

CHARACTER, COMMON-SENSE, AND EXPERTISE

Jonathan Webber

forthcoming in *Ethical Theory and Moral Practice* (2006 or 2007) – please cite publication

Abstract: Gilbert Harman has argued that the common-sense characterological psychology employed in virtue ethics is rooted not in unbiased observation of close acquaintances, but rather in the “fundamental attribution error”. If this is right, then philosophers cannot rely on their intuitions for insight into characterological psychology, and it might even be that there is no such thing as character. This supports the idea, urged by John Doris and Stephen Stich, that we should rely exclusively on experimental psychology for our explanations of behaviour. The purported “fundamental attribution error” cannot play the explanatory role required of it, however, and anyway there is no experimental evidence that we make such an error. It is true that trait-attribution often goes wrong, but this is best explained by a set of difficulties that beset the explanation of other people’s behaviour, difficulties that become less acute the better we know the agent. This explanation allows that we can gain genuine insight into character on the basis of our intuitions, though claims about the actual distribution of particular traits and the correlations between them must be based on more objective data.

1. INTRODUCTION

The idea that we should concern ourselves with developing good character is common in ethical discourse and proclaimed from a wide variety of meta-ethical positions. If moral goodness is primarily a matter of good character, of course, then it seems obvious that this should be our concern. If moral value attaches primarily to actions in respect of the intentions behind them or the consequences they have, on the other hand, then the development of good character might nevertheless be the best way to promote good action (see Nussbaum 1999, esp. § I; Trianosky 1997, esp. § 3). Philosophers from these different meta-ethical perspectives recommend fostering a range of character traits.

But there is a growing dissatisfaction with this consensus, rooted in a concern that the psychological picture involved has been arrived at by a flawed methodology. Philosophical talk of character should be grounded in the findings of experimental psychology, critics argue, but is instead usually based ultimately only on common-sense intuitions. Gilbert Harman, for example, claims that “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 and vices” (1999, p. 316). “Far too many moral philosophers have been content to *invent* the psychology or anthropology on which their theories depend”, write John Doris and Stephen Stich (2005, p. 114).

Doris and Stich, of course, are being provocative: they know very well that the philosophers they have in mind would claim that the psychology involved in their theories is not simply invented, but is rather rooted in common-sense intuitions that are themselves ultimately based on observation of the patterns in people’s behaviour. As well as one’s own observations, one’s intuitions are informed by reading other people’s observations in narrative and other forms, and of course by understanding the language that has gradually developed to reflect these observations.

But the challenge presented by Harman, Doris, and Stich is clear nonetheless: to show that the empirical grounding of the intuitive understanding of character involved in this area of ethical discourse is respectable. There are two ways in which philosophers have argued that it is not. One is by arguing that various aspects of the conception of character employed in philosophical ethics are incompatible with certain experimental data (Doris and Stich 2005, § II; see also Doris 1998, § I; Doris 2002, chs. 2-3; Harman 1999, pp. 316, 325-6). In response to this claim, philosophers have generally argued that it misconstrues the characterological claims that moral philosophers typically make and that these are perfectly compatible with the data cited (e.g. Kamtekar 2004; Webber 2006).

A second kind of argument against the empirical respectability of characterological moral philosophy, however, has not yet met with any response. Interwoven with his version of the first kind of argument, Harman has also presented a directly epistemological challenge to the intuitive conception of character, a challenge also grounded in empirical psychology. Research into trait-attribution has shown, according to this argument, that our ordinary characterological understanding of the people around us arises from a misleading heuristic and is consolidated by a cognitive bias. It is not simply our ability to correctly identify the character traits of those around us that is threatened by this point: the very idea of the nature,

interrelation, and development of character traits embodied in our intuitive characterological understanding of those around us is equally threatened.

If our characterological intuitions are necessarily mistaken, as Harman points out, then it may even be that there is no such thing as character as we intuitively understand it. We should therefore stop using this notion in philosophical ethical discourse, unless experimental data can show that there is indeed such a thing as character as we traditionally conceive of it. The traditional philosophical employment of intuitions, thought-experiments, and literary narratives to ground claims about character should be abandoned in favour of exclusive reliance on experimental psychology, as Doris and Stich urge.

We will see, however, that Harman is wrong to claim that our characterological intuitions are necessarily mistaken. We often make erroneous attributions of specific character traits, of course, but this fact is not best explained by the idea that common-sense trait-attribution is necessarily misleading. It is best explained by the contingent difficulties that often beset the attribution of traits, particularly their attribution to strangers and loose acquaintances. This allows that our attributions of traits to people we have known for quite some time can indeed be warranted, and that the common-sense understanding of the nature of character based on such acquaintance is also warranted. Ethical discourse need not rely exclusively on experimental data for its psychological claims.

2. EXPERIENCE AND EXPERIMENT

The characterological psychological claims involved in philosophical ethical discourse are generally concerned with the nature of character traits, the ways in which they develop, the ways in which they can conflict or harmonise, and the advantages and disadvantages of particular traits. In order to recommend that we develop certain traits and not others, these are all the aspects of character that philosophers need to be concerned with. Since their discipline is normative rather than descriptive, they need not be concerned with the actual distributions of particular traits or the actual correlations between different traits among populations. They may have such concerns, since they may recommend traits that are useful only because certain other traits are prevalent among the surrounding population, but this is not at the heart of the characterological ethical project, and indeed is incompatible with the traditional aim of recommending traits for everyone to adopt.

Philosophers have often understood claims about the nature, combination, and relative utilities of traits to be justified with reference to intuitions grounded ultimately in our accumulated experience of trying to explain and predict the behaviour of those around us. Experimental psychology is not often referred to in characterological ethical discourse, and the reason for this is that our own longitudinal acquaintance with our family, friends, colleagues, and neighbours is generally understood to provide all the psychology required. Some judgements about the nature and value of character traits may be mistaken, of course, but this is to be explained by their being made without sufficiently careful consideration. Intuition grounded in experience can be our guide.

