

"Unofficial"

Virtue, Virtue Skepticism, And the Milgram Studies

Deborah C. Zeller

Cleveland State University

~~enthymeme09@yahoo.com~~

enthymeme099@gmail.com

ABSTRACT: Virtue, the centerpiece of ancient ethics, has come under attack by virtue skeptics impressed by results of psychology experiments including Milgram's obedience studies. The virtue skeptic argues that experimental findings suggest that character structures are so fragile vis-à-vis situational factors as to be explanatorily superfluous: virtues and robust character traits are a myth, and should be replaced by situation-specific "narrow dispositions" (Gilbert Harman) or "local traits" (John Doris). This paper argues that the virtue skeptics' sweeping claims are ill-founded. First, blending Aristotelian and contemporary insights about virtue, I reach a decision about a reasonable, non-straw definition of "virtue" and of "character trait." Next, I argue that explanations give by Lee Ross and Richard Nisbett for the Milgram findings covertly invoke character traits. Reflection reveals that more robust, cross-situationally consistent traits are needed for explanation of subject behavior, and that it is reasonable to suppose that such traits were in place.

THE CONCEPT OF A VIRTUE IS AT A CROSSROADS: on the one hand, there has been an revival of Aristotelian virtue ethics in latter 20th century philosophy, and on the other hand a critique issuing from experimental psychology that calls into the question the very existence—or at least the human possibility of instantiation—of the virtues. In this paper I reinterpret the Milgram data via an analysis of the explanations for the Milgram data offered by Lee Ross and Richard Nisbett. The Milgram studies are often taken, as by Gilbert Harman and John Doris, to count against the existence of virtue. I maintain that they instead actually support the existence of certain stable and cross-situationally consistent character traits, including some that count as virtues. Virtues and character traits, I conclude,

are still viable concepts; indeed, human behavior cannot plausibly be explained without them.

Gilbert Harman was the first to draw philosophical attention to the possibility that virtues don't exist, beginning in 1999. In his boldest statement on the subject, Gilbert Harman asserts:

I myself think it is better to abandon all thought and talk of character and virtue. I believe that ordinary thinking in terms of character traits has disastrous effects on people's understanding of each other, on their understandings of what social programs are reasonable to support, and on their understandings of international affairs. I think we need to get people to stop doing this. We need to convince people to look at situational factors and to stop trying to explain things in terms of character traits. We need to abandon talk of virtue and character, not find a way to save it by reinterpreting it (Harman 2000, 224).

One set of experiments cited by both Gilbert Harman and John Doris in defense of their claim that virtues do not exist is the Stanley Milgram experiments of the 1960's. In these experiments, subjects were administered fake shocks that they had reason to suppose real to a confederate learner whenever he gave a wrong answer. The shocks began at 15 volts and increased at 15-volt intervals up to 450 volts. Only a little more than a third of the subjects made the four requests to stop shocking the victim necessary to terminate the experiment. Fully sixty-five percent of the experimental subjects went all the way up the scale of shocks to administer the most severe shock, despite the learner's screams, pounding on the wall, and final silence (Milgram, 13ff; Ross and Nisbett, 53-58).

Gilbert Harman and John Doris each place considerable weight on these experiments, maintaining they show that situational factors swamp the influence of any putative character traits, and to such a degree that, as explanatory constructs, character traits contribute so little as to be superfluous. Moreover, they both, and especially Harman as in the passage quoted, hold that belief in character traits is pernicious in that it leads to typing of people as bad and good, when in reality people act according to the situation.¹

The first task in exploring the skeptic's claim, it would seem, is to get clear about what sort of thing it is whose nonexistence the skeptic is alleging. According to Aristotle, the virtues are qualities that are constitutive of living well, for living well, or happiness, consists in activity in accordance with virtue. Virtues are, by definition, ways of performing one's function well, as a knife has virtue when it cuts well or the eye when it sees well. Because the specifically human function is reasoning, human beings are virtuous when they exercise their rationality well in action and contemplation (Aristotle, 1098a3-15).²

¹ On the perniciousness of belief in virtue, see Doris, 167-169.

