



OXFORD JOURNALS  
OXFORD UNIVERSITY PRESS

## Mind Association

---

Errors about Errors: Virtue Theory and Trait Attribution

Author(s): Gopal Sreenivasan

Source: *Mind*, New Series, Vol. 111, No. 441 (Jan., 2002), pp. 47-68

Published by: [Oxford University Press](#) on behalf of the [Mind Association](#)

Stable URL: <http://www.jstor.org/stable/3093787>

Accessed: 17/10/2010 07:18

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=oup>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



Oxford University Press and Mind Association are collaborating with JSTOR to digitize, preserve and extend access to *Mind*.

<http://www.jstor.org>

# Errors about Errors: Virtue Theory and Trait Attribution

GOPAL SREENIVASAN

This paper examines the implications of certain social psychological experiments for moral theory—specifically, for virtue theory. Gilbert Harman and John Doris have recently argued that the empirical evidence offered by ‘situationism’ demonstrates that there is no such thing as a character trait. I dispute this conclusion. My discussion focuses on the proper interpretation of the experimental data—the data themselves I grant for the sake of argument. I develop three criticisms of the anti-trait position. Of these, the central criticism concerns three respects in which the experimental situations employed to test someone’s character trait are inadequate to the task. First, they do not take account of the subject’s own construal of the situation. Second, they include behaviour that is only marginally relevant to the trait in question. Third, they disregard the normative character of the responses in which virtue theory is interested. Given these inadequacies in situationism’s operationalized conception of a ‘character trait’, I argue that situationism does not really address the proposition that people have ‘character traits’, properly understood. *A fortiori*, the social psychological evidence does not refute that proposition. I also adduce some limited experimental evidence in favour of character traits and distil two lessons we can nevertheless learn from situationism.

Philosophical attention has recently been drawn again to a body of empirical research in social psychology showing, among other things, that people are remarkably liable to be overhasty in their everyday attributions of character traits.<sup>1</sup> This particular liability is commonly referred to as the ‘fundamental attribution error’ (Ross and Nisbett 1991, p. 4); and the empirical research which exposes it belongs to the ‘situationist’ tradition in social psychology. What some philosophers now suggest, in effect, is that situationism’s empirical findings cast serious doubt on the idea that there is even such a thing as a character trait,<sup>2</sup> everyday attributions notwithstanding. Since philosophical accounts of the virtues—especially those inspired by Aristotle—often

<sup>1</sup> I have Harman (1999) and Doris (1998, in press) particularly in mind, but compare also Blackburn (1998, pp. 36–7) and Railton (1995, pp. 93–6). Essentially the same research is also discussed by Flanagan (1991), whose conclusions I find more congenial.

<sup>2</sup> Harman writes, for example, that it ‘seems that ordinary attributions of character traits to people are often deeply misguided and it may even be the case that there is no such thing as character, no ordinary character traits of the sort people think there are, none of the usual moral virtues

analyse a virtue as a species of character trait, the further implication on offer is that many philosophical accounts of the virtues are beset with false empirical presuppositions (Harman 1999, p. 319; Doris 1998, pp. 505–9). To the extent that philosophy simply reflects common sense here, a corrective lesson in situationism is equally held out for ordinary morality (Doris 1998, pp. 513–15). ‘Character based virtue ethics may offer a reasonable account of ordinary moral views. But to that extent, these ordinary moral views rest on error’ (Harman 1999, p. 327).

If the empirical evidence marshalled by situationism is well-founded, and if its philosophical advocates are correct, the revisions required in our theory of virtue are striking indeed—as they are doubtless meant to be. But are they really warranted? That is the question I wish to pursue. In so doing, I shall take the relevant empirical findings wholly for granted.<sup>3</sup> I shall also assume that a theory of virtue should conform to certain standards of empirical psychological adequacy, standards which would be violated if next to no one actually turned out to have a character trait in the relevant sense. My enquiry will therefore focus on the proper interpretation of the situationists’ experimental results from the standpoint of a philosophical theory of virtue.<sup>4</sup> What actually follows from these experiments? For the most part, I shall approach this issue constructively: I shall concentrate, that is, on a free-standing identification of the nature of the empirical constraint that social psychology places on moral philosophy.

I shall argue that the social psychological evidence does very little to undermine the virtue-theoretic assumption that certain people actually have character traits in the relevant sense. It even provides some positive evidence in favour of this assumption. On the whole, however, the psychological evidence does not really engage the assumption one way or the other,<sup>5</sup> at least not to date. I conclude that the empirical constraint faced by a theory of virtue is instructive, but not particularly

and vices’ and that ‘we must conclude that there is no empirical basis for the existence of character traits’ (1999, p. 316). Doris is also guarded, but less so: ‘To put things crudely, people typically lack character’ (1998, p. 506). Compare the title of Harman (2000).

<sup>3</sup> Within psychology, the situationists’ empirical data have been received with considerable scepticism by traditional personality theorists. However, situationists pointedly note that more careful research has nevertheless failed to overturn their basic findings. See, for example, Ross and Nisbett (1991, pp. 105–109); and Mischel and Peake (1982).

<sup>4</sup> My discussion of these experiments is largely based on the presentation in Ross and Nisbett (1991). Not only is theirs an accessible textbook, but Harman (1999) and Doris (1998) seem indebted to it as well. Henceforth, bare page references will be to this textbook.

<sup>5</sup> Up to a point, as we shall see, this is because the relevant sense of ‘character trait’ shifts somewhat with the context, as one moves from psychological debates between situationists and ‘personologists’ to philosophical theories of virtue.

difficult to satisfy. The question of what follows from all of this for common-sense morality I shall have to leave to one side.

1. It will be useful to begin by explaining some stock terms. How *character trait* itself is to be understood will, after a fashion, occupy our attention throughout. But we can take Harman and Doris as our initial points of reference. As Harman has it, character traits ‘are relatively long-term stable disposition[s] to act in distinctive ways’ (1999, p. 317). The most careful of Doris’s formulations is cast in terms of ‘virtue’ instead of ‘character trait’: ‘if a person possesses a virtue, she will exhibit virtue-relevant behavior in a given virtue-relevant eliciting condition with some markedly above chance probability  $p$ ’ (1998, p. 509).