Some philosophers have gone further and argued that reliance on intuitions grounded in longitudinal acquaintance with a few individuals over significant stretches of time is actually superior to the use of experimental psychology in this area (e.g. Kupperman 1991, pp. 162-4). This is because most psychological experimentation is latitudinal rather than longitudinal: it studies a number of people in a given situation, not a given person in a number of situations. The relevant kind of longitudinal study is rare, for both logistical and ethical reasons: professional psychologists need to be able to publish results frequently, the funding systems available to them often reflect this aim, and subjects are difficult to track across sustained periods of time; character is revealed most not by what people say but by what they do, particularly by what they do without being aware that psychologists are watching them, so longitudinal experiments would ideally require long-term secret surveillance of the public and private lives of unwitting subjects. If longitudinal information is required for understanding the nature, structure, and development of character, according to this argument, then this understanding will have to be grounded in informal longitudinal experience of particular individuals. This experience can be our own, but we can also gain characterological insights from the writings of others, particularly if they are unusually acute observers of human behaviour.

Harman's epistemological argument runs directly counter to this idea. He draws on experimental data to argue that our intuitions about the nature of character are not based on informal and unbiased observation, but are rather rooted in a misleading heuristic that he follows Lee Ross in calling the "fundamental attribution error" (Harman 1999; Ross 1977). His evidence of this is that the heuristic is the best explanation of our tendency, repeatedly demonstrated in a wide variety of experiments since the 1940s, to attribute a person's behaviour to uncommon traits had by that person, even when the person in fact behaved as most people would behave in that situation. He further suggests that our failure to recognise

the misleading nature of our trait-attributions is due to a second problem, known as “confirmation bias”: when we want to know whether an attribution is correct, we do not consider the person’s behaviour as a whole but look only for actions that would be consistent with the attribution, ignoring those that are not (1999, p. 325).

If this is right, then we cannot accept the idea that intuitions provide a reliable guide in this area at all, never mind one that is superior to experimental psychology. We must rather accept that our whole intuitive characterological understanding is based on an error and reinforced by a bias. Ethical theories motivated partly by intuitive characterological understanding are vitiated not by the obvious fact that we are often mistaken about which traits people in fact have, on this view, but rather by the purported fact that we could very well be mistaken about the existence and nature of traits altogether. Harman’s argument aims to impugn our entire set of intuitions about the explanation of behaviour in terms of traits that have a certain kind of structure, combine in certain ways, and develop in certain ways. If this argument is successful, it shows not that there is no such thing as character, but that our intuitive commitment to the reality of character should be ignored and issues concerning behavioural explanation should be settled entirely by reference to experimental psychology.

If Harman is right, therefore, the fact that the acceptability of characterological explanation just seems obvious to most people can be explained in a way that is compatible with there being no such thing as character as traditionally understood. We should not rely on our intuitions. But if Harman’s argument is mistaken, then the abandonment of reliance on our intuitions in favour of exclusive reliance on the findings of experimental psychology could well be inimical to ethical discourse. It would be the abandonment of a perfectly good source of insight into the relevant area of psychology, and its replacement with a methodology that seems to many to be poorly suited to uncovering the nature, structure, and development of character traits.

3. ERRONEOUS ATTRIBUTION

The issue hinges on the acceptability of Harman’s interpretation of the extensive experimental literature that shows much trait-attribution to be erroneous. In order to judge the acceptability of this interpretation, we need to clarify exactly what that literature reports. Stanley Milgram’s famous experiments concerning reactions to authority provide a good illustration, and one that Harman employs. The subject of the experiment is asked to administer a memory

test to someone the subject believes to be another volunteer, but in fact is not. Each time this 'learner' gives the wrong answer or no answer at all, the subject is to deliver an electric shock. These seem to start at 15 volts, and increase by 15 volts each time. In fact, of course, there are no shocks, but the behaviour of the 'learner' makes it seem as though there are: he responds from 75 volts upwards, first by groaning, then by complaining that they are becoming painful, then by refusing to go on, screaming, and remaining silent after 330 volts. The subject is instructed by the 'experimenter', a man wearing a technician's coat and holding a clipboard. If the subject questions the procedure, the 'experimenter' politely responds in ways that encourage compliance. The experiment ends either when the subject questions the procedure for the fifth time, or when the shock level has reached its maximum of 450 volts.

Milgram asked groups of psychiatrists, academic staff and graduate students in behavioural sciences, college sophomores, and middle-class adults to predict the results of this experiment were it to be performed on one hundred Americans of diverse ages and occupations. The various groups responded with remarkably similar answers: they predicted that only a pathological minority of one or two per cent would reach the maximum shock, that almost everyone would have refused to comply before reaching 300 volts, and that most would not go beyond 150 volts, when the 'learner' first explicitly requests that the experiment end. This experiment has been performed many times, however, and on average around sixty-five percent of subjects continue to administer the shocks all the way up to 450 volts, the majority go beyond 300, and almost all reach 150. (Summarised from Milgram 1974, chs. 2, 3, 4, and 6.)

The difference between the actual results and those commonly predicted is due to the tendency of those making the predictions to assume that only people with little or no regard for the pain of others would obey the 'experimenter', and to assume that such people are rare. They seem to overlook the possibility that the man in the technician's coat will command most people's obedience, or perhaps that most people will defer to his expertise, even when they are inclined against the actions he requires. When people try to predict the results of this experiment, they therefore think of the subjects who follow the instructions as displaying some unusual psychological trait, such as cruelty, rather than as responding to some situational feature that strongly influences most people. It is this tendency to exaggerate individual character differences that has been noticed in a wide variety of contexts and discussions of its nature have been central to social psychology since its inception in the 1940s (see Gilbert and Malone, 1995, pp. 22-24.)

