

1st Shearman Lecture: Naturalism in Moral

Philosophy

Gilbert Harman

Princeton University

Tuesday, May 19, 2009

1 Introduction

Naturalism in philosophy is a special case of a more general conception of philosophy. In this conception there is no special philosophical *method* and no special philosophical *subject matter*.

Consider some of the ways in which philosophy interacts with and is continuous with other disciplines.

Aesthetics is obviously pursued in philosophy departments and in departments of literature, music, and art. Monroe Beardsley, who wrote the most important survey of aesthetics in the 20th century, was one of the authors of the important paper, “The Intentional Fallacy,” a statement of a central aspect of the “New Criticism.”

More recently, Richard Wollheim (who may have invented the expression “minimalist art”) and Arthur Danto have had a significant influence on art theory and criticism. They themselves have been important critics.

Alexander Nehamas is another important contemporary figure, who is by the way a member of both the Philosophy Department and the Comparative Literature Department at Princeton.

Anthropology. Anthropologists are often involved with philosophy and philosophers have sometimes acted as anthropologists to study the moralities of one or another culture. Richard Brandt, lived with the Hopi in order to study their ethics. John Ladd lived with the Navaho in order to study their ethics. The anthropologist Dan Sperber is the same person as the philosopher Dan Sperber.

Economics. Recent figures include Robert Nozick, Amartya Sen, maybe John Rawls, David Gauthier, Allan Gibbard, John Broome, Philip Pettit, and many more. *Political theory* is of course a related example with many of the same players.

Linguistics is another very clear case. Philosophers were involved early in the development of generative grammar (e.g. Jerry Katz and Jerry Fodor). Many more wrote about Chomsky's ideas and argued with them (e.g. Paul Ziff, Hilary Putnam). Famously, at the end of the first chapter of *A Theory of Justice*, John Rawls suggested that generative grammar might be a good model for moral theory. Even earlier, Robert Nozick tried to sketch how that might work.

In recent years there has been philosophical interest in and interaction with developments in linguistics. And there has been much interdisciplinary research in semantics involving philosophers and linguists.

Psychology is another clear case. In his *Theory of Justice* Rawls suggested that an adequate moral theory had to be sensitive to developmental

psychology, especially in Piaget. Rawls' early work on justice in turn influenced the psychologist Lawrence Kohlberg's adaptation of Piaget.

Donald Davidson more or less regularly discussed rationality with psychologists like Amos Tversky and Daniel Kahneman, trying to get them to accept that there were limits on how irrational people could be interpreted to be.

J. L. Austin's study of excuses was influential on psychology studies of children's development by John Darley and his colleagues.

In recent years there has been considerable back and forth between psychologists and philosophers on many issues. Relevant philosophers include Daniel Dennett, Stephen Stich, and many younger people working in the general area of (real) moral psychology.

One important issue has concerned whether social psychology undermines ordinary conceptions of character traits and threatens certain forms of virtue ethics. But there are many other issues too.

Computer science. Artificial intelligence, machine learning, and related topics have been considered highly relevant to philosophy of mind. For example, the philosopher John Pollock studies epistemology by designing computer programs to simulate reasoning in accord with one or another set of epistemic principles.

Philosophy of science is another obvious example. Philosophers discussing the interpretation of quantum field theory may publish in physics journals.

I myself went into philosophy because it allowed me to pursue my own interests in issues in linguistics, artificial intelligence, and cognitive science.

My earliest publication was in linguistics.¹ Soon after that Donald Davidson and I organized workshops that brought linguists and philosophers together, including a notorious six week summer school in Irvine, California.

Later the psychologist George Miller and I started the Princeton University Cognitive Science Laboratory and an undergraduate program in Cognitive Studies. More recently, I have co-taught courses with faculty in linguistics, psychology, computer science, and engineering.

I do not mean to suggest that I am in any way special. Most of my colleagues at Princeton take a wide view of philosophy in one or another respect.

1.1 Naturalism

Philosophical naturalism is a special instance of the wider conception of philosophy, taking the *subject matters* and *methods* of philosophy to be continuous with the subject matters and methods of other disciplines, especially including the natural sciences. From a naturalistic perspective, productive philosophers are those who (among other things) produce fruitful more or less speculative theoretical ideas, with no sharp distinction between such theorizing by members of philosophy departments and such theorizing by members of other departments. (In my view, department boundaries are of interest only to administrators.)

