



Character, Common-Sense, and Expertise

Jonathan Webber

Esercizi Filosofici 1, 2006, pp. 15-32

ISSN 1970-0164

link: <http://www.univ.trieste.it/~eserfilo/art106/webber106.pdf>

CHARACTER, COMMON-SENSE, AND EXPERTISE

Jonathan Webber

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. But if moral value attaches primarily to actions in respect of the intentions behind them or the consequences they have, then the development of good character might be the best way to promote good action (see Nussbaum 1999: esp. § I; Trianosky 1997: esp. § 3). From these different perspectives, philosophers recommend a range of character traits. But there is a growing dissatisfaction with this consensus, rooted in a concern that the psychological picture involved is unwarranted. 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: 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: 114).

Doris and Stich, of course, are being provocative: they know very well that the philosophers they have in mind would deny that the psychology involved in their theories is simply invented. But the challenge presented by Harman, Doris, and Stich is to show that the understanding of character involved in this area of ethical discourse has a respectable empirical grounding. There are two ways in which philosophers have argued that it does not. One is to argue 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: 316, 325-6). Various philosophers have responded to this kind of argument by claiming that it misconstrues the characterological claims that moral philosophers typically make (e.g. Kamtekar 2004; Webber 2006).

Less attention has been paid to the second kind of argument against the empirical respectability of characterological moral philosophy, Harman's directly epistemological argument grounded in empirical psychology. Research into trait-attribution has shown, he claims, that the ordinary understanding of character employed in philosophical discourse arises from a misleading heuristic and is consolidated by a cognitive bias. If this argument is right, it shows at least that we should try to escape the influence of this heuristic and this bias by grounding our explanations of behaviour in the findings of experimental psychology.

This paper is concerned with this issue of the epistemology of character. Before Harman's argument can be assessed, it is important to be clear about just how moral philosophers formulate their psychological outlook: the next section explains the use of non-scientific observation and consideration of literary narratives for this purpose. Harman's argument against this methodology is elaborated in sections 3 and 4, and then criticised in section 5 for mis-interpreting the experimental data on which it is based. A better interpretation of that data is explained in section 6, and the final section shows how this better interpretation positively licenses the traditional philosophical methodology in certain areas of characterological psychology. Experimental research into trait-attribution itself therefore undermines the claim that theories of character should rely solely on experimental research.

2. Longitudinal acquaintance

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 project, and indeed is incompatible with the traditional project of recommending traits for everyone to adopt.

For Harman's argument to show that this ethical approach should rely on experimental psychology for its characterological claims, therefore, it would need to show that we are generally mistaken not about which traits people in fact have, but about what traits are, how traits develop, how they relate to one

another, and the impacts they have on our lives. We will see that Harman's argument itself does have these implications, but that the data he relies on is better interpreted in a way that does not.

Philosophers traditionally base their understanding of these aspects of character on consideration of thought-experiments and literary narratives. The thought-experiments are designed to elicit our intuitions about character that are themselves grounded in our accumulated experience of trying to explain and predict the behaviour of those around us. The insights they are taken to provide are therefore grounded ultimately in our own longitudinal acquaintance with our family, friends, colleagues, and neighbours. Of course, while some of our judgements about the nature and relative values of character traits may be mistaken, this is to be explained by their being made without sufficiently careful consideration.

Our experience of a few individuals over significant stretches of time has not only been used to justify the careful employment of intuitions in this area, but has also been used to argue that these intuitions are superior to the findings of experimental psychology in this area (Kupperman 1991: 162-4). This is because longitudinal experimental data is extremely rare. Most psychological experimentation is latitudinal, studying a number of people in a given situation. The reasons for this are partly logistical, partly ethical: professional psychologists need to be able to publish results frequently, the funding systems available to them reflect this aim, and anyway 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 ideally require long-term secret surveillance of the public and private lives of unwitting subjects.

