J Ethics (2009) 13:213–233
DOI 10.1007/s10892-009-9049-z

Against Moral Character Evaluations: The Undetectability of Virtue and Vice

Peter B. M. Vranas

Published online: 10 June 2009

Abstract I defend the *epistemic thesis* that evaluations of people in terms of their moral character as good, bad, or intermediate are almost always epistemically unjustified. (1) Because most people are *fragmented* (they would behave deplorably in many and admirably in many other situations), one's prior probability that any given person is fragmented should be high. (2) Because one's information about specific people does not reliably distinguish those who are fragmented from those who are not, one's posterior probability that any given person is fragmented should be close to one's prior—and thus should also be high. (3) Because being fragmented entails being *indeterminate* (neither good nor bad nor intermediate), one's posterior probability that any given person is indeterminate should also be high—and the epistemic thesis follows. (1) and (3) rely on previous work; here I support (2) by using a mathematical result together with empirical evidence from personality psychology.

Keywords Cross-situational consistency · Stanley Milgram · Moral character · Situationism · Stanford prison experiment

1 Introduction: The Epistemic Thesis and the Argument

Think about the following people: your mother, your father, your best friend, your worst enemy, your first love, the current U.S. president, Lincoln, Gandhi, Hitler, and Stalin. I presume that some of these people you hold in high esteem because you take them to be good people, while others you hold in contempt because you take them to be bad people. Such evaluations of people in terms of their moral character

P. B. M. Vranas (✉)
Department of Philosophy, University of Wisconsin-Madison, 5185 Helen C. White Hall, Madison, WI 53706, USA
e-mail: vranas@wisc.edu

as good, bad, or intermediate—for short, *moral character evaluations*—are so deeply entrenched in our everyday lives that it may be hard to imagine how we could avoid them. Yet avoid them we should—or so I will argue. In an epistemic sense of "should".

My main thesis in this paper—call it the "epistemic thesis"—is that moral character evaluations are almost always epistemically unwarranted, unjustified. This is not to deny that such evaluations are often *accurate* (unjustified beliefs can be true), so the thesis is not that no good, bad, or intermediate people *exist*. The epistemic thesis is in a way analogous to the claim that one is almost always unjustified in believing that one's lottery ticket will win—even if the ticket does happen to win. The thesis is not about *expressing* evaluations; it is about *making* (i.e., forming or holding) evaluations, even if they are never expressed. The thesis is about evaluations of (people in terms of) moral *character*, not of character *traits* like honesty or courage. Nor is the thesis about evaluations of *actions* as good or bad. If the epistemic thesis is true, then feelings and attitudes like those of esteem and contempt, which are based on moral character evaluations, are epistemically criticizable: similarly to the way in which one's fear of flying is epistemically criticizable if it is based on an unjustified belief that flying is much more dangerous than driving, one's esteem for a person is epistemically criticizable if it is based on an unjustified evaluation of the person as good.

The epistemic thesis should be now relatively clear, but what is my argument for this surprising thesis? Think again about the lottery analogy. (1) The numbers of most tickets will not be drawn, so my prior probability that the number of my ticket will not be drawn should be high. (2) My information about the number of my ticket (e.g., that it is an even number) should almost never significantly change the above probability (except, e.g., if I know that the lottery is rigged), so my posterior probability that the number of my ticket will not be drawn should almost always be high. (3) My ticket cannot win if its number is not drawn, so my posterior probability that my ticket will not win (should be at least as high as my posterior probability that its number will not be drawn, and thus) should also almost always be high. It follows that I am almost always unjustified in believing that my ticket will win. My argument for the epistemic thesis has a similar structure. First, I argue that most people are what I call *fragmented* (i.e., they would behave deplorably in many and admirably in many other situations), so that one's prior probability that any given person is fragmented should be high. Second, I argue that one's information about specific people (e.g., information about their behavior in various situations) should almost never significantly change the above probability, so that one's posterior probability that any given person is fragmented should almost always be high. Third, I argue that being fragmented entails being what I call *indeterminate* (i.e., being neither good nor bad nor intermediate), so that one's posterior probability that any given person is indeterminate (should be at least as high as one's posterior probability that the person is fragmented, and thus) should also almost always be high. It follows that one is almost always unjustified in evaluating a person as good, bad, or intermediate—and this is the epistemic thesis.

Let me state more precisely and more formally the above argument for the epistemic thesis:

	(P1) Most (i.e., the majority of) people are fragmented.
Thus (from P1):	(L1) For every person *p*, one's prior probability that *p* is fragmented should be high.
	(P2) For almost every person *p*, one's posterior probability (given one's information about *p*) that *p* is fragmented should be close to one's prior probability that *p* is fragmented.
Thus (from L1 and P2):	(L2) For almost every person *p*, one's posterior probability that *p* is fragmented should be high.
	(P3) Being fragmented entails being indeterminate.
Thus (from L2 and P3):	(L3) For almost every person *p*, one's posterior probability that *p* is indeterminate should be high.
Thus (from L3):	(ET) For almost every person *p*, one's posterior probabilities that *p* is good, that *p* is bad, and that *p* is intermediate should be low [and this amounts to the epistemic thesis].[1]

I hope it is clear that the argument is valid,[2] so it remains to defend its three premises: P1, P2, and P3. I have defended P1 and P3 in a previous publication (Vranas 2005); the main task of the present paper is to defend P2. In Sect. 2 I summarize my previous defense of P1 and P3; in Sect. 3 I defend P2; in Sect. 4 I address three objections to the epistemic thesis; in Sect. 5 I conclude by suggesting two directions for future research.

2 Review: Most People are Fragmented (P1), and Being Fragmented Entails Being Indeterminate (P3)

I call a person *fragmented* exactly if the person does or would behave *deplorably* (i.e., in a seriously blameworthy way) in an open list of actual or counterfactual situations and *admirably* (i.e., in a highly praiseworthy way) in another such open list. I have defended the claim that (P1) most people are fragmented by arguing that (1) there is an open list of situations in which most people would behave deplorably and (2) there is another open list of situations in which most people would behave admirably. In support of (1), I have argued that most people would behave deplorably in three (kinds of) experiments from social psychology: (a) the *obedience experiments* (Milgram 1974), in which most participants administered powerful (in fact fictitious) electric shocks to a screaming confederate, (b) the *Stanford Prison Experiment* (Zimbardo 2007; Zimbardo et al. 1973), in which most "guards" in a simulated prison maltreated the "prisoners", and (c) the *seizure experiments*

[1] If the epistemic thesis is true, then one's posterior probability that *p* is *not* good should be high, and then evaluations of people as not good (similarly, as not bad or as not intermediate) are almost always epistemically *justified*.

[2] One might argue that the move from P1 to L1 violates a version of the is/ought thesis because L1 is a normative claim but P1 is not. To avoid this problem, I can replace P1 with the (normative) claim that one's probability that most people are fragmented should be high (assuming one's background knowledge includes the evidence I adduce for P1). Compare: my probability that the numbers of most tickets will not be drawn should be high (assuming my background knowledge includes commonly known propositions about lotteries).

(Latané and Darley 1970), in which most participants failed to help a confederate who pretended to be the victim of an epileptic seizure and repeatedly asked for help. In support of (2), I have argued that most people would behave admirably in three other (kinds of) experiments: (a) the *electrocution experiments* (Clark and Word 1974), in which most participants helped an apparently electrocuted confederate at the risk of being electrocuted themselves, (b) the *theft experiments* (Moriarty 1975), in which most participants stopped a simulated theft, and (c) the *rape experiments* (Harari et al. 1985), in which most participants tried to stop a simulated rape.