Naturalism also often has an ontological or metaphysical aspect in supposing that the world is the natural world, the world that is studied by the

¹“Generative grammars without transformation rules: a defense of phrase structure,” *Language* 39 (1963).

the natural sciences, the world that is available to methodological naturalism. But the main naturalistic theme is methodological.

I am going to discuss certain prospects for naturalism in moral philosophy. I begin with metaphysical issues of the sort just mentioned, having to do with naturalistic reduction in ethics. I will then say something about some recent naturalistic methodological approaches in moral psychology, taking up character traits and virtue ethics today if there is time, discussing a possible analogy between linguistics and moral theory tomorrow, and saying what is wrong with feelings of guilt on Thursday.

2 Naturalistic Reduction

Naturalistic reduction in ethics attempts to locate the place of value in a world of (naturalistically conceived) facts.

In one view, goodness and evil and rightness and wrongness are not features that have a place in the naturalistic world as described by science. Naturalists who take this view either abandon ethics altogether or try to provide a nonfactual account of it.

Alternatively, naturalists might try to identify an act's being morally right or wrong, good or bad, just or unjust, etc., with certain natural properties of the act.

The most straightforward naturalistic reductive strategy appeals to the *supervenience* of the moral on the natural facts. Any change in what the agent ought morally to do requires a change in the (natural) facts of the case. This appears to imply that there is a more or less complex natural relation

between an agent, a possible act, and the agent's situation (conceived as a whole possible world) that holds when and only when the agent in that situation is morally permitted to do that act. The idea then is to identify the property of being what an agent is morally permitted to do in a given situation with the property of being a possible act for which this natural relation holds.

For example, suppose that act utilitarianism provided the correct account of what an agent is morally permitted to do. Given that supposition, the supervenience strategy identifies a possible act's being what an agent is morally permitted to do in a given situation with its being an act that maximizes utility in that situation.

More generally, the strategy identifies a possible act's being what an agent is morally permitted to do in a given situation with the holding of the relevant natural relation, whatever it is, which exists between agent, act, and situation if and only if the agent is morally permitted to do that act in that situation.

It is not a good objection that such an identification fails to capture the *meaning* of "morally permitted." To suppose that water can be identified with H_2O is not to say what the word "water" means as used by ordinary people.

It is true that the moral case raises a methodological issue for naturalism, since different moral theories disagree with each other and so offer incompatible naturalistic reductions. There are various versions of utilitarianism, social contract theory, virtue theory, Kantianism, and many others. Is there a naturalistically acceptable way to resolve disputes between these compet-

ing reductions by testing them against the world as competing scientific theories can be tested?

Instead of trying to answer this question directly, let us consider three kinds of naturalistic reduction, associated with theories of normative functionalism, response dependent theories, and social convention theories.

2.1 Normative Functionalism and Virtue Ethics

One kind of virtue ethics² appeals to a normative functionalism that seeks to derive normative results from assumptions about functions—about designed or natural functions, purposes, roles, etc. For example, the most important function of a clock is to keep time. Whether something is a clock depends on its function, not on what it is made of or what it looks like, as long as it can serve to indicate to an observer what the time is.

Furthermore, a clock can be evaluated in terms of its function. So, a *good* clock is one that keeps time accurately. That's what a clock *ought* to do. If it does not do so, something is *wrong* with it. The features of a good clock that contribute to its accurate functioning are *virtues* of the clock.

Bodily organs are also defined by their proper functions. A heart is an organ whose nature or function is to pump blood steadily. Lungs are organs that function in breathing. Whether something counts as a heart or lung is not a matter of its shape or what it is made of, but whether it has the relevant function. One that actually does so is to that extent a *good* heart or lung. A heart that fulfills its function poorly, by irregular pumping, or by

²Philippa Foot, *Natural Goodness* (Oxford: Clarendon, 2001); Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999).

leaking blood, is a *bad* heart and something is *wrong* with it. The *virtues* of a heart include steady pumping and not leaking.

People who have social roles have associated functions or purposes. A *good* teacher is one who teaches well, who enables students to learn. Something is *wrong* with a teacher whose students do not learn. *Virtues* in a teacher are those characteristics that enable the teacher's students to learn as well also they can. A teacher who cannot get students to learn is *not a good example of a teacher, not a real teacher*.

It is in the *nature* of human beings and certain other animate beings (bees and chimpanzees, for example) that they are social beings. A good human being has various *virtues*, like courage and compassion. A man lacking courage is *not a good example of a man, not a real man*.

