milgram experiment or the citizens in Nazi Germany who allowed and participated in the persecution of the Jews. They acted badly. Others who did not participate in the Milgram experiment and did not live in Nazi Germany did not act badly. What is needed, in this way of looking at things, is social support for virtuous actions.

This fits in with Sartre’s idea that in one sense you are what you have done and in another sense you are free to act the same or differently in the future, where it is irrelevant and perhaps indeterminate how you would have acted under other circumstances.

Around the time of my 1997 talk at Stanford, I read Peter Railton’s paper, “Made in the Shade: Moral Compatibilism and the Aims of Moral Theory,” which mentions (among other things) that results in social psychology may cast doubt on explanations of behavior that appeal to character traits. Railton refers to John Doris’ impressive dissertation, which became his even more impressive book, Lack of Character.

I gave a talk at Caltech in 1999 on my paper, “Moral philosophy meets social

psychology: virtue ethics and the fundamental attribution error”22, with John Doris as sympathetic but critical commentator. David Baltimore came to the talk and in discussion noted the relevance of our conversations in 1961 about Hannah Arendt’s New Yorker reports on the Eichmann trial and “the banality of evil.”

Later that year I presented the same paper to a meeting of the Aristotelian Society in London. During discussion Rosalind Hursthouse argued vigorously against almost everything I said. Her important book On Virtue Ethics23, came out shortly after. The book elegantly defends three theses, (1) that it is possible to use the idea of a virtuous character trait to explain other moral notions, (2) that moral motivation is best understood in terms of what motivates a virtuous person, and (3) that there is or may be an objective basis for a single set of human virtues of character.

I am not completely convinced by the first thesis—that it is possible to explain other moral notions in terms of virtue. In response to the objection that what is right (in an action-guiding) sense cannot always be identified as the choice that a perfectly virtuous human being would make in the circumstances, because sometimes a completely virtuous human being could never be in the relevant circumstances, Hursthouse argues that virtue ethics is still applicable, because it provides rules that can apply to such a case. However, even if virtue ethics can provide rules, it remains unclear how the rules provided handle this particular situation. Hursthouse says that every virtue of character yields a positive rule of action and every vice or defect of character yields a negative rule; so, virtue ethics allows for such rules as that one ought to tell the truth, one ought to keep

ones promises, one ought to be kind to others and and one should not act meanly, lie, or break promises. Where these simple rules conflict, Hursthouse proposes to “fine tune” them by considering what a virtuous human being would do in various circumstances. But this “fine tuning” cannot obviously yield the right rules for circumstances no virtuous human being could be in.

Hursthouse also notes that someone in a situation no virtuous person might reason using something that sounds like the terminology of virtue and vice. “Perhaps it would be callous to abandon A, but not to abandon B. Perhaps it would be more irresponsible to abandon A than to abandon B. . . . Then marrying A would be the morally right decision.” But in these remarks the vices of callousness and irresponsibility are characteristics of possible actions rather than character traits of the agent, so the fact that someone might use virtue concepts in this way does not show that there is a way to explain right action in terms of virtuous character traits.

Hursthouse defends her third thesis, that there is or may be an objective basis for a single set of human virtues of character, by observing that it is possible that there is a unique set of character traits that are natural to human beings and are such that, if everyone has them, it is generally true that an individual’s having them promises to contribute to that individual’s preservation, the preservation of the human species, the function of social groups to which the individual belongs, and the flourishing of that individual and others. If that possibility were realized, that set of character traits would be the objective set of human virtues in her approach.

One way for this to fail would be that a satisfactory outcome for people would require some human beings to have one set of character traits while others had a different set, as in Nietzsche’s master and slave moralities, and somewhat as
there are worker bees and queen bees. While Hursthouse thinks that this is a view within virtue ethics that needs to be taken seriously, she also thinks that we have not yet been given sufficient reason to give up on the existence of a single set of human virtues. (I strongly disagree with that!)