Let us use *honesty* as our organizing example of a character trait. Consider the following situations in which one might plausibly assess someone’s honesty: some change has been left on a table in an empty classroom and there is an opportunity to steal it; another child is going to get in trouble and there is an opportunity to avert this by lying; one is correcting one’s own test sheet in class and there is an opportunity to cheat, seemingly with impunity.<sup>6</sup> Call these respectively the *stealing* situation, the *lying* situation, and the *cheating* situation. We naturally imagine that an honest person would decline the opportunities presented in these situations. We suppose that honesty is or includes a reliable disposition not to steal, not to lie, and not to cheat.

Psychologists distinguish two dimensions along which someone’s disposition to honesty may be reliable. The dimension of *temporal stability* concerns the reliability of a character trait with respect to a specific situation relevant to the assessment of that trait—for example, with respect to our cheating situation (in the case of honesty). Temporal stability measures the correlation between one set of behavioural responses to the situation in question and another set of the same responses to the same situation—for example, the correlation between cheating on a classroom test last week and cheating on one this week.

By contrast, the dimension of *cross-situational consistency* concerns the reliability of a character trait across different specific situations relevant to the assessment of that trait—for example, with respect both to our cheating situation and to our stealing situation (in the case of honesty). Cross-situational consistency measures the correlation between a set of behavioural responses to one relevant situation and another set of behavioural responses to a second relevant situation—for example, the

<sup>6</sup> These situations describe some of the behavioural measures actually employed in a landmark study of honesty in school children (Hartshorne and May, 1928).

correlation between cheating on a classroom test and stealing change left on a table.

Before introducing the fundamental attribution error, we should register three points about these measures of trait reliability. To begin with, notice that both temporal stability and cross-situational consistency are objective behavioural measures. Each is based upon observations of someone else's overt behaviour in a carefully controlled situation. But the measures are also 'objective' in the further sense that the operative descriptions of the responses and the situations being studied—for example, 'stealing'—are fixed or coded by the observer. It is one of the hallmarks of situationism to privilege objective behavioural measures in the assessment of character traits, at the expense of various forms of subjective assessment favoured by traditional theorists of personality—for example, self-reports, peer evaluations, and personality assessment scales (pp. 100–109).

The next point is that the debate about character traits primarily concerns the dimension of cross-situational consistency. Situationists do not dispute that people's behaviour exhibits considerable temporal stability.

Stability coefficients—the correlation between two measures of the same behavior on different occasions—often exceed .40, sometimes reaching much higher. For example, Hartshorne and May (1928) found that the tendency to copy from an answer key on a general information test on one occasion was correlated .79 with copying from an answer key on a similar test six months later. (p. 101)

Temporal stability, however, is narrowly tied to fairly specific situations: Hartshorne and May's correlation of .79, for example, is not simply tied to cheating situations, as opposed to stealing or lying situations. It is tied to cheating on a classroom test. But we ordinarily suppose not only that an honest person can be relied upon not to cheat, on classroom tests or elsewhere, but also that he can be relied upon not to steal or to lie either. So the reliability of honesty as we ordinarily conceive it evidently extends into the dimension of cross-situational consistency. That is where the opposition from situationism will begin, as we shall soon see.

Finally, we should acknowledge the importance, in principle, of a distinction that I shall henceforth elide. A suitable regularity in someone's behaviour, whether a temporal stability or a cross-situational consistency, can always be distinguished from a corresponding character trait. Consider, for example, the claim that Jezebel has some temporally stable character trait. This goes beyond the more basic claim that

her behaviour exhibits a certain temporal stability, as it also offers a particular explanation for that temporal stability, namely, Jezebel's possession of a stable psychological disposition that produces such behaviour (Harman 1999, pp. 317–18; Doris 1998, p. 509). Since there might be alternative explanations for it, the fact that Jezebel's behaviour is temporally stable in a certain respect does not suffice to establish that she has the corresponding temporally stable trait. Similarly, in themselves, the situationists' observations of various temporal stabilities in people's behaviour leave room for competing explanations. Harman and Doris actually part company here. Doris explicitly embraces temporally stable traits (1998, pp. 507–8), whereas Harman underscores the availability of non-trait explanations for temporally stable behaviour (1999, p. 326).

Having acknowledged this distinction, I shall proceed to ignore it. In effect, I shall simply assume, given some suitable regularity in a person's behaviour, that the best explanation of it appeals to a corresponding trait. I do so solely for the sake of simplicity. As we shall see, the mainstay of situationism's critique consists in the claim that people's behaviour is not cross-situationally consistent. It denies the *explanandum*, and not the *explanans*, implicit in the assertion of a cross-situationally consistent trait. My simplifying assumption serves to isolate this denial, without having always to distinguish it. But nothing of substance rests on this assumption, as my aim is only to disarm a critique of character traits. I make no pretence of presenting a complete positive case for their existence.

We are now ready to consider the *fundamental attribution error*. In general terms, this may be described as an 'inflated belief in the importance of personality traits and dispositions, together with [a] failure to recognize the importance of situational factors in affecting behavior' (p. 4). While I shall be emphasizing the first half of this error ('an over-eager dispositionism'), it may be helpful at least to illustrate the second half ('an underdeveloped situationism', p. 130). A good illustration is provided by the experiment that Darley and Batson (1973) modelled on the parable of the Good Samaritan.

In the experiment, which both Harman (1999) and Ross and Nisbett (1991) adduce, Darley and Batson had students from the Princeton Theological Seminary prepare a short talk that was to be recorded in another building. The students also completed a questionnaire concerning the basis of their interest in religion. As they left for the other building, some students were told to hurry because they were late; others were told they had just enough time to get there; and yet others

were told they would arrive a little early. *En route*, each student encountered a man slumped in a doorway, who coughed twice and groaned. Which of the students stopped to help? There were three variables—content of assigned talk, religious orientation, and degree of lateness. But only the degree of lateness—that is, a situational variable—turned out to be of any significance. 63 percent of those running early stopped to help, as did 45 percent of those running just on time, whereas only 10 percent of those running late stopped to help. The suggestion is that, in considering the parable, we are apt not to recognize the extent to which the priest and the Levite may simply have been running late.