Various issues arise for views that attempt to derive moral assessments from functionalism. Do human beings have functions or purposes as part of their nature as human beings? Is the relevant function or purpose to lead a good life, or even the best life? Can this function or purpose be characterized naturalistically? Given competing views of the best life, is there a way of testing these views against the world in the way that scientific hypotheses can be tested?

I am pessimistic about this approach.

2.2 Response Dependent Theories and Social Convention Theories

Another rather different naturalistic approach identifies moral categories in terms of something about human responses to the consideration of pos-

sibilities, in the way in which colors are sometimes identified in terms of something about the responses of normal human perceivers.

In this approach, an act's being wrong might be identified with the dispositions of impartial unbiased sympathetic people to feel moral disapproval of the act on being made vividly aware of the facts of the situation.

David Hume and Adam Smith defend different versions based on different interpretations of sympathy. Hume has a tuning fork account of sympathy: Humean sympathy leads someone to vibrate in tune with others and feel similarly (if less intensely) what others are perceived to be feeling. This yields a utilitarian result. Since people would rather be happy than unhappy, they will favor situations in which there is more net happiness.

Smith objects that Hume's conception of sympathy cannot account for the fact that unhappy people crave sympathy and feel better when they receive it. Humean sympathetic vibrations would make an acquaintance of an unhappy person sympathetically unhappy and then the unhappy person would vibrate with the acquaintance's unhappiness, making the originally unhappy person even more unhappy. Since the sympathy of an acquaintance makes an unhappy person less unhappy, Hume is wrong about what sympathy is.

Smith observes that ordinary sympathy involves *approval*. If someone gets a minor bump and moans and complains, observers who are aware of the minor pains involved will *not* sympathize, because they will not approve of the complainer's reactions. According to Smith, people want sympathy because they want approval. Furthermore, in Smith's view, the relevant sort of approval tends to be an internalized reflection of community standards.

My desire for the approval of others leads me to imagine how they will react to me. I imagine being one of them to consider how I would react, in this way internalizing their standards. This yields a different view of morality from Hume's—one in which what counts as right or wrong is more heavily influenced by the conventional practices of one's society. Smith's theory, while response dependent, sees morality as more of a matter of social convention than Hume's does.

It is true that Hume takes social convention to be important for those aspects of morality having to do with justice: People are disposed to approve of those conventions that promote the general welfare. But for Smith social conventions affect approval and disapproval more directly.³

2.3 Worries about Relativism

I think that the most promising naturalistic reductions have relativistic implications. Adam Smith's is explicitly relativistic, because what captures one's sympathy is directly affected by local customs. The point generalizes to other response dependent theories to the extent that the relevant response, usually approval, is directly influenced by varying customs or personal values. And functionalist theories may have to suppose that moral conclusions are relative to one or another competing conception of the best life, the purpose of life, etc.

Any absolutist (non-relativist) reduction of morality faces the epistemological problem of showing how that conception of morality is better

³I say more about this difference between Hume and Smith in "Moral Agent and Impartial Spectator," in *Explaining Value*, Oxford, Clarendon Press, 2000.

supported than its competitors. The problem is that there are competing moral frameworks and no obvious way to test them against the world. Compare the dispute between Cardinal Belarmine and Galileo about whether the earth is at rest. Their dispute assumed that there is such a thing as being absolutely at rest. The correct resolution of their dispute is that motion is always relative. This conclusion is grounded in the fact that there is no empirical difference between competing views about what is at absolute rest.

Similarly, there appears to be no empirical difference that would resolve fundamental moral disagreements, which suggests that, from a naturalistic perspective, there may be no reason to believe in absolute right and wrong.⁴

Some might respond that to believe in moral relativism would be to accept moral nihilism, at least if one's initial conception of moral values is absolutist and not relativistic. But that would be like saying that to believe in the relativity of motion would be to give up on the idea that things move or are at rest!

Consider atheists who were brought up to believe that, when they said that something was wrong, they intended to be saying that it violated God's commands. They may have at one time firmly believed that, if God had not existed, nothing would have been morally prohibited. But now, having come to doubt that God exists, they continue to accept the same moral principles as before (at least as regards nonreligious matters) and instead have stopped believing that morality is the expression of God's will. In the same way, someone who is initially committed to moral absolutism who

⁴Gilbert Harman, "Moral Relativism," in Gilbert Harman and Judith Jarvis Thomson, *Moral Relativism and Moral Objectivity* (Cambridge, MA: Blackwell, 1996).

later decides that moral relativism is true can still accept (most of) the same moral principles as before.