² Many but not all contemporary philosophers follow Aristotle's lead in making virtue ethics teleological, that is, in taking virtues to conduce to happiness or flourishing, e.g., Hursthouse, 207-8; 248, MacIntyre and Wallace also hold that the virtues are teleologically related to the good for human beings; the good is conceived by MacIntyre as "achievement of the goods internal to practices" (*After Virtue*, 191) and by Wallace as "a social life informed by convention" (*Virtues and Vices*, 37). However, some virtue ethicists prefer an approach that makes virtues even more basic in ethics. Linda Zagzebski, for example,

For Aristotle, then, virtues have a relation to happiness. The question next arises what virtues are in themselves. Aristotle holds that virtue is a settled state, concerned with feelings, choice, and action, consisting in a mean between extremes. He names eleven virtues: courage, temperance, generosity, magnificence, a nameless virtue concerned with small honors, magnanimity (concerned with great honors), mildness, honesty, wittiness, friendliness, proper pride, and justice (Aristotle, 1107b1-1108b10).

Aristotle discusses some vices, but in general he is not concerned with any non-virtue character traits. Since it is character traits whose existence Harman and Doris deny, the notion of a character trait must be explicated. It appears that character traits are like virtues in being complex qualities of persons, settled—longstanding or enduring—states, but that they need not conduce to happiness or flourishing. Talkativeness, aggressiveness or passivity, and curiosity or cheerfulness are examples of character traits, and the latter might be virtues as well.

Turning to the definitions of virtue given by virtue skeptics, we note that there is an emphasis on the behavioral manifestation of a virtue—a concern undoubtedly derived from their attention to empirical psychology. Harman holds that two necessary features of character traits, if they were to exist, would be stability and cross-situational consistency, a combination that he as well as Doris term “robustness”: character traits “are supposed robust in the sense that they are relatively long lasting and are or would be exhibited in a variety of circumstances” (Harman 1999b, 3; Doris, 63). Both Harman and Doris acknowledge the existence of stable traits, that is, certain cross-temporal regularities in behavior, but they insist that these are too narrow, that is, too situation-specific to amount to character traits: Harman calls these narrow traits “narrow dispositions,” and Doris calls them “local traits.” Harman cites “roller coaster cowardice” as an example of a narrow disposition, and Doris cites “office party sociability,” and “wild animals physical courage” as examples of local traits (Harman 1999a, 318; Doris, 62, 66). Each denies the existence of robust traits (Doris, 63; Harman 1999b, 4).

Given that they acknowledge the *stability* of certain narrow traits, that is, stability over time, the linchpin of the virtue skeptic’s charge is cross-situational consistency. As Harman puts it, “character traits are *broad based* dispositions that help to explain what they are dispositions to do” (Harman 1999a, 318). Harman and Doris bring to light evidence that people are, in fact, not cross-situationally consistent in their behavior: they do not manifest characteristics across a wide range of circumstances. People tend to be honest in one situation, dishonest in another, and so on for all the other “virtues” (Doris, 64). Since cross-situational consistency is a requirement of robustness, there are no robust traits.

Now, cross-situational consistency is lightly touched on by Aristotle, who notes that courage may be exhibited in battle, poverty, and sickness (Aristotle, 1115b1-2). Aristotle also emphasizes that the person who suffers severe misfortune will

prefers what she calls a motivation-based virtue ethics approach, such that the virtuous agent and the virtuous agent’s virtuous motivations are ethically central, with notions of right action derived from them as well as the concept of happiness or the good (*Virtues of the Mind*, 80-83). However, in this paper I will assume an Aristotelian, or teleological, view of virtue, such that virtues are those states which constitute or promote human flourishing.

maintain a certain dignity despite his marred blessedness (Aristotle, 1100b25-33). Aristotle, then, did not think behavior was tied to situational factors, and thought that it would be exhibited in a range of situations. For Aristotle, virtues will indeed be robust traits in Harman's sense.