I call a person *indeterminate* exactly if the person, evaluated in terms of moral character, is neither good nor bad nor intermediate. So the claim that a person is indeterminate is a claim not to the effect that one lacks certain *information* about the person, but rather to the effect that the person lacks certain *properties*, namely the properties of being good, of being bad, and of being intermediate. I have defended the claim that (P3) being fragmented entails being indeterminate by arguing that it is entailed by the following claim: (P0) if two people are such that the first behaves (i.e., does or would behave) much better than the second in an open list of situations and the second behaves much better than the first in another such open list, then neither person is (overall) better than the other. In support of P0, note that if a first person is a model spouse and parent but a tyrannical employer whereas a second person is a tyrannical spouse and parent but a model employer, then it is natural to say that the two people are overall too different to be compared, or at least that neither of them is better than the other. It takes some work to show that P0 entails P3, but the derivation does not matter for my purposes in the present paper; I refer the reader to my previous publication (Vranas 2005) for details and for an examination of numerous objections, but I do not presuppose any familiarity with that publication in what follows.

If most people are fragmented and being fragmented entails being indeterminate, then *most people are indeterminate*. It is important to note, however, that the epistemic thesis does not immediately follow. This is because we might be able to *reliably distinguish* the minority of people who are (e.g.) good from the majority who are indeterminate. (As an analogy, although most people do not have green eyes, we can reliably distinguish the minority of people who have green eyes from the majority who do not, and thus we are often epistemically justified in believing that a person has green eyes.) It is the function of P2 to rule out the possibility that we might be able to make such reliable distinctions (and thus to rule out the possibility that virtuous and vicious people might be "detectable"—hence the title of this paper), so I turn next to my defense of P2.

3 Posterior Versus Prior Probability of Fragmentation

3.1 The Argument for P2

Recall that P2 is the following claim: for almost every person p, one's posterior probability (given one's information about p) that p is fragmented should be close to one's prior probability that p is fragmented. I will defend P2 via a defense of what I call the *Approximate Independence Condition*. This is roughly the claim that a person's behaviors in various situations should be considered approximately independent of each other; for example, learning that a given person behaves (i.e., does or would behave) deplorably in situations 1 and 2 should leave almost unaffected one's probability that the person behaves deplorably in situations 3 and 4. To formulate this condition more precisely, I need to introduce some notation. Consider a person p and a situation s, and let the random variable B_{ps} take the values -1, $+1$, and 0 if person p in situation s behaves deplorably, admirably, and neither deplorably nor admirably respectively. (The values -1, $+1$, and 0 are arbitrary: any three distinct numbers will do.) Then define:

Approximate Independence Condition For almost every person p, and for any S situations, the random variables B_{ps} ($s = 1,\ldots,S$) should be considered approximately (jointly) independent; i.e., their joint subjective probability distribution function should be approximately equal to the product of their marginal subjective probability distribution functions.

Given the above definition, here is my argument for P2:

(P4) If the Approximate Independence Condition is true, then P2 is true.

(P5) The Approximate Independence Condition is true.

Thus: (P2) For almost every person p, one's posterior probability (given one's information about p) that p is fragmented should be close to one's prior probability that p is fragmented.

The argument is clearly valid (by modus ponens), so it remains to defend its two premises. I defend P4 in Sect. 3.2, and P5 in Sects. 3.3 and 3.4.

3.2 A Defense of P4

Let the *Strict Independence Condition* be that, for almost every person p, and for any S situations, the random variables B_{ps} ($s = 1,\ldots,S$) should be considered exactly (jointly) independent; for example, one's probability that p behaves deplorably in a given obedience experiment given that p behaves *deplorably* in a given electrocution experiment should be the same as one's probability that p behaves deplorably in the obedience experiment given that p behaves *admirably* in the electrocution experiment. Let the *Symmetry Condition* be that the random variables B_{ps} should be considered identically distributed; for example, one's probability that p behaves deplorably in a given obedience experiment should be the same as one's probability that p behaves deplorably in a given electrocution experiment. These two conditions are clearly false, but they will be relaxed later on. For the moment I want to make the preliminary points that, when both conditions hold, the claim that one's posterior probability that p is fragmented should be close to one's prior probability (a) is still not trivial but (b) is nevertheless true.

Let F be the proposition that p is fragmented, and let the proposition E describe one's evidence (i.e., information) about p; for example, E may be the proposition that p behaves deplorably in situation s. One might think that, if the Strict Independence Condition holds, then trivially one's posterior and prior probabilities $P(F|E)$ and $P(F)$ should be equal: if learning that p behaves deplorably in s should leave unaffected one's probabilities that p behaves deplorably (or admirably) in any other situation, then should it not also leave unaffected one's probability that p is fragmented? This reasoning is fallacious: maybe learning that p behaves deplorably in s should affect one's probability that p is fragmented because it should increase one's estimate of the *number* of situations in which p behaves deplorably, even if it should leave unaffected one's probability that in any *specific* situation p behaves deplorably. As an analogy, if I know that 10 independent tosses of a fair coin took place, then learning that the coin came up heads in tosses 1 through 6 should increase (from five to eight) my estimate of the number of tosses among the 10 in which the coin came up heads, although it should leave unaffected my probability that the coin came up heads in any particular toss from 7 to 10. So more work is needed to show that, if the Strict Independence Condition holds, one's posterior probability that p is fragmented should be close to one's prior probability. This extra work is partly carried out by the following theorem.

Theorem *Consider S independent and identically distributed random variables $B_{p1},\ldots,B_{pS}$, each of which can take the values -1, 0, and $+1$ with probabilities p_D, p_N, and p_A respectively ($p_D + p_N + p_A = 1$). Let F be the event that more than σ_D of these variables take the value -1 and more than σ_A take the value $+1$. Let D_s be the event that B_{ps} takes the value -1. Then: $P(F|D_s) - P(F) = (1 - p_D)P($exactly σ_D of the remaining $S - 1$ variables take the value -1 and more than σ_A of them take the value $+1) - p_A P($exactly σ_A of the remaining $S-1$ variables take the value $+1$ and more than σ_D of them take the value $-1)$.*[3]

Given that S is very large (it is larger than σ_D and σ_A, which are very large because they correspond to *open lists* of situations), one's probability that *exactly* σ_D or σ_A variables take a specific value should be very small,[4] so both terms of the difference which gives $P(F|D_s) - P(F)$ should be very small, and so should be $P(F|D_s) - P(F)$. Of course $P(F|E) - P(F)$ may still be large if one's evidence E is not limited to the proposition (D_s) that p behaves deplorably in the *single* situation

