

# CHARACTER, CONSISTENCY, AND CLASSIFICATION

Jonathan Webber

University of Sheffield

*Mind* 115, no. 459 (July 2006)

– this is a draft – please cite publication –

**Abstract:** John Doris has recently argued that since we do not possess character traits as traditionally conceived, virtue ethics is rooted in a false empirical presupposition. Gopal Sreenivasan has claimed, in a paper in *Mind*, that Doris has not provided suitable evidence for his empirical claim. But the experiment Sreenivasan focuses on is not one that Doris employs, and neither is it relevantly similar in structure. The confusion arises because both authors use the phrase ‘cross-situational consistency’ to describe the aspect of character traits that they are concerned with, but neither defines this phrase and it is ambiguous: Doris uses it in one sense, Sreenivasan in another. Partly for this reason, the objections Sreenivasan raises fail to block the argument Doris provides. In particular, the most reliable data Doris employs, Milgram’s famous study of authority, is entirely immune to Sreenivasan’s objections. Sreenivasan has not shown, therefore, that Doris provides unsuitable evidence for his claim.

Many philosophers are currently concerned with the impact that data from certain experiments in the situationist tradition of social psychology should have on philosophical discussions of character and virtue. A flurry of recent publications respond to the claim made by John Doris, first in an article and then in more detail in a book, that virtue theory in philosophical ethics is rooted in a false empirical presupposition: virtues are generally defined as a subset of character traits as traditionally understood, but empirical evidence suggests that we do not have such character traits, so there cannot be any virtues, he argues (1998, pp. 505–7; 2002, chs. 2–3). The central theme of Gopal Sreenivasan’s recent paper in *Mind* is that Doris has not cited suitable empirical evidence for his view.<sup>1</sup> As we will see, however, the objections he raises simply do not apply to the data on which Doris rests his argument.

Sreenivasan focuses his discussion on a study of around eight thousand subjects between the ages of eight and sixteen observed in situations where they had the opportunity to cheat in tests, lie about whether they had cheated in those tests, fake a record of their athletic performance, steal a small amount of unattended money, and so on. The study found a high correlation between the proportion of a given group of

---

<sup>1</sup> Although this is the central theme, Sreenivasan’s paper also briefly claims that since virtue theorists need not hold virtues to be widespread, data purporting to represent the population as a whole is beside the point (2002, pp. 56–7). This seems to miss the point Doris makes, which is that the data shows that people do not really have *character traits* — virtuous or otherwise — as traditionally construed (see Webber 2006, Sect. 4). Sreenivasan also briefly considers the epistemological view that everyday trait-attribution is systematically erroneous, but puts this to one side as only indirectly relevant to the main issue (2002, pp. 51–5).

children behaving dishonestly in one situation and that doing so in another situation of the same type, but low correlations between the proportion of a given group of children behaving dishonestly in one type of situation and that doing so in any other type of situation. Those cheating in one test were likely to cheat in another, it seems, but not significantly more likely to pocket unattended change than those who did not cheat in the test.

This is not, however, a study on which Doris places much emphasis. One good reason for this is that, as Doris points out in his book, ‘children with developing personalities are plausibly thought to exhibit less behavioural consistency than fully formed adults’ (2002, p. 63). Showing that children do not have traits as traditionally construed would not show that the same is true of adults. Another good reason is that the degrees to which one is averse to lying, to cheating, and to stealing could be three distinct traits as these are traditionally understood. That the term ‘honesty’ is used of all three might simply reflect the connections between them: successful cheating requires deliberately giving the impression that one is not cheating; stealing seems to involve passing off someone else’s money or goods as one’s own and the intention to deny that this is so. Perhaps the full virtue of honesty involves understanding these connections and is thereby an integrated aversion to all three actions. But since one might nevertheless be averse to lying but not to cheating or stealing, for example, if one did not see these connections, low correlations between each type of aversion would provide no evidence against the existence of these distinct traits.