Cross-situational consistency does not seem to be an all-or-nothing thing; rather, it comes in degrees. It seems, therefore, that a decision about the degree of cross-situational consistency required for a virtue will be crucial for resolving the debate between skeptics and believers in virtue. Doris targets something called "globalism," which he attributes to Aristotle. Globalism has three tenets: 1) traits will be stable 2) they will be cross-situationally manifested, and 3) a person's character will be "evaluatively integrated," that is, the virtues will hang together, and the vices will as well (Doris 22-23). However, it is not evident that Aristotle or indeed any virtue ethicist requires the "global" degree of cross-situational consistency Doris attributes to the believer in virtue. It seems, rather, that believers in virtue can allow that a virtue can be present yet not be manifested in every conceivable manner on every conceivable occasion. For example, a person may manifest generosity in certain favored ways and not others, cultivate some manifestations of friendliness but not others, and even be courageous in nonparadigmatic ways (e.g., through retreating). The kind of truly globally cross-situationally consistent honesty associated with Kant's *On a Presumed Right to Lie* is taken by most as pathological, not even virtuous. And so the believer in virtue does not, and ought not, require global cross-situational consistency of a virtue, and the skeptic should not require it either.

And so, putting all of these insights together, we get the following definitions:

A virtue is a complex quality Q of a person, comprising patterns of thought and desire, which issues (in suitable virtue-apt circumstances C) in a tendency to actions of a certain character V, with considerable transtemporal stability and a modest (though not necessarily extensive or unlimited) cross-situational consistency; moreover, this quality conduces to human happiness or flourishing.

A character trait is a complex quality Q of a person, comprising patterns of thought and desire, which issues (in suitable circumstances C) in a tendency to actions of a certain character T, with considerable transtemporal stability and a modest (though not necessarily extensive or unlimited) cross-situational consistency. This quality may or may not promote human happiness or flourishing.³

Returning now to the Milgram experiments: I now argue that, in order to explain the failure of certain virtues without resorting to the unlikely explanation of cruelty, psychologists in their explanations of subject behavior have invoked *other* character traits including those of reliability, trust, cooperativeness, fairness, and even minimal compassion. I further argue that these features are character traits, at least, and probably even happiness-conducive traits or virtues on our definition. The invocation of character traits is implicit; the explanations are couched in terms of subject "needs," and certain laws of behavior, rather than in terms of outright traits. Nevertheless, when one carefully considers the explanations given for

³ For the general style of these definitions as well as the idea that virtues either are called for or function better in some circumstances rather than others I am indebted to Ernest Sosa's definitions of intellectual virtue in *Knowledge in Perspective*, pp. 138 and 140.

subject behavior, traits emerge: the needs and laws translate into stable, cross-situationally consistent individual character traits and virtues. Moreover, as I will show, it is impossible to take a minimalist or deflationary or "narrow disposition" or "local trait" view of the traits involved in subject behavior and preserve explanatory efficacy. Without a supposition of full-blooded, robust traits, the explanations cease to be explanatory.

My interest in arguing this way is the following: the claim that virtues don't exist is a very radical and sweeping claim. The basis for the virtue skeptics' claim is a rather small number of experiments that show the power of the situation and a corresponding failure of personality, including the Milgram experiments, and a few others that purportedly show a failure of cross-situational consistency. If, in fact, these same experiments yield unexpected evidence *for* the existence of character traits, then skeptics will have to look elsewhere for evidence that virtues do not exist.

I now turn to the psychologists' explanations. Lee Ross and Richard Nisbett consider the Milgram experiments in *The Person and the Situation*. They approvingly cite Milgram's own explanation for subject obedience, "the subject's implicit contract to do as one is told without asking why, faithfully serving the authority figure to whom one has willingly ceded responsibility" (Ross and Nisbett, 56). Milgram writes at length about a process called "the agentic shift," whereby an agent cedes responsibility for his own action to a higher-up. Milgram notes that social organizations can only be effective if individual volition is to some degree suppressed and hierarchy accepted. The individual cedes authority as the price to be paid for social efficiency: the social contract. On his view, every social interaction has a contractual nature; to defy the authority of the experimenter would be to break a contract (Milgram, 123ff).