[3] I will give a proof of the following simpler result (the proof of the theorem uses the same methods): *If Δ is the event that more than σ_D of the S variables take the value -1, then $P(\Delta|D_s) - P(\Delta) = (1 - p_D)P$ (exactly σ_D of the remaining $S - 1$ variables take the value -1)*. Let Δ_n be the event that exactly n of the S variables take the value -1; with an obvious change in terminology, Δ_n is the event of n "successes" in S "trials" of a Bernoulli process with success probability p_D and failure probability $q_D = 1 - p_D$. Clearly $\Delta = \cup_{n > \sigma_D} \Delta_n$, so (1) $P(\Delta|D_s) - P(\Delta) = \Sigma_{n > \sigma_D} [P(\Delta_n|D_s) - P(\Delta_n)]$. Now $P(\Delta_n D_s)$ is the probability of n successes in S trials *and* a success at the s-th trial, so it is the probability of $n - 1$ successes in $S - 1$ trials *times* the probability of a success at the s-th trial. So (2) $P(\Delta_n|D_s) = P(\Delta_n D_s)/P(D_s) = P(n - 1$ successes in $S - 1$ trials). Now (3) $P(\Delta_n) = P(n$ successes in S trials) $= P(n - 1$ successes in $S - 1$ trials and a success at the remaining trial) $+ P(n$ successes in $S - 1$ trials and a failure at the remaining trial) $= P(n - 1$ successes in $S - 1$ trials)$p_D + P(n$ successes in $S - 1$ trials)q_D. From (2) and (3) we get: $P(\Delta_n|D_s) - P(\Delta_n) = q_D[P(n - 1$ successes in $S - 1$ trials) $- P(n$ successes in $S - 1$ trials)]. So $P(\Delta|D_s) - P(\Delta) = \Sigma_{n > \sigma_D} q_D[P(n - 1$ successes in $S - 1$ trials) $- P(n$ successes in $S - 1$ trials)] $= q_D\{[P(\sigma_D$ successes in $S - 1$ trials) $- P(\sigma_D + 1$ successes in $S - 1$ trials)] $+ [P(\sigma_D + 1$ successes in $S - 1$ trials) $- P(\sigma_D + 2$ successes in $S - 1$ trials)]$+\cdots+$[$P(S - 1$ successes in $S - 1$ trials) $- P(S$ successes in $S - 1$ trials)]\} $= q_D P(\sigma_D$ successes in $S - 1$ trials).

[4] Applying the normal approximation to the binomial distribution, the probability that exactly σ_D out of $S - 1$ variables take the value -1 is at most $[2\pi(S - 1)p_D(1 - p_D)]^{-1/2}$.

s;[5] but given that in everyday life one's evidence about people is almost never so extensive as to encompass a number of situations approaching an open list, (a generalization of) the theorem suggests that $P(F|E) - P(F)$ should almost always be small (if the Strict Independence Condition and the Symmetry Condition hold).[6] One might object that we may have observed some of our intimates for decades. I reply that our observations are typically limited to a small number of recurring situations.

I conclude that, if the Strict Independence Condition and the Symmetry Condition hold, then one's posterior probability that any given person is fragmented should almost always be close to one's prior probability. Now let me relax the above two conditions. Relaxing the Symmetry Condition does not affect my argument: it becomes hard to state explicitly a theorem analogous to the above one (because the probabilities p_D, p_N, and p_A are replaced by situation-specific probabilities p_{Ds}, p_{Ns}, and p_{As}), but the essential point remains that $P(F|E) - P(F)$ should be small if E describes behavior in a *small* number of situations (relative to σ_D and σ_A). To relax the Strict Independence Condition, replace it with the Approximate Independence Condition. I can offer no rigorous argument to rule out the possibility that $P(F|E) - P(F)$ may become large when strict independence is replaced with approximate independence, but it would be strange if such a discontinuity existed. Perhaps more seriously, one might worry that replacing the Strict with the Approximate Independence Condition is small improvement: if the former is false, why should the latter be true? It is to this question that I now turn.

3.3 A Defense of the Approximate Independence Condition

An extensive literature in personality psychology suggests that the average correlation coefficients between people's behaviors in various situations are (positive but) low. Here are four examples (cf. Mischel 1968). (1) A massive series of studies was carried out in the late 1920s by H. Hartshorne, M. May, and their collaborators (Hartshorne and May 1928; Hartshorne et al. 1929, 1930). They subjected thousands of schoolchildren to a battery of tests of (e.g.) honesty and found correlations of on average about .26 (Hartshorne and May 1928, p. 383) between scores on these tests. In other words, a child who (e.g.) behaved much more dishonestly than average on a cheating test typically did not behave much more dishonestly than average on a stealing test. (2) Newcomb (1929) measured behaviors presumably reflective of introversion–extroversion in 51 "problem boys" and found an average correlation coefficient of .14 among these measures.

[5] Trivially, for example, in the extreme case in which E is the proposition that p behaves deplorably in more than σ_D and admirably in more than σ_A situations, E *entails* that p is fragmented; so $P(F|E)$ should be 1 even if $P(F)$ is not close to 1.

[6] It might be objected that one's evidence about a given person need not be limited to *behavioral* information, namely information on how the person behaves in various situations: one's evidence may also include *personological* information, namely information on how the person's character is judged by others (including those who know the person well and those who are considered expert judges of character) or information on the person's character based on a questionnaire. I reply that such judgments of character have been found to be inaccurate: their correlations with actual behavior are low (Mischel 1968).

(3) Dudycha (1936) examined the punctuality of 307 college students in six situations (coming to class, coming to an appointment, coming for breakfast, coming to church, etc.). He found an average correlation coefficient of .19 among punctuality scores in these situations. (4) For a more recent example, Peake (1982) monitored 63 undergraduates at Carleton College over a 10-week period for behaviors related to friendliness and conscientiousness. He found average correlation coefficients of .08 for conscientiousness and .05 for friendliness.

One might object that these results conflict with the commonsensical observation that we *can* reliably predict our friends' behavior. In reply, distinguish (as social and personality psychologists do) the *cross-situational consistency* from the *temporal stability* of behavior. The above results do not conflict with the claim that behavior is temporally stable, namely that people behave in more or less the same way when the *same* situation recurs. In fact, the average correlation coefficients between repeated measures of people's behavior in recurring situations are typically much higher than cross-situational correlation coefficients.[7] In everyday life we typically observe our friends in a limited number of recurring situations; this is why we can reliably predict their behavior. But in unusual situations our predictions would often go awry. This claim is also supported by another psychological literature which suggests that personality characteristics do not enable us to reliably predict who helps in bystander intervention experiments or who obeys in obedience experiments. Latané and Darley, for example, found that personality variables such as "[a]lienation, Machiavellianism, acceptance of social responsibility, need for approval, and authoritarianism did not predict the speed or likelihood of help" (1970, p. 116). In a more recent review of the helping literature by Piliavin et al. (1981, Chap. 8), only a couple of personality traits were found to consistently predict helping behavior. Similarly, in a comprehensive review of the obedience literature, Blass concluded that, although personality variables can predict obedience, "some of the findings are either contradictory or weak" (Blass 1991, p. 399; cf. Meeus and Raaijmakers 1995, p. 168; Modigliani and Rochat 1995, pp. 121–122).

Another objection to my argument for the Approximate Independence Condition is that a low positive correlation, or even a *zero* correlation, does not amount to independence. This is because the (Pearson product-moment) correlation coefficient measures the degree of *linear* dependence between two variables; the coefficient can be zero even when there is a perfect non-linear dependence. I reply that the Approximate Independence Condition, as formulated in Sect. 3.1, does not state that people's behaviors in various situations *are* approximately independent of each other; it states instead that they *should be considered* approximately independent.