More relevant for our purposes is the overeager dispositionism half of the fundamental attribution error. At bottom, this derives from a pair of mistakes that people make in their assessment of the evidence base required to attribute a character trait to someone. A character trait, it should be said explicitly, is being assumed here to be grounds for *predicting* how the person to whom the trait is attributed will behave in various particular situations (cf. Flanagan 1991, p. 279). This assumption derives from the fact that character traits are supposed, as we have seen, to explain (at least some of) their possessor's behaviour; and it is reflected in the evidence base required to warrant an attribution.

The pair of mistakes people make can be perspicuously represented as corresponding to the two dimensions of trait reliability we distinguished earlier. Take temporal stability first. Imagine that one day we observe Homer declining to steal the change on the table in the empty classroom. Perhaps he even reports the situation to the principal. One mistake people are liable to make is to conclude that Homer is an honest person, simply on the basis of his honest behaviour on this one occasion. They attribute the (temporally stable) *trait* honesty to Homer, that is, and take this trait to license very confident predictions that (e.g.) Homer will not pocket the change next time he gets the chance either (Kunda and Nisbett 1986, pp. 209–11). More generally, what people apparently fail to appreciate is that the reliability of their predictions depends upon the *number* of observations underlying their trait attributions. To warrant predictions about Homer of the order on which people typically make them, an attribution of honesty to him would have to be based on evidence aggregated over numerous observations.

Turn now to cross-situational consistency, and recall our single observation of Homer's honesty in the stealing situation. A second mis-

take people are liable to make is to predict, on this same basis and also very confidently, that (e.g.) Homer will not cheat on the classroom test next time he gets the chance. Not only is the honesty trait they attribute to Homer temporally stable, that is, but it is cross-situationally consistent as well (Kunda and Nisbett 1986, pp. 209–11). In part, the new prediction suffers from the same evidential weakness as the old one, since it is equally based on a single observation. However, it also suffers from a further weakness, which is connected to people's failure to distinguish between temporal stability and cross-situational consistency.

This failure is nicely brought out in a study on attributions of extroversion reported by Ross and Nisbett. Subjects were given information about the behaviour of certain target persons in three situations.

For some of the subjects, the behavioral information was drawn from just one context, either three academic settings or three social settings. For other subjects, the behavioral information was drawn from both contexts, either two academic settings and one social setting or vice versa ... Subjects were then asked to predict the target's behavior in an academic setting, a social setting, and an ambiguous setting. (pp. 129–30)

The study found that the subjects' predictions were completely insensitive to the situational specificity in their evidence base. Thus, while (e.g.) 'the trait-consistent inferences should be stronger for the ambiguous setting when the behavioral information came from academic and social settings alike than when it was confined to a single type of setting' (p. 130), that is not how the subjects actually predicted. More generally, then, something else that people apparently fail to appreciate is that the reliability of their cross-situational predictions depends upon the *distribution*, as we might put it, of the observations underlying the relevant trait attribution. To warrant the attribution of cross-situationally consistent honesty to Homer, one would require evidence of his honesty that was distributed across stealing situations, lying situations, and cheating situations.

2. Situationists allege that ordinary people—such as you and I—commit the fundamental attribution error left and right; and their philosophical advocates concur. We seriously and routinely underestimate, so it is said, both the number of observations and the distribution within them that is required to warrant the attribution of cross-situationally consistent character traits, such as honesty. Let us suppose that this is correct.

The first reservation I should like to enter is that, strictly speaking, the fundamental attribution error is *irrelevant* to the question of

whether anyone really has a character trait. The issue is whether there is, in fact, warrant to attribute traits and not whether the trait attributions that people commonly make are actually warranted. From the fact that people happen to add badly, it does not follow that there are no sums. Likewise with character traits.

A neat demonstration of this point is implicit in the material we have already reviewed. One source of the fundamental error in people's attributions was their disregard of the need to aggregate the behavioural evidence over a sufficient number of observations. This applied equally to their attribution of temporally stable traits. Recall, however, that the situationists themselves concede that there are temporally stable traits.<sup>7</sup> In one study of conscientiousness, for example, in which there were 19 different behavioural measures, observed between 2 and 12 times each, the mean temporal stability coefficient after aggregation was .65 (Mischel and Peake 1982, pp. 734–35). So here there is a clear warrant to attribute various temporally stable traits, even if elsewhere people attribute temporally stable traits without having acquired a warrant.

It is therefore at best misleading of Doris, for example, to say that 'trait attribution does not ground confident predictions of particular behaviors' (1998, p. 509). If we suppose that a trait has been attributed without warrant, it will come as no surprise to learn that the predictions it licenses are frequently confounded. The predictive inefficacy of everyday trait attributions does not itself tell us anything either about the predictive efficacy of warranted trait attributions or about the prospects for acquiring the relevant warrants.

Of course, there is a sense in which the fundamental attribution error is indirectly relevant to the question of whether there are character traits. It requires us to bracket one source of apparent evidence for the existence of traits, namely, our ordinary views about character, the ones subject to the error.<sup>8</sup> As Harman says, these views 'can be explained without supposing that there are such traits' (1999, p. 329). Still, none of this is strictly evidence against the thesis that character traits, even cross-situationally consistent ones, do exist.

<sup>7</sup> Here I elide the distinction between a suitable regularity in someone's behaviour and the corresponding trait.

<sup>8</sup> I do not say that our ordinary views about character *are* spoiled by the fundamental attribution error. To pronounce on the matter, we would have to enter into the question of the validity and precise bearing of situationism's empirical data, which I have simply granted for the sake of argument here. What I am saying is that *if* our ordinary views about character traits are spoiled by that error, then *we would* have to bracket the evidence they appear to offer, when considering whether character traits exist.