The question arises where such a need to keep contracts or commitments comes from. It is implausible to suppose that it arose *ex nihilo* within the experimental situation. Presumably, this need to keep commitments was quite general: for example, most of the obedient subjects (and probably most of the disobedients as well) probably also showed up at work at the expected times and stayed for the entire shift, came home from work at the expected times, attended sporting events as expected, etc. It is difficult to believe that the need to keep commitments was an isolated need that arose, *ex nihilo*, as a situation-specific goal in response to the Milgram experiment. Therefore, the experimentally manifested tendency of the majority of subjects to keep their commitment to participate in a psychological experiment, what one might call the general transpersonal tendency, must be undergirded by a robust intrapersonal tendency on the part of individual subjects. The majority of subjects would have been both cross-temporally and cross-situationally consistent in their commitment-abiding, or in other words, with reference to our definition, would have possessed the character trait of reliability. It would appear, in other words, that the explanation offered by Milgram and Ross and Nisbett is not compatible with the character trait-eliminativist position of Harman and Doris.

Consider another of Ross and Nisbett's explanations for subject obedience, namely, the problem of finding a suitable justification for stopping the experiment

at any particular time, given the stepwise progression. Subjects did not stop the experiment early because there seemed to be no good reason to stop at any given point in the middle after having gone as far as they had already:

It was only as the stepwise progression continued, and the shocks being administered reached alarming levels, that the teacher's [subject's] psychological dilemma became apparent. In a sense, the teacher had to find a rationale (one satisfactory to himself, to the experimenter, and perhaps even to the learner) that would justify his decision to desist now when he hadn't desisted earlier, a way to explain why it was illegitimate to deliver the next shock when it presumably had been legitimate to deliver one of only slightly lesser magnitude just moments before. Such a rationale is difficult to find (Ross and Nisbett, 56).

But the question arises why just any justification would not have done, *even given* the stepwise nature of the progression. In fact, a subject could have simply cited the stepwise progression as justification. For example, a subject might have said, "He's screaming more now, saying more—I want to quit." For some reason, most people could not advance this justification, and continued in the experiment. However, some subjects did, at least implicitly, make this very justification and quit during the middle of the experiment (Milgram, 35).

Ross and Nisbett's view that the stepwise progression of the shocks somehow explains the difficulty of quitting early implies that the majority of subjects were concerned to appear rational rather than emotional, and wanted to avoid attaching too much significance to subtle changes that they were told were not important. Perhaps emotionality was considered a luxury in the setting of an experiment that might have lasting educational psychology benefit. These subject concerns seem related to a general desire to be cooperative. Having cooperated up to a certain point, the subjects wanted to continue to cooperate, even at some cost to themselves and others. Cooperativeness seems related to reliability, but has the added feature of a willingness to sacrifice for the greater good. The great anxiety felt by many of the Milgram subjects suggests that they were weighing multiple incompatible values, for example, cooperation versus humanity toward a victim.

A situationist might counter that one could explain the cooperativeness in terms of a narrow disposition. However, again, it seems unlikely that such the subjects' cooperativeness was so situation-specific, such that a narrow disposition, e.g., "shocking-experiment-cooperativeness" could be present and operative in the novel experimental situation. One would expect, rather, that the subjects exhibited cooperativeness across a range of situations in their daily lives. It is reasonable to suppose, for example, that those subjects who were employees were willing to follow their employers' instructions on the job in order to achieve greater productivity or customer satisfaction, at some cost to themselves and their fellow-employees.