[7] There is a gray area between cross-situational consistency and temporal stability: depending on how coarse- or fine-grained one's individuation of situations is, in some cases one may consider a situation s' to be either (1) the same as a (recurring) situation s (so that the comparison between s and s' is relevant to temporal stability) or (2) different from—but very similar to—a situation s (so that the comparison between s and s' is relevant to cross-situational consistency). The average correlation coefficients between people's behaviors in various situations are higher if the situations are similar than if they are not (cf. Davies 1991, p. 528; Doris 2002, p. 64; Hartshorne and May 1928, p. 384). So for the Approximate Independence Condition to hold I adopt a coarse-grained individuation of situations: I consider (what one might consider to be) very similar situations to be identical.

To see the difference, take an analogy. Consider a pseudo-random number generator: a device that uses a deterministic algorithm to generate numbers whose distribution looks random. If I do not know (and I have no way of finding out) the algorithm that the device uses, then I should consider the distribution as random, although the distribution is not *in fact* random. Similarly, I claim that *given the state of the art* in prediction methods, the low correlations have the consequence that people's behaviors in various situations should be considered approximately independent; I cannot rule out the possibility that future prediction methods will enable one to reliably distinguish those who are fragmented from those who are not.

A third objection to my argument for the Approximate Independence Condition is that the correlation coefficients, low as they are, are *not sufficiently low* to require considering people's behaviors in various situations as even approximately independent. Funder and Ozer (1983), for example, argue that some typical correlation coefficients in personality psychology are of about the same magnitude as the correlation coefficients that correspond to some typical social psychological experiments (cf. Richard et al. 2003; Sarason et al. 1975; Swan and Seyle 2005, p. 159). In Milgram's obedience experiments, for example, the correlation between experimenter proximity and obedience was calculated by Funder and Ozer to be .36 (1983, p. 110). I have three points in reply. First, I am not making a *comparative* claim about the relative importance of personality characteristics versus situations when I defend the unpredictability of behavior in unusual situations. Second, quoting Nisbett (1980), Funder and Ozer (1983) took correlation coefficients with values up to .40 to be typical of personality research; but as we saw above, the average correlations found by Hartshorne and May (1928), Newcomb (1929), Dudycha (1936), and Peake (1982) were much lower than .40. Third, and most important, Ross and Nisbett (1991, pp. 109–115; see also Ross and Thomas 1986, 1987) argue extensively that average correlation coefficients of almost .20 result in only negligible improvements in the predictability of behavior.

3.4 Epstein's Objection from Aggregation

In a series of publications, Epstein (1977, 1979a, b, 1980, 1983a, b, 1984, 1986) has adduced certain considerations that might be taken to provide an objection to my argument for the Approximate Independence Condition. Epstein basically argues that, although the average correlation coefficient for B_{ps} is indeed low, this is only to be expected given measurement errors; the average correlation coefficient for Y_σ will be high, where each Y_σ is an *aggregate* measure of behavior over a large number of situations (i.e., Y_σ is the average of B_{ps} over all s in the set σ of situations).[8] Epstein's considerations can be divided into two parts: an intuitive a priori argument, and a more rigorous argument based on the Spearman-Brown formula. I will address these two arguments in turn; then I will examine empirical evidence relevant to aggregation.

[8] One could equivalently define Y_σ as a sum (rather than an average): the correlation coefficients would remain unchanged.

3.4.1 Epstein's Intuitive Argument

Epstein writes (1979b, p. 1102):

> It is no more reasonable to assess the stability of nonintellective behavior by correlating single observations than it is to assess the stability of intellective behavior by correlating single items in an intelligence test. Thus, for statistical reasons alone, the low correlations cited as evidence against the existence of stable response dispositions can be discounted.[9]

To take another analogy, if we measure the *lengths* of all objects in a series by using an imprecise instrument, then any two series of measurements are expected to correlate only slightly; but if we take repeated measurements and average them, then any two series of average measurements are expected to correlate highly. In terms of the theory of measurement, we can say that $B_{ps} = T_{ps} + E_{ps}$, where B_{ps} is the measured score, T_{ps} is the "true score", and E_{ps} is the measurement error (e.g., Gulliksen 1950). The average Y_σ of B_{ps} over all s in σ will be approximately equal to the average of T_{ps} if (as one standardly assumes) the average of E_{ps} is approximately zero. But each T_{ps}, and thus also the average of T_{ps}, is just t_p, the length of object p; so the correlation coefficient between Y_σ and $Y_{\sigma'}$ (for two different sets σ and σ' of measurement-situations) will be approximately equal to the correlation coefficient between t_p and t_p, namely 1.

The problem with the above argument is its reliance on the assumption that T_{ps} is constant across s. This assumption seems reasonable when it is *length* that we are measuring (after all, if we take standard precautions, we do not expect the length of an object to change each time we make a measurement), but the corresponding assumption when we are measuring *behavior* results in a circular argument. Opponents of cross-situational consistency maintain that the true scores T_{ps} and $T_{ps'}$ are *not* equal because person p does *not* behave in the same way in situations s and s' (even when we abstract from questions of measurement error). So in the context of the cross-situational consistency debate Epstein is not entitled to assume that T_{ps} and $T_{ps'}$ are equal, and thus his argument fails.

3.4.2 Epstein's Use of the Spearman-Brown Formula

Consider two sets of situations, σ and σ', consisting of K situations each. If the variance of B_{ps} is constant across all s in σ and σ', and if the correlation coefficients $r_{ss'}$ are the same for all pairs of situations within σ, within σ', and between σ and σ', then, according to the Spearman-Brown formula: $r_{\sigma\sigma'} = Kr_{ss'}/[1 + (K - 1)r_{ss'}]$, where $r_{\sigma\sigma'}$ is the correlation coefficient between the averages Y_σ and $Y_{\sigma'}$ (e.g., Gulliksen 1950, p. 78). It follows that, no matter how small $r_{ss'}$ is, $r_{\sigma\sigma'}$ goes to 1 as K goes to infinity. Does it follow that we can get correlation coefficients as high as we want by aggregating behavior over sufficiently large numbers of situations?

[9] Epstein is talking here about the *temporal stability*, not about the *cross-situational consistency* of behavior (see Epstein 1983b), but an analogous argument (as Epstein himself points out) might be given about cross-situational consistency.

Originally, Epstein (1979b) did not appeal to the Spearman-Brown formula but gave (in addition to empirical evidence) only the intuitive argument I examined above (in Sect. 3.4.1). In fact, when Mischel and Peake (1982) disparaged Epstein's claim that aggregation increases correlation coefficients as nothing more than what would be expected from the Spearman-Brown formula, Epstein replied that the Spearman-Brown formula holds only "when standard assumptions are met that are rarely met in real-life situations";[10] "thus it is important to empirically investigate the effects of different forms of aggregation on the reliability and validity of real-life data" (1983b, p. 180). Later on, however, Epstein himself used the Spearman-Brown formula (e.g., 1986, p. 1203), and claimed that, "given some degree of true relation to begin with, it is possible with sufficient aggregation to obtain a correlation of 1.00" (1986, p. 1204). In fact, Epstein has gone down in the literature as "championing" the use of the Spearman-Brown formula (Ross and Nisbett 1991, p. 108); moreover, Ross and Nisbett claim that "in a sense, Epstein's argument was purely statistical and beyond dispute" (1991, p. 107).