The terms in which this tendency is discussed can be misleading. It is often claimed that we overemphasise dispositions or traits and underemphasise situations. This leads to the objection that these so-called situational forces cannot influence behaviour directly, but only in concert with the subject's dispositions to respond to them in certain ways. Milgram's experiments do not reveal that there are no dispositions, the objection runs, only that we are commonly very strongly disposed towards obedience or deference (e.g. Athanassoulis, 2000, p. 217).¹ Social psychologists, of course, are not making so elementary a mistake as to assume that situational forces such as authority influence behaviour directly. They are not trying to distinguish actions that manifest dispositions from those that respond to situations.

Their language rather reflects the view that the *relevant* explanation of an agent's behaviour refers to the agent's dispositions only where these are uncommon and characteristic of the agent, but refers to the situational feature to which they are responding when it is one to which people generally respond. To say 'she ran away because there was a lion on the loose' is to give a perfectly good explanation even without mentioning her fear of lions, whereas saying 'he ran away because there were buttercups' is not, unless it is added or already understood that he has an unusual and powerful aversion to buttercups. For this reason, social psychologists have come to describe explanations referring to uncommon traits as 'dispositional', those referring to or implying common traits as 'situational'. An explanation of the behaviour of the subjects of the Milgram experiment in terms of obedience or deference is therefore 'situational', whereas one in terms of cruelty is 'dispositional'.

Despite its name, therefore, a 'situational' explanation therefore need not refer directly to a feature of the immediate situation. "One cannot see, smell, taste, or hear 'audience pressure', which exists only in the mind of the public speaker", for example, and such 'situational' factors as social norms and parental threats are "temporally or spatially removed from the behavioural episodes they constrain" (Gilbert and Malone, 1995, p. 25). A 'dispositional' explanation, on the other hand, will refer the trait of having what is considered to be an uncommon motivation, such as cruelty, or to the uncommon trait of possessing a

¹ Milgram himself describes his results in terms of a disposition towards obedience (1974, pp. 1-2 and 42-3). Some thinkers find it implausible to postulate a widespread tendency towards obedience, since people clearly disobey rules all the time in our society. Perhaps we can explain this by saying that people disobey when they think they will not be detected, a condition that does not hold in Milgram's experiment. But even if this response is unacceptable, we could still agree with John Sabini and Maury Silver (2005, pp. 550-1) that the subjects' behaviour is to be explained in terms of the character trait of deference to expertise.

common motivation to an unusually high degree, such as being extremely generous or exceptionally honest. We should understand the terms ‘situational’ and ‘dispositional’ in these technical senses, and take care not to confuse these with their ordinary senses.

Harman is aware of this. He correctly explains the data as showing that “ordinary observers wrongly infer that actions are due to distinctive character traits” (1999, 323). On the basis of this observation, he provides two arguments against the use of characterological explanation. One is based on the claim that if people did have traits as traditionally construed, and as construed by most philosophers concerned with character, then we would expect subjects in the experiments he discusses to behave in differing ways, reflecting their differing, distinctive, characteristic traits. Since we instead find a striking uniformity of behaviour, he argues, experiments that might have supported the idea that behavioural differences between people are due to differences in their dispositions have found no evidence that this is the case and rather suggest that they are due to differences in their situations (1999, pp. 316, 325-6).

If this is right, he argues, then moral philosophy should embrace the idea that behaviour is best explained and predicted by reference to situational features rather than dispositional ones (1999, pp. 324-330; see also Harman 2000, p. 223). This argument has met with the response that the data Harman cites can only impinge on *empirical* claims about the distribution of character traits, and hence does not impinge on *ethical* claims about the traits we ought to strive to develop (Athanasoulis 2000; Kamtekar 2004). It might be added (adapting Webber 2006) that the data is perfectly compatible with the idea that behaviour issues from character traits, and that further data provides positive evidence in favour of this idea, so long as this idea is correctly understood. A third response might question the premise that common-sense expects varied responses to the situation: the people interviewed by Milgram, after all, did predict a somewhat uniform response, as we have seen.

But this aspect of Harman’s discussion is not our concern here. We are concerned with his second, directly epistemological argument. This is grounded in his preferred explanation of our tendency to explain behaviour in terms of uncommon traits. Decades of empirical research has indeed shown that we have this tendency, and if Harman’s interpretation of this data is right then, as we will see in more detail in the next section, we should reject the use of characterological intuitions in the explanation of behaviour, and rely instead solely on experimental psychology.

4. HARMAN'S INTERPRETATION

The idea that common-sense trait-attribution is often mistaken should certainly give philosophers who rely on it pause for thought, but whether it impugns our intuitive characterological understanding of behaviour depends on exactly how and why trait-attribution tends to go wrong. While social psychologists have agreed for decades that we have a misleading tendency to attribute distinguishing characteristics to people rather than explain their behaviour in terms of common traits, they have provided over that time a wide variety of explanations of this phenomenon. This tendency of ours has been described as “something of a stray puppy that no one could quite get rid of but whose owner no one could seem to track down” (Gilbert and Malone, 1995, p. 24). Harman emphasises one strand of thought about this tendency that has featured in various theories attempting to explain it:

“Where we distinguish figure from ground, we pay more attention to figure and less to ground and we try to explain what happens in terms of features of the figure rather than features of the ground. Typically, the actor is figure and the situation is ground, so we seek an explanation of the action in features of the actor in the foreground rather than in features of the background situation” (1999, p. 325).