Philosophers do not rely solely on their own intuitions, of course, but also consider the characterological ideas of various salient authors. These can be divided between those who present their ideas in literary form, such as Homer, Shakespeare, or Dostoyevsky and those who present their ideas in prose both more concise and more direct, such as Hume, Rochefoucauld, or Nietzsche. In considering these kinds of writings, philosophers do not simply make immediate intuitive judgements of plausibility or implausibility, but present argued assessments that draw on insights from other writings but also and ultimately on their own experience. This involves not only remembering acquaintances' patterns of behaviour, but might also involve trying to see new patterns in that behaviour, or understand the existing patterns in a new light. These assessments can also be made in dialogue with previous assessments of the same ideas, which expands still further the body of experience that they draw on.

Two influential examples of this kind of careful consideration involve literary narratives. Martha Nussbaum (1990: ch. 4), for example, reads Henry

James's novel *The Golden Bowl* as recommending the trait of responding to situations through sophisticated attention to their details rather than through stock responses to salient general features, and as recommending this not on the grounds that it will ensure that one's life will go well, but on the grounds that the trait equips one as best as possible to face the vicissitudes of life. Bernard Williams (1993) puts a very different kind of literary narrative to a very different use. Careful consideration of the works of Aeschylus, Euripides, Homer, and Sophocles, he argues, highlights the positive value of accepting responsibility for actions that flow from our characters, regardless of whether we could in any real sense have done otherwise, and of regulating our behaviour by adopting a proper sense of shame with regard to our less noble motivations and actions. Nussbaum and Williams take these literary narratives to embody useful characterological insights grounded in their authors' own longitudinal acquaintances.

The difference between presentations of characterological ideas in literary narrative and their presentation in other forms is not merely stylistic. As Nussbaum (1990, ch. 1) has argued, literary narratives can present ideas about the structure and development of character that cannot be presented in other forms. There are two aspects of narrative that ground this unique ability: its attention to the fine-grained detail of situations, and its attention to the agent's subjective understanding of situations. These are, of course, related. The fine-grained detail of a situation can include long-term relationships between the various agents in that situation, and this can include the emotional attitudes each agent has towards the others, the aims and projects each agent pursues, each agent's experiences of their relationships with the other agents in the situation, and the ways in which each imagines these relationships developing. The difference between the actual details of the situation and those noticed and taken into account by an agent might also be illuminating, as might the differences between the ways two agents construe the same situation. As Nussbaum points out, presenting this wealth of detail about a situation and the agents within it just is presenting a complex narrative, however well or poorly written.¹

Literary narratives can also provide readers with new longitudinal acquaintances for consideration. Since fictional characters may be rooted in cultures different from the reader's own, moreover, they may represent acquaintances the reader would never have made, or at least were unlikely to know well, in reality. Of course, fictional characters may have unrealistic

¹ Nussbaum presents this point along with the related idea that our ethical judgement of the behaviour of literary characters also depends on our understanding of the fine-grained detail of their situations, and of their aims, values, emotional attachments, and imaginative hopes, and so is more nuanced than any judgement available on the basis of thought-experiments. This claim about the moral utility of literary narrative can be separated from the point relevant to our concern: that literary narratives are uniquely apt for presenting ideas about the nature and integration of various character traits and the behaviour they lead to in certain types of situation.

behavioural traits and so could potentially mislead the reader into formulating a false understanding of character. But readers are free to draw their own conclusions from literary narratives, however much authors might try to direct them. They can draw these conclusions partly by considering the narrative in relation to their experience of other real and fictional characters. They can also discuss literary characters with a wide variety of other readers, each with equal access to their behaviour patterns, and since literary characters have no feelings to hurt and no interests to protect, they can be discussed without compunction. Consideration of literary characters therefore has advantages over consideration of personal experience in the formulation and development of characterological ideas.