Ross and Nisbett cite as further explanation of the subjects' obedience the fact that the progression began from a fairly innocuous action, a small shock (Ross and Nisbett, 56). Now, this explanation implies that had the experiment started with a non-innocuous action—say, a severe shock which produced loud screams—more

subjects would have stopped the experiment early. For the explanation to count as explanatory, one must accept a suppressed premise that most people do not greet sudden severe pain with equanimity. This explanation implies the existence of yet another character trait among the majority: a certain minimal compassion that revolts at an abrupt, severely harmful action. Presumably, the revulsion to abruptness stems, at least in part, from a sense of justice: if the learner had received a severe shock after only one wrong answer, violating any correlation between success and freedom from punishment, then the subject's sense of justice would have been outraged. A certain minimal compassion and sense of justice, both character traits, are implied by Ross and Nisbett's explanation.

An explanation in terms of a narrow disposition, such as "innocuous beginning tolerance" is implausible, because it is itself in need of explanation. For why would subjects have innocuous beginning tolerance as opposed to innocuous beginning *intolerance*? Ross and Nisbett's explanation that an innocuous beginning passed below many subjects' radar only makes sense on the supposition that there was some sort of radar in place. This radar, I submit, were the character traits of minimal compassion and minimal justice. It is very probable that at least many subjects exhibited these qualities cross-situationally, refraining from harming their fellows abruptly or severely in a wide range of situations, and refraining from punishing others severely for minor infractions. A certain stable and wide-ranging minimal compassion and sense of justice, amounting to character traits, are implied by Ross and Nisbett's explanation.

Finally, Ross and Nisbett cite the difficulty of ending the experiment as another reason for the high rate of obedience. The subject had to make four requests for release in order to actually be released from the experiment (Milgram, 21). Ross and Nisbett hypothesized that had there been a "panic" button allowing immediate release from the experiment, virtually none of the subjects would have been fully obedient (Ross and Nisbett, 57).

The question arises why it was difficult for the subjects to make those four requests. Suppose a subject's compassion were aroused by the recorded screams of the confederate victim. Why didn't this compassionate subject simply calmly ask to be released four times, and ignore the experimenter's encouragements to continue? Why would the second, third, and fourth requests be any more difficult than the first, such that the number of requests required is *explanatory* of the fact that so many subjects obeyed?

Presumably, trust entered the picture. Subjects trusted the experimenters' reassurances, some of which were explicit and some of which were implicit in the experimenters' requests that the subject continue the experiment. It is impossible to negotiate everyday life well without a certain modicum of trust. The person who deposits money in a bank trusts that that money will be taken care of. It would be pathological to constantly question the bank tellers as to where they put the money, etc.; or to wonder whether the receipt they give is genuine, to demand to see the computer screen of one's account to make sure it matches the receipt, and so on. We must trust authority in order to function well in the world. Trust is an important character trait, a virtue, and a basic one, like reliability and cooperativeness. However, like reliability and cooperativeness, trust is a character trait that can and

should be overridden in certain circumstances. For the reason noted above, namely, the implausibility of an situation-specific narrow disposition being present in the novel experimental setting, it is likely that the subjects who manifested trust did so both stably and cross-situationally, making trust a character trait and virtue, rather than a narrow disposition.

In summary, Ross and Nisbett's explanations of the obedient's behavior raise the question why they are *explanations* of the obedient's behavior rather than simply re-descriptions of what in fact happened. In explanation, an individual instance is subsumed under a larger rubric, that is, a more robust tendency. For each of Ross and Nisbett's explanations, what is genuinely explanatory, that is, the general rubric, is not merely the transpersonal lawlike tendency of behavior, which itself demands explanation, but rather a stable and cross-situationally consistent intrapersonal entity, namely, a character trait. Keeping a commitment to participate in an experiment would not have been particularly pressing unless subjects had a general tendency to keep their commitments, i.e., were generally reliable. The problem of justifying one's quitting early would not arise unless people were generally cooperative. The innocuous beginning would not be explanatory of obedience unless people had minimal compassion and a sense of justice. The four requests would have been no bar to early quitting unless people were generally trusting.