This interpretation is unacceptable, as we will see in section 5, because the purported tendency to explain events in terms of features of the figure rather than the ground simply cannot do the explanatory work required of it, and because there is anyway no good experimental evidence correlating the relative salience of the agent with the perceived importance of dispositional factors in explaining the agent's behaviour. But first, it is important to see why this idea about the salience of the agent, were it correct, should lead us to question our ideas about the nature, structure, and development of character, and to abandon the use of our intuitions as evidence in favour of these ideas.

To see this, consider a different moral one might draw from the data. It could be taken to show not that characterological explanation is itself mistaken, but that the generalisations that we make about people's characters are often mistaken. The fact that the results of Milgram's experiment are surprising could be taken to show that we generally assume that the demands of compassion will weigh more heavily with most people than will the demands of obedience or deference to the experimenter, and that we are wrong to assume this

(Athanasoulis, 2000, pp. 217-8). If this interpretation is right, then the data shows only that our intuitions do not provide a good guide to the distribution of particular traits across the population. This would not be very surprising: our understanding of character is presumably based on intimate longitudinal acquaintance with very few individuals and a much less intimate acquaintance with more people, but still a very small sample, and this should not be expected to provide a good guide to the population.

If this alternative interpretation is correct, then we should rely on latitudinal experimental data for information about the distribution of traits, but can retain our intuitive general understanding of the nature and development of character and of the relative values of different traits. Peter Goldie interprets the data in this way. He takes it to show that we are too ready to ascribe traits on the basis of meagre evidence, but that we do not make such mistakes when we have ample evidence. We make false assumptions about people we do not know, or do not know well, but nonetheless develop a more nuanced and precise picture of individuals the more we observe their behaviour (Goldie 2000, p. 166; 2004, pp. 52-69). If this is right, then informal longitudinal acquaintance could still be a reliable source of insights into the nature of character.

Harman's interpretation of the data leads to the more radical conclusion that even our understanding of those closest to us is likely to be mistaken. Longitudinal acquaintance does not lead to detailed understanding of the causes of an individual's behaviour, but to an increasingly complicated illusion. Our tendency to focus on the salient agent rather than the less salient situational features leads us to construct complicated distinctive characters to explain the behavioural differences between those closest to us, when in fact those differences are due to the differences between the situations in which they find themselves.

This is not just to say that we might think that those we know well are more distinctive than they in fact are, but to say that we might be wrong in thinking that they possess any character traits at all. Perhaps proper consideration of the situational factors involved in their behaviour will lead us to abandon the whole project of explaining their behaviour in terms of character altogether. This interpretation therefore leads to the wholesale rejection of common-sense as a grounding for characterological psychology, as Harman rightly points out, rather than to the restrictions on the use of common-sense required by the less radical interpretation.

Goldie (2000, p. 166) argues against Harman's interpretation on the grounds that our ability to predict someone's behaviour increases the more time we spend with that person. This is best explained by our increasing understanding of that person's character, he claims. But Harman's position is easily defended from this criticism: this increase can be explained

equally well by our increasing knowledge of the details of the kinds of situations our acquaintance typically acts in, even if we are unaware that this is what we increasingly know. When someone's behaviour surprises us, on this account, this is because some situational feature is novel or unusual for that person, or at least is so in our experience, and not because of an abrupt alteration of that person's character. Goldie is right that any acceptable interpretation of the data must be compatible with the fact that we seem to get better at predicting a person's behaviour the more time we spend with that person. But Harman's position is compatible with that fact.

5. SALIENCE AND ATTRIBUTION

There are better reasons to reject the interpretation that Harman endorses. This interpretation relies, as we have seen, on the ideas that the agent is more salient than the situation, is figure to its ground, and that we tend to explain events in terms of properties of the more salient figure. These two points together are supposed to explain our tendency to attribute character traits when explaining behaviour. As an interpretation of the data on erroneous attribution this is simply unacceptable, for two distinct reasons.

The first and simplest is that it has lost sight of the tendency in need of explanation. As we have seen, the relevant literature shows not that we cite character traits in explanations when behaviour is not due to character traits, but rather that we have a misleading tendency to cite *uncommon* traits *characteristic* of the agent rather than more common dispositions that do not mark that agent out. In the case of the Milgram experiment, what needs to be explained is why people tend to think that anyone reaching the maximum shock level must be unusually cruel or lacking in compassion rather than ordinarily obedient or deferential. To say that we naturally explain behaviour in terms of properties of agents does not answer this question, since it does not explain our apparent preference for uncommon rather than common properties. It does not explain why we tend to think, in advance of knowing the actual data, that only a pathologically cruel minority would continue to the maximum shock level in Milgram's experiment, because it is perfectly compatible with our considering those who would do this to have a common trait of obedience or deference.

Harman's preferred interpretation of the data, therefore, could only explain a preference for dispositional over situational explanations if we understand these terms in their ordinary sense, rather than in the technical sense outlined in section 3. What is required is an

interpretation that explains our preference for dispositional explanations in the technical sense, our preference for citing uncommon character traits. In the next section, we will see that there is an interpretation available that succeeds in explaining this, and which is therefore superior to Harman's. That alternative explanation, moreover, does not require us to abandon intuition as a source of insights into the nature, structure, and development of character.

The second reason why Harman's interpretation is unacceptable is that there simply is no significant evidence in favour of the idea that we tend to explain behaviour in terms of properties of the salient figure rather than the less salient ground. Harman supports his acceptance of this idea by referring to a well-known social psychology textbook written by Lee Ross and Richard Nisbett, *The Person and The Situation* (1991). A very brief passage of this book cites three experiments in favour of the idea that "what you *attend* to is what you *attribute* to" (p. 140). Harman does not go on to consider these experiments himself. As we shall see, they do not provide the evidence that Ross and Nisbett proclaim.