While the truth of the Spearman-Brown formula is a mathematical fact which is beyond dispute, I take the conclusion that with sufficient aggregation one can obtain a correlation coefficient of 1 to provide a reductio of an objection (to my argument for the Approximate Independence Condition) that uses the Spearman-Brown formula. So I agree with Epstein's (1983b) original response to Mischel and Peake (1982): I take the Spearman-Brown formula to be inapplicable to many real-life situations because the assumption that all $r_{ss'}$ are equal is violated. To take a concrete example, suppose all situations in σ concern honesty in school situations, whereas all situations in σ' concern honesty in party situations. It is then reasonable to expect that $r_{ss'}$ for pairs of situations within σ (or within σ') will be higher than $r_{ss'}$ for pairs of situations between σ and σ' (because the two situations in each pair of the former kind will be more similar to each other than the two situations in each pair of the latter kind will be similar to each other). Assuming, for example, that the former correlation coefficients are all equal to .3 and the latter are all equal to .1, one can compute $r_{\sigma\sigma'}$ by using the following formula, of which the Spearman-Brown formula is a special case: $r_{\sigma\sigma'} = Kr_{ss'\text{-}between}/[1 + (K - 1)r_{ss'\text{-}within}]$ (e.g., Gulliksen 1950, p. 77). So as K goes to infinity $r_{\sigma\sigma'}$ goes to $r_{ss'\text{-}between}/r_{ss'\text{-}within} = .1/.3 = 1/3$ for my example. One can thus go beyond Epstein's original vague appeal to the non-satisfaction of the assumptions behind the Spearman-Brown formula: the above considerations provide specific reasons for believing that the specific assumption of equality between correlation coefficients fails, and this failure renders unsound the above argument for the conclusion that with sufficient aggregation one can obtain a correlation coefficient of 1.[11]

[10] Epstein did not specify what these standard assumptions are, nor did he explain why they are rarely met in real-life situations.

[11] It is also worth noting that the above discussion was about prediction of *aggregate* behavior (over a number of situations) from a measure of *aggregate* behavior (over a number of other situations). When one tries to predict behavior in a *single* situation, using a measure of *aggregate* behavior would be of much less help even if the assumptions behind the Spearman-Brown formula were true. This is because, under these assumptions, the correlation coefficient $r_{s\sigma}$ between B_{ps} and Y_σ (as opposed to the correlation

3.4.3 Empirical Evidence Concerning Aggregation

Although references to Hartshorne and May (1928, namely *Studies in the nature of deceit*, the first volume of *Studies in the nature of character*) are a staple of the debate about aggregation and cross-situational consistency, references to Hartshorne et al. (1930, namely *Studies in the organization of character*, the third volume of *Studies in the nature of character*) are much less frequent. But in the latter work the authors explicitly examine correlations between aggregate measures: they compute aggregate scores for (a) 23 deception (honesty) tests, (b) 5 cooperation tests, (c) 4 inhibition tests, and (d) 5 persistence tests, and then they compute correlation coefficients for pairs of these aggregate measures. These coefficients are again low, ranging from .049 to .361 (Hartshorne et al. 1930, p. 151). Epstein (like many others) neglects to mention these important results when he claims that "in almost all cases the correlations were based on single items of behavior" (Epstein 1979b, p. 1102).[12]

This completes my defense of the Approximate Independence Condition, and thus of the premises of the argument for the epistemic thesis.

4 Objections to the Epistemic Thesis

In this section, I address three objections to the epistemic thesis: the objection that people seldom make moral character evaluations and thus the epistemic thesis is uninteresting (Sect. 4.1), the objection that people cannot avoid making moral character evaluations and thus the epistemic thesis conflicts with the ought-implies-can principle (Sect. 4.2), and the objection that *comparative* moral character evaluations are epistemically justified and thus noncomparative ones also are (Sect. 4.3).

4.1 The Triviality Objection

The triviality objection contests not the *truth* of the epistemic thesis, but rather the *importance* of the thesis. The objection infers that the epistemic thesis is unimportant, or trivial, from the premise that people seldom actually make moral character evaluations. (As an analogy, on the basis of the premise that nowadays hardly anyone believes in witches, one might dismiss as trivial the thesis that we are always epistemically unjustified in believing that someone is a witch.) In support of

Footnote 11 continued
coefficient $r_{\sigma\sigma'}$ between $Y_{\sigma'}$ and Y_{σ}) is given by the following formula: $r_{s\sigma} = K^{1/2} r_{ss'}/[1 + (K - 1) r_{ss'}]^{1/2}$; thus $r_{s\sigma}$ goes to $r_{ss'}^{1/2}$, not to 1, as K goes to infinity (e.g., for $r_{ss'} = .16$, $r_{s\sigma}$ goes to .4).

[12] In their response to Epstein, Mischel and Peake (1982, p. 731) note: "far from overlooking reliability, virtually all of the classic, large-scale investigations of cross-situational consistency (e.g., Dudycha, 1936; Hartshorne & May, 1928; Newcomb, 1929) routinely employed behavioral measures aggregated over repeated occasions". Mischel and Peake (like many others), however, fail to note that (as I just said) Hartshorne et al. (1930) also employed behavioral measures aggregated over *situations* (not just over *occasions*).

the premise that people seldom actually make moral character evaluations, one might argue that talk of character occurs mostly in special contexts like recommendation letters; in everyday discourse, people seldom use expressions like “he is a man of good character”. In reply, I will contest both the premise and the inference underlying the objection.

Concerning first the premise, note for a start that even if people seldom *express* moral character evaluations, they may still often *make* such evaluations but refrain from expressing them because expressing them would be socially undesirable: telling you to your face that you are bad or good is often insulting or ingratiating respectively, and telling you that a third person is good or bad may sound moralistic and is often in bad taste. Moreover, the questions of how often people express and how often people make moral character evaluations are empirical ones, so I collected some relevant data. I asked 189 introductory psychology students the following four questions:

1. How often does it happen that **you** evaluate someone *in conversation* in terms of their moral character (in other words, you *say to someone* something *like* “she is a good [or bad] person”)?
2. How often does it happen that **you** evaluate someone *in your mind* in terms of their moral character (in other words, you *say to yourself* something *like* “she is a good [or bad] person”, but you *don’t* go on to voice this thought)?
3. How often does it happen that **people** evaluate someone *in conversation* in terms of moral character? In other words, how often do you *hear people saying* something *like* “she is a good [or bad] person”?
4. How often does it happen that **people** evaluate someone *in their mind* in terms of moral character? In other words, how often do people *say to themselves* something *like* “she is a good [or bad] person”, but they *don’t* go on to voice this thought?

The results, reported in Table 1, suggest that it is not rare for people to express moral character evaluations, and that it is common for people to make such evaluations. For example, the most frequent responses to question 2 were “somewhat frequently” and “frequently”, and the mean response was between these two. I conclude that the premise of the triviality objection is implausible.

Concerning now the inference underlying the objection, I submit that the epistemic thesis is important *regardless* of whether moral character evaluations are made rarely or frequently, because such evaluations sometimes have the following important functions. First, they underlie *esteem* and *contempt* (cf. Sect. 1): sometimes we esteem or despise people because we evaluate them as having a good or bad character.[13] Second, they underlie *praise* and *blame* (cf. Mandelbaum 1955, p. 178): sometimes we praise or blame people not (just) for performing good or bad actions, but (also) for having a good or bad character. Third, they underlie

[13] It is true that esteem is sometimes *non-moral* (I can esteem you as a tennis player) or moral but *partial* (I can esteem you for your courage and still overall despise you), and that such kinds of esteem are not based on moral character evaluations. But it is also true that esteem is sometimes moral and *global* (I can esteem you for your character, as a good person), and my point in the text is that *this* kind of esteem is based on moral character evaluations.