Hence, to pursue our question, we actually have to put the fundamental attribution error as such to one side. What we need to consider instead is whether a warrant to attribute cross-situationally consistent traits *can be* acquired, as we know it can for temporally stable traits (cf. note 7). Here we come to the heart of the situationists' case, which consists in the fact that empirical investigation of cross-situational consistency coefficients has found them to be almost uniformly low.<sup>9</sup> As Ross and Nisbett summarize it,

the average correlation between different behavioral measures specifically designed to tap the same personality trait (for example, impulsivity, honesty, dependency, or the like) was typically in the range between .10 and .20, and often was even lower ... Virtually no coefficients, either between individual pairs of behavioral measures or between personality scale scores and individual behavioral measures, exceeded the .30 'barrier'. (p. 95)

Since we have been using honesty as our central example, let me also report that, according to Hartshorne and May (1928), the average correlation between stealing and lying was .13, between stealing and cheating .13, and between lying and cheating .31. Overall, the average correlation between any given pair of the various behavioural measures of honesty they studied was .23. Once again, I shall simply suppose that the data are correct.

If we make two further assumptions, it can be shown that Hartshorne and May's data exclude any warrant for attributing cross-situationally consistent honesty to our friend Homer. One assumption is that .23 is also the overall average correlation between any given pair of behavioural measures of honesty *in Homer's own case*. The other assumption is that Hartshorne and May's behavioural measures *properly* operationalize the character trait *honesty*. Recall that a warrant to attribute cross-situationally consistent honesty to Homer requires that we collect behavioural evidence of his honesty that is distributed across the various situations, including the stealing, lying, and cheating situations. Granted our assumptions, it will not be possible to collect the requisite evidence.

To see this, begin by considering the average correlation between (e.g.) Homer's stealing behaviour and his lying behaviour. Say it is .13, in line with the first assumption. This implies that repeated observations of Homer's behaving honestly were recorded in at most one of the stealing and lying situations. Had repeated observations of his behaving honestly been recorded in both situations, the average correlation

<sup>9</sup> In fairness to Doris, I should say explicitly that he adduces this evidence also and gives it due emphasis. Cf. Harman (1999, p. 326).

would obviously have been higher than .13. A similar conclusion applies to most pairs of Hartshorne and May's behavioural measures, given that in Homer's case the average correlation between any given pair of measures is only .23. Repeated observations of Homer's behaving honestly must therefore have been limited to a relatively small number of the various situations. It follows that the evidence of Homer's behaving honestly does not satisfy the required distributive constraint.

Otherwise put, it follows that *Hartshorne and May's data* do not warrant attributing cross-situationally consistent honesty to Homer. The conclusion we are looking for, however, is that *no data* warrant this attribution—that any such attribution would be false. We therefore need some guarantee, in effect, that Hartshorne and May's data are the only data that matter. Hence the second assumption: that Hartshorne and May's behavioural measures properly operationalize the character trait honesty. New data about Homer might be collected on the basis of somewhat different behavioural measures; and these new data might yield rather higher average consistency coefficients. If so, they may well warrant attributing some cross-situationally consistent trait or other to Homer. Yet, granted the second assumption, there is no possibility that the trait we may thereby be warranted in attributing to Homer is the trait *honesty*. If everyone is in Homer's position, we may conclude that no one's honesty is cross-situationally consistent.

I do not know whether this is the sort of argument that Harman and Doris have in mind. Neither offers to explain how the situationists' data lead to the conclusion that there are no cross-situationally consistent character traits (cf. Doris 1998, pp. 506–7). Since the argument we have just rehearsed actually delivers the conclusion at issue (at least in the case of honesty), it may be instructive to see where it goes wrong. I have reservations about each of its two further assumptions. While I shall concentrate on the second of these assumptions, let me start by entering a reservation briefly about the first.

Notice that something like the first assumption is needed in order to discuss the attribution of cross-situationally consistent traits to particular individuals. That is because the results of the psychological research are reported at the level of a population aggregate. The consistency coefficient of .13 between lying and stealing, for example, is an average over *all* the children whom Hartshorne and May studied. By itself, however, this figure does not exclude there being *some* individuals for whom the correlation between the stealing and lying situations was much higher; and likewise for the overall average coefficient of .23 for any pair of behavioural measures of honesty. Of course, the figures do

exclude there being many people whose own overall average consistency coefficient is much higher than .23. In this way, the presentation of the results reflects psychology's interest here in how most people behave.

My reservation concerns the transposition of this interest to the case of a theory of virtue. If a theory of virtue does indeed take 'most people' within its scope, then the possibilities left open by the situationists' data need not be excluded. The claim that there are no character traits, or even next to none, may thus be regarded as an innocent exaggeration. On the other hand, if a theory of virtue only applies in the first instance to 'some people', then its empirical presuppositions are *not* falsified unless it is really true that next to no one's (virtue) traits are cross-situationally consistent. Yet the further assumption necessary to reach this conclusion takes us well beyond the existing data.

The scope of a theory of virtue presumably varies with the theory. Still, I mention this reservation because, in my own view, the correct theory of virtue is a theory of what Aristotle called full virtue, which only some people need have.<sup>10</sup> These people are models of virtue and ordinary people will only approximate them in varying degrees, including zero. Since I cannot defend that view here, I do not insist on the second reservation.

3. My fundamental reservation concerns the assumption that Hartshorne and May's behavioural measures properly operationalize the character trait honesty. The assumption has the merit of forcing us explicitly to confront the question of how honesty, *as a cross-situationally consistent character trait*, is to be understood. In general terms, the traits in which virtue theory is interested may be conceived as dispositions to respond in a distinctive way *to* a certain sort of situation.<sup>11</sup> Recall, for example, the core of Doris's formulation: a virtuous person 'will exhibit virtue-relevant behavior in a given virtue-relevant eliciting condition' (1998, p. 509). This means that in order to articulate our understanding of a particular trait, such as honesty, we have to specify which responses are to count as 'honesty-relevant behaviour', *as well as* to specify which situations are to count as 'honesty-eliciting' situations.