It is hardly surprising, then, that philosophical moral discourse concerned with character development often employs literary narratives not only to illustrate claims but also to support them. Thought-experiments have their place in eliciting our intuitions, but literary narratives can also help to formulate the understanding of character those intuitions express. Given that this discourse is essentially grounded in our longitudinal acquaintance with our nearest and dearest, and since this kind of longitudinal acquaintance is not really available in any other way, it seems that abandoning this methodology in favour of exclusive reliance on experimental psychological data would be inimical to the project of discovering the kinds of character traits that best equip us for the vicissitudes of life. But such limitation would be necessary and indeed beneficial, of course, were the longitudinal acquaintance that grounds this method to be worthless as a guide to the nature and development of character and the relative values of particular traits.

3. Erroneous attribution

Harman's argument against the use of common-sense characterological psychology in ethical discourse makes precisely this claim: that our longitudinal acquaintance with those around us does not consist in drawing conclusions about the sources of their behaviour on the basis of observation. He draws on experimental data to argue that our conclusions about character are drawn on the basis of a misleading heuristic that he follows Lee Ross in calling the «fundamental attribution error» (Harman 1999; Ross 1977). This heuristic is claimed to explain our tendency, repeatedly demonstrated in a wide variety of experiments since the 1940s, to attribute a person's behaviour to uncommon dispositions had by that person, even when the person in fact behaved as most people would behave in that situation. He 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 fair, 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 (see Ross 1977: 325). If he is right about this, then it seems that our understanding of character based on longitudinal acquaintance is unreliable, and so we should indeed abandon the traditional philosophical discussion of thought-experiments and literary narratives in favour of analysing the results of psychological experiments.

In order to judge whether his interpretation of the literature on erroneous trait-attribution is correct, we need to clarify the data embodied in that literature. 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 deference, 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 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: 22-24).

The terms in which this is discussed can be misleading. It is often described as showing 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: 217).² 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 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 do 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 we add (or it is already understood) that he has a strange pathological fear of 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

² Milgram himself describes his results in terms of a disposition towards obedience (1974: 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, then we could still agree with Sabini and Silver (2005: 550-1) that the subjects' behaviour is to be explained in terms of the character trait of deference to expertise.

threats are «temporally or spatially removed from the behavioural episodes they constrain» (Gilbert and Malone 1995: 25). We should understand the terms «situational» and «dispositional» in these technical senses, and take care not to confuse these with their ordinary senses.

Harman 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 a two-part argument against the use of characterological explanation. One argument 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 traits. Since we instead find a striking uniformity of behaviour, 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 (Harman 1999: 316, 325-6).

If this is right, he argues, then moral philosophy should not rest on common-sense intuitions about character, but should instead embrace the idea that behaviour is best explained and predicted by reference to situational features rather than dispositional ones (Harman 1999: 324-330; see also 2000: 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.

But this aspect of Harman's discussion is not our concern here. We are concerned with the second, epistemological part of his argument, 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, then we should reject the use of thought-experiments and literary narratives as resources in the philosophical discussion of character, 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 should lead us to reject the traditional philosophical methodology in this area depends on

exactly how and why trait-attribution tends to go wrong. Once we understand this, we can understand whether it is possible to guard against such errors, and if so how. 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: 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. (Harman 1999: 325)

Harman supports this claim by referring to an influential textbook of social psychology, which in turn cites three studies in support of its claim that «what you attend to is what you attribute to» (Ross and Nisbett 1991: 140). We will see in section 5, however, that these studies simply do not support the claim, that there is good experimental reason to reject the claim, and that the claim anyway could not do the explanatory work required of it. But first, it is important to see why this idea about the salience of the agent, were it correct, should lead us to abandon the use of thought-experiments and literary narratives in philosophical discussions of character.

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. 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 were correct, then it would show that we should rely on latitudinal experimental data for information about the distribution of traits, but it would not undermine our general understanding of

the nature of character and 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: 166; 2004: 52-69). If this is right, then informal longitudinal acquaintance could still be a reliable source of insights for philosophers concerned with character.