One of these studies, write Ross and Nisbett, "showed that an actor's behaviour was attributed less to his environment when the environment was stable than it was when it was in motion" (1991, p. 140). This is somewhat tendentious. The agent chose an artwork — a black-and-white etching — from a selection presented either by photographs on a display board or by a videotape that panned across each work, and both agents and observers were asked to rate the impact of "the situation (e.g. the lighting, the laboratory equipment, the method of presentation)" on the decision-making, and to rate separately the impact of dispositional factors (Arkin and Duval, 1975, p. 432). The experiment did show that both agents and observers gave higher mean scores for the relevance of the cited situational factors when the artwork was chosen from a video than when it was chosen from a static presentation (p. 434). But an analysis of the ratings of the relevance of dispositional factors in the two situations "revealed no significant main effects or interactions" (p. 434). For this reason, Daniel Gilbert and Patrick Malone write that this experiment "is not directly relevant to the salience explanation" of the tendency towards dispositional attribution (1995, 31).

Ross and Nisbett do not cite it as direct evidence of this claim, of course, but as evidence that situational attribution is more likely when the situation is more salient, which they intend to support the more general idea that attribution tracks salience. Gilbert and Malone go on to argue that the experiment does not show this, because in order to do so the dispositional attributions would have to decline as the agent's salience decreased in proportion to the increased salience of the environment (1995, 31). In this, Gilbert and Malone are mistaken. There is no reason why salience need be relative in this way. We could

instead accept that in one condition only the agent was salient, whereas in the other both agent and environment were salient, as indeed the experimenters point out (Arkin and Duval, 1975, p. 430).

It does not seem, however, that the salience of the environment was manipulated at all. When Ross and Nisbett describe the experiment as involving the environment being either “stable” or “in motion”, they distort the experiment significantly. Most of the environment was stable in both conditions. The only difference was that the agent was watching moving pictures on a television screen in one case, looking at a display board in the other. Why should a television screen that the agent is clearly watching be more salient to an observer than a display board that the agent is clearly inspecting? Indeed, why should the screen also be more salient *to the agent* than the display board?

There is available a perfectly good explanation of the data that has nothing to do with salience. In one condition, the agents were free to inspect the artworks in any way they chose, whereas in the other the information they could glean about the artworks, and the order in which they could glean that information, was determined by the content of the videotape. The fact that both agent and observer thought that the mode of display had more influence over the outcome when it was a video than when it was a display board can be explained by this difference: the video constrained the information available to the agent in ways that the display board did not.

Ross and Nisbett say no more about their citation of this experiment as evidence of their view than is quoted above. They simply report the experimenters’ own conclusion (Arkin and Duval, 1975, p. 434), a conclusion which we have seen to be unwarranted. The authors of the other two studies Ross and Nisbett cite, however, draw justified conclusions that directly contradict the contention they are cited as evidence for. Ross and Nisbett say nothing in support of their divergent reading of these studies, and we will see that their reading is mistaken.

One is the study they describe as showing that “when an observer watches actors A and B interact but can see A better than B, causal attributions about the outcome are made more to A than to B” (1991, p. 140). This involved a conversation between agents A and B with observers seated either behind A with a clear view of B, behind B with a clear view of A, or to the side with a clear view of both. Observers were asked to rate the relative causal contributions each agent made to the tone and direction of the conversation, and this study did indeed find that those behind A attributed more causality to B than to A, those behind B attributed more causality to A than to B, and those to the side attributed causality equally to

the two participants (Taylor and Fiske, 1975, pp. 441-2). This does indeed support the general claim that Ross and Nisbett make, that “what you *attend* to is what you *attribute* to” (1991, p. 140). But Ross and Nisbett intend this general claim to support the more specific claim that attending to a particular individual leads to explaining an event in terms of the *dispositions* of that person.

The study in question did ask further questions about whether each agent’s causal contribution to the conversation reflected such dispositions as talkativeness, friendliness, and nervousness. The study found that “there were no significant effects or trends” in relation to these questions (Taylor and Fiske, 1975, p. 442). The experimenters emphasise that “although viewpoint markedly influenced who was seen as the causal agent in the interaction, this bias did not extend to an interpretation of why each agent behaved as he did” (p. 442). A variation, reported in the same article, involved some subjects being explicitly instructed to attend to a particular participant, and found that these subjects were “no more likely to see his behaviour as dispositionally based than were subjects who were not told to attend to any participant in particular” (p. 443). This study cited by Ross and Nisbett in favour of their general claim, therefore, actually presents evidence against the more specific claim that they intend the general claim to support.

Ross and Nisbett cite one more study, which they describe as having “found that the actor’s behaviour was attributed less to his situation when he was brightly illuminated or moving than it was when he was poorly illuminated or stationary” (1991, p. 140). This is again a somewhat selective presentation of the results. The article they refer to reports five experiments. In each experiment, observers were shown video recordings of two agents in conversation. In the first experiment, one agent was more brightly illuminated than the other, and in the second, one agent was in a rocking chair while the other was still. Observers of these experiments were asked to rate the explanation of each agent’s behaviour on a single scale with dispositional factors at one end and situational factors at the other. The results were as Ross and Nisbett report.

In the third experiment, one actor’s shirt was boldly patterned where the other’s was plain. In the fourth and fifth experiments, the agents were accompanied by two other silent individuals. In the fourth experiment, one agent wore a shirt different in colour from the other three people. In the fifth, one agent was of a different sex to the other three. In these three experiments, observers were asked to rate the importance to each agent’s behaviour of dispositional factors and situational factors separately, rather than on a single scale. The third experiment found that the increased salience of one agent led to a marginally significant

decrease in situational attribution, but no significant alteration in the dispositional attribution, for that agent's behaviour. Experiments four and five seem to provide evidence for the reverse of the general claim made by Ross and Nisbett: the *salient* agent's behaviour was attributed more to *situational* factors than the non-salient agent's, though again there was no significant alteration in dispositional attribution.