<sup>10</sup> Doris concedes this possibility (1998, pp. 511–12). He has something to say against it, but I shall not pursue the matter here. I defend a theory of virtue along these lines elsewhere, in my *Emotion and Moral Judgement* (in preparation).

<sup>11</sup> This may well be only a subset of the traits studied by psychologists. For example, it could be argued that extroversion, talkativeness, and dependency—all studied extensively by situationists (pp. 96–100)—do not fit this conception. That was one of my reasons for focusing on honesty.

Three important issues intersect here, each of which cuts against the assumption under examination.

To begin with, there is the issue of *whose* specification is to count. From the standpoint of the aspiration to predict a subject's behaviour, it would seem that the subject's own specifications should be the ones to count. Suppose, for example, that Homer believes in 'finders keepers' and so does not consider pocketing some stray change to count as *stealing*. Then he will regard pocketing the change as perfectly consistent with his conventionally honest behaviour in, say, the cheating situation. For their part, however, Hartshorne and May count *not* taking the change as the 'honest' response in this situation. Since the behavioural measures in their study are 'objective', as we have already noted, it is their specifications of honesty which are used to score the subjects. Accordingly, Homer will be scored as inconsistent across the stealing and lying situations. This brings out a point that has been well put by Daryl Bem, namely, that

[t]he traditional inference of inconsistency is not an inference about individuals, but a statement about a disagreement between a group of individuals and an investigator over which situations and behaviors may properly be classified into a common equivalence class. (1983, p. 214)

Ironically, situationists themselves are at pains to emphasize what amounts to the same point. Ross and Nisbett call it the *principle of construal*, and describe it as one of the fundamental tenets of social psychology.

The impact of any 'objective' stimulus situation depends upon the personal and subjective meaning that the actor attaches to that situation. To predict the behavior of a given person successfully, we must be able to appreciate the actor's construal of the situation. (p. 11)

It follows, however, that failure to predict a person's behaviour on the basis of 'objective' behavioural measures—that is, low consistency coefficients between such measures—is not always good evidence that the person's behaviour is actually inconsistent across the situations in question. It also depends on the extent to which the subject and the observer agree about the relevance of the trait specifications which the behavioural measures incorporate. Moreover, to accept this is decidedly not to sever the important link between warranted trait attribution and prediction. It is simply to acknowledge the practical difficulties of meaningfully predicting individual behaviour, which are fairly severe (cf. pp. 163–68).

Next there is the issue of *degrees* of relevance to the requirements of honesty. Certain behavioural measures, one might reasonably think,

will be of greater relevance to the assessment of honesty than others. Suppose we accept, for example, that stealing, lying, and cheating are all centrally relevant cases of the sort of behaviour excluded by honesty. There is still a question about the relevance of the more concrete situation–response pairs used to exemplify these cases. Consider Hartshorne and May’s stealing situation again. If I do not believe in ‘finders keepers’, I may well accept that pocketing stray change is, strictly, stealing it. Nevertheless, it hardly seems a central or paradigmatic case of theft. Shoplifting, for example, strikes me as a much clearer case, bearing in mind that the subjects in this study were children. By contrast, changing one’s test answers as one corrects them from the answer key does seem like a fairly central case of cheating—this was the behavioural measure, recall, with the temporal stability coefficient of .79. A reasonable argument could thus be made to substitute a shoplifting scenario for the stray change pocketing scenario within the ‘stealing’ situation, while leaving the cheating situation as it stands; and this quite independently of any boost that might result to Homer’s average consistency coefficient between the two situations.

Finally, there is the issue of the *normative sensitivity* of the specification. Even if we have a concrete situation–response pair that is centrally relevant to the assessment of honesty, not every apparently equivalent variation on our concrete specification will preserve its central relevance. That is because, in the case of the traits that interest a theory of virtue, the responses in question are not meant to be responses simply to the situation as such. Rather, they are meant to be responses to *some reason* for action present within the situation. As a couple of examples should make clear, apparently equivalent variations on a concrete situation may preserve its identity under some mechanical description without preserving the significance of the reason for action present within it.

Consider first another of Hartshorne and May’s situations, the lying situation. Let us stipulate that ‘knowingly communicating a falsehood with the deliberate intention to mislead’ specifies a behavioural response that is a paradigm case of lying. In that sense, Hartshorne and May’s behavioural measure of lying clearly appears to be a paradigm case. Yet recall the particular situation with which they pair this response: it is one in which the intention to mislead serves to achieve a genuine good, namely, preventing another child from getting into trouble. The question therefore arises whether this good suffices to justify the communication of the falsehood. Assume for the moment that it does suffice. Then, obviously enough, the reason not to lie present

within 'the lying situation' is not a decisive reason. It seems to me, however, that it belongs to a *paradigm* case of lying that the reason not to lie is there a decisive reason for action. If I am right, it follows that lying to prevent another child from getting into trouble is not a paradigm case of lying after all, although it satisfies the stipulated description. Of course, I could be wrong. But that does not really matter, since Hartshorne and May's lying situation would still not be a good *behavioural* measure of honesty, unless we wanted to define a model of honesty as someone who behaves 'honestly' even when the balance of reasons stands against it.

These conclusions have been reached under the assumption that lying can be justified (and is, here). We could rescind this assumption. But since it articulates an eminently comprehensible view, the effect would only be to transform the case into one in which a subject and the observer might well disagree on this score. So the value of Hartshorne and May's lying situation as a behavioural measure of consistency is spoiled either way.