2.4 Naturalism as a response to evolutionary debunking

Recent attempts to debunk nonutilitarian moral intuitions appeal to possible explanations of the intuitions in terms of evolution by natural selection. The claim is that the relevant moral intuitions result from factors having nothing to do with their truth, namely tendencies to develop whatever intuitions might help to get one's genes into following generations.

While some theorists have argued that such evolutionary explanations debunk intuitions that conflict with utilitarianism, leaving utilitarianism unchallenged as the correct normative view,⁵ others say that, if the explanations debunk nonutilitarian intuitions, they also debunk intuitions that appear in support of utilitarianism.⁶

Naturalism offers a response to these debunking arguments. Consider a corresponding attempt at evolutionary debunking of color perception and how one might respond to that. One response would be to identify colors as response dependent properties, determined by how we perceive them. If that is right, an evolutionary account of how we happen to have the color experience is not a debunking account. Similarly if what is right or wrong is response dependent in one or another of the ways already considered, an evolutionary account of moral intuitions is not a debunking account⁷.

⁵Peter Singer, "Ethics and Intuitions," *The Journal of Ethics* (2005); Joshua Greene, "The Secret Joke of Kant's Soul" in W. Sinnott-Armstrong, ed., *Moral Psychology 3: The Neuroscience of Morality*, Cambridge, MA: MIT Press, 2008

⁶Guy Kahane, "Evolution and Debunking Arguments," *Nous*, forthcoming.

⁷Sharon Street, "A Darwinian Dilemma for Realist Theories of Value," *Philosophical*

This is all I am going to say about a possible naturalistic reduction of morality. I now turn to an example of methodological naturalism.

2.5 Brief Sketch of Recent Skepticism about Character Traits

2.5.1 Philosophy—Sartre

Sartre (1956) describes ways in which one may present oneself to others as being a certain sort of person. In one of his examples, a waiter presents himself as a waiter by as it were acting the part of a waiter. More generally, Sartre argues that, wanting to be, or at least to appear to others to be, a person of a certain sort or character, one often acts the part of a person of that sort of character. Sartre denies that people have fixed characters in the sense that they actually are in themselves any of the sorts of people they present themselves as being. People merely pretend, sometimes even to themselves, to be one sort of person rather than another.

Sartre takes such pretense to oneself to be a paradigm instance of what he calls “bad faith”. (The basic pretense to oneself, according to Sartre, is the pretense that one lacks free will and cannot do otherwise.) One may have actually acted bravely, or cowardly, in a friendly or unfriendly way, etc. in the past and one may have been a brave, cowardly, friendly, unfriendly, or whatever person on various past occasions. But that does not mean that one now *is* such a person. Because of one’s free will, nothing that is now settled can make it the case that one *is* such a person.

Studies 127 (2006).

2.5.2 Sociology—Goffman

Goffman (1959) powerfully develops a related idea in his classic empirical study of “The Presentation of Self in Everyday Life,” leaving it open whether there is any “real” or “true” self beyond various presentations of self.

2.5.3 Social Psychology

Taking a somewhat different approach, social psychologists study how a person’s situation can affect what the person does, in ways that appear to conflict with ordinary thinking about personality and character. There is a vast relevant literature, one aspect of which Kunda (1999) summarizes as follows:

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 (p. 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 (p. 443).

Even slight variations in the features of a situation can lead

to dramatic shifts in people's behavior (p. 499).

2.5.4 Personality Psychology

Personality psychology studies the ways ordinary people think about personality and character traits, which is to be distinguished from studying the truth about personality and character traits.

It is uncontroversial that there is usually a difference between the study of ordinary conceptions of a given phenomenon and the study of the phenomenon itself. We distinguish between folk or common-sense physics, which is studied by certain psychologists, from physics, which is studied by physicists; these are both interesting subjects, but they are different. Similarly, there is a clear difference between the study of conceptions people at a certain time had about witches and witchcraft and the study of what was actually true about people who were taken to be witches and phenomena thought to be witchcraft. We distinguish the study of how people conceive of God from theology. We distinguish between the study of doctors' views about good medical treatment and an investigation into what sorts of treatment are actually effective. We distinguish interviewers' conceptions of the value of interviewing from whether interviews actually improve selection processes. In the same way, there is a clear conceptual difference between what people generally think about character and personality and what is actually the case; the study of what people think about character and personality is part of the study of folk psychology and is not the same as a study of character and personality.