Harman's interpretation of the data leads to the more radical conclusion that even our understanding of those closest to us is erroneous. 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 characters to explain the behavioural differences between those closest to us, when in fact these differences are due to the differences between the situations in which they find themselves. This interpretation therefore leads to the wholesale rejection of common-sense characterological psychology, as Harman rightly points out, rather than to the restrictions on its use required by Goldie's less radical interpretation. Harman's rejection of common-sense brings with it a rejection of the use of thought-experiments and literary narratives, because our judgements about these will result from this illusion, as will the ideas their authors want to present.

In the next section, we will see that Harman's interpretation of the data is not viable. The idea that philosophical ethical discourse should abandon the use of thought-experiments and literary narratives in favour of exclusive reliance on experimental data therefore requires a different interpretation of the literature on trait-attribution. As we will see in section 6, however, there is a perfectly plausible interpretation of that literature available that does not have such radical consequences. This better interpretation in fact licenses certain uses of thought-experiments and literary narratives in discussions of virtue and character, as we will see in the final section. The new interpretation of the literature on trait-attribution required to support the exclusive use of experimental data in philosophical discussions of character must therefore be shown to be superior to this less radical interpretation.

5. Harman's errors

Our ability to predict someone's behaviour increases the more time we spend with that person, argues Goldie (2000: 166), and since this is best explained by our increasing understanding of that person's character, Harman's interpretation

of the literature on trait-attribution must be wrong. Harman's position is easily defended from this criticism: our increasing ability to predict an individual's behaviour is explained equally well by our increasing knowledge of the details of the kinds of situations they find themselves in, even if we are unaware that this is what we increasingly know. When someone's behaviour surprises us, moreover, this is because some situational feature is novel or unusual for that person, or at least is so in our experience of that person, and not because of an abrupt alteration of their character. If Harman's position can be shown to be wrong on other grounds, of course, then the way is clear for accepting Goldie's interpretation of our increasing understanding of those around us. Indeed, any acceptable interpretation of the data must be compatible with the fact that we are better at predicting a person's behaviour the more time we spend with that person. But this fact does not undermine Harman's position.

There are better reasons to reject the interpretation of the data on attribution 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. This interpretation seems unacceptable, for two distinct reasons.

The first is that a careful survey of the experiments cited in the social psychology textbook to which Harman refers has shown that they do not support the claim that we tend to explain behaviour in terms of properties of the salient figure, and that they even present some evidence against this claim (Gilbert and Malone 1995: 30-32). The first study cited by the textbook (McArthur and Post 1977) is claimed there to have «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» (Ross and Nisbett 1991: 140). Some of the results of this study indeed suggest this, and some suggest that increasing the salience of certain aspects of the situation increases the likelihood that an observer would explain the behaviour in situational terms. But most of the results seem to show exactly the reverse: that increasing the salience of the agent makes situational attribution more likely. From this mixed outcome, the authors concluded 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: 534).

The second study (Arkin and Duval 1975) is claimed to have «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» (Ross and Nisbett 1991: 140). This is a somewhat distorted presentation of the results. The agent chose an artwork from a selection presented either on a display board or on a video-

tape 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: 432). The experiment did show that both agents and observers gave higher mean scores for the relevance of situational factors when the artwork was chosen from a video than when it was chosen from a static presentation (Arkin and Duval 1975: 434). But an analysis of the ratings of the relevance of dispositional factors «revealed no significant main effects or interactions» (Arkin and Duval 1975: 434). The increase in the salience of the situation, by presenting moving rather than static images to choose from, therefore made no significant difference to the observer's rating of the importance of dispositional factors in the agent's choice.