If we focus only on dispositional attribution, then these five experiments do not seem consistent, so the study overall seems to show that there is no conclusive evidence of the relation between salience and attribution. If we focus on situational attribution, however, we find consistency. Notice that in the fourth and fifth experiments, the fact that three people are wearing the same shirt or are of the same sex makes that group as a whole more salient (as well as making the fourth person more salient) than if all four had been wearing the same shirt or were of the same sex. Increased salience of the agent, in experiments four and five, is accompanied by increased salience of the environment. Situational attribution therefore increases with the salience of the environment in these experiments, as the experimenters point out (McArthur and Post, 1977, p. 528).

This could be taken as evidence in favour of the general claim Ross and Nisbett make, that “what you *attend* to is what you *attribute* to” (1991, p. 140). But the experiments do not present evidence in favour of their more specific claim, endorsed by Harman, that the salience of the agent is responsible for dispositional attributions. The only reason why the first two experiments of this study, the only two cited by Ross and Nisbett, appear to support that more restricted claim is that observers were asked to rate the relative importance of dispositional and situational factors on a single scale, so the increase in situational attribution was necessarily accompanied by a decrease in dispositional attribution. The other three experiments, which use separate scales, show that situational attribution can increase without a corresponding decrease in dispositional attribution. As the experimenters put it, this shows that “situational and dispositional attributions are not psychological reciprocals of one another” (p. 530). We should discount the apparent evidence of the first two experiments in favour of dispositional attribution tracking salience, therefore. The experimenters are right to conclude that “being physically conspicuous or responding to relatively inconspicuous environmental cues does not seem sufficient to have a significant influence on attributions of behaviour to dispositional causes” (McArthur and Post, 1977, p. 534).

The three studies Ross and Nisbett cite in support, ultimately, of the idea that dispositional attribution is a result of agent salience do not support this idea, then, and indeed the authors of two of these three studies explicitly and rightly say so. The above analysis of

these experiments, moreover, brings to light a deeper problem with Harman's implicit reliance on them as evidence in favour of his interpretation of the data on erroneous attribution. That data shows that we tend to cite uncommon traits in behavioural explanations rather than citing common ones or the situational features to which common ones respond. Common and uncommon traits are, you might say, "psychological reciprocals" of one another. Why postulate an uncommon trait like cruelty of some unknown person if we have already ascribed to them a common one like obedience that explains the behaviour in question?

So it seems that the last study discussed cannot be tracking the kind of situational and dispositional attributions relevant to the data on erroneous attribution. The same can be said for the study involving agents choosing artworks: the increase in situational attribution is not accompanied by a decrease in dispositional attribution in that study either, as we have seen. Indeed, the definitions of 'situational' factors given to the subjects in these studies do not fit the technical definition outlined in section 3. In the study that asked observers about the behaviour of agents in conversation, 'situational' factors were defined as "factors such as being in an experiment, the getting acquainted situation, the topics of conversation, and the way the other participant behaved" (McArthur and Post, 1977, p. 523). In the study concerned in which agents were choosing artworks, as we have seen, subjects were asked about the impact of "the situation (e.g. the lighting, the laboratory equipment, the method of presentation)" on the decision-making (Arkin and Duval, 1975, p. 432). Even if these experiments had shown that the salience of the agent underlies a preference for dispositional as opposed to situational explanations of that agent's behaviour in the ordinary sense of these words, therefore, they would not have shown that salience explains our preference for uncommon traits rather than common ones: they would not have explained the data on erroneous attribution.

6. THE ROOTS OF ERRONEOUS ATTRIBUTION

Harman's preferred interpretation of the data on erroneous attribution is therefore unacceptable. The data shows that we tend to attribute uncommon traits rather than common ones. Harman's preferred explanation cannot explain this, but could explain at most why we prefer to explain behaviour in terms of properties of the agent rather than properties of the agent's environment. The evidence offered in favour of that interpretation also only

investigates this related but distinct claim, and anyway does not find in favour of it. So there does not seem to be good reason to believe that trait-attribution is grounded in the purported fundamental attribution error.

The data on erroneous attribution is itself robust, however. A wide range of experimental data, including Milgram's much-replicated experiment, strongly supports the idea that we often make mistakes when understanding or predicting people's behaviour in terms of character traits. If the traditional philosophical reliance on intuitions in understanding the nature, integration, structure, development, and value of character traits is justified, then this data on erroneous attribution needs to be interpreted and explained in a way that does not impugn all of the common-sense understanding of character those intuitions express. It remains to be shown, that is, that the data on erroneous attribution is no threat to the claim that our understanding of character itself — rather than our understanding of the characters of particular individuals — is grounded ultimately in observation rather than in erroneous heuristics and biases. This requires that the data on erroneous attribution is compatible with the idea that we can observe the behaviour of close longitudinal acquaintances in ways that avoid the errors manifested in that data.

The interpretation offered by Gilbert and Malone (1995) allows for this, and has the additional advantage of explaining why we become better at predicting a person's behaviour the more time we spend with that person. The first step Gilbert and Malone take is to abandon the assumption pervasive in earlier literature, and clearly present in the interpretation Harman favours, that all of the relevant data should be explained as the effect of a single cognitive tendency. Instead, they provide four basic problems that beset trait-attribution, and argue that any given instance of erroneous trait-attribution results from one or more of these problems.

The first problem is simply that the observer may not be aware of the relevant situational features. This problem has two parts. First, some aspects of the situation may be hidden from the observer, and these may include features that would influence most people to behave in a certain way. We have already seen that relevant 'situational' features might include audience pressure, social norms, and parental threats. Other relevant features of the situation may not be strictly invisible, but might escape notice nonetheless "because the cues that evoke behaviour are both subtle and powerful" (p. 25). In a variation of Milgram's experiment in which the man in the technician's coat was replaced with someone who seemed to be another volunteer, the percentage of people reaching 450 volts was not sixty-five but twenty (Milgram, 1974, pp. 93-97). The technician's coat and clipboard therefore confer a

certain authority to which many people respond, but their significance need not be noticed by the observer.