It happens that personality theory or personality psychology is in pretty

bad institutional shape as a scientific discipline. Funder (2001), one of the few remaining defenders of personality psychology, reports that personality psychology has collapsed as a serious academic subject. He regrets

the permanent damage to the infrastructure of personality psychology wreaked by the person-situation debate of the 1970s and 1980s. . . . [O]ne reason for the trend . . . for so much personality research being done by investigators not affiliated with formal programs in personality may be that there are so few formal programs to be affiliated with. The graduate programs in personality psychology that were shrunken beyond recognition or even abolished during the 1970s and 1980s have not been revived (213).

Personality psychology has been concerned with characterizing ordinary folk conceptions of personality. Social psychology is concerned with the accuracy of these conceptions. To the extent that one is interested in the truth and accuracy of claims about character and personality, one needs to consult social psychology, not personality psychology. (There is further discussion of this point in Doris, 2002, pp. 67-75).

2.5.5 Political Theory

Hardin (1995) 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.

2.5.6 More Recent Philosophy

Flanagan (1991) 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, arguing that the challenge from this research does not succeed in undermining ordinary views about character.

Railton (1997) mentions (among other things) how results in social psychology may cast doubt on explanations of behavior that appeal to character traits. Doris (2002) provides an elaborate account of the psychological literature and its potential philosophical morals.

2.5.7 Are There Broad Stable Character Traits?

Are there any character traits of the sort that people ordinarily attribute to others, involving broad and counterfactually stable dispositions of the relevant sorts? While it seems obvious at first that there are, this obviousness may simply be due to our regularly making a “fundamental attribution error.” That is, we explain an action by appeal entirely to features of an agent’s character, overlooking the relevance of subtle aspects of the agent’s perceived situation.

Sabini et al. (2001) argue in response 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 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 of a sort that people ordinarily attribute to others.

By the way, Sabini and Silver (2005) interpret me (in Harman, 1999, and 2000) as claiming that “the psychological data show that people do not have characters, in the sense required by virtue ethics.” (Tucker, 2004, interprets me in a similar way.) I 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 in the effectiveness of their version of

psychotherapy merely on the basis of their personal experience.

Kamtekar (2004) 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 for me the point at 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. If so, as I believe, then I think there are reasons to worry about certain versions of virtue ethics.

2.6 Character Traits and Virtue Ethics

There are many versions of virtue ethics. Some are concerned with character traits. Others only with virtues of actions, with no commitment to virtuous or unvirtuous character traits. Of versions that refer to virtuous character traits there are differences in how such traits are conceived. Some versions take character traits to be broad and counterfactually stable; others may treat them as mere regularities in behavior. Of versions that take character traits to be broad and counterfactually stable, some require that people actually have such traits while other versions treat relevant traits as ideals that may or may not be attained or even attainable by actual people. Empirically based skepticism about character traits is only relevant to some versions of virtue ethics.

2.6.1 Virtue Ethics with Traits as Merely Actual Regularities

Consider those forms of virtue ethics that take character traits to have a counterfactual aspect, so that whether one has a certain trait depends not only on how one reacts in actual situations but also on 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 a “learner” in the Milgram (1974) experiment, so the results of that experiment show that hardly anyone is really virtuous.

Merritt (1999) argues that there is another, Humean, conception of character traits involving how a person 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 any implications for 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, whereas citizens in England or Connecticut were not horrible people in the same way even if 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.

2.6.2 Virtue Ethics without Character Traits

Thomson (1996, 1997, 2001) defends a related view of virtue ethics. Terms for virtues are to be 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. We don't have to say anything about character. What is needed, in this way of looking at things, is social support for virtuous actions.

This fits 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.

2.6.3 Virtue Ethics with Robust Traits

On the other side, Hursthouse (1999) describes a form of virtue ethics with robust character traits. She 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.

However, she appears to conflate virtues of character with virtues of actions when she defends her claim that it is possible to explain other moral notions in terms of virtue. Consider the objection that the right thing to do cannot always be identified as what a perfectly virtuous human being would do in the circumstances, because sometimes a completely virtuous human being could never be in the relevant circumstances (Harman, 1973).