But there is a deeper reason why Doris does not use these studies. His argument rests on the claim that behaviour is strongly influenced by variations in situation that we might not even notice and that neither common-sense nor philosophical tradition considers motivationally significant. In the paper to which Sreenivasan responds, he emphasises an experiment in which fourteen of sixteen people who found a dime in a phone booth subsequently helped a woman to collect the papers she had dropped, whereas only one of the twenty-five people who had not found a dime did so (1998, p. 504; see also 1998, p. 507 and 2002, p. 30). He also refers to two further experiments. One is the famous Obedience Experiment in which, at the behest of a polite experimenter, almost all volunteers delivered apparently dangerous electric shocks to someone seeming to be another volunteer, but far fewer did so at the behest of another volunteer, and it seems that very few would do so without any request at all. The other is the Samaritan Experiment in which the likelihood of seminarians at Princeton stopping to help a distressed stranger was shown to vary significantly with their degree of hurry (1998, p. 510; see also 2002, pp. 33–4, 39–51).

It is unfortunate that Doris showcases the dime experiment, since attempts to replicate the results have yielded wildly differing data (see Miller 2003, Appendix). Similar experiments do seem to show more reliably that minor situational variations have an impact on behaviour, but these experiments have all involved relatively trivial scenarios such as tipping waiters or helping with surveys, and there is no evidence that minor variations affect the likelihood of helping in more serious scenarios, such as when someone appears to be in significant distress (see Sabini and Silver 2005, pp. 539–40).

We can therefore take these experiments to show only that we should not judge character on the basis of such minor actions.

But it is clear nonetheless that the experiment that Sreenivasan focuses on differs fundamentally from those Doris employs in his argument. Both authors use the phrase ‘cross-situational consistency’ to describe a feature of traits as traditionally construed, and both rightly claim that the experiments they cite investigate this feature. Neither author explicitly defines this feature, however, and the phrase is ambiguous. In one sense, a trait is cross-situationally consistent if it is manifested across situations that differ in respect of the kind of feature inviting behaviour that manifests that trait, so that behaving honestly in a situation in which one could cheat and in a situation in which one could pocket unattended change would manifest this kind of consistency. Sreenivasan’s discussion focuses on this kind of consistency.

Doris, on the other hand, is not concerned with this. He is arguing against the view that behaviour is regulated by character traits, where each character trait is a relatively stable disposition to respond to a certain situational feature in a certain way. Compassion, on this view, is the trait of generally trying to avert or alleviate other people’s distress when in a situation in which one can do so. Cowardice is the trait of shrinking from apparent danger or difficulty. Doris takes this view to be part of the conception of character prevalent in common-sense and in philosophical ethics, and wants to show that it is mistaken, and hence that the ethical aim of developing the right kind of character is misplaced. In order to show this, he argues that whether you behave compassionately is not a simple matter of there being someone in distress and your

having the trait of compassion, but varies with such situational features as the presence or absence of authority figures and whether or not you are in a hurry.

What Doris is trying to show, therefore, is that we do not possess traits that are cross-situationally consistent in a second sense. He is not concerned with consistency of behaviour across situations that differ in respect of the situational feature inviting the trait-relevant behaviour, but consistency across situations in which this feature remains constant but other features vary. He is not concerned with compassionate responses to different kinds of distress exhibited by people with different relationships to the agent, for example, but with compassionate responses to the same kind of distress of a stranger when in a hurry and when not in a hurry, or when there is an authority figure issuing instructions and when there is not. His argument is that the traditional notion of character, the notion employed in philosophical virtue theory, should lead us to expect people to behave consistently in this sense, and that experimental data shows that this expectation is not met. The experiments involving schoolchildren that dominate Sreenivasan's discussion do not investigate this kind of consistency at all.

This vitiates Sreenivasan's argument against Doris, which consists of three objections to the way in which the character trait of honesty is operationalised in the experiments involving schoolchildren. If these are to undermine the argument Doris has presented, they must be applicable to the experiments he relies on, and we will see that they are not: the first simply cannot be applied to any experiments concerned with cross-situational consistency in the sense in which Doris is interested; applying the second to the data Doris employs requires also applying the third; and the third cannot be applied

to the most widely replicated experiment he cites, and which he dwells on at length in his book, the Obedience Experiment.