Table 1 Data on the triviality objection

Question no.	1 Never	2 Almost never	3 Rarely	4 Somewhat rarely	5 Somewhat frequently	6 Frequently	7 Very frequently	8 Almost always	No response	Mean response
1	3	20	30	46	*54*	26	5	5	0	4.33
2	0	5	10	25	*53*	*53*	34	8	1	5.45
3	1	5	13	45	*56*	46	21	1	1	5.01
4	0	2	3	19	42	*66*	39	14	4	5.84

explanations of behavior (cf. Sturgeon 1989, p. 443): sometimes we explain good or bad actions by claiming that the agent has a good or bad character. [I am not claiming that such explanations are *accurate*. Similar points can be made about *predictions* of behavior (cf. Upton 2009).] Fourth, they underlie *advice* (cf. the earlier point about recommendation letters): sometimes we advise others to hire or to avoid hiring a person by claiming that the person has a good or bad character. Similarly, sometimes we decide to associate with or to avoid certain people because we evaluate them as having a good or bad character; as a practical example, the official U.S. naturalization requirements include "good moral character" (U.S. Citizenship and Immigration Services 2004). These considerations incidentally bolster my previous point that moral character evaluations are not seldom made, but my present point is different: even if such evaluations are seldom made and thus seldom have the above functions, those functions are so important that the importance of a thesis which purports to undermine them—as the epistemic thesis does—can hardly be denied. I conclude that the triviality objection fails.[14]

4.2 The Objection from Ought-Implies-Can

In stark contrast to the triviality objection, the objection from ought-implies-can relies on the premise that moral character evaluations are *inevitable*: We cannot avoid making them, because—as psychological research on "spontaneous trait inferences" suggests—we routinely make them spontaneously, namely without any intention to make them (and often without even being aware that we have made them). (Cf. Bargh 1989; Uleman 1987; Uleman et al. 1996.) The objection relies also on a second premise, namely a version of the ought-implies-can principle: we have no epistemic obligation to believe (or to avoid believing) what we cannot believe (or cannot avoid believing). From these two premises it follows that we have no epistemic obligation to avoid making moral character evaluations. But according to a third premise, the epistemic thesis entails that we do have an epistemic obligation to avoid making moral character evaluations. It follows that the epistemic thesis is false. In reply, I will contest the first two premises of the objection.

Concerning the first premise, I note for a start that I am not aware of any research supporting the claim that we spontaneously evaluate people in terms of moral *character* (as opposed to character *traits*). Nevertheless, let me accept this claim for the sake of argument, since it sounds plausible that "[w]e look at a person and immediately a certain impression of his character forms itself in us" (Asch 1946, p. 258). Still, the claim that moral character evaluations are *inevitable* is a far cry from the claim that such evaluations are routinely spontaneous. Phobias are also

[14] According to a variant of the triviality objection, the epistemic thesis is uninteresting *for philosophers*, because philosophers (e.g., Plato and Aristotle) believe that virtue is rare (cf. DePaul 2000) and thus seldom make moral character evaluations. I do not see much force in this criticism; as an analogy, even if a novel refutation of creationism is uninteresting *for biologists*, it may still be important if there are many creationists among non-biologists. Moreover, even if philosophers believe that virtuous people in the sense of *moral exemplars* are rare, they need not believe that virtuous people in the sense of *morally good* (as opposed to *morally excellent*) people are rare; and even if philosophers believe that morally good people are rare, they may still believe—especially if they hold a version of the "reciprocity of the virtues" thesis (Irwin 1988, p. 61; cf. Badhwar 1996)—that we can reliably detect the rare good people.

spontaneous, and yet they can often be eliminated (cf. Lindemann 1996; Rachman 1990). Moreover, some evidence suggests that spontaneous trait inferences are "controllable, but usually uncontrolled" (Uleman et al. 1996, p. 234), and according to a review "the evidence shows that automatic processes can be influenced by the perceiver's motives and goals" (Blair 2002, p. 243). Finally, the existence of individual and cultural differences in spontaneous trait inferences (Uleman et al. 1996, pp. 262–266)[15] lends further support to the possibility that we might be able to eliminate spontaneous moral character evaluations. (Eliminating them might be very hard, but who said that "complying" with our epistemic obligations should be easy?) I conclude that the first premise of the objection from ought-implies-can is implausible.[16]

Consider how the second premise, which is a version of the ought-implies-can principle in terms of epistemic obligations. What is an epistemic obligation? If one accepts the third premise, then it is reasonable to understand the claim that we have an epistemic obligation to avoid believing a given proposition (e.g., the proposition that a given person is good) as the claim that, given our evidence, our posterior probability that the proposition is true should be low; in other words, as (something like) the claim that our evidence makes the proposition unlikely. But on this understanding the second premise is false: whether our evidence makes a proposition unlikely does not depend on whether we can avoid believing the proposition. For example, my evidence may make a proposition unlikely even though, due to a neurological defect, I cannot avoid believing the proposition. I can grant that in such a case, in some *other* sense of "epistemic obligation" which makes such obligations depend not only on our evidence but also on our abilities, I have no epistemic obligation to avoid believing the proposition, and a corresponding epistemic version of the ought-implies-can principle is true.[17] But the corresponding version of the third premise is false: if our epistemic obligations depend not only on our evidence but also on our abilities, then the epistemic thesis (which is a claim about our evidence but not about our abilities) does not entail that we have an epistemic obligation to avoid making moral character evaluations. I conclude that the objection from ought-implies-can fails.

[15] On individual differences in moral character evaluations see: Dweck (1996), Dweck et al. (1993, 1995), and Gervey et al. (1999). On cultural differences in trait attributions see: Choi et al. (1999), Doris (2002, pp. 105–106), and Nisbett (2003, Chap. 5).

[16] Moreover, even if *forming* moral character evaluations turns out to be inevitable, I see no reason to suppose that *holding* such evaluations is inevitable. As an analogy, even if in an optical illusion we are unable to avoid forming spontaneously the erroneous belief that one of two lines is longer than the other, we are able to discard this belief once we measure the two lines and find them to be of equal length (although the one may still *look* longer than the other). So a restricted version of the epistemic thesis, in terms of *holding* (rather than forming or holding) moral character evaluations, escapes the objection from ought-implies-can. But I see no need to retreat to such a restricted version even if the first premise of the objection turns out to be true, since in the text I go on to argue that the second premise of the objection is false.

[17] Indeed, I believe I can defend such an epistemic version of the ought-implies-can principle by adapting my published defense (Vranas 2007) of a non-epistemic version of the principle.

4.3 The Objection from Comparative Evaluations

In Sect. 3 I argued that the typical values of cross-situational correlation coefficients are so low that they do *not* justify claims like the following: given that a person behaves admirably (or deplorably) in a given situation, one should be confident that the person also behaves admirably (or deplorably) in any other situation. Ross and Nisbett (1991, pp. 116–118), however, argue that the typical values of cross-situational correlation coefficients, low as they are, do justify claims like the following: given that a first person behaves admirably and a second person behaves deplorably in a given situation, one should be much *more* confident that the first person rather than the second behaves admirably, and that the second person rather than the first behaves deplorably, in any other situation.[18] But this seems to support the claim that *comparative* moral character evaluations, for example to the effect that a first person is *better* (i.e., has a better moral character) than a second, are often epistemically justified, and this claim is the premise of the objection from comparative evaluations. From this premise it seems to follow that *non*comparative moral character evaluations, for example to the effect that a person is good, are sometimes epistemically justified (and thus that the epistemic thesis is false); for example, given that a person is better than the great majority of people, one should be confident that the person is good. In reply, I will contest both the premise and the inference underlying the objection.