Another way for her favored approach to fail would be for it to turn out that no distribution of character traits will promote the flourishing of all human beings. Hursthouse argues that we do not have to accept the conclusion that human beings are in this sense just a “mess,” because, “When we look, in detail at why so many human beings are leading, and have led, such dreadful lives, we see that occasionally this is sheer bad luck, but characteristically, it is because either they, and/or their fellow and adjacent human beings, are defective in their possession and exercise of the virtues on the standard list.” She adds in a footnote, “I suppose that one of the reasons we find it so hard to come to terms with the Holocaust is that pre-Nazi German society looks so like our own at the same period, and we are forced to the unpalatable conclusion that if it happened there because of lack of virtue in its members, we must have been similarly lacking and might have gone the same way” (264).

Now, it seems to me that what happened in Nazi Germany and in Bosnia, Somalia, etc. taken together with results in social psychology about the relative explanatory importance of individual character versus the situation in which a human being is placed, indicate that the very natural human tendency to think in terms of character traits leads us in the wrong direction. To the extent that we are interested in improving the lot of mankind it is better to put less emphasis on moral education and on building character and more emphasis on trying to arrange social institutions so that human beings are not placed in situations in which they will act badly.
Hursthouse agrees with the need to set up the right social institutions. So, perhaps the best way to think of her program in this respect is to claim that there are attainable institutions that would, if in place, encourage in participants the development of the relevant character traits, where these traits would tend to sustain and be sustained by the institutions. Presumably, this would require Maria Merritt’s second conception of character traits involving only actual rather than counterfactual regularities in the way people act, feel, etc. And I still think it would be best simply to replace thought and talk about virtuous character traits with thought and talk about virtuous acts and other responses.

My negative thoughts about character traits and virtue ethics derive from considering developments in social psychology indicating that judgments about character result from a fundamental attribution error. Not all psychologists accept these developments. Personality theorists do not, for example. That may be because personality theorists are concerned in the first instance to spell out ordinary conceptions of character and personality, that is, to articulate folk theories of personality without taking seriously the possibility that the folk theories are wrong.

4 Some Critics

Psychologists John Sabini, Michael Siemann, and Julia Stein argue that “The Really Fundamental Attribution Error in Psychological Research” is the error of thinking that there is a fundamental attribution error. More precisely, they argue that it is unclear how exactly to characterize the relevant error involved. But they agree that research in social psychology shows that observers often do wrongly attribute character traits to actors on the basis of inadequate behavioral

evidence. This leaves it unclear whether we have any reason to believe there are character traits.

In a paper in *Ethics*, Sabini and Maury Silver\(^{25}\) interpret me as claiming that “the psychological data show that people do not have characters, in the sense required by virtue ethics.” (Chris Tucker interprets me in a similar way.\(^{26}\)) I would prefer to say that the data show that people often wrongly attribute character traits to actors on the basis of inadequate evidence and that it is unclear there is any more reason for us to believe in character traits on the basis of our personal experience than for psychotherapists to believe their version of psychotherapy is effective on the basis of their personal experience.

In an excellent paper in *Ethics* Rachana Kamtekar\(^{27}\) observes “that the character traits conceived of and debunked by situationist social psychological studies have very little to do with character as it is conceived of in traditional virtue ethics.” But to me the real issue is not about social psychology’s conception of character traits. It is about whether the results cast doubt either on ordinary views about character traits or on the conception of character in one or another version of virtue ethics.

5 Final Unsystematic Remarks

To repeat, I do not think that social psychology demonstrates there are no character traits, either as ordinarily conceived or as required for one or another version

---

of virtue ethics. I do think that social psychology tends to undermine one’s confidence that it is obvious there are such traits.

In thinking about this issue, it is important to distinguish giving an account of the folk conception of character traits from giving an account of what is actually true. It seems to me that defenders of virtue ethics are best understood as trying to give an account of the folk conception.

The word “virtue” and the phrase “virtue ethics” can be interpreted in various ways and it is important to be clear about how it is being used in a particular context. For one thing, virtue or character as a fleeting feature of an act must be distinguished from virtue or character as an enduring characteristic of a person. I am more inclined to believe that there are virtuous and vicious acts than that people have virtuous or vicious characters.

Furthermore, virtues or character traits of a person might be interpreted as having counterfactual implications or they might be interpreted as limited to actual world regularities. There is the issue that Sartre (among others) raises: Given human freedom, can someone really have character traits of the first sort with counterfactual implications? I am inclined to agree with Sartre in answering that question in the negative.