Sreenivasan's first objection is that experimenters record inconsistent behaviour where the agent might see no inconsistency at all. The experimenters classify pocketing stray change as dishonest, for example, where many people — presumably including some of the subjects — think it acceptable to keep whatever one finds. So where the experimenters record inconsistency when a child does not cheat on a test but does pocket unattended change, there may not be any inconsistency here from the child's point of view (Sreenivasan 2002, p. 58). This objection relies on the two situations involving different situational features understood by the experimenters — but perhaps not by the subjects — as inviting a response manifesting the same trait. So it cannot be applied to experiments in which the situational feature inviting the trait-relevant behaviour does not vary. If the subject agrees that delivering electric shocks to a stranger who does not give the right answer to your question displays less compassion for that stranger than would be displayed by refraining from delivering those shocks, then they will agree that this is true regardless of whether or not an authority figure instructs you to deliver the shocks.

Of course, they might think that one *ought to* deliver those shocks when instructed to do so by some authority figure, but *ought not* do so otherwise, but if so then they simply find the same compassionate behaviour to be wrong in one case and right in the other: in their view, one ought to be cross-situationally inconsistent in the sense that interests Doris. This thought, that one can be inappropriately compassionate, leads to the

objection that Doris fails to understand the kind of consistency required for virtue, but this objection will not be discussed here (see Miller 2003, Sect. IV; Webber 2006, Sect. 6). What is important here is that it seems plausible to suppose that those subjects who delivered shocks under instruction when they would not have delivered them otherwise would agree that this shows an inconsistency of compassionate action. Indeed, as we will see, such subjects did generally view their own actions as regretfully lacking in compassion.

Sreenivasan's second point is that the different ways of being dishonest may not be equally paradigmatic of their class. While changing one's answers as one is marking one's own test seems fairly paradigmatic of cheating, pocketing loose change does not seem paradigmatic of stealing, he argues, so we should not look for correlations between these two kinds of behaviour but between that kind of cheating and a more paradigmatic example of stealing, such as shoplifting (Sreenivasan 2002, pp. 58–9). In order to apply this point to the Obedience Experiment, one would need to argue that delivering dangerous electric shocks to a fellow volunteer is paradigmatic of a lack of compassion when at the behest of another volunteer or at no behest at all, but not when instructed by an experimenter.

To do so would be to adapt Sreenivasan's third point, that whether or not a particular incidence of lying is paradigmatic is debatable. In a paradigm case of lying, he claims, 'the reason not to lie is a decisive reason for action', but it is not obvious whether the reason not to lie is decisive when lying would lead to some perceived good, such as helping a friend. Whether we should classify such a case as paradigmatic of lying

should therefore be decided not by the experimenter alone but by the agreement of experimenter and subject.

In comparing a situation in which lying would help a friend with one in which it would not, Sreenivasan has introduced into his discussion for the first time the kind of cross-situational consistency that interests Doris. This enables him to levy this objection at the use Doris makes of the Samaritan Experiment. If ‘the fact that one is in a hurry can defeat the reason to help someone in distress’, he writes, then ‘failure to help someone in distress is perfectly compatible (in that case) with the trait of compassion’. The subject might think that in this case hurry makes the manifestation of compassion inappropriate (2002, pp. 59–60). So if Sreenivasan’s third point is right, then whether helping someone while hurried is a paradigm case of compassion is also something that should be agreed between experimenter and subject.