Concerning first the premise, for the sake of argument let me grant a claim much stronger than what is supported by Ross and Nisbett's considerations, namely the claim that, given the evidence that a first person behaves admirably and a second person behaves deplorably in a given situation, one is justified in believing that the first person is better than the second. Still, it does not follow that comparative moral character evaluations are *often* (as opposed to *sometimes*) epistemically justified, because we do not *often* have evidence of the above kind. Consider three kinds of situations. (1) In situations propitious to deplorable behavior, as in the obedience experiments (Sect. 2), most people behave deplorably but hardly anyone behaves admirably. (Refraining from administering powerful electric shocks was obligatory, not admirable.) (2) In situations propitious to admirable behavior, as in the theft experiments (Sect. 2), most people behave admirably but hardly anyone behaves deplorably. (Refraining from stopping the "thief" was permissible, not deplorable.) (3) In most everyday life situations, as when a beggar asks for money, hardly anyone behaves admirably or deplorably. (Recall that admirable and deplorable behaviors are *extreme*: they are *highly* praiseworthy and *seriously* blameworthy respectively.) So evidence of the above kind, namely that a first person behaves admirably and a second person behaves deplorably in the same situation, is not often available. (And a perhaps more frequently available kind of evidence, namely that a first person behaves *better* than a second in a small number of situations, has not been shown by Ross and Nisbett to justify the claim that the first person is better than the second.) I see thus no reason to accept the premise of the objection from comparative evaluations.

[18] More formally, letting E be the proposition that p_1 behaves admirably in s_1 and p_2 behaves deplorably in s_2 (in symbols, and with a slight change in notation from Sect. 3.2, E is Ap_1s_1 & Dp_2s_1), the claim is: for any s_2, $P(Ap_1s_2|E)$ should be much higher than $P(Ap_2s_2|E)$ and $P(Dp_2s_2|E)$ should be much higher than $P(Dp_1s_2|E)$.

Concerning now the inference underlying the objection, I question the link from the justifiability of comparative to the justifiability of noncomparative moral character evaluations. Why, for example, accept the claim that one should be confident that a person is good given that the person is better than the great majority of people? The only at least somewhat plausible way I see to support this claim is by deducing it from (something like) the following two claims: (1) given that a person is better than the great majority of people, one should be confident that the person is better than at least one good person, and (2) anyone who is better than at least one good person is good. The second claim seems uncontroversial, but the first claim is questionable. To see the problem with the first claim, consider an example. Suppose that 70% of people (and thus *most* people, in accordance with Sect. 2) are indeterminate, and that 70% of the remaining people (so 21% of all people) are intermediate. Then at most 9% of people are good, and even given that a person is better than 80% (and thus than the *great majority*) of people, one need not be confident that the person is better than at least one good person. The general point is that we just do not know what the percentage of good people is; it may be so small that, even given that a person is better than the great majority of people, one need not be confident that the person is better than at least one good person. In response one might replace in (1) "the great majority" with a percentage very close to 100%, for example 99.9%. But I do not see why we should be confident that even .1% of people are good; and even if we should, the claim that we are justified in believing that a person who is better than 99.9% of people is good is compatible with the epistemic thesis that moral character evaluations are *almost* always epistemically unjustified. I conclude that the objection from comparative evaluations fails.

A variant of the objection starts with the same premise—that comparative moral character evaluations are often epistemically justified—and concludes not that the epistemic thesis is false, but rather that the thesis is uninteresting because comparative evaluations are *all we care about*. (So this is also a kind of triviality objection.) I already have a reply, since I have contested the premise, but in addition I contest the claim that comparative evaluations are all we care about. For example, a recommendation letter that says "this is the best philosopher to have ever graduated from our department" need not move us: the department may be mediocre. But what about a letter that says "this is the best philosopher currently on the market"? Such a letter may indeed move us (if we believe the statement), but this is presumably because we assume that at least some philosophers currently on the market are good. If we learn that this assumption is false, we may well decide not to hire anyone this year, because we may well care about hiring a philosopher who is *good*, not (just) *better* than every other currently available philosopher. Similar points apply to evaluations of people in terms of moral character (rather than philosophical ability or performance).

5 Conclusion

To conclude I would like to suggest two directions for future research.

(1) According to the epistemic thesis, moral character evaluations are almost always *epistemically* unjustified. But are such evaluations *pragmatically* unjustified

as well? Given the important functions of moral character evaluations that I listed in Sect. 4.1 (sometimes such evaluations underlie esteem and contempt, praise and blame, explanations of behavior, and advice), one might argue that for pragmatic reasons we should keep our practice of engaging in such evaluations even if they are almost always epistemically unjustified. In response I believe I can adduce even pragmatic considerations against moral character evaluations; I have sketched such considerations elsewhere (Vranas 2005, p. 30), but elaborating them is a project for a future occasion.

(2) The epistemic thesis is *negative*: we should *not* evaluate people as good, bad, or intermediate (in an epistemic sense of "should"). Following Doris (1996, p. 131, 1998, p. 507, 2002, p. 25), I have adumbrated elsewhere (Vranas 2005, p. 30) the *positive* thesis that we may engage in *local* evaluations of people in light of their behavior in relatively restricted ranges of situations. Formulating this thesis precisely, and defending it by means of a rigorous argument, is another project for a future occasion.

Clearly, the prospects of the above two projects in no way affect the success of my project in the present paper, which was mainly to argue that the epistemic thesis is true.

Acknowledgments My work on this paper was supported by a fellowship from the National Endowment for the Humanities. I am grateful to many people, especially David Brink, Stephen Darwall, Allan Gibbard, and Peter Railton, for help with my doctoral dissertation in philosophy, on a considerably modified part of which this paper is based. Thanks also to Julia Driver for comments, to Gopal Sreenivasan, Chris Tucker, and Jennifer Wright for discussion, and to my mother for typing the bulk of the paper. A version of this paper was presented at a conference on Virtue Ethics and Moral Psychology (University of Denver, October 2005).

References

Asch, S. 1946. Forming impressions of personality. *Journal of Abnormal and Social Psychology* 41: 258–290.

Badhwar, N. 1996. The limited unity of virtue. *Noûs* 30: 306–329.

Bargh, J. 1989. Conditional automaticity: Varieties of automatic influence in social perception and cognition. In *Unintended thought*, ed. J. Uleman, and J. Bargh, 3–51. New York: Guilford Press.

Blair, I. 2002. The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review* 6: 242–261.

Blass, T. 1991. Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology* 60: 398–413.

Choi, I., R. Nisbett, and A. Norenzayan. 1999. Causal attribution across cultures: Variation and universality. *Psychological Bulletin* 125: 47–63.

Clark III, R., and L. Word. 1974. Where is the apathetic bystander? Situational characteristics of the emergency. *Journal of Personality and Social Psychology* 29: 279–287.

Davies, S. 1991. Evidence of character to prove conduct: A reassessment of relevancy. *Criminal Law Bulletin* 27: 504–537.

DePaul, M. 2000. Character traits, virtues, and vices: Are there none? In *Proceedings of the 20th world congress of philosophy: Vol. 9. Philosophy of mind*, ed. B. Elevitch, 141–157. Bowling Green: Philosophy Documentation Center.