The third study (Taylor and Fiske 1975) is claimed to have shown 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» (Ross and Nisbett 1991: 140). Whilst this is true, it is off the point. The experimenters also concluded that the salient participant's behaviour «was not seen as indicative of his dispositions, nor was his partner's behaviour seen as situationally based» (Taylor and Fiske, 1975: 442). In versions of the experiment where some subjects were explicitly instructed to attend to a particular participant, the experimenters 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» (Taylor and Fiske, 1975: 443). A later variant of the experiment (Ellis and Holmes 1982) found that participants in the conversation are no more likely to describe their interlocutor's behaviour in dispositional terms if they attend to the interlocutor than if they do not.

The experimental data cited in support of Harman's interpretation of the data on trait-attribution, therefore, fails to support it, and might even be taken as evidence against it. The second reason to reject this interpretation is that it anyway seems to have lost sight of the tendency in need of explanation. Even if it were true that we focus on the salient agent, this would fail to explain why we tend to cite *uncommon* dispositions *characteristic* of the agent rather than more common dispositions. 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 simply 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 seems that this explanation understands the ideas of dispositional and situational explanations in terms of the ordinary meaning of their labels, rather than in their technical sense outlined earlier. We can explain our preference for dispositional

explanations in the technical sense, moreover, if we employ a different interpretation of the data, as we shall see in the next section.

6. The roots of erroneous attribution

Daniel Gilbert and Patrick Malone (1995) have offered an interpretation of the experimental data concerning trait-attribution that seems to account for all of the scenarios in which observers have been shown to mistakenly explain behaviour in terms of uncommon traits characteristic of the agent, but which allows that our characterological understanding of those people with whom we are well acquainted over long stretches of time is largely accurate. This interpretation would also explain why we become better at predicting a person's behaviour the more time we spend with that person. It would meet, that is, the constraint that was drawn out of Goldie's criticism of Harman at the beginning of the last section. If this interpretation is correct, therefore, reliable characterological intuitions can indeed be grounded in non-scientific longitudinal acquaintance, and at least some uses of thought-experiments and literary narratives in philosophical characterological discourse are therefore justified.

The first step Gilbert and Malone take is to abandon the assumption pervasive in earlier literature, and clearly present in the interpretation favoured by Harman, 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» (Gilbert and Malone 1995: 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: 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: 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, such as obedience or deference (see Gilbert and Malone 1995: 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: 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 discussing one of a set of topics. 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: 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, though as we will see in the next section this analysis of the data sets certain constraints on the range of those 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 an unreliable heuristic or to be systematically misleading for some other reason.

7. Conclusions

The view that we should abandon the use of thought-experiments and literary narratives in philosophical ethical discussions of character is therefore unsupported by the experimental investigation of trait-attribution. This data does not show, as Harman claims it does, that our longitudinal acquaintance with our family, friends, colleagues, and neighbours involves a «fundamental attribution error» that explains all of our erroneous trait-attributions by showing trait-attribution itself to be mistaken. It is better interpreted as showing that there are various ways in which our attributions can and do go wrong, ways that can be avoided.

This is the truth behind Goldie's view that 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 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.

The experimental data certainly does seem to indicate, however, that we should not rely on our intuitions for our understanding of the distribution and correlation of traits. This information should be drawn from latitudinal experiments. But our intuitions can indeed embody expertise about the nature and development of character, and about the relative values of certain traits, an expertise drawn from longitudinal acquaintance with a variety of real individuals. The data provides no reason, moreover, why this expertise should not also be drawn from discussion of literary characters. Careful consideration of thought-experiments, literary characters, and real cases can elicit intuitions that manifest this expertise. For reasons discussed earlier, our expertise can be increased by consideration and discussion of literary narratives as well as by experience of real cases.