This objection cannot be pressed against the Obedience Experiment, however, since there is ample evidence that subjects and experimenters are in agreement here. Most of the subjects did not obey the authority figure blithely, but displayed such symptoms of stress and nervous tension as sweating, shaking, stammering, muttering, and even crying (Milgram 1974, Ch. 12). Those who later said that they had behaved as one should behave under the circumstances were very few and far between (Milgram 1974, Ch. 5).<sup>2</sup> When the experiment was explained to diverse groups of people and they were asked to predict how they themselves would behave, all said they would quit before reaching 300 volts, and most said they would quit at around 150 volts, when the learner

---

<sup>2</sup> For the symptoms of stress, see also Milgram’s film of the experiment, *Obedience*.

first started to complain (Milgram 1974, Ch. 3). When asked to predict how one hundred Americans of diverse ages and occupations would behave, people generally gave the same predictions, and claimed that only a pathological one or two percent would reach the maximum shock level (Milgram 1974, Ch. 3).

It seems that there usually is agreement between experimenter and subject, as there is among the population at large, that compassion is inconsistent with reaching the maximum shock level in the Obedience Experiment, or even with going much beyond 150 volts. Even if Doris were to disregard those few subjects who disagreed that compassion would be appropriate on this occasion, he would still be able to mount his argument that the lack of compassion shown by the majority of the remainder results from situational features other than that inviting compassionate behaviour, and hence that our behaviour is not cross-situationally consistent in the sense that interests him.

The Obedience Experiment is all the data Doris needs. If people really do have character traits as these are construed in philosophical virtue theory, his argument runs, then we should expect people to respond to the same kind of stimulus in the same kind of way except in situations where they judge other factors to make that response inappropriate. We should expect that those people who would refuse to administer electric shocks in the absence of an authority figure encouraging them to do so would also refuse in the presence of an authority figure encouraging them to do so, unless they judge that the authority figure's instructions render manifesting compassion for the victim inappropriate in this situation. Since most of the subjects of the Obedience Experiment do not judge the instructions to render compassionate behaviour

inappropriate, and yet fail to behave compassionately when under instruction, the expectation grounded in the notion of character employed in virtue ethics is not met. None of the points Sreenivasan makes in his discussion of the experiments concerned with honesty among schoolchildren can be pressed against this argument.

This is not to say that Doris draws the right conclusion from the Obedience Experiment, of course. It is indeed questionable whether he has correctly understood the notion of character employed in virtue ethics in the first place. It is questionable, that is, whether that notion should lead us to expect ‘cross-situational consistency’ in his sense of the phrase. Perhaps the traditional notion of character should not lead us to expect subjects who would not issue the shocks in the absence of an authority figure to *either* refuse to issue shocks in the presence of an authority figure *or* claim that the authority figure’s instructions make such behaviour inappropriate in this situation. But that is another matter entirely (see Webber 2006). It remains nonetheless that Sreenivasan’s argument focuses on the wrong sense of the phrase ‘cross-situational consistency’, and hence fails to block the argument that Doris presents. It is also clear, moreover, that the points Sreenivasan makes do not show the Obedience Experiment to be inappropriate empirical grounding for discussions of the nature of character and virtue.<sup>3</sup>

---

<sup>3</sup> I am grateful to two anonymous reviewers for *Mind* and to the editor for their comments on earlier drafts.

## REFERENCES

Doris, John 1998: 'Persons, Situations, and Virtue Ethics'. *Noûs*, 32, no. 4, pp. 504-30.

————— 2002: *Lack of Character: Personality and Moral Behaviour*. Cambridge: Cambridge University Press.

Milgram, Stanley 1965: *Obedience*. Film. State College PA: The Pennsylvania State University Media Sales. Reissued on VHS 1993. Also issued on VHS by British Universities Film and Video Council, 1991.

————— 1974: *Obedience to Authority: An Experimental View*. New York: Harper and Row.

Miller, Christian 2003: 'Social Psychology and Virtue Ethics'. *The Journal of Ethics* 7, no. 4, pp. 365-392.

Sabini, John and Maury Silver 2005: 'Lack of Character? Situationism Critiqued'. *Ethics* 115, no. 3, pp 535-562.

Sreenivasan, Gopal 2002: 'Errors About Errors: Virtue Theory and Trait Attribution'. *Mind* 111, no. 441, pp. 47-68.

Webber, Jonathan 2006: 'Virtue, Character and Situation'. *Journal of Moral Philosophy* 3, no. 2, pp. 195-216.