The second part of this problem concerns the agent's view of the situation. The subjects of Milgram's experiment labour under various constraints, some of which alter their construal of the behavioural options available. They are not prevented from walking out of the experiment at any time, and their protests are met only with calm and polite requests to continue. Yet these requests can make it seem to them that obedience or deference will engender the good will of the experimenter and defiance may be humiliating. In order to fully understand the subject's situation, therefore, we need to understand how the subject construes that situation, not how we ourselves construe it, and this might be very difficult in many cases. There is strong evidence to suggest that observers tend to assume that the way they see the situation is the way the agent sees it. This tendency prevents observers from understanding which situational features the agent is responding to, and hence from judging accurately whether responding to those features reflects an uncommon or a common disposition (Gilbert and Malone, 1995, pp. 26-7).

The second problem is that observers may have unrealistic expectations of the behaviour of individuals in a given situation. This can operate independently of the first problem, or in conjunction with it. Our expectations are usually grounded in our limited experience of those around us or even in how we imagine we ourselves would respond. Whether we rely on our limited experience or on our imagination, this 'availability heuristic' is useless as a guide to the distribution of character traits and the correlations between them across the population. We are therefore very poor at judging the probability that an unknown agent will behave one way rather than another. Behaviour that matches our expectation will thereby be taken to reflect a common trait, behaviour that does not match it will be taken to reflect an uncommon trait. It is because people like to think that they themselves would refuse to issue electric shocks to a complaining volunteer that they mistakenly expect Milgram's subjects to do likewise, and for this reason assume that issuing the shocks would indicate an uncommon trait like cruelty rather than a common one like obedience or deference (see Gilbert and Malone, 1995, pp. 27-8).

The third problem is that one's expectations might unduly influence one's perception of the observed behaviour. This problem can operate independently of the previous two, or in conjunction with either or both. Consider the classic study, much discussed in psychological literature on attribution, in which subjects were shown essays that either supported or opposed the Cuban president, Fidel Castro. They were told either that the author was free to choose

which position to take, or that the author had been instructed to defend a particular position as an exercise in debate training. Subjects told that the author had chosen which position to defend tended to infer from the essays that their authors had strong pro- or anti-Castro attitudes. Subjects told that the author was not free to make this choice still tended to infer pro- or anti-Castro attitudes, though weaker ones (Jones and Harris, 1967). Why do subjects assume that the second kind of speech manifests its author's attitudes? Having assumed that the instructions of the debate coach would be followed, these subjects expected to find certain sentiments in the speech, which caused them to focus on these elements of the speech at the expense of their context. Their perception of these speeches was therefore distorted: the pro-Castro sentences stood out, causing the readers to take the speeches to be more strongly partisan than in fact they were, and this made them seem more strongly partisan than was necessary to follow the instruction, and so indicative of a partisan attitude (Gilbert and Malone, 1995, pp. 28-9).

The fourth and final problem is that we explain behaviour by first attributing any trait that will explain it, however unusual that trait may be, and then correcting that attribution as we gain more information about the situation and the way the agent perceived the situation. Gilbert and Malone cite a wide range of psychological literature to support this model of the attribution process. The first stage is spontaneous, but the second requires some amount of thoughtful deliberation. So the second stage is far more susceptible than the first to interruption or impairment by competing cognitive demands. Unwarranted dispositional attributions are therefore more likely when the observer is engaged in other tasks than when the observer is free to spend time correcting the initial attribution. This effect has been demonstrated in an experiment in which subjects were shown film of a woman nervously engaged in conversation with a stranger. The sound had been removed, but subtitles indicated the topic under discussion. Some topics were mundane, others could reasonably be expected to induce anxiety. Subjects who had been asked to rehearse a series of word strings while watching this film tended to describe the woman as an anxious person regardless of which topic she was discussing, whereas subjects who were free to concentrate on the film were less likely to describe her as an anxious person when she was discussing topics likely to induce anxiety than when she was discussing mundane topics. It seems that the subjects were likely to first categorise her as anxious, and then revise this in the light of the subtitles so long as they were not distracted from doing so (Gilbert and Malone, 1995, p. 29).

These four difficulties account for all the observed trait-attribution errors in the experimental literature, and they do so in a way that explains our predilection towards

unwarranted trait-attributions without claiming or implying that trait-attribution itself is unwarranted. Our longitudinal acquaintance with the individuals around us may ground reliable characterological intuitions. The challenge to those who advocate the exclusive use of experimental data in this area is clear: to find reliable evidence of a common error of attribution that cannot be explained in the terms proposed by Gilbert and Malone, and that is best explained in a way that shows informal longitudinal acquaintance with individuals to involve a bias or an unreliable heuristic or to be systematically misleading for some other reason.

7. CONCLUSIONS

The view that we should abandon our characterological intuitions and rely solely on experimental psychology for our understanding of behaviour, therefore, is not supported by the experimental data on erroneous trait-attribution. That data does not show that we could very well be mistaken about the existence and nature of traits altogether. Harman's epistemological argument that aims to impugn our entire set of intuitions about the explanation of behaviour in terms of traits that have a certain kind of structure, combine in certain ways, and develop in certain ways fails. There is no *fundamental* attribution error that impugns attribution itself, there is only a range of attribution difficulties that account for the ways in which attribution can go wrong.