Harman's discussion of erroneous trait-attribution might, however, seem to raise a different epistemological objection to the use of thought-experiments and literary narratives 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 judgements of literary narratives for plausibility and relevance or our reactions to thought-

experiments be rooted in the «confirmation bias» that Harman mentions, regardless of whether there is any such thing as «fundamental attribution error»? The two certainly seem to be distinct. Nothing in this paper so far has shown that our reactions to fictional scenarios might not be shaped by a tendency to look for confirming instances of a claim rather than for both confirming and disconfirming ones. There might, indeed, be further cognitive biases that call into question the whole process of garnering intuitions from experience. We can judge experimental reports, on the other hand, in a reliable and uncontroversial way: by assessing their experimental and statistical methods.

There is much that could be said about this general worry concerning the application of expertise built up from experience, but two closing comments sketching a defence of this process should help to assuage it. First, it is important to remember that philosophers do not simply present their immediate knee-jerk reactions to thought-experiments or literary narratives, but instead present carefully considered and detailed assessments that actively try to seek out and accommodate intuitions that might appear to favour opposing views. Second, the assessment of thought-experiments and literary narratives is a collective enterprise whose participants vary in background and experience and hence in the kinds of people with whom they are or 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 the philosophical discussion of thought-experiments and literary narratives.

References

- Athanassoulis, N., *A Response to Harman: Virtue Ethics and Character Traits*, «Proceedings of the Aristotelian Society», 100, 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*, «Nous», 32, 1998, pp. 504-530.
- Doris, J., *Lack of Character: Personality and Moral Behaviour*, Cambridge University Press, Cambridge 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 University Press, Oxford 2005.
- Ellis, R.J., and Holmes, J.G., *Focus of Attention and Self-Evaluation in Social Interaction*, «Journal of Personality and Social Psychology», 43, 1982, pp. 67-77.
- Gilbert, D.T., and Malone, P.S., *The Correspondence Bias*, «Psychological Bulletin» 117, 1995, pp. 21-38.
- Goldie, P., *The Emotions: A Philosophical Exploration*, Clarendon, Oxford 2000.
- Goldie, P., *On Personality*, Routledge, London 2004.

- Harman, G., *Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error*, «Proceedings of the Aristotelian Society», 99, 1999, pp. 315-331.
- Harman, G., *The Nonexistence of Character Traits*, «Proceedings of the Aristotelian Society», 100, 2000, p. 223-225.
- Jones, E.E., and Harris, V.A., *The Attribution of Attitudes*, «Journal of Experimental Social Psychology», 3, 1967, pp. 1-24.
- Kamtekar, R., *Situationism and Virtue Ethics on the Content of Our Character*, «Ethics», 114, 2004, pp. 458-491.
- Kupperman, J., *Character*, Oxford University Press, Oxford 1991.
- McArthur, L.Z., and Post, D.L., *Figural Emphasis and Person Perception*, «Journal of Experimental Social Psychology» 13, 1977, pp. 520-535.
- Milgram, S., *Obedience to Authority: An Experimental View*, Harper and Row, New York 1974.
- Nussbaum, M.C., *Virtue Ethics: A Misleading Category?*, «Journal of Ethics», 3, 1999, pp. 163-201.
- Nussbaum, M. C., *Love's Knowledge: Essays on Philosophy and Literature*, Oxford University Press, Oxford 1990.
- Ross, L., «The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process», in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, vol. 10, Academic Press, New York 1977.
- Ross, L., and Nisbett, R., *The Person and the Situation: Perspectives of Social Psychology*, McGraw-Hill, New York 1991.
- Sabini, J., and Silver, M., *Lack of Character? Situationism Critiqued*, «Ethics», 115, 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, 1975, pp. 439-445.
- Trianosky, Gregory Velazco, «What Is Virtue Ethics All About?», in D. Statman (ed.), *Virtue Ethics: A Critical Reader*, Edinburgh University Press, Edinburgh 1997.
- Webber, J., *Virtue, Character and Situation*, «Journal of Moral Philosophy », 3, 2006, pp. 193-213.
- Williams, B., *Shame and Necessity*, University of California Press, Berkeley 1993.