Now consider a second example. Suppose we stipulate that 'helping someone in distress' specifies a situation–response pair that constitutes a paradigm case of compassion, a different character trait. It does not follow that a failure to help someone in distress always contra-indicates compassion, or always indicates inconsistency in an otherwise compassionate person. It does not follow because, like the reason not to lie, the reason to help someone in distress can be defeated. Whenever that reason is defeated, failure to help someone in distress is perfectly consistent (in that case) with the trait of compassion. I should have thought, for instance, that the fact that one is in a hurry can defeat the reason to help someone in distress. Naturally, it depends upon a comparison of various factors, such as the importance of what one is hurrying for and the nature of the victim's distress. I do not know how these factors should be assessed in the particular case of Darley and Batson's Good Samaritan experiment: they themselves flag the possibility that 'conflict' explains some of the failures to stop (1973, p. 108). Whatever the right assessment,<sup>12</sup> the point to notice here about their experiment is that its adequacy as a behavioural measure of compassion *depends* on

<sup>12</sup> In a subsequent variant on this experiment, Batson *et al.* (1978) found that the importance of what the subject is hurrying for did, indeed, make a significant difference to the results. Half the new subjects were told that their performance of the task awaiting them in the second building was 'of vital importance' to the experimenter, while the other half were told that theirs was 'not essential'. 70 percent of those in a hurry, but for something of little importance, *stopped to help* (versus 10 percent of those hurrying for something important). 50 percent of those not in a hurry, but still on their way to something important, stopped to help (versus 80 percent of those neither in a hurry nor on their way to something important). I thank Peter Vranas for this reference.

the subject's degree of lateness *not* defeating the reason to help. By contrast, its power to highlight the predictive relevance of the degree of lateness does not depend on the status of the reason to help.

Let me summarize the importance of these three issues. We can do so in either of two ways. From a critical perspective, each issue lays the ground for a principled objection to the relevance of various of Hartshorne and May's behavioural measures to the nature of honesty as a cross-situationally consistent trait. Together these objections vitiate the assumption that Hartshorne and May's behavioural measures properly operationalize the character trait honesty. As this assumption is a premiss of the argument we rehearsed to the conclusion that no one's honesty is cross-situationally consistent, we may therefore reject that argument.

From a constructive perspective, consideration of the same issues suggests certain desiderata our behavioural measures must satisfy if we are to be able adequately to investigate the cross-situational consistency of someone's honesty. Some of these desiderata arise in connection with a theory of virtue's commitment to empirical adequacy—its commitment to there being people who actually possess the character traits it describes. Others arise in connection with a theory of virtue's commitment to moral-theoretical adequacy—its commitment to the character traits it describes being genuine virtues. It will be useful to consider the two sorts of desiderata separately.

Let us begin with the empirical desiderata, which have been the focus of our discussion. We have been taking it as given all along that a character trait must secure a proper licence to predict behaviour. A theory of virtue has no interest, moreover, in a character trait that is not responsive to reasons. We should therefore restrict our attention to a certain genus of character traits—traits that are both responsive to reasons and capable of properly licensing predictions. The cross-situationally consistent character traits described by any theory of virtue will belong to this genus. Our previous discussion enables us to identify three *generic requirements* a behavioural measure must satisfy in order properly to operationalize a character trait of this kind. For a given character trait,

- (i) each behavioural measure must specify a response that represents a central or *paradigm case* of what that trait requires;
- (ii) the concrete situation each specifies must *not* have any features that *defeat the reason* on account of which that trait requires the response in question; and

(iii) the *subject and the observer must agree* on these characterizations of the specified responses and situations.

If a set of behavioural measures fails these requirements, then the character trait it operationalizes is not even minimally eligible to be a virtue. For example, behavioural measures that fail requirement (ii) do not operationalize the responsiveness to reasons (or normative sensitivity) that manifestly belongs to any virtue. Likewise, behavioural measures that fail requirement (iii) operationalize a trait that cannot properly license predictions,<sup>13</sup> as situationists themselves concede (p. 11).

Now there are various ways in which an empirically adequate theory of virtue may nevertheless be defective. Not every character trait minimally eligible to be a virtue is in fact a virtue. As the phrase suggests, minimum eligibility is a necessary, but not a sufficient condition. The further requirements on being a virtue arise in connection with a theory of virtue's commitment to moral-theoretical adequacy. I shall illustrate them briefly to put the generic requirements on behavioural measures of cross-situational consistency in some perspective.

On the one hand, there are various *virtue-specific requirements*. These have to do with how a given theory of virtue understands the specific nature of individual virtues, such as honesty. For example, a theory of *honesty* will either uphold or repudiate the principle of 'finders keepers'. Suppose it upholds that principle, thereby departing from Hartshorne and May. Suppose, indeed, that it does so in order to conform to the third generic requirement—perhaps, so far as honesty is concerned, everyone who is cross-situationally consistent qualifies as consistent only on the supposition that finders are keepers. In that case, this theory of honesty may well turn out to be empirically adequate, in the sense of describing a character trait that people actually possess. Yet, for all that, it may still be false. Whether it is depends, among other things, on whether 'finders keepers' is a correct principle; and that is a *separate* question. I emphasize this because it brings out the crucial point that the requirement of agreement between subject and observer in no way limits or conditions the moral substance of a given virtue. If pocketing stray change *is* stealing, it does not cease to be stealing when subjects disagree. Nor does the generic requirement (iii) imply otherwise.

<sup>13</sup> Of course, the agreement required between subject and observer has to be a suitably limited one: it has, as it were, both to enable prediction and yet to leave room for it at the same time. The third requirement is therefore likely to prove a somewhat delicate matter in practice (pp. 163–68), even though the general idea behind it is clear enough in principle.

On the other hand, there are also various *theory-specific requirements*, as we might call them, on being a virtue. These have to do with how a given theory of virtue understands the moral nature of virtue in general. For example, a theory will take some position on the question of the unity of the virtues; and this will influence the account it gives of 'individual' virtues, such as honesty. We encountered elements of this influence in our discussion of normative sensitivity, where my assumptions about certain conflicts between reasons reflected the presumption that the virtues are *not* strongly unified.<sup>14</sup> If they were, the requirements of honesty could not conflict with the requirements of beneficence, let alone be defeated by them. More to the present point: if the virtues were strongly unified, no person with a cross-situationally consistent character trait would have a *virtue* unless the behavioural measures across which her trait was consistent incorporated the requirements of *every* virtue.