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 could apply to 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 then observes that someone in a situation no virtuous person could be in 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!

Now consider Hursthouses third thesis, that there is or may be an objective basis for a single set of human virtues of character. She observes 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 take a different view in Harman, 1983, 1996.)

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 Merritt’s second conception of character traits involving only actual rather than counterfactual regularities in the way people act, feel, etc.

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, because ordinary thinking about character traits is such a mess. As Hardin points out, it blocks rational thought about international issues. In domestic politics it leads to a concern with the alleged character

of candidates instead of consideration of policies.

2.7 Conclusion about Character Traits

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. But I do think that results in social psychology 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. (Perhaps 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 they are being used in a particular context. Virtue or character as a fleeting feature of an act must be distinguished from virtue or character as an enduring characteristic of a person. There is more reason to believe that there are virtuous and vicious acts than to believe 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. It is more reasonable to believe in character traits of the latter sort than to believe in character traits of the former sort.

Finally, there is the issue that Sartre (among others) raises: Given human freedom, can someone really have character traits of the first sort with their counterfactual implications? I am inclined to agree with Sartre in answering this question in the negative, but I have not tried to discuss the point here.

3 Final remarks

To sum up, I have tried to say something about contemporary naturalistic approaches to morality. I started by describing more traditional issues of naturalistic reduction. I then said something about character traits and virtue ethics from a naturalistic methodological point of view. Next time I will discuss the possibility of a linguistic analogy for moral theory. The following day I will argue for the unimportance to moral theory of guilt feelings, as these feelings are ordinarily conceived.

4 Bibliography

Doris, J. (2002). *Lack of Character*. New York: Cambridge University Press.

Flanagan, O. (1991). *Varieties of Moral Personality*. Cambridge, MA: Harvard University Press.

Funder, D. C., (2001). Personality. *Annual Review of Psychology*, 52: 197-221.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday Anchor.

Harman, G., (1983). "Human Flourishing, Ethics, and Liberty," *Philosophy and Public Affairs* 12 pp. 307-322.

Harman, G., (1996). "Moral Relativism," in Harman and Thomson (1996), pp. 1-64.

Harman, G., (1999). Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error." *Proceedings of the Aristotelian Society*, 99: 315-31.

Harman, G., (2000). The nonexistence of character traits.” *Proceedings of the Aristotelian Society*, 100: 223-6.

Harman, G. and Thomson, J. J. (1996). *Moral Relativism and Moral Objectivity* (Cambridge, MA: Blackwell, 1996).

Hardin, R. (1995). *One For All: The Logic of Group Conflict*. Princeton, NJ: Princeton University Press.

Hursthouse, R. (1999). *On Virtue Ethics*. Oxford: Oxford University Press.

Kamtekar, R. (2004). “Situationism and Virtue Ethics on the Content of Our Character,” *Ethics* 114, 458-491.

Kunda, Z. (1999). *Social Cognition: Making Sense of People*. Cambridge, MA: MIT Press.

Merritt, M. (1999). *Virtue Ethics and the Social Psychology of Character*. Ph.D. Dissertation, University of California, Berkeley, 1999.

Milgram, S., (1974). *Obedience to Authority*. New York: Harper and Row.

Railton, P. (1997). “Made in the Shade: Moral Compatibilism and the Aims of Moral Theory,” *On the Relevance of Metaethics: New Essays in Metaethics*, ed. J. Couture and K. Nielsen, *Canadian Journal of Philosophy Supplementary Volume* 21, pp. 79-106.

Sabini, J., Siemann, M., and Stein, J. (2001). “The Really Fundamental Attribution Error in Psychological Research,” *Psychological Inquiry* 12, pp. 1-15.

Sabini, J. and Silver, M. (2005). “Lack of Character? Situationism Critiqued,” *Ethics* 115, pp. 535-562.

Sartre, J. P. (1956) *Being and Nothingness*, translated by Hazel E. Barnes. New York: Philosophical Library.

Thomson, J. (1996). "Evaluatives and directives," in Harman and Thomson (1996), pp. 125-154.

Thomson, J. (1997). "The right and the good," *Journal of Philosophy* 94 (1997), 273-298.

Thomson, J. (2001). *Goodness and Advice*. Princeton, NJ: Princeton University Press.

Tucker, C., (2004). "Harman vs. Virtue Theory: Do Character Traits Explain Behavior?" *Southwest Philosophy Review*, 21, pp. 137-146.