This better interpretation of the literature on erroneous attribution explains why the more time we spend with someone, the better we become at predicting their behaviour: we do not need to rely on the availability heuristic to judge the probability that a close friend has a certain trait, but can judge from our experience of their behaviour; we do not have to assume that the aspects of the situation salient to our friends are those salient to ourselves, since our experience of their behaviour and discussions with them help us to see how they see the situation; we can classify observed behaviour in the context of previous observed behaviour rather than in the context of the situation alone; and we can take the time to revisit and revise our attributions in the light of new information. Of course, we may still make mistakes by failing to notice all the important aspects of a given situation, but this is hardly a reason to suggest that we do not and cannot understand the people closest to us. We make errors about our friends, it is true, but we need not.

The experimental data on trait-attribution certainly does seem to indicate, however, that we should not rely on our intuitions for our understanding of the distribution of traits and the correlations between them. We are usually wrong about how most people are. This information should be drawn from not from our common-sense, but from more objective sources including psychological experiments but also other records of behaviour such as sociological surveys and crime statistics. But our intuitions can indeed embody expertise about the nature and development of character, and about the relative values of certain traits, an expertise grounded in longitudinal acquaintance with a variety of real individuals. The writings of careful informal observers of the behaviour of their closest acquaintances could also provide us with such insights.

Harman's discussion of erroneous trait-attribution might, however, seem to raise a different epistemological objection to the reliance on intuition in this area, a different epistemological reason to prefer experimental data. This would concern not the process of attribution itself, but the process of drawing intuitions about character from one's experience of longitudinal acquaintance with various individuals. Might our intuitions themselves involve the confirmation bias that Harman mentions, or some other bias, irrespective of how trait-attribution should be understood? Might the process by which intuitions are garnered from our characterological expertise be subject to a bias that invalidates the use of intuitions in this area? Experimental reports can be judged in a reliable and uncontroversial way, by assessing their experimental and statistical methods. How can we assess our intuitions for bias?

There is much that could be said about this general worry concerning the application of common-sense expertise built up from experience, but two closing comments sketching a defence should help to assuage it. First, it is important to remember that the employment of intuition is not simply the presentation of immediate knee-jerk reactions. Traditional philosophical discourse rather involves carefully considered and detailed assessments of ideas that actively try to seek out and accommodate intuitions that might appear to favour opposing views. Second, this is a collective enterprise whose participants vary in background and experience and hence in the kinds of people with whom they have been acquainted over significant periods of time. So long as these participants are sufficiently open-minded and sufficiently critical, then it seems that they can indeed learn about the nature of character traits, the ways in which they develop, the ways in which they can conflict or harmonise, and the practical advantages and disadvantages of particular traits through traditional philosophical discussion of intuitions rooted in informal, personal acquaintance.

ACKNOWLEDGEMENTS

This paper has benefited greatly from discussion at the University of Trieste. I am grateful to Marina Sbisà for organising that event. I am also grateful for very helpful comments from two anonymous referees for this journal.

REFERENCES

- Athanassoulis, N., A Response to Harman: Virtue Ethics and Character Traits, *Proceedings of the Aristotelian Society* 100 (2) (2000), pp. 215-221.
- Arkin, R. M., and Duval, S., Focus of Attention and Causal Attributions of Actors and Observers, *Journal of Experimental Social Psychology*, 11, pp. 427-438.
- Doris, J., Persons, Situations, and Virtue Ethics, *Noûs* 32 (4) (1998), pp. 504-530.
- Doris, J., *Lack of Character: Personality and Moral Behaviour*. Cambridge: Cambridge University Press, 2002.
- Doris, J., and Stich, S., As a Matter of Fact: Empirical Perspectives on Ethics, in F. Jackson and M. Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press, 2005.
- Gilbert, D. T., and Malone, P. S., The Correspondence Bias, *Psychological Bulletin* 117 (1) (1995), pp. 21-38.
- Goldie, P., *The Emotions: A Philosophical Exploration*. Oxford: Clarendon, 2000.
- Goldie, P., *On Personality*. London: Routledge, 2004.
- Harman, G., Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error, *Proceedings of the Aristotelian Society* 99 (3) (1999), pp. 315-331.
- Harman, G., The Nonexistence of Character Traits, *Proceedings of the Aristotelian Society* 100 (2) (2000), p. 223-225.
- Hursthouse, R., Virtue Theory and Abortion, *Philosophy and Public Affairs* 20 (3) (1991), pp. 223-246.
- Jones, E. E., and Harris, V. A., The Attribution of Attitudes, *Journal of Experimental Social Psychology* 3 (1) (1967), pp. 1-24.
- Kamtekar, R., Situationism and Virtue Ethics on the Content of Our Character, *Ethics* 114 (3) (2004), pp. 458-491.

- Kupperman, J., *Character*. Oxford: Oxford University Press, 1991.
- McArthur, L. Z., and Post, D. L., Figural Emphasis and Person Perception, *Journal of Experimental Social Psychology* 13 (6) (1977), pp. 520-535.
- Milgram, S., *Obedience to Authority: An Experimental View*. New York, Harper and Row, 1974.
- Nussbaum, M. C., 'Virtue Ethics: A Misleading Category?', *Journal of Ethics* 3 (3) (1999), pp. 163-201.
- Ross, L., The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process, in L. Berkowitz (ed.), *Advances in Experimental Social Psychology* vol. 10. New York: Academic Press, 1977.
- Ross, L., and Nisbett, R., *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill, 1991.
- Sabini, J., and Silver, M., Lack of Character? Situationism Critiqued, *Ethics* 115 (3) (2005), pp. 535-562.
- Taylor, S. E., and Fiske, S. T., Point-of-View and Perceptions of Causality, *Journal of Personality and Social Psychology* 32 (3) (1975), pp. 439-445.
- Trianosky, G. V., 'What Is Virtue Ethics All About?', in D. Statman (ed.), *Virtue Ethics: A Critical Reader*. Edinburgh: Edinburgh University Press, 1997.
- Webber, J., Virtue, Character and Situation, *Journal of Moral Philosophy* 3 (2) (2006): 195-216.