A theory of virtue aspires, then, to both empirical adequacy and moral-theoretical adequacy. Since these are independent desiderata, it is at least possible that empirical adequacy can only be secured at the expense of moral-theoretical adequacy. But that is not the possibility with which we have been primarily concerned.<sup>15</sup> Situationism raises the more basic question of whether empirical adequacy can be secured for any theory of virtue at all: does anyone actually *have* a character trait some theory of virtue describes (quite apart from whether the traits people have, if any, are genuine virtues)? The generic requirements I have articulated aim to clarify what it takes to answer this question properly.

What a theory of virtue presupposes, in my view, is that there is some non-trivial number of people whose honesty is cross-situationally consistent across a range of behavioural measures that satisfy all three generic requirements; and likewise for such other virtues as it recognizes.<sup>16</sup> A theory of virtue committed to some such presupposition is vulnerable to empirical falsification. I submit, however, that no such theory of virtue has as yet actually been falsified. In fact, so far from fal-

<sup>14</sup>I defend this presumption, along with other theory-specific requirements, in my *Emotion and Moral Judgement* (in preparation). Nothing in the earlier discussion rests upon it.

<sup>15</sup>Nor, for that matter, do I think we have been given any reason to believe that it will be realized. Still, the possibility described in the text would suffice to undermine any theory of virtue. So there is a sense, interestingly, in which the position taken by situationism's philosophical advocates is more radical than it needs to be.

<sup>16</sup>I do not mean to suggest that a theory of virtue must recognize the virtue of honesty as we have been discussing it. For my part, I doubt that there is a single character trait that responds to the reasons not to steal, not to lie, *and* not to cheat.

sifying this presupposition, Hartshorne and May's study does not even put it to the test.<sup>17</sup>

4. I should like to emphasize that there is no need to read my conclusions as implying that the psychologists' experiments were ill-designed. Nor should they be so read. It is perfectly possible to conceive of a 'character trait' as licensing predictions simply on the basis of 'objective' behavioural measures, measures which furthermore exhibit no normative sensitivity and which may be drawn from the widest range of measures minimally associated with the trait by common sense. Indeed, this is the conception that situationists impute to their opponents in the psychological debate, the traditional personality theorists or personologists (pp. 90–94; Mischel and Peake, 1982). On the assumption that this is a fair portrayal of personology, let us call this the *personologist's conception* of a character trait. While it is not endorsed by the situationists themselves (p. 163), the personologist's conception is nevertheless a perfectly legitimate target; and it is certainly contradicted by the empirical findings of Hartshorne and May, among others. Their experiments can thus be read as refuting personology on its own terms.

As we have seen, however, virtue theory has a variety of principled grounds for rejecting the personologist's conception of a character trait. Hence, it is not refuted by refuting that conception. There is accordingly a sense in which the confrontation between situationism and a theory of virtue over the existence of 'character traits' operates at cross-purposes. I am afraid that this crossing of purposes must be laid at the door of situationism's philosophical advocates.

It is also worth recognizing that virtue theory has no special claim on the principled grounds we have observed for rejecting the personologist's conception. Neither does philosophy. There is nothing to prevent psychologists from affirming the three corresponding generic requirements on behavioural measures of cross-situational consistency. To some extent, moreover, they have already done so. Earlier I mentioned that social psychology's principle of construal affirms the basic point which motivates the third requirement. It turns out, in addition, that various situationists also affirm the point which motivates the first requirement (e.g. Mischel and Peake, 1982).<sup>18</sup>

<sup>17</sup> A similar conclusion applies to the other studies adduced by Ross and Nisbett (1991), though it bears repeating that the traits there in question are not really of the sort in which virtue theory is interested. See note 11.

The present relevance of this ecumenical obeisance lies in some empirical work undertaken by these fellow-travelling psychologists. Their results encourage virtue theory in its expectation that cross-situational consistency coefficients calculated on the basis of *adequately* specified behavioural measures would be rather higher than the coefficients we have encountered thus far (cf. note 12). I shall call attention to two studies in particular, both of the trait conscientiousness. In one study, Charles Lord (1983) found that an individual's cross-situational consistency in conscientiousness was significantly higher when the pair of situations (behavioural measures of conscientiousness) in question was regarded as similar *by the individual* himself or herself.<sup>19</sup> This result supports the expectation generated by our third generic requirement.

In a second study, to which I referred previously, Mischel and Peake (1982) divided their behavioural measures of conscientiousness according to whether the measures were more prototypical or less prototypical of the trait conscientiousness, as rated by the subject population as a whole.<sup>20</sup> They found that individuals who perceived themselves as highly consistent in their conscientiousness had much higher mean temporal stability coefficients, as a group, on the more prototypical measures (.71), than did individuals who perceived themselves as less consistent in their conscientiousness (.47). The stability coefficients of the two groups did not differ significantly on the less prototypical measures, nor is cross-situational consistency reported as varying significantly with self-perceived consistency (regardless of the prototypicality of the behavioural measures). Mischel and Peake conclude that the perception of consistency in behaviour is rooted in the temporal stability of an individual's behaviour with respect to some bundle of measures—but not necessarily many—that are highly prototypical for a given trait (1982, pp. 750–52).

a particular waiter who 'didn't do anything consistently—except seek to maximize tips', meaning that '[w]ith families, he was warm and homey ... [w]ith adolescents on dates, he was haughty and intimidating', and so on (p. 164). While this recognizes the broader phenomenon of teleological sensitivity, it also reinforces the impression that situationism's sense of 'inconsistency' lies at cross-purposes with a theory of virtue.

<sup>18</sup> There is even some recognition of the issue of normative sensitivity. Ross and Nisbett relate an anecdote from a study concerned with waiters' strategies for maximizing tips. They remark on

<sup>19</sup> Interestingly, the result depended upon the method used to assess the subject's view of the similarity between situations. Simply asking the subject how similar various pairs of situations seemed, for example, did not produce the result.

<sup>20</sup> Note that when Lord (1983) employed his previously successful methods of similarity assessment at this same level—the aggregate level, that is, rather than that of the individual concerned—he found that they then failed to produce his result.