Ross and Nisbett's explanations, in my view, make reference to five virtues or character traits which are very basic and may be considered fairly widespread: reliability, cooperativeness, basic compassion and basic justice, and trust. One need not, therefore, invoke a character defect such as cruelty to explain the Milgram results. One can even invoke what are normally good character traits, and simply note that in the particular context, people showed a lack of judgment, failing to note that compassion should trump the other traits.⁴ This result is in consonance with the view that evil is banal—that it occurs among good people. It is impossible to explain subject behavior via narrow dispositions, and, even if one grants that narrow dispositions are sometimes explanatory, it is incredible that people would have had already-evolved narrow, experiment-specific dispositions ready, like cards to be played, in the novel experimental context. Rather, subject reactions in the disorienting experimental situation reflected their most general patterns of social interaction; and it is unsurprising that reliability, cooperativeness, trust, and minimal fairness and compassion were evinced, for these are the qualities needed to live in a peaceful society, as Plato, Hobbes and Hume have pointed out in their stories of the evolution of justice or government. These qualities are therefore widespread in the population. And as I have shown, on an individual level it is reasonable to suppose that these qualities were both stable and—more importantly in view of the skeptics' claims-- demonstrated with sufficient cross-situational consistency in the subjects' lives to meet the definitional requirements for character traits and virtues.

⁴ I thank Alan Kim for this insight.

Therefore, one should not conclude from these experiments, as the virtue skeptic does, that character traits and virtues do not exist. If the psychologists' explanations for the results are on-target, and I believe they are, then character traits do exist. As I have shown, one must invoke some robust character traits to explain why people failed to exhibit other character traits such as empathy during the Milgram experiments. I conclude that the virtue skeptic has overstated his case.

Finally, I would like to say just a few words about the charge that virtue-talk is pernicious, and misleads us about our fellow human-beings. From a metaphysical point of view, it is evident to me that human beings, like other existents, have qualities; from a human point of view, it is apparent to me that people differ in their character traits. So virtue- and trait-talk has, in my view, truth on its side. I acknowledge that situationism can provide insight into our moral fragility and malleability, and perhaps make us more tolerant and forgiving: surely good moral goals. However, on the other side, I think that virtue-skeptical thinking risks blinding us to what virtue there is. Aristotle thought that the *phronimos* was the centerpiece of morality; Owen Flanagan rightly thinks that there are many ways of living a moral exemplary life, many sorts of *phronimoi* (Flanagan, 1-10). To be overly skeptical about their existence might be to risk losing a chance to learn from them.

Acknowledgements

I would like to thank The University of Memphis Philosophy Department, and in particular Timothy Roche, Nancy Simco, John Tienson, Deborah Tollefsen, as well as David Henderson and Alan Kim, for valuable discussions of drafts of this paper. I would also like to thank Jim Montmarquet for discussions on this topic.

Works Cited

- Aristotle. 1999. *Nicomachean Ethics*. Terence Irwin, tr., Second Edition. Indianapolis: Hackett.
- Doris, John. 2002. *Lack of Character*. Cambridge: Cambridge University Press.
- Flanagan, Owen. 1991. *Varieties of Moral Personality*. Cambridge: Harvard University Press.
- Harman, Gilbert. 1999 a. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society*, 99.
- Harman, Gilbert. 1999 b. "Virtue Ethics without Character Traits," Harman's website, 10/10/2003 (<http://www.princeton.edu/~harman/Papers/Thomson.html>)
- Gilbert Harman, 2000. "The Nonexistence of Character Traits," *Proceedings of the Aristotelian Society* 100. 224.
- Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.
- MacIntyre, Alasdair. 1984. *After Virtue*. Notre Dame: University of Notre Dame Press.

- Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. New York: Harper & Row.
- Ross, Lee and Nisbett, Richard E. 1991. *The Person and the Situation*. New York: McGraw Hill.
- Sosa, Ernest. 1991. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press.
- Wallace, James. 1978. *Virtues and Vices*. Ithaca: Cornell University Press.
- Zagzebski, Linda Trinkaus. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.