Doris, J. 1996. People like us: Morality, psychology, and the fragmentation of character. Doctoral dissertation, University of Michigan.

Doris, J. 1998. Persons, situations, and virtue ethics. *Noûs* 32: 504–530.

Doris, J. 2002. *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press.

Dudycha, G. 1936. An objective study of punctuality in relation to personality and achievement. *Archives of Psychology* 29: 1–53.

Dweck, C. 1996. Implicit theories as organizers of goals and behavior. In *The psychology of action: Linking cognition and motivation to behavior*, ed. P. Gollwitzer, and J. Bargh, 69–90. New York: Guilford Press.

Dweck, C., C. Chiu, and Y. Hong. 1995. Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry* 6: 267–285.

Dweck, C., Y. Hong, and C. Chiu. 1993. Implicit theories: Individual differences in the likelihood and meaning of dispositional inference. *Personality and Social Psychology Bulletin* 19: 644–656.

Epstein, S. 1977. Traits are alive and well. In *Personality at the crossroads: Current issues in interactional psychology*, ed. D. Magnusson, and N. Endler, 83–98. Hillsdale: Erlbaum.

Epstein, S. 1979a. Explorations in personality today and tomorrow: A tribute to Henry A. Murray. *American Psychologist* 34: 649–653.

Epstein, S. 1979b. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology* 37: 1097–1126.

Epstein, S. 1980. The stability of behavior: II. Implications for psychological research. *American Psychologist* 35: 790–806.

Epstein, S. 1983a. Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality* 51: 360–392.

Epstein, S. 1983b. The stability of confusion: A reply to Mischel and Peake. *Psychological Review* 90: 179–184.

Epstein, S. 1984. The stability of behavior across time and situations. In *Personality and the prediction of behavior*, ed. R. Zucker, J. Aronoff, and A. Rabin, 209–268. San Diego: Academic Press.

Epstein, S. 1986. Does aggregation produce spuriously high estimates of behavior stability? *Journal of Personality and Social Psychology* 50: 1199–1210.

Funder, D., and D. Ozer. 1983. Behavior as a function of the situation. *Journal of Personality and Social Psychology* 44: 107–112.

Gervey, B., C. Chiu, Y. Hong, and C. Dweck. 1999. Differential use of person information in decisions about guilt versus innocence: The role of implicit theories. *Personality and Social Psychology Bulletin* 25: 17–27.

Gulliksen, H. 1950. *Theory of mental tests*. New York: Wiley.

Harari, H., O. Harari, and R. White. 1985. The reaction to rape by American male bystanders. *Journal of Social Psychology* 125: 653–658.

Hartshorne, H., and M. May. 1928. *Studies in the nature of character: Vol. 1. Studies in deceit*. New York: Macmillan.

Hartshorne, H., M. May, and J. Maller. 1929. *Studies in the nature of character: Vol. 2. Studies in service and self-control*. New York: Macmillan.

Hartshorne, H., M. May, and F. Shuttleworth. 1930. *Studies in the nature of character: Vol. 3. Studies in the organization of character*. New York: Macmillan.

Irwin, T. 1988. Disunity in the Aristotelian virtues. In *Oxford studies in ancient philosophy: Supplementary volume*, ed. J. Annas, and R. Grimm, 61–78. Oxford: Clarendon Press.

Latané, B., and J. Darley. 1970. *The unresponsive bystander: Why doesn't he help?*. New York: Appleton-Century-Crofts.

Lindemann, C. (ed.) 1996. *Handbook of the treatment of the anxiety disorders*. Northvale: Jason Aronson.

Mandelbaum, M. 1955. *The phenomenology of moral experience*. Glencoe: Free Press.

Meeus, W., and Q. Raaijmakers. 1995. Obedience in modern society: The Utrecht studies. *Journal of Social Issues* 51: 155–175.

Milgram, S. 1974. *Obedience to authority: An experimental view*. New York: Harper and Row.

Mischel, W. 1968. *Personality and assessment*. New York: Wiley.

Mischel, W., and P. Peake. 1982. Beyond déjà vu in the search for cross-situational consistency. *Psychological Review* 89: 730–755.

Modigliani, A., and F. Rochat. 1995. The role of interaction sequences and the timing of resistance in shaping obedience and defiance to authority. *Journal of Social Issues* 51: 107–123.

Moriarty, T. 1975. Crime, commitment, and the responsive bystander: Two field experiments. *Journal of Personality and Social Psychology* 31: 370–376.

Newcomb, T. 1929. *Consistency of certain extrovert-introvert behavior patterns in 51 problem boys*. New York: Columbia University, Teachers College, Bureau of Publications.
Nisbett, R. 1980. The trait construct in lay and professional psychology. In *Retrospections of social psychology*, ed. L. Festinger, 109–130. Oxford: Oxford University Press.
Nisbett, R. 2003. *The geography of thought: How Asians and Westerners think differently … and why*. New York: Free Press.
Peake, P. 1982. Searching for consistency: The Carleton student behavior study. Doctoral dissertation, Stanford University.
Piliavin, J., J. Dovidio, S. Gaertner, and R. Clark III. 1981. *Emergency intervention*. New York: Academic Press.
Rachman, S. 1990. The determinants and treatment of simple phobias. *Advances in Behaviour Research and Therapy* 12: 1–30.
Richard, F., C. Bond Jr., and J. Stokes-Zoota. 2003. One hundred years of social psychology quantitatively described. *Review of General Psychology* 7: 331–363.
Ross, L., and R. Nisbett. 1991. *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.
Ross, L., and E. Thomas. 1986. Notes on regression, aggregation, and the statistics of personal prediction. Unpublished manuscript, Stanford University.
Ross, L., and E. Thomas. 1987. The statistics of cross-situational consistency and behavioral predictability. Unpublished manuscript, Stanford University.
Sarason, I., R. Smith, and E. Diener. 1975. Personality research: Components of variance attributable to the person and the situation. *Journal of Personality and Social Psychology* 32: 199–204.
Sturgeon, N. 1989. Moral explanations. In *Ethical theory: Classical and contemporary readings*, ed. L. Pojman, 437–448. Belmont: Wadsworth. (Originally published 1984.)
Swan Jr., W., and C. Seyle. 2005. Personality psychology's comeback and its emerging symbiosis with social psychology. *Personality and Social Psychology Bulletin* 31: 155–165.
Uleman, J. 1987. Consciousness and control: The case of spontaneous trait inferences. *Personality and Social Psychology Bulletin* 13: 337–354.
Uleman, J., L. Newman, and G. Moskowitz. 1996. People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology* 28: 211–279.
Upton, C. 2009. The structure of character. *The Journal of Ethics* (this issue).
U.S. Citizenship and Immigration Services. 2004. *A guide to naturalization*. Washington, DC: U.S. Citizenship and Immigration Services.
Vranas, P. 2005. The indeterminacy paradox: Character evaluations and human psychology. *Noûs* 39: 1–42.
Vranas, P. 2007. I ought, therefore I can. *Philosophical Studies* 136: 167–216.
Zimbardo, P. 2007. *The Lucifer effect: Understanding how good people turn evil*. New York: Random House.
Zimbardo, P., C. Haney, C. Banks, and D. Jaffe. 1973. The mind is a formidable jailer: A Pirandellian prison. *New York Times Magazine*, April 8, 38–60.