This result supports the conjecture that adhering to our first generic requirement would boost average consistency coefficients, at least for subjects who perceive themselves as highly consistent in the relevant respect. Anyone who has high (positive) temporal stability coefficients on each of two behavioural measures of a given trait will perform have a high cross-situational consistency coefficient as between those same measures. It follows that, within the range of the particular highly prototypical measures that made up their respective individual ‘trait bundles’, Mischel and Peake’s self-perceived high consistency subjects also had high cross-situational consistency coefficients for conscientiousness.<sup>21</sup> Moreover, if the measures in a subject’s trait bundle correspond to *his or her own* ratings of high prototypicality, there will be nothing *post hoc* about that bundle.

In closing, let me simply state two lessons that a theory of virtue can learn from situationism. They constitute the empirical constraint under which moral philosophy labours here. First, cross-situationally consistent character traits will be narrower than imagined. Certainly, they will be considerably narrower than the range of behavioural specifications permissibly associated with a given trait by common sense (cf. Mischel 1990, p. 131). For any given virtue trait, it is an empirical question how wide the widest bundle of paradigmatic behavioural measures is that is consistently instantiated by an actual person; and it is another empirical question how many individuals exhibit the trait with *that* range.<sup>22</sup> These persons will turn out to be our models of the relevant virtue. Second, the models for a given virtue will have to be nominated relative to some ‘normal background’ range of situations. It cannot be assumed that, in literally *any* novel situation, the behaviour of an actual model—even a reliable one—will be consistent with his or her paradigm performances. How to specify this restriction is a task to be discharged by particular theories of virtue.

I believe these lessons are both instructive and readily assimilated. They are also very much in line with the conclusions reached by Owen Flanagan (1991, pp. 279–81 and 290–92), who discussed this literature a decade ago. Harman (1999, p. 326) criticizes Flanagan for not citing

<sup>21</sup> Mischel and Peake do not themselves draw this conclusion, but then they do not focus on the restricted range given by a subject’s trait bundle. Rather, their consistency coefficients are calculated over the wider range given by *all* the measures of conscientiousness rated as more prototypical by their subject population. However, in view of our third generic requirement, this wider range has no special authority (cf. note 20). It is merely a vestige of the personologist’s conception.

<sup>22</sup> It is a theoretical question, and a neglected one, what the relation is between this range of paradigmatic behavioural measures and all the other behavioural measures *properly* associated with the same virtue.

empirical evidence in support of these conclusions. This gives the impression that some evidence already stands against them, evidence which is in need of a rebuttal. Yet, as we have seen, that impression does not withstand scrutiny. Moreover, there *is* some empirical evidence to cite in favour of cross-situationally consistent character traits, properly conceived (cf. also note 7). Or, to be precise, there is evidence that there *would be* evidence for their existence, if only we looked in the right place.<sup>23</sup>

*Department of Clinical Bioethics*  
*Warren G. Magnuson Clinical Center*  
*Building 10, Room 1C118*  
*National Institutes of Health*  
*Bethesda, MD 20892-1156*  
 USA  
*gsreenivasan@nih.gov*

GOPAL SREENIVASAN

## References

- Batson, C. Daniel, Pamela J. Cochran, Marshall F. Biederman, James L. Blosser, Maurice J. Ryan, and Bruce Vogt 1978: 'Failure to Help When in a Hurry: Callousness or Conflict?' *Personality and Social Psychology Bulletin*, 4, pp. 97–101.
- Bem, Daryl J. 1983: 'Toward a Response Style Theory of Persons in Situations', in Monte M. Page (ed) *Nebraska Symposium on Motivation*, 1982, volume 30. Lincoln: University of Nebraska Press.
- Blackburn, Simon 1998: *Ruling Passions*. Oxford: Oxford University Press.
- Darley, John M. and C. Daniel Batson 1973: "From Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behavior'. *Journal of Personality and Social Psychology*, 27, pp. 100–108.
- Doris, John 1998: 'Persons, Situations, and Virtue Ethics'. *Noûs*, 32, pp. 504–530.
- in press: *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

<sup>23</sup> For helpful comments, I am grateful to Paul Benacerraf, Sarah Buss, Peter Vranas, and two anonymous referees. I am especially indebted to John Doris and Gilbert Harman for instructive exchanges. An earlier version of this paper was given as a philosophy colloquium at Washington University in St. Louis. I thank the audience for a stimulating discussion.

- Flanagan, Owen 1991: *Varieties of Moral Personality*. Cambridge, MA: Harvard University Press.
- Harman, Gilbert 1999: 'Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error'. *Proceedings of the Aristotelian Society*, 99, pp. 315–331.
- 2000: 'The Nonexistence of Character Traits'. *Proceedings of the Aristotelian Society*, 100, pp. 223–226.
- Hartshorne, Hugh and Mark A. May 1928: *Studies in the Nature of Character*, vol. 1. *Studies in Deceit*. New York: Macmillan.
- Kunda, Ziva and Richard E. Nisbett 1986: 'The psychometrics of everyday life'. *Cognitive Psychology*, 18, pp. 195–224.
- Lord, Charles G. 1982: 'Predicting Behavioral Consistency From an Individual's Perception of Situational Similarities'. *Journal of Personality and Social Psychology*, 42, pp. 1076–1088.
- Mischel, Walter 1990: 'Personality Dispositions Revisited and Revised: A View after Three Decades', in Lawrence A. Pervin (ed) *Handbook of Personality: Theory and Research*. New York: Guilford Press.
- Mischel, Walter and Philip K. Peake 1982: 'Beyond Déjà Vu in the Search for Cross-Situational Consistency'. *Psychological Review*, 89, pp. 730–755.
- Railton, Peter 1995: 'Made in the Shade: Moral Compatibilism and the Aims of Moral Theory'. *Canadian Journal of Philosophy*, supp. vol. 21, pp. 79–106.
- Ross, Lee and Richard E. Nisbett 1991: *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill.
- Sreenivasan, Gopal in preparation: *Emotion and Moral Judgement*.